

# Perspectives on Distributed Computing: Thirty People, Four User Types, and the Distributed Computing User Experience

---

by

L. Childers, L. Liming, and I. Foster

Distributed Systems Laboratory

Mathematics and Computer Science Division, Argonne National Laboratory

Computation Institute

University of Chicago

September 30, 2008





















UChicago ►  
Argonne<sub>LLC</sub>



















A U.S. Department of Energy laboratory managed by UChicago Argonne, LLC



## Table of Contents

<b>1</b>	<b>EXECUTIVE SUMMARY.....</b>	<b>5</b>
<b>2</b>	<b>INTRODUCTION .....</b>	<b>6</b>
<b>3</b>	<b>INTERVIEW SUMMARIES: AN INTEGRATED VIEW.....</b>	<b>8</b>
3.1	GOALS.....	9
3.2	ISSUES.....	11
3.3	SATISFACTION POINTS .....	13
<b>4</b>	<b>CHARACTERIZING USERS BY THEIR INTERACTIONS WITH TECHNOLOGY .....</b>	<b>14</b>
4.1	CHARACTERIZATION METHOD .....	14
4.2	 TYPE 1: THE HPC SCIENTIST .....	16
4.3	 TYPE 2: THE HPC DOMAIN-SPECIFIC DEVELOPER.....	21
4.4	 TYPE 3: THE GENERAL-PURPOSE HPC INFRASTRUCTURE PROVIDER .....	33
4.5	 TYPE 4: THE GENERAL-PURPOSE HPC TECHNOLOGY DEVELOPER .....	42
<b>5</b>	<b>RECOMMENDATIONS.....</b>	<b>48</b>
5.1	FOR DEVELOPERS OF DISTRIBUTED COMPUTING TECHNOLOGY .....	48
5.2	FOR FURTHER STUDY .....	53
	<b>ACKNOWLEDGMENTS .....</b>	<b>55</b>
	<b>APPENDIX A STUDY METHODOLOGY .....</b>	<b>56</b>
	<b>APPENDIX B THE INTERVIEWEES.....</b>	<b>59</b>
	<b>APPENDIX C INTEGRATED SUMMARY DATA (FOR SECTION 3) .....</b>	<b>62</b>
C.1	USER GOALS .....	62
C.2	USER ISSUES .....	70
C.3	SATISFACTION POINTS.....	100
	<b>APPENDIX D THE INTERVIEWS .....</b>	<b>106</b>
D.1	 THE SCIENTISTS' HAPPINESS IS MY MAIN MEASURE OF SUCCESS .....	106
D.2	 TROUBLESHOOTING REQUIRES KNOWLEDGE ABOUT SOFTWARE INTERNALS .....	111
D.3	 THE GRID IS A BLACK BOX TO ME .....	120
D.4	 THE REASON MY TASKS ARE SO TIME-CONSUMING IS FAILURE.....	124
D.5	 PERFORMANCE IMPROVED FROM DAYS TO SECONDS .....	135
D.6	 THE GRID IDEA IS GREAT, BUT THERE ARE BARRIERS TO MAKING IT WORK TODAY .....	145
D.7	 IF YOU ADD UP ALL THE TOOLS YOU DON'T GET A GOOD USER ENVIRONMENT .....	152
D.8	 I AM TRYING TO UNDERSTAND WHERE GRID COMPUTING ADDS VALUE.....	160
D.9	 GLOBUS ENABLES MORE SCIENCE .....	168
D.10	 SOLVING PROBLEMS IS EASY ONCE YOU HAVE ALL THE DATA .....	180
D.11	 THE DIFFICULTY IS NOT THAT THINGS BREAK, BUT DETECTING THAT SOMETHING IS BROKEN ..	191
D.12	 THE RIGHT APPROACH IS TO BE HIGHLY COLLABORATIVE WITH DOMAIN SPECIALISTS .....	199
D.13	 WE PLAY A STRONG BRIDGE ROLE IN CONNECTING PEOPLE WITH TECHNOLOGY .....	202
D.14	 I START WITH MICROBENCHMARKS AND FOLLOW-UP WITH REAL APPLICATIONS.....	216

D.15		WE PROVIDE MECHANISMS FOR SHARING VIDEO, AUDIO AND APPLICATIONS.....	229
D.16		WE ASSUME A WORLD WHERE LIGHTPATHS CAN BE SCHEDULED BETWEEN COMPUTERS.....	236
D.17		GRAM2 IS KEPT ALIVE BY THE NEED TO INTEROPERATE WITH EUROPEAN EXPERIMENTS .....	241
D.18		WE PROVIDE AN APPLIANCE FOR EACH CLUSTER THAT ACTS AS A PARALLEL HEAD NODE .....	250
D.19		I WOULD LIKE SITES TO SERVE 100,000 OR MORE END-USERS PER WEEK .....	255
D.20		WE CAN PROVIDE OUR USERS WITH FRESH DATA MORE FREQUENTLY BECAUSE OF THE GRID...	267
D.21		WE WORK TO ENABLE DISCOVERY, ACCESS AND SYNTHESIS OF DISTRIBUTED DATASETS.....	273
D.22		OUR GOAL IS TO BRING ATTRIBUTE-BASED AUTHORIZATION TO THE GRID .....	281
D.23		IT WOULD TAKE FOREVER FOR A BIOLOGIST TO GET THIS MACHINERY WORKING .....	288
D.24		THE VAST MAJORITY OF PEOPLE WHO COULD USE SUPERCOMPUTING ARE EXCLUDED .....	295
D.25		THE DIVERSITY OF THE SYSTEMS WE RUN ON IS A PROBLEM .....	303
D.26		OUR GOAL IS TO MAKE IT EASIER TO TROUBLESHOOT GRID APPLICATIONS .....	311
D.27		THE END GOAL IS TO AUTOMATICALLY DETECT NETWORK ANOMALIES .....	318
D.28		THE PRODUCTION WORTHINESS OF INFRASTRUCTURE IS OF THE UTMOST IMPORTANCE.....	326
D.29		OUR FRAMEWORK MUST ADAPT TO CHANGING CONDITIONS FROM THE PROBLEM & THE GRID	334
D.30		THE SCRIPTABLE INTERFACES AT VARIOUS SITES ARE NOT CONSISTENT .....	338

# 1 Executive Summary

This report summarizes the methodology and results of a user perspectives study conducted by the *Community Driven Improvement of Globus Software (CDIGS)* project. The purpose of the study was to document the work-related goals and challenges facing today’s scientific technology users, to record their perspectives on Globus software and the distributed-computing ecosystem, and to provide recommendations to the Globus community based on the observations. Globus is a set of open source software components intended to provide a framework for collaborative computational science activities.

Rather than attempting to characterize all users or potential users of Globus software, our strategy has been to speak in detail with a small group of individuals in the scientific community whose work appears to be the kind that could benefit from Globus software, learn as much as possible about their work goals and the challenges they face, and describe what we found. The result is a set of statements about specific individuals’ experiences. We do not claim that these are representative of a potential user community, but we do claim to have found commonalities and differences among the interviewees that *may be* reflected in the user community as a whole. We present these as a series of hypotheses that can be tested by subsequent studies, and we offer recommendations to Globus developers based on the assumption that these hypotheses are representative.

Specifically, we conducted interviews with thirty technology users in the scientific community. We included both people who have used Globus software and those who have not. We made a point of including individuals who represent a variety of roles in scientific projects, for example, scientists, software developers, engineers, and infrastructure providers.

The following material is included in this report:

- A summary of the reported work-related goals, significant issues, and points of satisfaction with the use of Globus software
- A method for characterizing users according to their technology interactions, and identification of four user types among the interviewees using the method
- Four profiles that highlight points of commonality and diversity in each user type
- Recommendations for technology developers and future studies
- A description of the interview protocol and overall study methodology
- An anonymized list of the interviewees
- Interview writeups and summary data

The interview summaries in Section 3 and transcripts in Appendix D illustrate the value of distributed computing software – and Globus in particular – to scientific enterprises. They also document opportunities to make these tools still more useful both to current users and to new communities.

We aim our recommendations at developers who intend their software to be used and reused in many applications. (This kind of software is often referred to as “middleware.”) Our two core recommendations are as follows. First, it is essential for middleware developers to understand and explicitly manage the multiple *user products* in which their software components are used. We must avoid making assumptions about the commonality of these products and, instead, study and account for their diversity. Second, middleware developers should engage in different ways with different kinds of users. Having identified four general user types in Section 4, we provide specific ideas for how to engage them in Section 5.

Feedback is appreciated; comments can be sent to [childers@mcs.anl.gov](mailto:childers@mcs.anl.gov).

## 2 Introduction

*We are at the very beginning of time for the human race. It is not unreasonable that we grapple with problems. But there are tens of thousands of years in the future. Our responsibility is to do what we can, learn what we can, improve the solutions, and pass them on. – Richard Feynman*

### **Purpose of the Study**

The purpose of the User Perspectives project is to document the work-related goals and challenges facing today's scientific technology users, to record their perspectives on Globus software and the distributed-computing ecosystem, and to provide recommendations to the Globus community based on those observations. The ultimate goal is to help developers of distributed computing technology better address the needs of the scientific community.

### **Relationship to Prior Work**

Several user studies have looked at scientific computing users and their needs. For instance:

In 2003, Fox and Walker produced a gap analysis for the UK e-Science Programme<sup>1</sup>. They interviewed eighty scientists engaged in e-Science projects in the UK, several European and U.S. e-Science projects, several companies involved in producing Grid middleware, and several organizations that were potential users of Grid middleware. Their report provided a classification of proposed Grid building blocks, identified six “styles” of grids, identified functionality and support pieces missing in the current e-Science programme, and proposed a development plan for filling those gaps.

In 2004, Newhouse and Schopf conducted interviews with twenty-five applied science and middleware groups in the U.K.<sup>2</sup>. They repeated this in 2007 (with Richards and Atkinson) with forty-five interviews with U.K. project members, a workshop, and an online survey<sup>3</sup>. In each study, interviewees were asked about the middleware functionality they had tried, what their applications needed at the present time, and what they felt they would need in the near future. The reports from these studies summarized recurring themes in the responses and proposed corresponding development plans.

In 2006, Zimmerman and Finholt conducted a user requirements workshop for the U.S. TeraGrid aimed at assessing the relationship between TeraGrid's development program and user requirements<sup>4</sup>. At the workshop, twelve invited TeraGrid users provided information on the computational and organizational requirements of their scientific enterprises and the contributions they desired from TeraGrid. The workshop report summarized the user priorities, proposed markers for measuring scientific impact, and discussed anticipated scientific breakthroughs and relevant barriers.

Also in 2006, the Research School of Systems Engineering at Loughborough University conducted a human factors audit of eight selected projects in the U.K. e-Science Programme<sup>5</sup>. The audit team reviewed written materials and conducted interviews with subject matter experts, end users, project developers, and project managers from each project. Data was collected on the scientific application details, technical environments, tasks, users, physical environments, and user types in each project. Based on experiences from these eight projects, the report provided a set of “best practices” in several areas for future projects.

<sup>1</sup> June 30, 2003, Report UKeS-2003- 01. [http://www.nesc.ac.uk/technical\\_papers/UKeS-2003-01/index.html](http://www.nesc.ac.uk/technical_papers/UKeS-2003-01/index.html)

<sup>2</sup> Cluster Computing Journal 10, no. 3, September, 2007

<sup>3</sup> Proceedings of the UK All Hands Meeting, Sept 2007

<sup>4</sup> TeraGrid User Workshop Final Report. Collaboratory for Research on Electronic Work, School of Information, University of Michigan

<sup>5</sup> Human Factors Audit of Selected Projects in the U.K e-Science Projects, issue 3, August 2006

This report differs from these previous studies in several important ways. First, none of these previous studies has published transcripts from their interviews. The goal of our study was not to produce incontrovertible conclusions, but rather to provide a rich set of data on which others can build their own analysis and conclusions. To that end, we are publishing the transcripts of our interviews, with permission from interviewees and in accordance with institutional review board guidelines. Second, our treatment of the data collected in the interviews has not been limited to summarization and description. The analysis in Section 4 of this report uses the data to identify a grouping strategy in which members of each group have similar relationships to technology. The recommendations in Section 5 go further and use the groupings to propose specific strategies for engaging with members of each group. Third, as described in Appendix A, we did not design this study with a particular outcome in mind, such as to develop a plan or a support strategy. Rather, we designed an open-ended interview script and then used qualitative data analysis to tag the transcripts, identify patterns, and examine them in detail. The result – identifying and profiling four user types based on technology interaction patterns – was not what we expected to accomplish at the study’s outset, but, as described in Section 5, we see many uses for this result.

### ***How to Read This Report***

In the remainder of the document we summarize the thirty user interviews (Section 3), categorize the users (Section 4.1), present user profiles for the types of users found in the interview data (Sections 4.2-4.5), and provide recommendations based on our observations (Section 5). Four appendices provide details about the methodology, the interviewees, and the data used for the main body of the report. The interviewees’ words make up the heart of this work; Appendix C (the integrated summary data) and Appendix D (the interviews) include compelling stories, rich in detail and insights. Table 1 briefly summarizes how readers can peruse this report to gain the information they desire.

**Table 1: Report contents**

	<b>Findings</b>	<b>Further Information</b>
<b>No detail</b>	<ul style="list-style-type: none"> <li>- Section 1, Executive Summary</li> <li>- Section 2, Introduction</li> <li>- Section 5, Recommendations</li> <li>- Section 6, Acknowledgements</li> </ul>	<ul style="list-style-type: none"> <li>- Appendix A, Study Methodology</li> <li>- Appendix B, The Interviewees</li> </ul>
<b>Minimal detail</b>	<ul style="list-style-type: none"> <li>- Section 3, Interview Summaries</li> <li>- Section 4, Characterizing Users</li> </ul>	
<b>Greater detail</b>		<ul style="list-style-type: none"> <li>- Appendix C, Summary Data</li> </ul>
<b>Greatest detail</b>		<ul style="list-style-type: none"> <li>- Appendix D, The Interviews</li> </ul>

### 3 Interview Summaries: An Integrated View

This section summarizes the work-related goals, issues, and satisfaction points expressed by the thirty users listed in Appendix B. We do not claim that the viewpoints of the people we interviewed are representative of the user community as a whole. We do claim, however, that these summaries represent the viewpoints expressed in the thirty interviews.

In the following sections, we use “mind maps” to visualize the ideas discussed in the interviews. In each figure the third, or outer, tier contains each unique idea expressed in one or more of the interviews. The second, or intermediate, tier provides one level of generalization, linking several related ideas. Figures 1-3 demonstrate how to interpret a single idea from each of the summary pictures.



Figure 1: "Some people aim to apply science to practical problems."

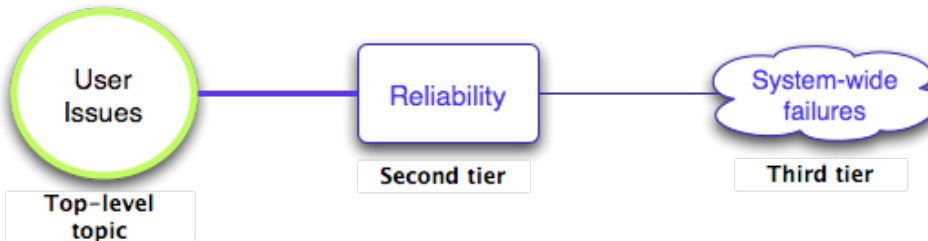


Figure 2: "Some people experienced system-wide failures."

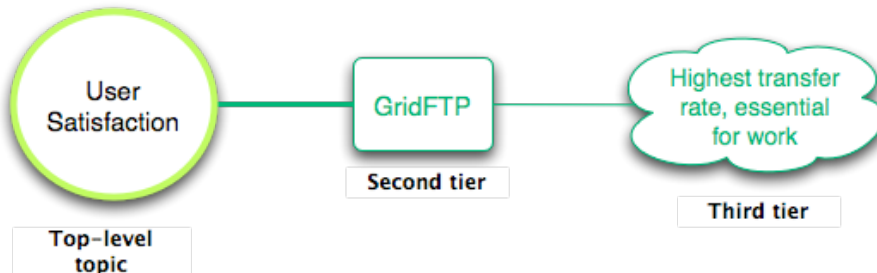


Figure 3: "Some people said that GridFTP provides the highest transfer rate available and that this is essential for their work."



### 3.1 Goals

During their interviews, participants discussed what they are trying to accomplish in their work on a single project of their choosing. Figure 4 provides a visual summary of reported work-related goals, with the rectangular boxes representing a high level of generalization. The summary data for this figure is in Appendix C.1; the summarization method is described in Appendix A.



**Figure 4: Interviewee goals**

Although many scientific disciplines are represented in Figure 4, the summarization yields four top-level goals: conduct and promote scientific research, satisfy user requirements for systems used to do science, expand the community that can use a specific resource, and expand the resources available to a specific community.

Figure 4 can be divided in half, both horizontally and vertically. Comparing the top half with the bottom of Figure 4, we see the top is characterized by goals that are expressed relative to people (social goals) and the bottom is characterized by goals that are expressed relative to technology (technical goals). Comparing the left half with the right, we see that the left-hand goals are expressed in terms of supporting current users and technology (operational goals) and the right-hand goals aim to create new users or technology (development goals).

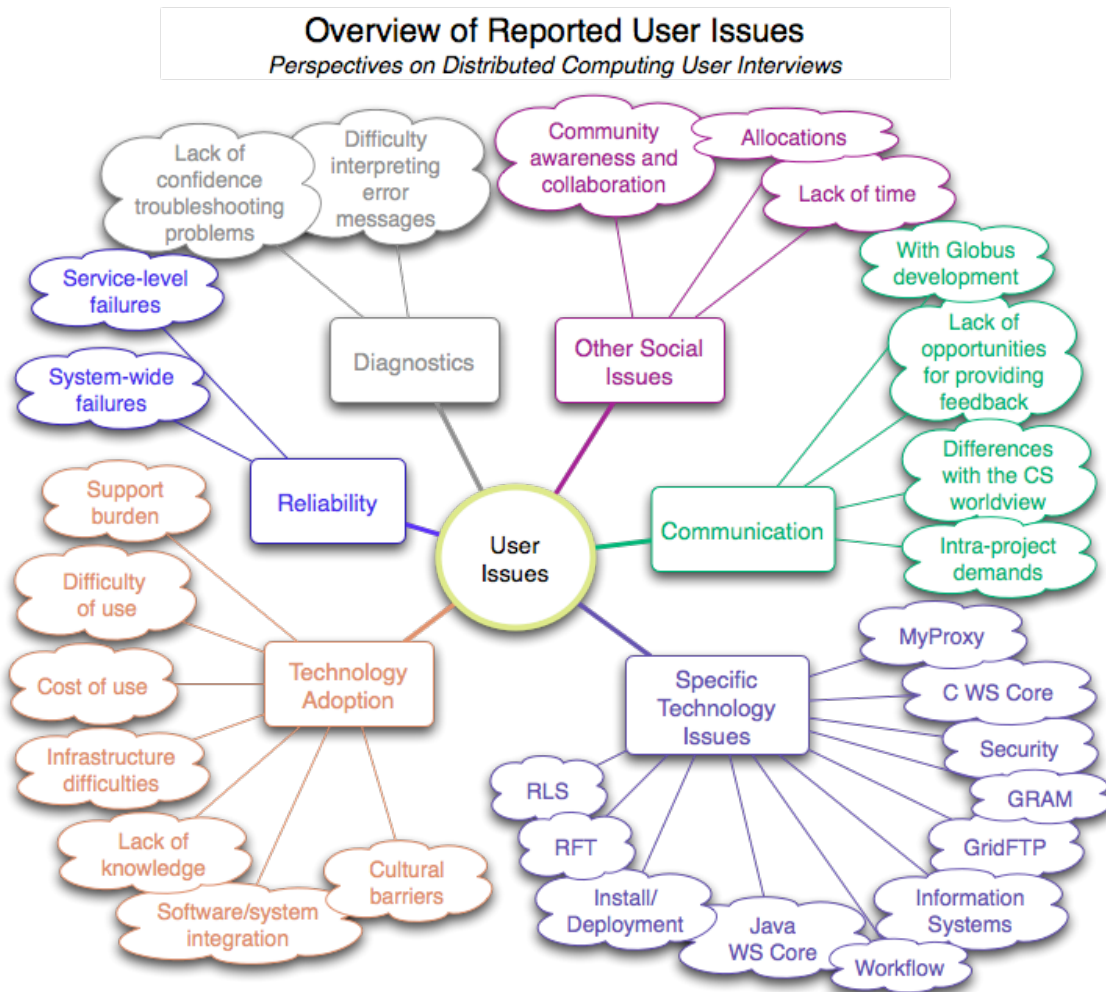
Table 2 shows the goals organized along these lines. This organizing framework shows a diversity of goals among the people we interviewed along multiple axes: social vs. technical and operations vs. development. When explaining the benefits of distributed computing technologies to potential users, we should speak in terms relevant to each of these goal types.

**Table 2: A second perspective on interviewee goals**

<b>Social Operation Goals</b>	<b>Social Development Goals</b>
<ul style="list-style-type: none"> <li>• Extend scientific understanding</li> <li>• Apply science to practical problems</li> <li>• Eliminate barriers to scientific investigation</li> <li>• Build a case for continued financial support</li> </ul>	<ul style="list-style-type: none"> <li>• Make scientific data accessible to more potential users</li> <li>• Make scientific applications accessible to more potential users</li> <li>• Make computation services accessible to more potential users</li> <li>• Make scientific colleagues accessible to more potential users</li> </ul>
<b>Technical Operation Goals</b>	<b>Technical Development Goals</b>
<ul style="list-style-type: none"> <li>• Establish and maintain system stability</li> <li>• Establish uniform diagnostic mechanisms that satisfy debugging needs in dynamic systems</li> <li>• Establish and employ security mechanisms that support dynamic, inter-organizational collaboration</li> <li>• Provide compatibility with existing system components</li> <li>• Establish provisioning mechanisms that efficiently satisfy varying demand</li> </ul>	<ul style="list-style-type: none"> <li>• Federate institutional computing resources</li> <li>• Aggregate cross-institutional resources</li> <li>• Run existing scientific applications at higher resolutions</li> <li>• Enable scientific applications that require coordinated use of multiple systems</li> <li>• Provide computing systems to scientific users</li> </ul>

### 3.2 Issues

During the interviews participants discussed several issues or problems they encounter that slow or stop the pursuit of their goals. This topic was explicitly addressed in the interviews in two contexts: one was a general query without attention drawn to any specific technology, and the other was with a specific focus on the Globus software components that the interviewee indicated they use. Users also spontaneously described issues as background information for other answers. Figure 5 provides an integrated view of the issues expressed by the thirty interviewees. The source data for this figure is in Appendix C.2; our summarization method is described in Appendix A.



<http://www.mcs.anl.gov/~childers/perspectives/>

**Figure 5: User issues**

The *Specific Technology Issues* category contains issues reported by the thirty users that are relevant to a particular component. Issues applying to multiple components are distributed among the other categories. One striking detail illustrated by Figure 5 is that many types of user issues

are not specific to a single component. In fact, the idea that user concerns extend beyond the boundaries of any specific technology is seen again and again in the interview data. This finding is the key motivator for our general recommendation – “Broaden the focus of component-centric development” – described in Section 5.1.

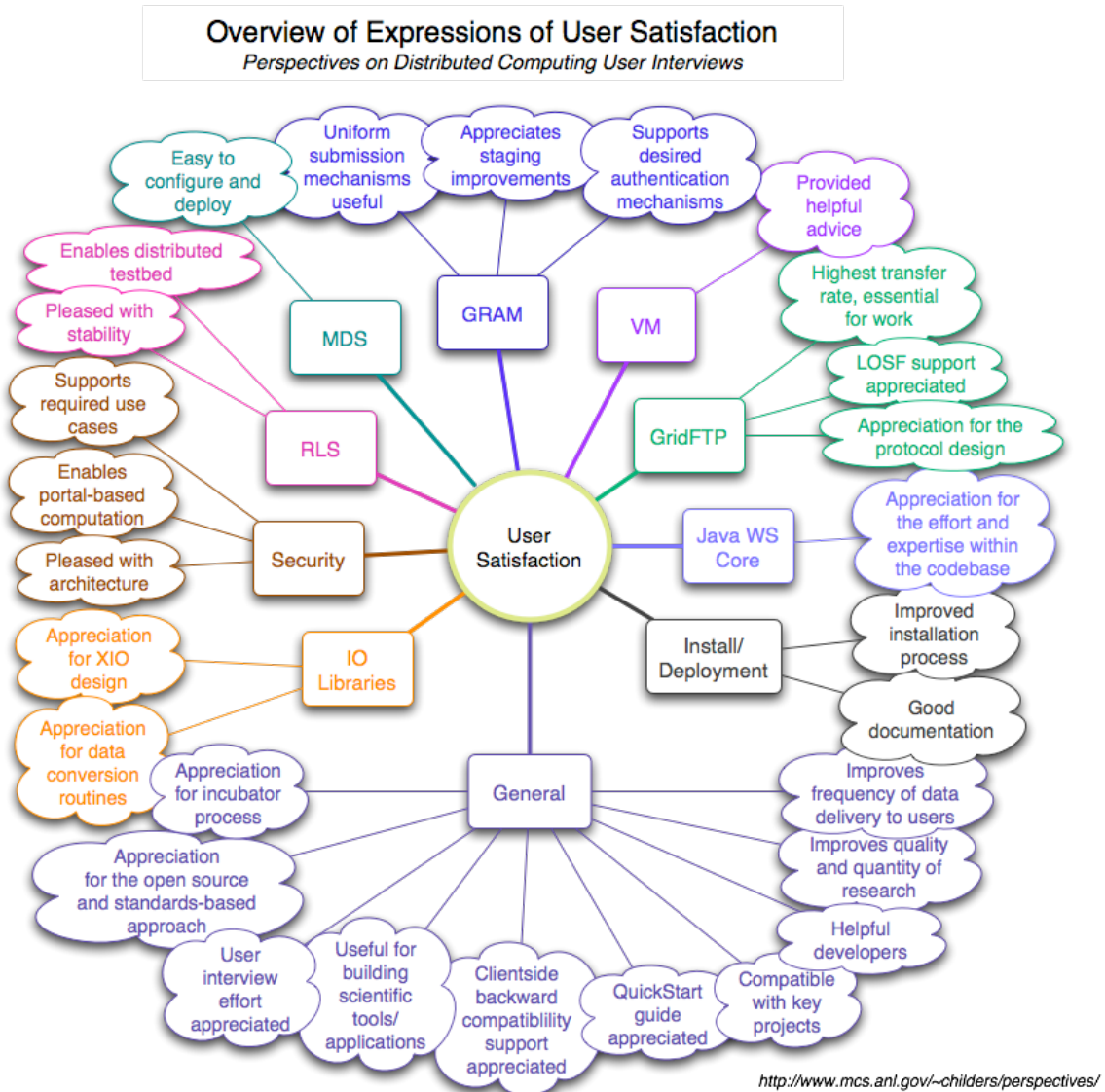
In Figure 5, we note that the issue types *Technology Adoption*, *Reliability*, *Diagnostics*, and *Communication* relate to long-term use of the software. The fact that interviewees mentioned these types of issues indicate they are trying to make long-term commitments to using distributed computing software.

In the *Other Social Issues* category, interviewees identified problems with the way their colleagues and service providers are making use of these technologies and the impact on their work. This can also be seen in the *Communication issues*→*Intra-project demands* category and the *Technology Adoption*→*Cultural barriers* category. These kinds of issues show the early phases of disciplinary transformations, when early adopters must “convert” their colleagues to a new mode of work.

We see another theme in the interview data, namely, that a number of the interviewees have encountered issues caused by the nature of distributed systems. In such systems, where many components are used together as a whole, a problem with one component can propagate throughout the system. Issues such as *Software/system integration*, *Infrastructure difficulties*, *System-wide failures*, and *Lack of confidence troubleshooting problems* suggest this experience.

### 3.3 Satisfaction Points

During the interviews we asked people why they use a Globus software component instead of other technologies. Though the purpose of the question was to better understand the motivation for component use, responses sometimes included expressions of satisfaction about the component. We also invited general comments at the end of the interview, and many users spontaneously discussed the value received from distributed computing tools. These satisfaction points are summarized in Figure 6. The source data for this figure is in Appendix C.3; the summarization method is described in Appendix A.



**Figure 6: Expressions of satisfaction**

What we find most interesting in this summary are the expressions of satisfaction about Globus and distributed computing in general (the *General* category). These enumerate some of the reasons interviewees have decided that distributed computing and Globus in particular are worth using in their work. They represent important benefits we must take care to preserve.

## 4 Characterizing Users by Their Interactions with Technology

Based on the assumption that users with similar technology interactions may have similar needs, we broke the thirty interviews into smaller groups to examine the user experience more closely. This section begins with a description of the method used to group users, and concludes with four composite profiles that describe each of the four groups found in the data. In Section 5 we recommend several ways to use the results. We note that there are many other ways to analyze the interview data, and we welcome other efforts to do so.

### 4.1 Characterization Method

#### *Interactions with Technology*

We created six technology interaction categories by crossing three broad interaction types (“develop,” “integrate,” and “use”) with two broad technology categories (“domain-specific technology” and “general-purpose high-performance computing technology”). The following definitions are used:

- **Domain-specific technology** includes domain-specific portals, applications, tools, libraries, and deployed systems.
- **General-purpose HPC technology** includes general-purpose distributed computing clients, services and tools, and general-purpose HPC deployments.
- People **develop** technology if they are involved in the creation of new tools, services and/or machine deployments.
- People **integrate** technology if they integrate existing technology into a larger system.
- People **use** technology if they make use of existing tools, services and/or machine deployments.

Table 3 shows the breakdown of technology interactions as described in the user interviews. Note that the sequence and timing of interactions are not represented.

**Table 3: Technology interactions reported in the interviews**

	Domain-Specific Technology			General-Purpose HPC Technology		
	Develop	Integrate	Use	Develop	Integrate	Use
User1	X	X			X	X
User2	X	X			X	X
User3	X	X			X	X
User4	X		X			X
User5	X	X			X	X
User6	X		X			X
User7		X		X	X	X
User8	X		X			X
User9	X	X			X	X
User10		X		X	X	X
User11		X		X	X	X
User12	X	X			X	X
User13		X		X	X	X
User14				X	X	X
User15		X		X	X	X

	Domain-Specific Technology			General-Purpose HPC Technology		
	Develop	Integrate	Use	Develop	Integrate	Use
User16		X		X	X	X
User17		X		X	X	X
User18		X		X	X	X
User19	X	X			X	X
User20	X	X			X	X
User21	X	X			X	X
User22				X	X	X
User23	X	X			X	X
User24	X	X			X	X
User25	X	X			X	X
User26				X	X	X
User27	X	X			X	X
User28				X	X	X
User29	X	X			X	X
User30	X	X			X	X

Table 3 shows that all users interviewed interact with technologies in at least three ways. Nearly all users report the need to integrate technology in order to accomplish their goals, and most report the need to integrate both domain-specific and general-purpose technologies.

### **Technology Interaction Clusters**

From the data we identify four technology interaction patterns, suggesting four distinct user types, as shown in Table 4.

**Table 4: Four user types and their interactions with technology**

	Domain-Specific Technology			General-Purpose HPC Technology		
	Develop	Integrate	Use	Develop	Integrate	Use
HPC Scientist	X		X			X
HPC Domain-Specific Developer	X	X			X	X
General-Purpose HPC Infrastructure Provider		X		X	X	X
General-Purpose HPC Technology Developer				X	X	X

Detailed profiles of each of the four user types are presented in Sections 4.2-4.5. The profiles provide an evidence-based view of key aspects of the distributed computing user's experience. Arguably, additional types not captured by the interviews exist in the community. We hope this report lays a useful foundation for further study.

## 4.2 ▼ Type 1: The HPC Scientist

Three of the thirty interviewees work on projects in which they develop domain-specific technology and use both domain-specific and general-purpose technology. Table 5 shows an overview of the technology interactions reported in the interviews. This group represents one of four technology interaction clusters, or user types, found in the interview data. We refer to this group as “HPC scientists” for the purposes of this report<sup>6</sup>.

In this section we present a composite profile of the HPC scientist user type. The profile is a distillation of key ideas from three interviews, with an emphasis on work-related goals and challenges. To broaden its relevancy beyond the three users, we highlight the abstractions behind the specifics and use details from the three interviews for illustrative purposes. The interview data underlying the profile can be found in Appendix D.4, D.6, and D.8, respectively.

**Table 5: HPC scientist technology interactions**

	HPC Scientist					
	Domain-Specific Technology			General-Purpose HPC Technology		
	Develop	Int	Use	Dev	Int	Use
<b>User4</b>	computational astrophysics codes		visualization tools, computational astrophysics codes			ssh, HDF5, bbFTP, GridFTP, GSI, MPI, 400TB storage, national HPC centers, csh, bourne shell, Fortran90, C, C++
<b>User6</b>	lattice gauge computations		lattice gauge computations, log files, lattice files, graphing and analysis tools			ssh, scp, SRM-copy, globus-url-copy, GSI, MPI, TotalView, tape archives, national HPC centers, perl, shell scripts, C, C++
<b>User8</b>	QCD configurations, MILC code		MILC code, graphing tools			ssh, scp, uberFTP, MyProxy, MPI, tape archives, national HPC centers, C, assembler, csh, bash, perl

### **HPC Scientist Overview**

The HPC scientists we interviewed run large-scale simulations at national computing centers on behalf of domain-specific communities. The scientists both develop and use domain-specific code. The simulation codes are the result of many person-years of development and are conservatively tended, with implementation changes introduced incrementally. Like all of the people we interviewed, the HPC scientists report using general-purpose technology in support of their work. Much of the general-purpose technology the scientists use is deployed and maintained by HPC facility staff at the national computing centers. The scientists run their simulations over several months on some of the largest machines in the world. These simulations can produce prodigious amounts of data consisting of several large files or numerous small files. The scientists move simulation results between facilities for further analysis and archival purposes. While their interest in distributed technology is grounded in their need to move data, the overriding focus of

<sup>6</sup> The authors chose the name “HPC scientists” as an arbitrary label to describe this group of three users. No relationship with people beyond this report who might call themselves HPC scientists is implied.



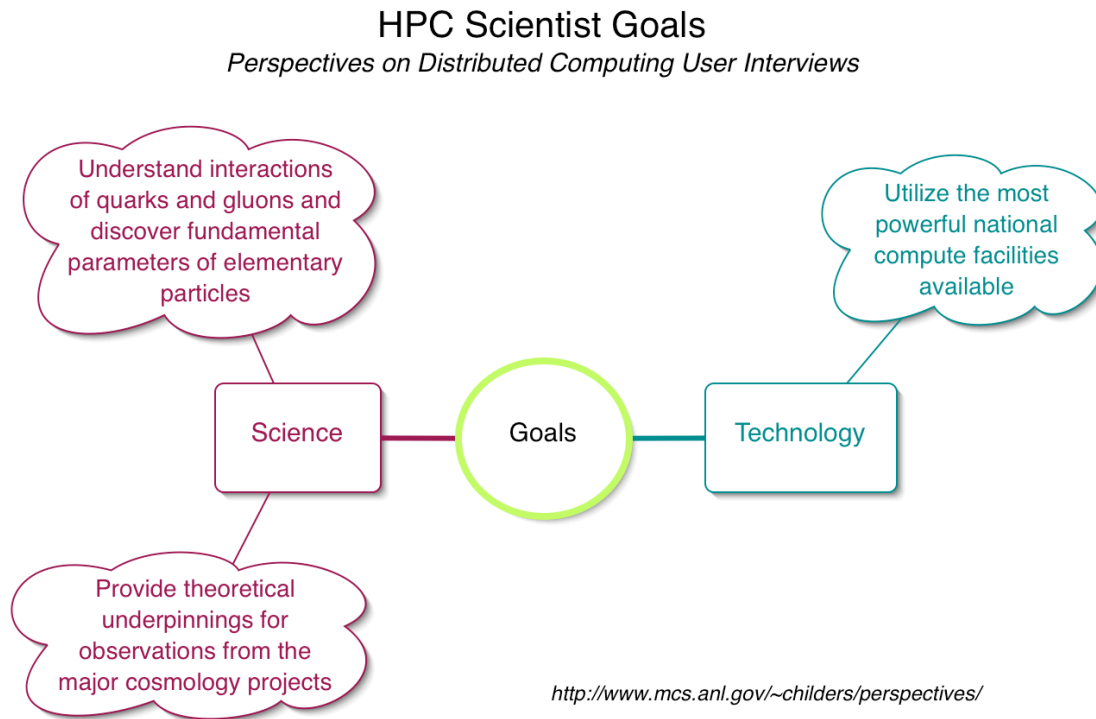
these users is to conduct pioneering calculations; technology is of interest to the extent that it helps further domain-specific goals.

### HPC Scientist Goals

Comparing the goals of the three HPC scientists with the integrated goals depicted in Figure 4, we find that the HPC scientist goals fall within two categories: *Conduct and promote scientific research*→*Extend scientific understanding* and *Expand the resources available to a specific community*→*Run existing scientific applications/codes at higher resolutions*. Comparing the scientists' goals with the integrated goals shown in Table 2, we see that they fall into the *Social Operation* and *Technical Development* quadrants.

Figure 7 depicts a summary that includes only those goals reported by the three HPC scientists we interviewed. Scientific achievements and resource usage come to the fore:

- Understand interactions of quarks and gluons and discover fundamental parameters of elementary particles<sup>7</sup>
- Provide theoretical underpinnings for observations from the major cosmology projects<sup>8</sup>
- Utilize the most powerful national compute facilities available<sup>9</sup>



**Figure 7: HPC scientist goals**

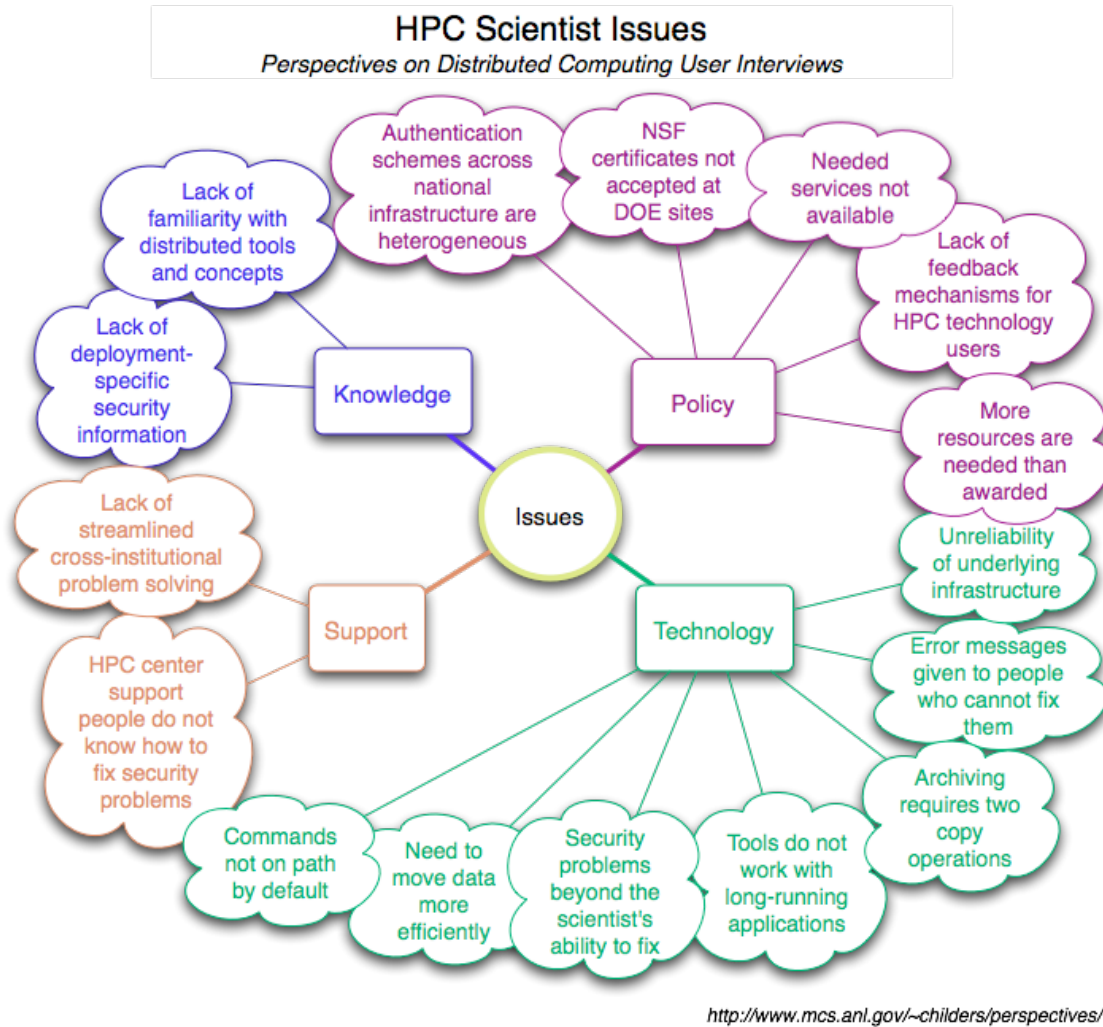
<sup>7</sup> Users 6, 8

<sup>8</sup> User 4

<sup>9</sup> User 4

### HPC Scientist Issues

Figure 8 shows a summary highlighting only those issues discussed by the three HPC scientists interviewed. Several issues directly relate to using, or trying to use, technology deployments listed in Table 5. Other issues, such as the lack of feedback mechanisms, are independent of any particular technology. More details, with references to the users corresponding to each cloud, follow the figure. Specific context for the issues can be found in the interviews in Appendix D.



**Figure 8: HPC scientist issues**

Because their domain-specific applications are so compute-intensive, funding agencies control the HPC scientists' use of the national HPC centers. In coordination with colleagues in their community, the scientists we interviewed submit formal proposals to resource allocation programs such as the DOE Innovative and Novel Computational Impact on Theory and Experiment (INCITE) program and the NSF Large Resource Allocations Committee (LRAC). Since the scientists' applications run at the leading edge of the computing power provided by the national computing facilities, one issue that can be encountered by this user is that more resources

are needed than are awarded<sup>10</sup>. The HPC scientists we interviewed all use multiple resources in pursuit of their goals. Several issues in this user profile arise from their need to move data between these resources.

During simulation runs the unreliability of the underlying infrastructure can prove problematic for the user<sup>11</sup>. Nodes crash, and jobs fail. Even typically stable services like filesystems can fail under the heavy loads generated by this type of user. Moving files between facilities is another source of failures. One scientist reported that dealing with technology failures is the most time-consuming aspect of his work. Further concerns were expressed that the failures at petascale will render these systems even less usable.

A lack of feedback mechanisms providing access to the people who fund and maintain these systems is seen to undermine scientific productivity<sup>12</sup>. An interviewee asserted that machine architecture decisions are made without the input of HPC scientists, resulting in systems that are often not well suited for their applications.

Simulation runs controlled by HPC scientists can take several months to complete. One problem that can arise is that general-purpose HPC technologies sometimes do not work well with long-running applications<sup>13</sup>. A user noted that tools with graphical user interfaces are not appropriate for the way he works. Usability problems with security tools were also mentioned as problematic in this context. One scientist felt that his approach to problem solving is of a different type entirely from the “myriad little processes” that seem to be the focus of today’s computer scientists.

The HPC scientists routinely manage and move data across the national networks. The need to move data more efficiently<sup>14</sup> stirs the scientists’ interest in newer distributed computing tools such as GridFTP. The interviews suggest new tools would be adopted if they were readily available, easy to use, reliable and offered clear advantages over current practice. One interviewee suggested wrapping tools in a script that can run in the background; the hypothetical script would execute a long-running job in one place, transfer the output, and run a subsequent job in another place.

According to the interviews, the HPC scientists’ unfamiliarity with some distributed computing concepts and tools can present barriers to their use<sup>15</sup>. For example, the HPC scientists reported problems in their attempts to use GridFTP. It is not always clear where to find documentation; and once the documentation has been found, the scientist is confronted with acronyms and unfamiliar concepts. Moreover, the information in the documents is organized in small chunks and is not necessarily arranged in an order the HPC scientist understands.

In addition to conceptual difficulties, hypothetical questions about credentials used for authentication in TeraGrid and other national infrastructures indicate a need for additional deployment-specific security information<sup>16</sup>:

- Which type of certificate<sup>17</sup> should I get for site X?
- How is a certificate used to authenticate<sup>18</sup> at site Y?
- How do I get my distinguished name<sup>19</sup> registered and linked to each of my user accounts at the various sites I need to use?

---

<sup>10</sup> Users 4, 8

<sup>11</sup> Users 4, 6, 8

<sup>12</sup> User 4

<sup>13</sup> Users 4, 8

<sup>14</sup> Users 4, 6, 8

<sup>15</sup> Users 4, 6, 8

<sup>16</sup> Hypothetical questions drawn from discussions with user 8

<sup>17</sup> <http://www.globus.org/toolkit/docs/4.2/4.2.0/security/key/security-key-concepts.html#security-key-certificates>

<sup>18</sup> <http://www.globus.org/toolkit/docs/4.0/security/key-index.html#s-security-key-mutualauthentication>

- Should I use [the security tool] gx-map or gx-request?

When using general-purpose technologies the HPC scientists also encounter a variety of difficulties with the underlying infrastructure. One interviewee mentioned that the Globus commands sometimes aren't on his path by default<sup>20</sup>, requiring that extra time be spent tracking things down. Another reported that NSF certificates are not accepted at DOE sites<sup>21</sup>. One user said that the heterogeneity of authentication schemes across the DOE centers can be quite disruptive<sup>22</sup>. Another infrastructure difficulty arises when needed services are not available<sup>23</sup>. Capabilities like Reliable File Transfer might be quite helpful to some of these scientists, yet the service is not perceived to be widely deployed.

Some common cross-site technology interactions could be better streamlined. One user reported that archiving simulation output can require two manual copy operations<sup>24</sup>: from the simulation site to the archive site, followed by a second copy within the archive site from disk to tape. Though technologies exist that are designed to help with this problem, they are not always in working order.

A second type of streamlining problem relates to user support. If a problem occurs during a data transfer, the support staffs at the two endpoints sometimes do not communicate well with each other<sup>25</sup>. The HPC scientists we interviewed appear to have separate relationships with each center based on their resource allocations (as opposed to being members of a cross-site VO). Thus, the HPC scientist can find himself negotiating with multiple technical support systems and user support staffs to fix a problem, with the system administrators communicating through the scientist rather than with each other directly.

Troubleshooting general-purpose technology problems is not always easy for the HPC scientists. Messages issued by Globus services are difficult to interpret and do not help HPC scientists respond effectively to problems<sup>26</sup>. The scientists generally lack the time to learn new troubleshooting techniques for general-purpose technologies. Further complicating the problem, one HPC scientist asserts that a significant number of HPC center staff members don't know how to fix Grid security problems on their own<sup>27</sup>. This is a problem for the HPC scientists in particular because HPC center support people are primary providers of the scientists' technical support.

The scientists themselves lack the time and expertise to install, configure, and maintain unfamiliar software, and so the cost of using the newer tools is perceived as high. Security in particular gives them trouble: when things go wrong with security, it's beyond the scope of these scientists to fix on their own<sup>28</sup>.

## Recommendations

In Section 5 we present several recommendations for developers of distributed technology; at least two recommendations would directly benefit HPC scientists. For example, in Section 5.1 we recommend that component developers *broaden the focus of component-centric development*. This would result in usability improvements for the HPC scientist in the areas of documentation, security and fault handling. We also recommend *providing a data movement*

---

<sup>19</sup> [http://www.numi.fnal.gov/offline\\_software/srt\\_public\\_context/GridTools/docs/glossary.html#dn](http://www.numi.fnal.gov/offline_software/srt_public_context/GridTools/docs/glossary.html#dn)

<sup>20</sup> User 8

<sup>21</sup> User 4

<sup>22</sup> User 8

<sup>23</sup> User 4

<sup>24</sup> User 6

<sup>25</sup> User 8

<sup>26</sup> Users 4, 6

<sup>27</sup> User 4

<sup>28</sup> Users 4, 6

**product for HPC scientists who routinely move data.** This would enable scientists to move their data from resource to resource with increased efficiency and reliability, reducing the need for human intervention. See Section 5 for more details on these and other recommendations.

### 4.3 Type 2: The HPC Domain-Specific Developer

Fifteen of the thirty interviewees work on projects in which they develop and integrate domain-specific technology and integrate and use general-purpose technology. Table 6 shows an overview of the technology interactions reported in their interviews. This group represents the second of four user types found in the interview data. We refer to this group as “HPC domain-specific developers”<sup>29</sup>.

In this section we present a composite profile of the HPC domain-specific developer. The profile is a distillation of ideas from the fifteen interviews, with an emphasis on work-related goals and challenges. To broaden its relevancy beyond the fifteen users, we highlight abstractions with details from the interviews used for illustrative purposes. The interview data underlying the profile can be found in Appendix D.

**Table 6: HPC domain-specific developer technology interactions**

	HPC Domain-Specific Developer					
	Domain-Specific Technology			General-Purpose HPC Technology		
	Develop	Integrate	Use	Dev	Integrate	Use
<b>User1</b>	workflows	scientific code, AFNI			R, MATLAB, Octave, GRAM, GridFTP	Swift workflow tool, Eclipse, Subversion, CVS, UNIX tools, bash, awk, Python
<b>User2</b>	meteorology portal, workflows	meteorological data, tools and legacy codes			GRAM, GridFTP, GSI, MyProxy, RFT, RLS, MPI, MPICH-G2, TeraGrid	Java CoG kit, PURSE, GridSphere, GPEL, GFac, WebMDS, perl, python, jython, shell scripts, Java, Fortran
<b>User3</b>	analytical framework for neuroscientists	brain image data, Monte Carlo simulations, tools			R, 250 CPU cluster, 4TB storage, national HPC facility	Swift workflow framework, database, perl, shell script, awk, ced
<b>User5</b>	Patient-Centric Authorization Model, Grid Interface Service, Grid Book	medical picture databases, medical images, radiology workstations, DICOM library			GSI, MyProxy, GridFTP, RLS, GridShib, Shibboleth, MDS4	Java WS Core, Java CoG kit, GridShib for GT, Java

<sup>29</sup> “HPC domain-specific developer” is an arbitrary label used to describe this group of fifteen people in the pool of thirty interviews.

	HPC Domain-Specific Developer					
	Domain-Specific Technology			General-Purpose HPC Technology		
	Develop	Integrate	Use	Dev	Integrate	Use
<b>User9</b>	infrastructure for analyzing gravitational waves	metadata service, interferometer data, analysis pipelines			GSI, GSI-OpenSSH, MyProxy, RLS, GridFTP, Condor, tape archives, community HPC facilities	Python Globus, Simple CA, Java WS Core, C WS Core, python, Java, C
<b>User12</b>	access to electro-physiological and neuro-physiological data	magnetic resonance imaging and electro-encephalography data, visualization and analytical tools			many processors, large filesystems, data archives	Matlab, R, SPSS, Java, Java++, C++
<b>User19</b>	distributed rendering client	scientific data and computer graphics models, POV-Ray			GSIFTP, MyProxy, GRAM, Condor, national HPC facilities, institutional HPC facilities, local end-user resources	WebStart, UML, Java, GSI-OpenSSH, Autojar
<b>User20</b>	system for high-throughput analysis of genomes and metabolic reconstructions, algorithms, workflow plans, bioinformatics portal	biology information repositories, biological data, bioinformatics tools			Globus, Condor, national HPC centers, institutional HPC facilities	perl, VDL, FTP, GridFTP
<b>User21</b>	ecological data warehouse, Ecological Metadata Language-compatible data exploration tools	data collection sites linked by community network, ecological data			10s of terabytes datastore, quad core blade servers	LDAP server, Java, PHP, perl, HTTP
<b>User23</b>	genomics portal, site selector	protein sequences, bioinformatic tools			Condor, GSI certificates, RLS, GRAM, GridFTP, Oracle database, national HPC centers	VDS, perl

	HPC Domain-Specific Developer					
	Domain-Specific Technology			General-Purpose HPC Technology		
	Develop	Integrate	Use	Dev	Integrate	Use
<b>User24</b>	nanotechnology tool development environment, nanotechnology simulation tools	3D data rendering farm, nanotechnology simulations			Perl, Python, Fortran, C, C++, MATLAB, MPI, Condor, X11 (X Window System), institutional HPC facilities, national HPC centers	Condor-G, LAMP (Linux, Apache, MySQL, PHP)
<b>User25</b>	nanotechnology application framework	nanotechnology simulations			Condor, GSIFTP, Condor-G, globusrun-based OSG probe script, Condor Stork, GSI-OpenSSH, scp, gridProxy, vomsProx, PBS, MPI, C, Fortran, Fortran90, Perl, institutional HPC facility, national HPC centers	Tcl, Python, bash shell
<b>User27</b>	Mechanisms for providing and exchanging network data	network data capture feeds, data mining algorithms and tools, files of network data			R, Globus, institutional HPC facility, national HPC centers	Python, perl, C, C++, database, shared filesystem
<b>User29</b>	water distribution simulation optimization framework	sensor data, EPANET simulation code			MPI, GT4, national HPC centers	Python, C, Java
<b>User30</b>	hydraulics and water quality simulations, optimization component, custom scripts	hydraulics information, visualization tool			MPI, Java CoG Kit, GridFTP, schedulers, institutional HPC facility, national HPC centers	bash shell, Python, C, Java, MATLAB

### ***HPC Domain-Specific Developer Overview***

The HPC domain-specific developers we interviewed design and build systems composed of both domain-specific and general-purpose HPC technology. These developers are familiar with domain-specific and general-purpose HPC technology concepts though they may not be experts in both areas. High-level work includes understanding domain-specific requirements and translating them to a distributed computing context. Detailed work entails integrating existing and developing new technology to support the scientific inquiry. Most HPC domain-specific developers integrate components like GridFTP into their systems for use by others, as opposed to

using them directly. A variety of general-purpose technologies are used to facilitate integration. HPC domain-specific developers help bridge the worlds of high-performance computing and domain science. They can be technology trailblazers, potentially transforming the way science is conducted in their domain.

### ***HPC Domain-Specific Developer Goals***

Comparing the goals of this subset of users with the integrated group of thirty depicted in Figure 4, we see that HPC domain-specific developer goals fall into each of the four top-level categories.

In the *Conduct and promote scientific research* category, the domain-specific developers report goals in every subcategory:

- Extend scientific understanding
- Apply science to practical problems
- Eliminate barriers to scientific investigation
- Build a case for continued financial support

In the *Expand the community that can use a resource* category, they report three goals:

- Make scientific applications accessible to more potential users
- Make scientific data accessible to more potential users
- Make computation services accessible to more potential users

In the *Expand the resources available to a specific community* category, they again report three goals:

- Provide computing systems to scientific users
- Aggregate cross-institutional resources
- Run existing scientific applications/codes at higher resolutions

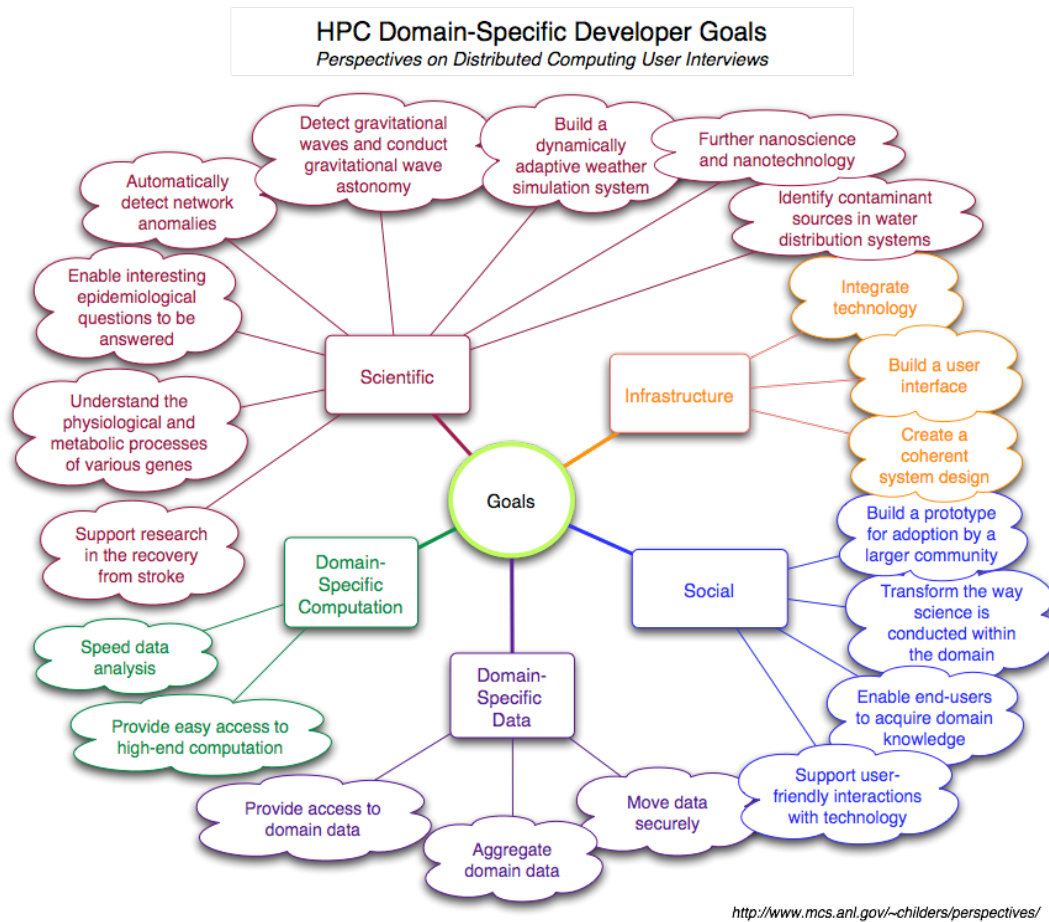
In the *Satisfy user requirements for systems used to do science* category, they report four goals:

- Establish provisioning/allocation mechanisms that efficiently satisfy varying demand
- Establish and employ security mechanisms that support dynamic, inter-organizational collaboration
- Establish and maintain system stability
- Provide compatibility with existing system components

Comparing this group with the view of integrated goals shown in Table 2, we see that the goals reported by the fifteen HPC domain-specific developers fall within all four quadrants: *Social Operation*, *Social Development*, *Technical Operation*, and *Technical Development*.

Figure 9 summarizes the HPC domain-specific developer goals. The diversity of goal types is clearly shown, with a variety of domain-specific goals reported in addition to infrastructure and social goals. Detailed information about the goals, with references to the users corresponding to each cloud, follows the figure.





**Figure 9: HPC domain-specific developer goals**

### ***Scientific Goals***

Nearly all of the HPC domain-specific developers referred to a specific domain goal during their interviews:

- Support research in the recovery from stroke<sup>30</sup>
- Understand the physiological and metabolic processes of various genes<sup>31</sup>
- Automatically detect network anomalies<sup>32</sup>
- Enable interesting epidemiological questions to be answered<sup>33</sup>
- Detect gravitational waves and conduct gravitational wave astronomy<sup>34</sup>
- Build a dynamically adaptive weather simulation system<sup>35</sup>
- Identify contaminant sources in water distribution systems to better apply remediation measures<sup>36</sup>
- Further nanoscience and nanotechnology<sup>37</sup>

<sup>30</sup> User 12

<sup>31</sup> Users 20, 23

<sup>32</sup> User 27

<sup>33</sup> User 5

<sup>34</sup> User 9

<sup>35</sup> User 2

<sup>36</sup> Users 29, 30

<sup>37</sup> User 24

Facilitating the publication of peer-reviewed papers was reported as a success measure of more than one HPC domain-specific developer. This reinforces the idea that the HPC domain-specific developer plays a crucial role in the integration of general-purpose technologies (e.g., a file transfer service or community authorization service) into the domain context.

### ***Domain-Specific Computation Goals***

The computational goals of the HPC domain-specific developers interviewed fall into two groups. One group pursues high-end computation goals in order to speed data analysis<sup>38</sup>. Specific examples include speeding the processing of brain data, minimizing the time spent analyzing huge volumes of genomic data, and identifying contaminants in a water distribution system as quickly as possible. The second group is interested in providing easy access to high-end computation<sup>39</sup>. Examples include enabling users to run sophisticated meteorological models on high-end resources, making it easier for data miners to interactively run their algorithms, and putting nanotechnology simulations into the hands of people who need them but otherwise wouldn't have access. A given HPC domain-specific developer may pursue none, one, or both types of these computational goals.

When addressing a problem such as speeding the analysis of domain data, the HPC domain-specific developer must understand how concepts such as targeted job types<sup>40</sup>, mutual authentication<sup>41</sup>, or delegation<sup>42</sup> apply. Even after distributed computing concepts are understood, additional time and expertise are needed to work out the specifics of the conceptual approach (e.g., determine the implementation details needed to support a domain-specific virtual organization). The HPC domain-specific developer plays a key role in two activities:

- *Translating domain goals to technological concepts*: developing an overall conceptual approach by merging domain-specific goals with distributed computing concepts
- *Translating technology concepts to practice*: designing and building a concrete technical solution based on an overall conceptual approach

### ***Domain-Specific Data Goals***

The HPC domain-specific developers describe three types of data-related goals in the interviews. One is providing access to domain data<sup>43</sup>. Examples include providing access to ecological data that is distributed and heterogeneous, and providing access to results of computationally intensive analyses. A second type of data-related goal is aggregating domain data<sup>44</sup>. One interviewee works to integrate distributed datasets to enable the creation of new synthetic products; another works to aggregate patient information so it can be shared among multiple healthcare providers; yet another user integrates data from remote sensors into an optimization and simulation framework. The third type of data-related goal reported during the interviews is moving data securely<sup>45</sup>.

The interviews suggest that the HPC domain-specific developer is generally concerned with the management, organization and movement of data. This is somewhat different than the HPC scientists, who manipulate their data as part of the scientific inquiry in addition to moving and organizing it.

---

<sup>38</sup> Users 1, 20, 21, 23, 30

<sup>39</sup> Users 2, 9, 19, 23, 24, 25, 27

<sup>40</sup> <http://www.globus.org/toolkit/docs/4.0/execution/key/>

<sup>41</sup> <http://www.globus.org/toolkit/docs/4.0/security/key-index.html>

<sup>42</sup> <http://www.globus.org/toolkit/docs/4.0/security/key-index.html#s-security-key-delegation>

<sup>43</sup> Users 3, 12, 21

<sup>44</sup> Users 5, 21, 29

<sup>45</sup> User 5

### **Infrastructure Goals**

In addition to domain-specific goals, HPC domain-specific developers reported three types of infrastructure goals. The first is to integrate technology. In contrast to the HPC scientists' direct use of components like GridFTP, many HPC domain-specific developers we interviewed integrate third-party components into their higher-level domain-specific products. Table 6 shows the wide variety of technologies integrated by HPC domain-specific developers. Specific integration goals mentioned in the interviews include the following:

- Integrating legacy domain-specific codes and applications<sup>46</sup>
- Integrating compute systems with science models and instruments<sup>47</sup>
- Leveraging existing general-purpose tools<sup>48</sup>
- Accommodating domain device incompatibilities<sup>49</sup>
- Enabling the local PC to participate as a resource in the workflow<sup>50</sup>

The second type of infrastructure goal reported in the interviews is to build a user interface<sup>51</sup>. Some HPC domain-specific developers build interfaces to shield their users from complex or unfamiliar technologies. Examples include “dashboards” for collaborators, domain-specific portals and other Web-based applications.

The third type of infrastructure goal observed in the interview data is to create a coherent system design<sup>52</sup>. Examples include the desire to implement a service-oriented architecture, build WSRF-compliant services, develop a systemwide security model and provide for future extensibility of the system. Users also pursue specific design goals relating to scalability and reliability.

### **Social Goals**

The HPC domain-specific developers we interviewed mentioned four social goals. One goal discussed by several developers was a desire to support user-friendly interactions with technology<sup>53</sup>. Examples include reproducing control mechanisms that the scientist currently uses, and implementing interactions in ways that the user can understand.

A second social goal mentioned by many HPC domain-specific developers is to enable end-users of the system to acquire domain knowledge<sup>54</sup>. As part of this, the domain-specific technology might support only one step in a larger scientific workflow, such as to host simulations prior to manufacture in order to help identify design flaws. On the other hand, the domain-specific technology being built might support the full end-to-end scientific inquiry.

A third social goal discussed in the interviews is to transform the way science is conducted within the domain<sup>55</sup>. An example is the HPC domain-specific technology developer who works to bring the concept of “simulate first, build later” to the field of nanotechnology.

The fourth social goal mentioned by the HPC domain-specific developers is to build a prototype system in the hope that a larger community will adopt it<sup>56</sup>. Such is the case for the water distribution project, which is attempting to promote adoption at the municipal and federal levels.

---

<sup>46</sup> Users 1, 2

<sup>47</sup> Users 2, 9, 30

<sup>48</sup> Users 1, 2, 3, 5

<sup>49</sup> User 5

<sup>50</sup> User 19

<sup>51</sup> Users 2, 5, 9, 19, 21, 23, 24, 25

<sup>52</sup> Users 2, 5, 9, 19, 21, 30

<sup>53</sup> Users 2, 3, 5, 9, 19, 24, 25

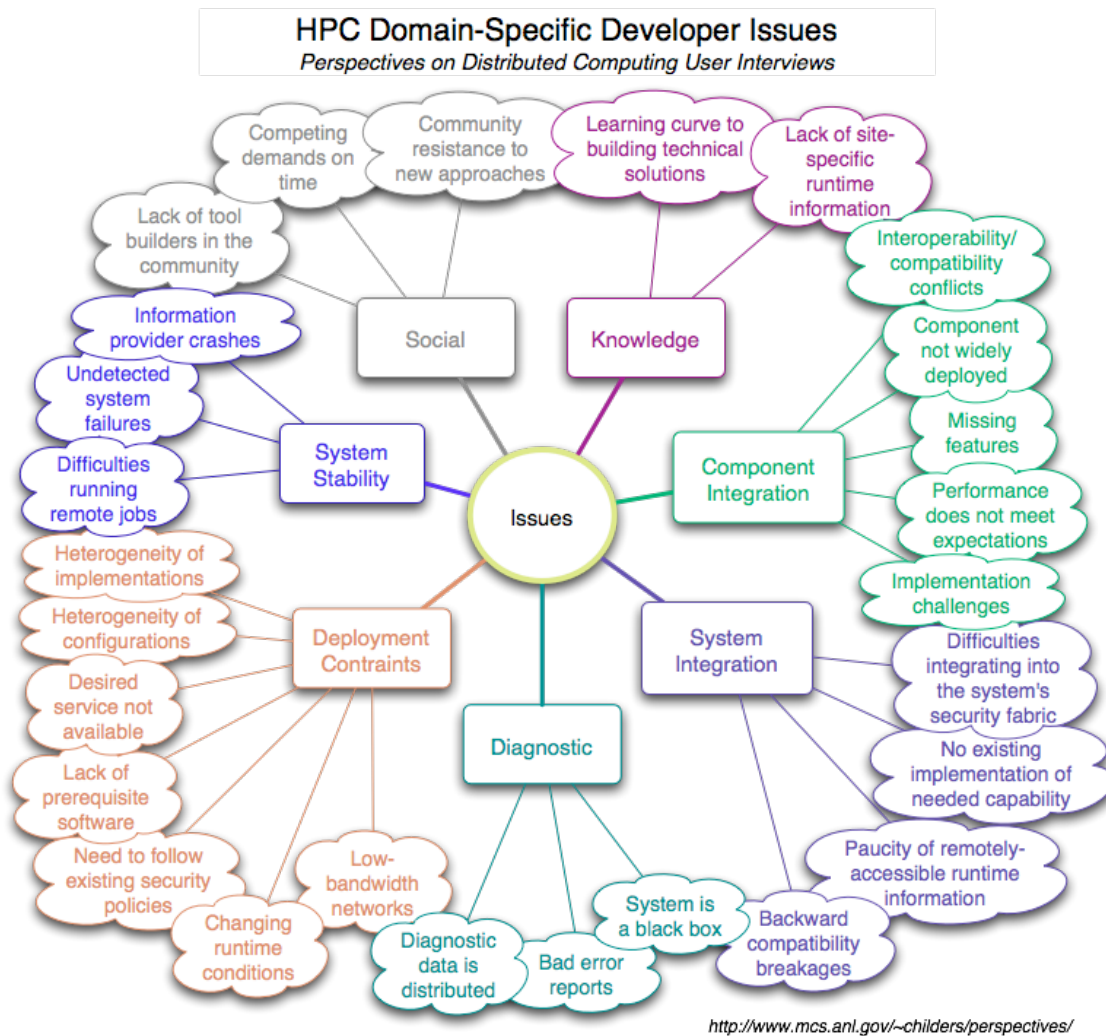
<sup>54</sup> Users 1, 3, 5, 9, 12, 20, 21, 24, 25, 27

<sup>55</sup> Users 5, 21, 24, 29

<sup>56</sup> Users 5, 29

### HPC Domain-Specific Developer Issues

Figure 10 summarizes those issues discussed in the interviews of the fifteen HPC domain-specific developers. Most of the issues reported arise from integrating (as opposed to directly using) the general-purpose technologies listed in Table 6. Problems are often described from the client-side view, as many of the developers in this group build systems that serve as clients for remotely-maintained services. More information about domain-specific developer issues, with references to the users corresponding to each cloud, follows the figure. Additional context can be found in their respective interview writeups in Appendix D.



**Figure 10: HPC domain-specific developer issues**

### ***Need to Overcome Social Barriers***

Some HPC domain-specific developers bring novel techniques to their domain by building systems that leverage new general-purpose HPC technology. Such users can encounter resistance in their communities when introducing new approaches<sup>57</sup>, triggering a need to educate and communicate benefits. We note that the HPC domain-specific developers serve not only as users of general-purpose HPC technology but also as advocates for it. Those assuming support responsibilities as a side effect of their advocacy may in turn rely heavily on general-purpose HPC technology developers for support.

HPC domain-specific developers can struggle to meet the many competing demands on their time<sup>58</sup>. Specific stresses mentioned in the interviews include juggling multiple projects, needing to build an end-user community in addition to developing the technology, and maintaining effective communications between distributed partners.

Interviewees also discussed a need to encourage development of tools and supporting technologies within the community<sup>59</sup>.

### ***Lack of Knowledge***

Many HPC domain-specific developers face a learning curve in how to transform science goals into technical solutions<sup>60</sup>. This issue especially affects those trained in disciplines other than computer science, as domain specialists can lack required technical training. Service-level and multiservice training materials are scarce and domain-specific case studies are not documented in detail. This situation, combined with a lack of general technical support mechanisms for science users, means that many HPC domain-specific developers face this learning curve on their own.

A second type of knowledge-related problem discussed is the HPC domain-specific technology developer who reports a lack of site-specific runtime information<sup>61</sup>. One user noted it can be difficult to discover information about resources outside his control, such as determining which environment settings are needed to interact successfully with the remote site.

### ***Component Integration Issues***

Several HPC domain-specific developers discussed challenges associated with third-party component integration. The far right column of Table 6 includes technologies that HPC domain-specific developers use to facilitate their integration work:

- Clientside APIs, used to interact with pre-existing remote services<sup>62</sup>
- Service development kits, used to build and host services<sup>63</sup>
- Workflow tools, used to encapsulate implementation details of the scientific workflow<sup>64</sup>
- Scripting languages, used to implement simple workflows<sup>65</sup>

These technologies are key to the successful component integration experience and should be taken into account when considering the technology requirements of these users.

---

<sup>57</sup> Users 5, 19

<sup>58</sup> Users 1, 9, 23, 27, 29

<sup>59</sup> Users 19, 24

<sup>60</sup> Users 2, 3, 9, 12, 19, 21, 24

<sup>61</sup> User 1

<sup>62</sup> User 2

<sup>63</sup> Users 5, 9, 12, 19, 24, 27, 29, 30

<sup>64</sup> Users 1, 2, 3, 20, 23

<sup>65</sup> Users 1, 2, 3, 20, 21, 23, 25, 29, 30

The summary of component integration problems reported in the interviews of the HPC domain-specific developers provides component developers with a list of key user concerns:

- The desired component conflicts with existing elements of the system<sup>66</sup> (not interoperable with another component, incompatible with a third-party library, machine architecture conflicts, etc.).
- The component is not widely deployed on remote resources<sup>67</sup> (so the domain-specific developer cannot depend on being able to use it at runtime).
- Desired component features are not implemented<sup>68</sup> (insufficient logging controls, lack of security hooks, etc.).
- The component does not perform as expected<sup>69</sup> (e.g., updates to the data in an information service happen less frequently than desired).
- Implementation issues hinder integration<sup>70</sup>
  - Extensive application-level changes are required to integrate.
  - Component APIs offered in an language unfamiliar to the user.
  - The user interface for the component is the wrong type (i.e., programmatic interface available, but users don't code).
  - Advanced configurations of the component are not well understood.
  - Component error messages are misleading or vague, making it difficult to implement automatic responses.

### ***System Integration Issues***

The HPC domain-specific developers also discussed several system-level integration issues. Such problems extend beyond the boundary of any one component. Interviewees reported four main types of such problems.

The first type involves difficulties integrating domain-specific code into the Grid security fabric<sup>71</sup>. First some context on this topic: secure component-to-component communication requires sharing interoperable and compatible security settings. While many general-purpose components include the ability to configure security without changing any code, domain-specific tools and legacy applications may not. Depending on the system design, building a secure distributed domain-specific application may require the domain-specific developer to learn a new programming model and modify working code. This requirement can seem a high barrier to some HPC domain-specific developers. Managing large, secure deployments can also pose problems; one interviewee reported that generating and distributing certificates for multiple end-users across many resources is a significant challenge.

The second type of system integration issue mentioned in the interviews involves missing technology. Many domain-specific developers use third-party code in their products, which is good news for general-purpose technology developers. Sometimes, however, no existing implementation of a needed capability can be found<sup>72</sup>, increasing the implementation burden of the HPC domain-specific developer. Specific missing capabilities mentioned in the interviews include automatically receiving notifications when resources becomes available, and being able to dynamically add new resources to a running job without restarting.

---

<sup>66</sup> Users 5, 9, 19

<sup>67</sup> User 1, 23

<sup>68</sup> User 9, 29

<sup>69</sup> Users 19, 23, 25, 30

<sup>70</sup> Users 2, 9, 19, 24, 25, 27

<sup>71</sup> Users 9, 19, 27

<sup>72</sup> Users 9, 19, 20, 23, 29

The third type of system integration issue that HPC domain-specific developers may face is a paucity of remotely accessible runtime information about the infrastructure they are integrating<sup>73</sup>. More information is needed to enable appropriate choices to be made about remote interactions at runtime. Example information mentioned in the interviews includes the version of a service, a description of the service's role in the remote system, its runtime dependencies, planned downtimes, current configuration settings and current load.

The fourth type of integration issue is disruption caused by feature changes or backward compatibility breakages in updates of third-party technologies<sup>74</sup>. We note that while some HPC domain-specific developers may be able to anticipate the impact of third-party technology workplans on their systems and respond accordingly, others may lack the time or technical grounding to fully track and understand the plans. Effectively engaging HPC domain-specific developers in the product planning process may require the general-purpose technology developer to reach out and explain workplans in terms of their impact on existing domain-specific systems.

### ***System Stability***

Addressing system stability issues is a key focus of some HPC domain-specific developers. One user expressed concern about the potential for information service failures to undermine the system's stability. When information providers go down, work stops<sup>75</sup>. Another information service-related problem is the delivery of false data<sup>76</sup>, such as when a remote job fails but goes unreported.

Out of a desire for stability, the HPC domain-specific developer may be interested only in using production-quality components<sup>77</sup>. There is a tension here, as it is not easy for general-purpose technology developers to provide builders of complex applications with software that is guaranteed to be stable. Domain-specific data and interaction patterns are difficult for general-purpose technology developers to emulate on their own. Moreover, deployment-specific problems can be a destabilizing factor outside of the component developer's control.

Some interviewees reported difficulties getting remote jobs to run reliably<sup>78</sup>: services crash, file systems fill up, certificates expire, applications fail, and so forth. One interviewee noted that end-users generally do not understand how to deal effectively with such failures. Some HPC domain-specific developers suggested making service reliability a top development priority. Another developer speculated that many failures might be eliminated if the service is hosted on an adequately sized machine. To facilitate the systems design process, one interviewee suggested that component developers publish load limits for their technologies.

### ***Diagnostics***

A key issue facing many HPC domain-specific developers is diagnosing problems at runtime. Lack of familiarity with the technical minutiae of deployed technologies and lack of access to diagnostic data across the entire system means much of it is a black box<sup>79</sup>. This frustrates attempts to fix problems efficiently.

---

<sup>73</sup> Users 2, 19, 23

<sup>74</sup> User 20, 23

<sup>75</sup> Users 9, 23

<sup>76</sup> User 23

<sup>77</sup> Users 2, 20

<sup>78</sup> Users 2, 23, 24, 25

<sup>79</sup> Users 1, 19, 25

The difficulties identifying problem sources are exacerbated by poorly documented error codes and misleading error messages<sup>80</sup>. One interviewee observed that effective troubleshooting can require an understanding of the implementation details of the component. This would seem an impractical requirement, given the time constraints and domain-specific focus of the HPC domain-specific technology developer.

Another problem users encounter is that diagnostic data is spread across multiple files<sup>81</sup> (syslog, messages.log, etc.) Hence, it can be difficult to build a complete picture of a problem. Even once a comprehensive view is assembled, log data generally does not suggest how to fix problems, but instead contains implementation-specific details.

The HPC domain-specific developers offered several ideas for improving the troubleshooting process. One suggestion is to provide mechanisms that capture diagnostic information at the time a problem occurs. A second suggestion is to provide remotely accessible diagnostic interfaces that enable multiple people (i.e., the system administrator, the HPC domain-specific developer, the end-user) to debug problems together. One user suggested that the community adopt a more centralized structure for administration of services; under such an arrangement experts could be called upon to quickly identify and respond to problems.

### ***Work within Deployment Constraints***

Some HPC domain-specific developers work on projects that provide them with complete control over the resources they use. Others work on projects that leverage shared resources over which they have limited control. Those who use shared resources must often work within the constraints of the deployed infrastructure. According to the interviews this situation can pose challenges for the HPC domain-specific developer.

One such challenge arises when a desired service is not available on the shared resource<sup>82</sup>. An example is the user who reports not having access to a GridFTP client and so uses *scp* to move data. In some cases the user may in fact have access to the technology, but not know how to invoke it or to become authorized to use it. In other cases the service may be deployed, but it is the wrong version. This can result in subtle compatibility problems, such as when a bug is fixed at one site but not at another.

A second deployment constraint involves the lack of prerequisite software for the application on the remote resource<sup>83</sup>. On local resources, a single copy of software prerequisites is often installed for all the end-users to share. When applications are run on remote resources, project-specific prerequisites must sometimes be deployed specially for each user. The burden of software prerequisite setup can fall to the HPC domain-specific developer, as opposed to the resource owner. One domain-specific developer had to develop new operational procedures for his application (invocation scripts, user documentation, etc.) to accommodate the relocation of the prerequisites. This class of work adds to a barrier that must be overcome to enable remote interactions.

HPC domain-specific developers also encounter problems because of the heterogeneity of implementations providing similar capabilities<sup>84</sup>. One interviewee expressed frustration about the diversity of resource allocation mechanisms in his pool of remote resources. In his case project

---

<sup>80</sup> Users 1, 2, 19

<sup>81</sup> Users 25

<sup>82</sup> Users 19, 20, 29, 30

<sup>83</sup> Users 1, 27

<sup>84</sup> User 25, 30



staff must test their job submission code individually at each remote site in order to ensure proper functioning of their application. The diversity of parallel libraries across the sites is also reportedly a problem area, as are the scripting interfaces on the various remote resources.

Users may also encounter problems managing the heterogeneity of remote configurations<sup>85</sup>. For example, if several GridFTP servers used by a project sit behind differently configured firewalls, the HPC domain-specific developer needs to do extra client-side work to accommodate the configuration differences. Interactions between issues can compound problems. For instance, relatively simple configuration conflicts may not be addressed efficiently if the error messages describing them are misleading or difficult to interpret.

Some HPC domain-specific developers encounter issues trying to move data across distributed infrastructure. For example, getting data in and out of large-scale facilities can be a challenge, especially when end-users are connected to lower-bandwidth networks<sup>86</sup>. Another interviewee was prevented from implementing his preferred approach for moving application data (streaming via direct connections with worker nodes) because of the need to work within HPC center security policies<sup>87</sup>. Changing runtime conditions<sup>88</sup> are also an issue for the domain-specific HPC developer. Interviewees discussed a need to assess and react to system load and other runtime characteristics in order to achieve desired performance.

## Recommendations

In Section 5 we present several recommendations for developers of distributed technology that would benefit HPC domain-specific developers. For example, we recommend that developers of general-purpose technologies both *broaden the focus of component-centric development* and *enlist the aid of HPC domain-specific developers to translate generic technology concepts into domain-specific concepts*. Adoption of these recommendations would result in the identification of key technology requirements of domain-specific developers, as well as improved documentation and testing for existing technologies on which the domain-specific developers depend. See Section 5 for detailed information on these and other recommendations.

## 4.4 ● Type 3: The General-Purpose HPC Infrastructure Provider

Eight of the thirty interviewees work on projects in which they integrate domain-specific technology and develop, integrate and use general-purpose technology. Table 7 shows an overview of the technology interactions reported in the interviews. This group represents the third of four technology interaction clusters, or user types, found in the interview data. We refer to this group as “general-purpose HPC infrastructure providers”<sup>89</sup>.

In this section we present a composite profile of the general-purpose HPC infrastructure provider. The profile is a distillation of key ideas from the eight interviews with an emphasis on work-related goals and challenges. To broaden its relevancy beyond the eight users, the profile

<sup>85</sup> User 2

<sup>86</sup> User 19

<sup>87</sup> User 30

<sup>88</sup> Users 2, 19, 30

<sup>89</sup> “General-purpose HPC infrastructure providers” is an arbitrary label denoting the group of eight users reporting this technology interaction pattern.

highlights the abstractions behind the specifics; details from the eight interviews are used for illustrative purposes. The interview data underlying the profile can be found in Appendix D.

**Table 7: General-purpose HPC infrastructure provider interactions**

	HPC Infrastructure Provider					
	Domain-Specific Technology			General-Purpose HPC Technology		
	Dev	Integrate	Use	Develop	Integrate	Use
<b>User7</b>		applications		portal, metascheduler, client tools package	MPI, TotalView, GSI, TAGPMA CA, GRAM4, GridFTP, globus-url- copy, uberFTP, GSI- OpenSSH, MyProxy, grid-proxy-init, Condor- G, MDS, GRMS, GridWay, community HPC centers	pacman, VDT, GridSphere, GridPort, perl, python, bash, Java C, C++
<b>User10</b>		scientific codes		test suites, national HPC facility services	Unclassified Nuclear Engineering data, scp, GridFTP, PVFS, MPICH, MPI-IO, HPSS, Cobalt Scheduler, 128,000 cores, 64 terabytes RAM, 5 petabytes disk, 100 petabyte tape archive, 768 10Gigabit Ethernet ports	Karajan, bcfg2, lperf, bash, python, shell scripts, C, Java
<b>User11</b>		scientific models		test probes, job submission mechanism, example wrapper scripts	job input/output files, test results, log files, GSI, GridFTP, RFT, MDS4, Condor-G, GRAM, Condor, classads, ReSS, VOMS, LRMs, PVFS, MPI, OSG	perl, python, shell, C, C++, Java
<b>User13</b>		applications		portal, monitoring layer, projectwide scheduler, security architecture, client stack	GridFTP, MDS, GRAM, GSI, MyProxy, GSI- OpenSSH, Java WS Core, Condor-G, GRMS, GridWay, GUMS, PRIMA, VOMRS, VOMS, TAGPMA CA, community HPC centers	workflow tools, VDT
<b>User15</b>		scientific applications, visualization data, remote instruments		collaboration framework, shared applications	virtual meeting spaces, user data, video data, audio data, community mailing lists, bug tracking system, schedulers, community servers, national HPC facilities	Python, PHP, C, C++, drupal, intaller toolkits, GDB, Visual Studio, DebugView
<b>User16</b>		high-resolution visualizations, atmospheric simulations		scalable display infrastructure, high-speed IO system	tiled displays, TeraGrid, community HPC facilities, ROCKS distribution	Python, C++

HPC Infrastructure Provider						
Domain-Specific Technology				General-Purpose HPC Technology		
Dev	Integrate	Use	Develop	Integrate	Use	
User17		job requests		jobmanager- condor emulator, test jobs	GridFTP, GRAM, MyProxy, GSI, VOMS, GUMS, Security Authorization Service, Grid LAN (GSI-based), MPI LAN (Kerberos- based), OSG software stack, institutional HPC facilities, national HPC centers	C
User18		applications		portal, benchmarks	HPC clusters, MATLAB, Abacus, C, C++, FORTRAN, MPICH for GigE, MVAPICH for InfiniBand, Open MPI	Tomcat, Globus Toolkit, Java, GridSphere, mySQL, shell, perl

### General-Purpose HPC Infrastructure Provider Overview

General-purpose HPC infrastructure providers develop and maintain infrastructure for use by the HPC scientific community. They integrate a variety of general-purpose HPC technologies (second column from the right in Table 7) and work with their users to translate domain-specific applications to a multi-use distributed computing context. The general-purpose HPC infrastructure providers we interviewed interact with domain-specific data and codes in a generic way, as files and processes, not in a domain-aware fashion. In terms of their development work, some interviewees build their systems from scratch, while others add new interfaces or tools to existing infrastructure. In either case the infrastructure providers are experts in general-purpose HPC technology, and they employ their expertise to help HPC scientists and HPC domain-specific developers manage their applications and data. As maintainers of some of the most powerful systems in the world, the general-purpose HPC infrastructure provider serves as a key enabler of high-end scientific computing.

### General-Purpose HPC Infrastructure Provider Goals

Comparing the goals of the eight general-purpose HPC infrastructure providers with those of the thirty users as a whole (depicted in Figure 4), we see they fall into all four top-level categories.

In the category *Conduct and promote scientific research*, the general-purpose HPC infrastructure providers work to

- Eliminate barriers to scientific investigation, and
- Build a case for continued financial support.

In the category *Expand the community that can use a resource*, the providers

- Make scientific applications accessible to more potential users,
- Make computation services accessible to more potential users, and
- Make scientific colleagues accessible to more potential users.

In the category *Expand the resources available to a specific community*, they

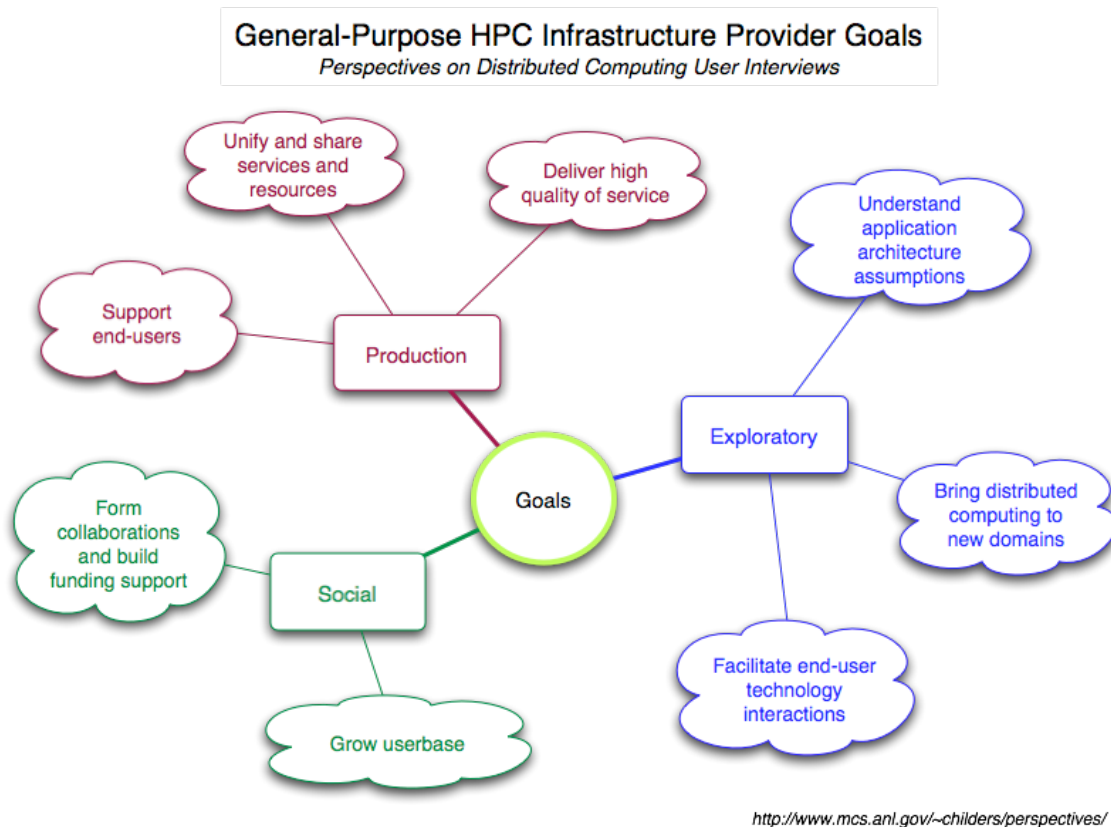
- Provide computing systems to scientific users, and
- Federate institutional computing resources.

In the category *Satisfy user requirements for systems used to do science*, they

- Establish and employ security mechanisms that support dynamic, inter-organizational collaboration, and
- Establish and maintain system stability.

Comparing the general-purpose HPC infrastructure provider goals with the integrated goals shown in Table 2, we see that they fall within all four quadrants: *Social Operation*, *Social Development*, *Technical Operation*, and *Technical Development*.

Figure 11 summarizes those goals reported by the eight general-purpose HPC infrastructure providers. Production, exploratory, and social goals come to the fore. Details about the goals, with references to the users who correspond to each cloud, follow the figure.



**Figure 11: General-purpose HPC infrastructure provider goals**

### ***Production-related Goals***

Unsurprisingly, the HPC infrastructure providers we talked to discussed a need to deliver a high quality of service<sup>90</sup> to their end-users. Interviewees mentioned the need to provide an operational system and to minimize failures. One user plans to implement redundancy for key services in his system. Another plans to replicate key application data for backup and high availability.

Another production-related goal of the HPC infrastructure provider is to unify and share services and resources<sup>91</sup>. Examples include unifying local resources of a particular type, such as distributed storage or multiple clusters by binding them through common interfaces, and sharing institutional resources with a larger community, such as the Open Science Grid. Other example unification goals are software-focused: providing a common authentication and authorization infrastructure or pooling use of licensed software applications. Additional examples are related to sharing work-related data, such as end-user video, audio, and scientific data.

The third production-related goal discussed by the general-purpose HPC infrastructure providers relates to end-user support<sup>92</sup>. An example is helping to integrate and debug domain-specific applications, in addition to providing the usual IT services. Other examples include facilitating the advancement of science in research and education and lowering the expertise requirements for use of the general-purpose HPC technologies.

### ***Exploration-related Goals***

One exploratory goal involves efforts to bring distributed computing to new application domains<sup>93</sup>. Meeting this goal can involve the identification and implementation of new infrastructure requirements.

Another exploratory goal is to understand how changing fundamental assumptions about the system affects application architectures and the user experience<sup>94</sup>. An example is the effort to understand what would happen if lightpaths could be scheduled between distributed computers in the same way that jobs today can be scheduled on a compute resource.

The third exploratory goal involves facilitating end-user interactions with technology<sup>95</sup>. One interviewee stated that he works to provide an environment in which distributed people can interact as if they are colocated, an area associated with multiple research topics. Other key end-user interactions involve remote instruments.

### ***Social Goals***

Some general-purpose HPC infrastructure providers we interviewed discussed their efforts to develop an initial set of end-users and recruit new end-users<sup>96</sup> for their system. One interviewee also mentioned the need to form collaborations in the community and build support for further funding<sup>97</sup>.

---

<sup>90</sup> Users 7, 10, 13

<sup>91</sup> Users 15, 17, 18

<sup>92</sup> Users 13, 17

<sup>93</sup> Users 7, 13

<sup>94</sup> User 16

<sup>95</sup> User 15

<sup>96</sup> Users 7, 11, 13

<sup>97</sup> User 7

### General-Purpose HPC Infrastructure Provider Issues

Figure 12 summarizes the issues discussed by the eight general-purpose infrastructure providers that make up this group. The interviewees reported problems integrating the technologies listed in Table 7 into their systems, with discussions generally addressing both the client-side and server-side perspectives. Other issues reflect the group's need to support the users of their infrastructure, as well as to interoperate with other systems. More detailed information, with references to the interviewees corresponding to each cloud, follows the figure. Specific context can be found in their respective interview writeups in Appendix D.



**Figure 12: General-purpose HPC infrastructure provider issues**

### ***End-User Support Issues***

General-purpose HPC infrastructure providers spend significant time on technical support of their users<sup>98</sup>. An infrastructure provider we interviewed observed that, taken as a whole, Grid software does not provide an easy-to-use operating environment for end-users. He asserted that getting users running on the Grid should be as easy as getting them running on a cluster.

In addition to technical issues, infrastructure providers also find themselves having to address cultural issues<sup>99</sup>. Whether helping the end-users overcome a fear of losing control of their data, getting them to trust unfamiliar security mechanisms, or persuading them to embrace a new way of working, the HPC infrastructure provider can find himself quite busy supporting users on a variety of nontechnical topics.

### ***Maintenance Issues***

The general-purpose HPC infrastructure providers also discussed several issues related to systems maintenance.

In order for the infrastructure provider to work effectively, more documentation is needed about the general-purpose HPC technologies they are asked to deploy<sup>100</sup>. According to the interviews, documentation is needed on the following topics: engineering information, details on the component design, and information on basic distributed computing concepts (What is a client? What is a server?). A concern was also raised about the difficulty finding documentation for older versions of components that are still in use, such as GRAM2.

More specific engineering information would be useful, according to one interviewee: How big should the machine be to host the component? How fast do the drives need to be? What should the network connectivity look like? What security infrastructure (or other software) will be needed to support the service?

Needed design and implementation information includes details that reveal the component developer's design assumptions. Examples include where and how job state is maintained in GRAM and a description of the concurrency locking strategies for managing job state.

In addition to needing more documentation, a major issue involves keeping deployments in a coherent state<sup>101</sup>. General-purpose HPC infrastructure providers, especially those with deployments consisting of thousands of nodes, face problems keeping software versions up-to-date on all the machines, maintaining cross-component compatibility, and ensuring the integrity of software configurations. One interviewee noted that these types of problems fall outside the responsibility of component developers because they are the result of choices made by the infrastructure provider.

Two additional maintenance-related issues were discussed in the interviews. HPC general-purpose infrastructure providers may find that deployment packaging is too coarse-grained<sup>102</sup>,

---

<sup>98</sup> Users 7, 13

<sup>99</sup> User 13

<sup>100</sup> Users 10, 17, 18

<sup>101</sup> Users 10, 13, 15, 17

<sup>102</sup> User 15

thus increasing to an undesired degree the size of their own distributions. The final issue reported in the interviews is a technology's lack of platform support<sup>103</sup>.

### ***Service-Level Issues***

Service-level issues for general-purpose HPC infrastructure providers can take various forms, such as when a service does not work as anticipated<sup>104</sup>. In this category interviewees reported features as being buggy, not fully implemented, not scaling to the degree needed, or having other poor performance characteristics.

Another category of service-level problems found in the interview data involves a service's lack of desired features or public interfaces<sup>105</sup>. Specific features mentioned in the interviews include guaranteed delivery of notifications, dynamic IP address handling and the need for service redundancy. One user suggested that if desired features like redundancy are not provided, the component developer should provide hooks to facilitate custom development. An example of a public interface issue is the case where GUI clients are desired but are not available.

The final service-level issue mentioned in the interviews involves the assertion that errors in client-side configurations are a significant source of end-user errors<sup>106</sup>. We note the intractability of this problem in situations where the general-purpose HPC infrastructure provider does not control client deployments.

### ***Systemic Issues***

The general-purpose HPC infrastructure providers we interviewed describe a variety of issues involving interactions across multiple services, resources, and sites.

Configuring and deploying GSI<sup>107</sup> can be a challenge for the general-purpose HPC infrastructure provider. One interviewee noted that information on how to establish large-scale security deployments is lacking. Configuring security can be difficult, as can setting up a certificate authority<sup>108</sup>.

General-purpose HPC infrastructure providers may be required to support multiple identity management systems<sup>109</sup>. For example, an end-user might simultaneously belong to an OSG virtual organization, EDUCAUSE, and TeraGrid, in addition to his home institution.

General-purpose HPC infrastructure providers who wish to interoperate with other sites can encounter missing, incompatible, misconfigured, or poorly supported remote services<sup>110</sup>. These types of problems can be intractable, particularly when the provider must support multiple distributed projects simultaneously.

The general-purpose HPC infrastructure provider must also manage problems triggered by distributed use cases<sup>111</sup>. Examples of such issues include network congestion due to the need to move high volume data between sites, varying constraints on network bandwidth across the entire system, and local system failures resulting from remotely initiated high IO loads.

---

<sup>103</sup> User 15

<sup>104</sup> Users 7, 17

<sup>105</sup> Users 11, 13, 17

<sup>106</sup> User 17

<sup>107</sup> Users 7, 10, 13, 15

<sup>108</sup> [http://en.wikipedia.org/wiki/Certificate\\_authority](http://en.wikipedia.org/wiki/Certificate_authority)

<sup>109</sup> User 13

<sup>110</sup> Users 10, 11

<sup>111</sup> Users 10, 15, 17, 18



## Diagnostics

Detecting site failures<sup>112</sup> can be a challenge for the general-purpose HPC infrastructure provider. Once such failures are detected, identifying the causes of problems can be difficult, at least in part because the source of trouble often needs to be inferred<sup>113</sup>. One infrastructure provider commented, “solving problems is easy once you have all the data.” Yet much of the information needed to efficiently diagnose troubles is not exposed via remote interfaces.

Even with the benefit of privileged local access, pinpointing the source of job failures at a site can be difficult. Information about a single job can span multiple log files<sup>114</sup>, each of which must be tracked down. Temporary files that might be useful for diagnosing problems may be deleted<sup>115</sup>. Moreover, error messages and codes may not accurately reflect underlying problems<sup>116</sup>.

We note a strong overlap of concerns in the area of diagnostics when comparing this user with the HPC domain-specific developer. One key difference is that the HPC domain-specific developer may have a more limited view of server-side data than the general-purpose HPC infrastructure provider.

## Social Issues

A wide variety of social challenges were mentioned during the interviews. For example, general-purpose HPC infrastructure providers can spend significant time coordinating with colleagues via meetings and email when projects span institutional boundaries<sup>117</sup>. As noted by the HPC domain-specific developers we interviewed, such communication burdens can be heavy.

Another social challenge mentioned by the general-purpose HPC infrastructure providers is tracking independent software development efforts<sup>118</sup>. Those who do not track developments run the risk of reinventing code. However, those who reuse code rather than write their own must then manage the bugs they encounter in the third-party code<sup>119</sup>.

The general-purpose HPC infrastructure provider may need to fight a perception that computer scientists are merely technicians for domain scientists<sup>120</sup>. In such cases, attention and time are required to establish a relationship that advances the work of both parties.

General-purpose HPC infrastructure providers also discussed challenges associated with the funding process<sup>121</sup>, such as low award rates, the requisite long wait before learning whether the award has been granted, and the need to decompose research priorities into fundable subsets.

Another issue discussed in the interviews is that establishing new trust relationships across distributed systems can be time-consuming<sup>122</sup>. This type of work can require manual setup of portal accounts or allocation negotiations on multiple remote resources.

One general-purpose HPC infrastructure provider mentioned a need to educate local IT staff about distributed computing practices<sup>123</sup>. Topics include preventing servers from being shut down despite the fact that passwords are not changed every ninety days.

---

<sup>112</sup> User 11

<sup>113</sup> Users 10, 11

<sup>114</sup> User 17

<sup>115</sup> User 17

<sup>116</sup> User 11

<sup>117</sup> User 11

<sup>118</sup> Users 13, 16

<sup>119</sup> Users 13, 15

<sup>120</sup> Users 13, 16

<sup>121</sup> Users 15, 16

<sup>122</sup> Users 7, 13, 18

<sup>123</sup> User 13

## Recommendations

In Section 5 we present several recommendations for developers of distributed technology, at least two of which would benefit general-purpose HPC infrastructure providers. For example, in Section 5.1 we recommend that developers of general-purpose HPC technology *broaden the focus of component-centric development*. This would result in improved diagnostic support and documentation on advanced component configurations. We also recommend *working with general-purpose infrastructure providers to refine requirements for reliability and multiuse deployments*. This would ultimately result in key usability and feature enhancements that would benefit infrastructure providers, as well as performance improvements for their end-users. See Section 5 for the discussion of these and other recommendations.

## 4.5 Type 4: The General-Purpose HPC Technology Developer

Four of the thirty interviewees work on projects in which they develop, integrate, and use general-purpose technology. Table 8 shows an overview of the technology interactions reported in the interviews. This group represents the last of the four technology interaction clusters, or user types, found in the interview data. We refer to this group as “general-purpose HPC technology developers”<sup>124</sup>.

In this section we present a composite profile of general-purpose HPC technology developers based on the four interviews. The profile is a distillation of key ideas from the interviews with an emphasis on work-related goals and challenges. To broaden its relevancy beyond the four users, we highlight the abstractions behind the specifics; details from the four interviews are used for illustrative purposes. The interview data underlying the profile can be found in Appendix D.

**Table 8: General-purpose HPC technology developer interactions**

	General-Purpose HPC Technology Developer					
	Domain-Specific Technology			General-Purpose HPC Technology		
	Dev	Int	Use	Develop	Integrate	Use
<b>User14</b>				Grid message passing interface (MPICH-G2)	MPI, grid-proxy-init, globusrun, globusrun- ws, Rendezvous Service, GridFTP, globus IO, XIO, Globus data conversion library, Reliable Blast UDP, UDT	C

<sup>124</sup> The authors chose the name “general-purpose HPC technology developer” as an arbitrary label to denote the four users who have this technology interaction pattern.

	General-Purpose HPC Technology Developer					
	Domain-Specific Technology			General-Purpose HPC Technology		
	Dev	Int	Use	Develop	Integrate	Use
<b>User22</b>				Technology enabling attribute-based authorization (GridShib)	Shibboleth, Java WS Core, GridShib CA	Java, PHP, OpenSAML and Shibboleth libraries, Globus Java Authorization Framework, Ant, shell scripts, GPT
<b>User26</b>				best practices for troubleshooting, logging collection mechanism	syslog-ng, OSG	python, OSG integration test bed
<b>User28</b>				VO services	GRAM2, SRM and dCache, gLExec	Java, C, Maven, VDT nightly builds, XACML protocol, Tomcat

### **General-Purpose HPC Technology Developer Overview**

The general-purpose HPC technology developer creates software, specifications and guidelines for use by other developers, scientists, and infrastructure providers. General-purpose HPC technology developers identify fundamental abstractions and interaction patterns in distributed systems and build products that can be applied in multiuse deployments and across a variety of science domains. Whether directly or indirectly, general-purpose HPC technologies help insulate domain-specific developers from the complexities of high-end distributed machinery, and help infrastructure providers support and participate in multi-institutional distributed systems.

### **General-Purpose HPC Technology Developer Goals**

Comparing the goals of the four general-purpose HPC technology developers with those of the thirty users as a whole (depicted in Figure 4), we see that they fall into two of the four top-level categories.

In the category *Expand the resources available to a specific community*, the general-purpose HPC technology developers work to

→ Enable scientific applications that require coordinated use of multiple systems.

In the category *Satisfy user requirements for systems used to do science*, they work to

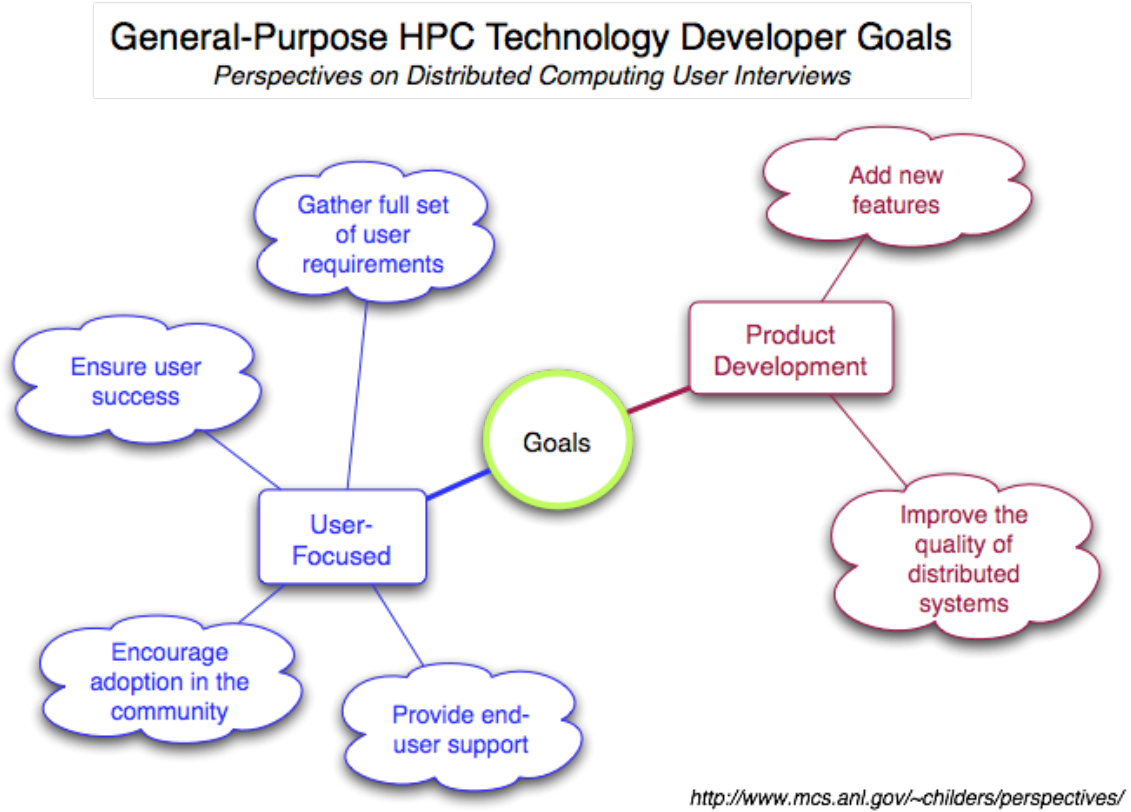
→ Establish and employ security mechanisms that support dynamic, interorganizational collaboration;

→ Establish uniform diagnostic mechanisms that satisfy debugging needs in dynamic systems; and

→ Acknowledge the requirements of all users.

Comparing the general-purpose HPC technology developer goals with the integrated view of goals shown in Table 2, we see they fall into two quadrants: *Technical Operation* and *Technical Development*.

Figure 13 summarizes those goals reported by the four general-purpose HPC technology developers. In this view, user-focused and product development goals come to the fore. Details of the goals, with references to the users corresponding to each cloud, follow the figure.



**Figure 13: General-purpose HPC technology developer goals**

### ***Product Development Goals***

One type of product development goal reported in the interviews is the desire to add new features<sup>125</sup>. Example features include enabling group and role-based access to resources, enabling interoperability among independently-developed technologies, and defining public interface guidelines for a community “best practices” document.

A second type of product development goal pursued by general-purpose HPC technology developers is to improve the quality of distributed systems<sup>126</sup>. Specific goals mentioned during the interviews include reducing the number of failed jobs, producing production-ready infrastructure, and making it easier to troubleshoot distributed applications.

<sup>125</sup> Users 22, 26, 28

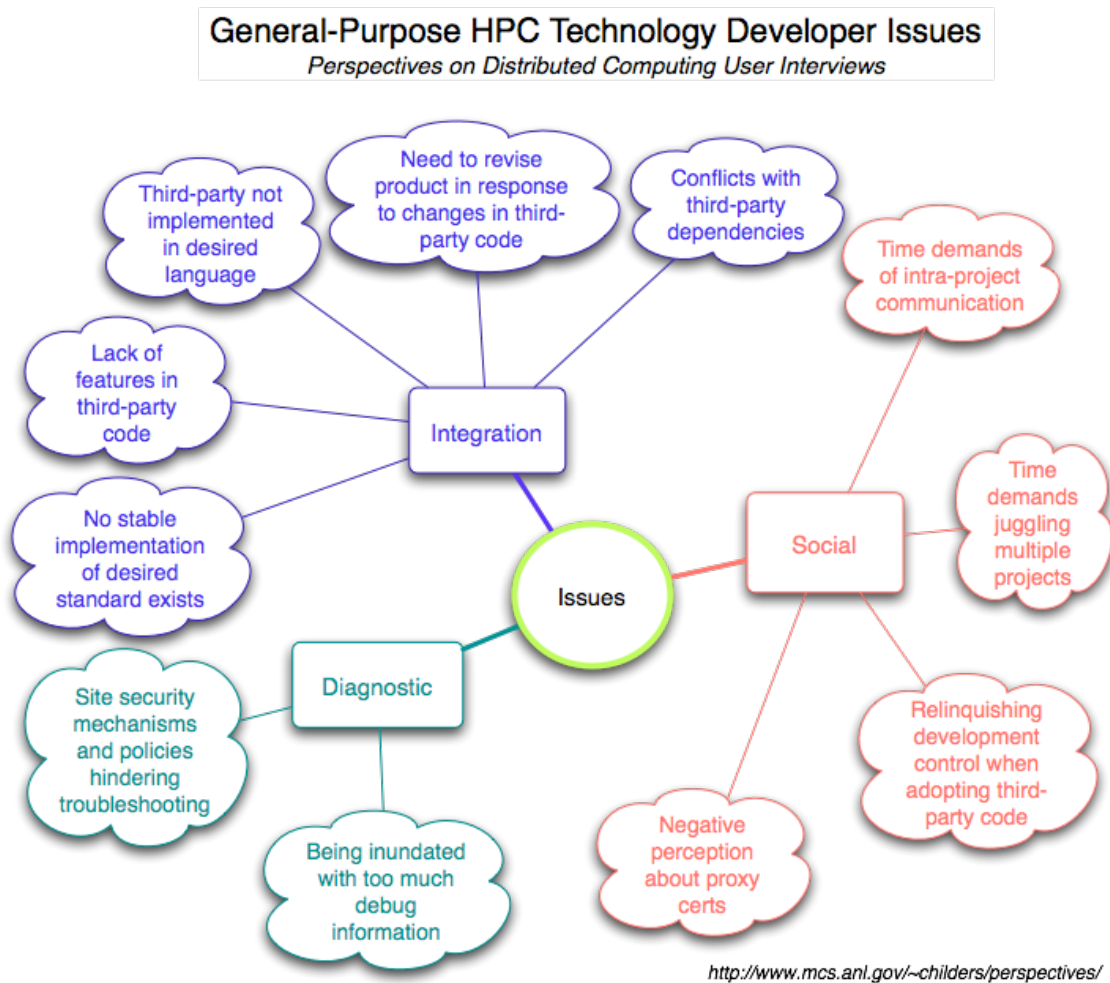
<sup>126</sup> Users 26, 28

### User-Focused Goals

The user-focused goals mentioned by the general-purpose HPC technology developers we spoke with include encouraging adoption in the community<sup>127</sup>, providing end-user support<sup>128</sup>, gathering the full-set of user requirements<sup>129</sup>, and ensuring that end-users succeed in their work<sup>130</sup>.

### General-Purpose HPC Technology Developer Issues

Figure 14 summarizes the issues reported by the four general-purpose HPC technology developers. The interviewees discussed design and compile-time integration problems, as opposed to runtime integration problems. More detailed information, with references to the interviewees corresponding to each cloud, follows the figure. Specific context can be found in their respective interview writeups in Appendix D.



**Figure 14: General-purpose HPC technology developer issues**

<sup>127</sup> Users 26, 28

<sup>128</sup> User 28

<sup>129</sup> User 28

<sup>130</sup> User 14

### ***Integration Problems***

General-purpose HPC technology developers who incorporate existing services or libraries into their work can encounter conflicts with third-party dependencies<sup>131</sup> such as Java runtime versions. Developers may find that a needed third-party library is not implemented in the desired programming language<sup>132</sup>. Sometimes multiple language implementations may be available but behave differently, such as is the case with the C and Java GSI implementations.

Another integration issue faced by general-purpose HPC technology developer is the need to revise their work in response to changes in third-party code on which they depend<sup>133</sup>. Developers must then decide whether to adapt to the changes or base their work on an old version, allowing it to fall out of sync with mainstream development.

The interviewees also expressed concern about the lack of desired features in third-party code<sup>134</sup>. Examples include usability and administration-related features such as monitoring capabilities. If the third-party vendor does not support such features, the general-purpose HPC technology developer must implement them himself.

The general-purpose HPC technology developers wishing to base their code on emerging standards also reported that they may find that no stable or accepted implementation of the desired standard exists<sup>135</sup>.

### ***Diagnostic Troubles***

Firewalls and security policies can hinder troubleshooting efforts<sup>136</sup>. For example, in trying to track down problems, developers may need to examine log files. However, this can be problematic when log files contain sensitive information; stripping out sensitive information from logs can undermine efforts to identify sources of runtime problems.

A second diagnostic problem described in the interviews that can hamper problem identification is being inundated with too much debug information<sup>137</sup>. One interviewee described a situation in which so much information is broadcast in response to an error (from thousands of concurrently executing processes all experiencing the same problem) that it is impossible to understand and manage; debugging tools for multithreaded code running at large scale would be helpful.

### ***Social Challenges***

General-purpose HPC technology developers who produce products used by the large distributed technology projects may find that communicating with fellow project members can be time-consuming<sup>138</sup>. The management and coordination of intra-project development tasks can produce hundreds of email messages and several conference calls per week. Work on multiple projects can also create time management pressures<sup>139</sup>.

For the general-purpose HPC technology developer, dependence on third-party code can save significant development time. However, depending on third-party code requires the developer to

---

<sup>131</sup> User 22

<sup>132</sup> Users 22, 28

<sup>133</sup> Users 14, 22

<sup>134</sup> Users 14, 28

<sup>135</sup> User 28

<sup>136</sup> User 26

<sup>137</sup> User 14

<sup>138</sup> Users 26, 28

<sup>139</sup> User 22

relinquish control over some aspects of his product development, in particular when and how changes are made to the third-party code<sup>140</sup>.

Another social issue discussed in the general-purpose HPC technology developer interviews was the negative perception outside the Globus community about proxy certificates<sup>141</sup>. This can be a barrier to adoption that the developer must work to overcome.

### **Recommendations**

In Section 5 we present several recommendations for developers of distributed technology, including the recommendation that they ***broaden the focus of component-centric development***. This would result in improved component tests and documentation and would perhaps decrease the burden of user support (though ultimately it might increase the burden if the user-base increased.) We also recommend ***partnering with other general-purpose technology developers to build and document interesting multi-component examples***. This would serve to highlight the usefulness of each component and generate instructive material for all parties. See Section 5 for more details on these and other recommendations.

---

<sup>140</sup> Users 14, 22, 26

<sup>141</sup> User 22

## 5 Recommendations

This section provides recommendations based on analysis of the interview data.

### 5.1 For Developers of Distributed Computing Technology

As the profiles in Section 4 show, the goals of distributed computing technology users extend well beyond the bounds of any single technology. Yet Grid middleware development is organized around providing users with functional components<sup>142</sup>. Developers specify, implement, and make components available for later assembly into products that satisfy specific use cases. A key advantage of the component approach is that it facilitates code reuse: a component can be combined with domain-specific code and other components to create higher-level products that perform large tasks. We see many examples of this in the interviews, with components such as RLS, GridFTP, and GRAM employed in a wide variety of domain-specific applications. Yet the developer who works solely inside a component-centric bubble runs the risk of missing key user requirements.

#### ***Recommendation 1: Broaden the focus of component-centric development***

The crux of this recommendation (and all that follow) is that understanding the broader context of component usage is essential to building useful, usable components. We recommend that component developers address five key needs, highlighted in bold italics.

Product requirements are derived from the users' needs: a goal or set of goals that users of the product are trying to realize. ***Understanding component usage in relation to the larger product requirements*** will help the developer write better tests for the component, based on realistic use cases; write better documentation, using terminology and concepts familiar to the user; and refine component requirements, using concrete ideas to augment the usual abstractions.

Note that a single set of product requirements can be satisfied by multiple system designs, as multiple implementations can provide similar results. Therefore, product requirements alone are insufficient to understand component-level requirements. Component developers must also work to ***understand the end-to-end system design of the product***. This understanding will yield requirements related to the component's public interface, such as interoperability, fault handling, and logging.

Component developers should actively manage the contexts in which the components are employed. If conflicting requirements exist, the developers will need to ***devise a strategy for supporting them***. One such approach is to design the component in such a way that incompatible functionality is encapsulated in pluggable modules. Another approach is to provide specialized versions of the component for each requirement set. Still another is to explicitly and noisily declare one or more requirement sets to be unsupported.

The component developer should also ***ensure that dependent elements in the system are readily available***. Developers should track the dependencies of each product on components from other sources and should monitor the availability of those components, for example by negotiating agreements with other development teams or contributing to other development products.

Component developers should fully ***document the types of products in which the component is intended to work***. Developers should not leave it to the user to infer the component's intended use or the intended context for applying it. Useful descriptive information includes a description of the product's users, a summary of the user needs met by the product, a list of use cases and the

---

<sup>142</sup> Many example components can be found at <http://www.globus.org/toolkit/docs/4.2/4.2.0/>



role the component plays in each use case, and an overview of an example system design and the key requirements provided by the component.

### ***Recommendation 2: Develop unique approaches for engaging different user types***

Figure 4 shows considerable diversity in what each of the thirty people interviewed aims to do using distributed computing tools. Clearly, a single approach to outreach, training, and user support will not serve all of these kinds of people equally well. Within each user type, however, we *can* take similar approaches to outreach, training, and support. The reason is that the user types are defined by the way members interact with technology. Technology developers should choose types of users to work with and devise unique approaches for the engagements.

Elaborating on this general recommendation, we discuss four specific recommendations, one for each user type: (1) provide a data movement product for HPC scientists who routinely move data; (2) enlist the aid of HPC domain-specific developers to translate generic technology concepts into domain-specific concepts; (3) partner with general-purpose infrastructure providers to refine requirements for reliability and multiple-use deployments; and (4) partner with other general-purpose technology developers to build and document interesting multicomponent examples. Again, we use bold italics to highlight points.

#### ***1. Provide a data movement product for HPC scientists who routinely move data***

Section 4.2 profiles the three HPC scientists we interviewed. The primary application they see for distributed computing tools is to move data between the systems where they perform simulation tasks, data analysis tasks, and the rest of their work. The HPC scientists are first and foremost scientists, not information technology specialists. Of the four user types identified in this report, the HPC scientists had the narrowest scope of interest in distributed computing capabilities.

Data movement is a task that many HPC scientists need to perform routinely. The scientists we interviewed were adamant that it is a time-consuming source of frustration for them. They feel that they need to “baby-sit” file transfer tools to ensure that data movement failures – which are frequent enough to become routine – are dealt with without losing valuable time. The scientists use various manual techniques to reduce the impact of failures and to assure themselves that the data has been transferred accurately. It is not only fault recovery that is a challenge. Diagnosing the cause of errors is also reported as a time-consuming issue.

An end-to-end product aimed at these users must ***move large volumes of data*** of varying numbers of files and varying file sizes. The solution must ***automatically restart*** in the event of transient failure. The solution must be able to ***diagnose common configuration issues*** at either end of the transfer (e.g., security misconfigurations, expired certificates) and provide diagnostics in layman’s terms. Most important, the product must reliably work with the installed systems at a wide range of HPC centers – multiagency, multi-institutional, international. The HPC scientist should not need to install, configure, or tune the solution: it should be ready to use. Nor should the HPC scientist need to tell the HPC center when the solution is not working: it should be monitored by center staff and kept working at all times. Consider the case where the GridFTP deployments at two centers do not recognize a common certificate authority. The scientists expect the performance of the solution to be high, utilizing networks and I/O interfaces to their capacity, but this is a secondary concern relative to reliability and automation: slower, foolproof transfers are preferable to faster, unreliable ones. A well-documented guarantee of data consistency and the method used to assure it is a critical requirement.

**Training materials for the HPC scientist must be practical and focused** on things that the scientist can and should be doing himself, not things that the IT support staff should be doing for him. A primary issue reported by HPC scientists is the need to move data from system to system because no single system meets all of their needs. Training materials for these scientists should clearly address how to use an end-to-end data movement solution, rather than how to build and deploy the solution, unless it is trivial to deploy. Additional issues that HPC scientists reported include having to personally deal with failures (e.g., restarting failed file transfers) and having to debug security issues (e.g., expired certificates). Training materials for this user type should provide practical suggestions for how to employ automated fault recovery mechanisms such as RFT and how to quickly interpret the nature of a security error and who to talk to in order to resolve each kind of error. Effective training materials depend, of course, on having in place a robust, reliable, easy-to-use, end-to-end data movement product on the major HPC systems used by HPC scientists.

## **2. Enlist the aid of HPC domain-specific developers to translate generic technology concepts into domain-specific concepts**

Section 4.3 profiles HPC domain-specific developers. These are people who both use and integrate general-purpose HPC technologies and also develop and integrate domain-specific technologies. Because of their role as mediators between scientists and technology, their goals are often of the form “Use [x technology] to facilitate [y scientific goal].” They mention social issues, but technological issues dominated among those we interviewed.

**Component developers should seek out domain-specific developers** as both a source of knowledge about the domain’s needs and an entry point or gateway to the users within the domain. Convincing the domain-specific developers of the value of a technology is likely to be an important step toward being given the chance to serve that domain’s needs. We found that two of the common activities pursued by HPC domain-specific developers are translating domain goals into technological concepts and translating concepts into practice. Domain-specific developers are the people who marshal information technologies to meet the needs of domain practitioners. They have feet in both the science domain and the domain of general-purpose technology, and translating between them is a key part of their job. The presence or absence of such domain-specific developers typically indicates the readiness of a domain for distributed systems technologies. Without such experts, the general-purpose technology developers will have to assume the translation burden normally shouldered by the domain-specific developers.

HPC domain-specific developers are interested in technology components that they can easily integrate with domain-specific systems to make those systems more powerful, easier to maintain, more flexible, or more robust. **Distributions should be in the form of software developer kits or modular packages** rather than preintegrated applications or suites. Components should offer the ability to customize behavior even at the expense of being complex, because the developer will take advantage of the configurability and hide the complexity from users.

**Documentation and training for these users should emphasize application programmer interfaces and service interfaces** for these components rather than direct user interfaces such as command line or graphical user interfaces. Included in such documentation should be examples of how the component has been integrated in other scientific systems and the benefits provided to those systems – robustness, performance, improved flexibility.

**HPC domain-specific developers should be provided diagnostic and status data** from many layers of their systems, including components provided by others, in a central location for debugging purposes. Error reports should be machine-readable and easy to decode.

***HPC domain-specific developers should be able to configure components to respond in specific ways to faults***, hiding low-level faults from the user of the systems they develop, and reporting problems only when the system cannot automatically recover. The recovery strategies may vary widely from application to application, however, based on the application-level requirements. Component suppliers should work closely with HPC domain-specific developers to understand how the developers believe that faults should be handled and provide ways for the developers to configure the components accordingly.

### **3. Partner with general-purpose infrastructure providers to refine requirements for reliability and multiple-use deployments**

Section 4.4 profiles general-purpose HPC infrastructure providers. These are people who use, integrate, and develop general-purpose technologies and who integrate – but neither develop nor use – domain-specific technologies. Operational and technical matters dominate their goals and issues, with considerable emphasis on providing stable, usable, and reliable technological capabilities to their users. Social elements do arise, however, because of the integration function the infrastructure providers perform for others.

General-purpose infrastructure is of key concern because it is one of the two places where scientific end-users encounter distributed systems technology. Domain-specific infrastructure is the other, and domain-specific infrastructure may also leverage general-purpose infrastructure. Infrastructure providers are also a source of stringent reliability requirements because they – and their users – highly value service reliability. General-purpose infrastructure must support multiple uses and users, which leads to another set of interesting requirements.

General-purpose infrastructure providers typically do not use the technologies they provide to scientists, so they are not likely to have firsthand knowledge of their users' requirements. Instead, they are interested in technology components that they can integrate to make their general-purpose systems more powerful, easier to maintain, more flexible, or more robust. It is not uncommon for them to have limited development resources, in which case they seek out third-party products. They value products that are configurable and flexible enough to adapt to their existing systems; a modular design is considered an advantage. They want components that can be customized both at deployment time by the infrastructure provider and at runtime by the end user. Runtime configurability is particularly important because the system will be used by a diverse set of end-users with unique needs. General-purpose infrastructure providers also develop general-purpose technologies, so they may avoid other implementations of components that they have developed themselves. See the "Develop" column of Table 7 for a list of these components.

***Documentation and training targeted at infrastructure providers should emphasize end-to-end use scenarios that illustrate the benefits of the products*** to both end-users and infrastructure providers. Include examples of how the component has been integrated in other scientific systems and the benefits provided to those systems (robustness, performance, improved flexibility, etc.). Deployment and configuration details are of equal importance, particularly the host system technical requirements and scaling limitations. Direct user interfaces (command line or graphical user interfaces) and service interfaces must be documented, but this information will most likely be passed directly on to end-users rather than used by the infrastructure providers themselves.

***Special attention should be given to diagnostics*** for general-purpose infrastructure providers. Local log files should be well organized. The data relevant to a given failure must be straightforward to find, ideally in a single location. Error messages must be clear and accurately identify any related system problems. Temporary files that may have contributed to failures must be preserved. The infrastructure providers noted that they often do not have control over the clients their users are using and noted that client version mismatches or configuration issues,

including security configuration, account for a nontrivial number of the issues. An easier means of flagging the use of “bad” clients would be an obvious way to help.

Infrastructure providers also find that their systems become elements of larger application networks and that failures at other sites in such networks or along the communication links may cause real or apparent failures in their local system. As in the client case, providers do not have control over these peer systems or the software used on them. Therefore, ***a means of flagging interactions with misconfigured or “known-to-be-bad” systems should be provided for diagnosing failures.*** Infrastructure providers also note that simply detecting failures in their deployed services can be difficult. Usage reports that record failures, log failure data, and notify infrastructure providers would be helpful.

#### **4. Partner with other general-purpose technology developers to build and document interesting multicomponent examples**

Section 4.5 profiles general-purpose HPC technology developers. These are people who use, integrate, and develop general-purpose technologies but who do not work with domain-specific technologies. The goals expressed by these developers are a mixture of technical development, with a strong focus on new or improved features, and social operations, such as encouraging user adoption and helping users to meet their goals. Interesting is the notable absence of technical operations goals, such as maintaining deployed systems, and of social development goals, such as identifying new types of users. Issues expressed by the general-purpose technology developers mainly involved integrating technologies developed by others, but there was a strong element of social and organizational issues, such as difficulties obtaining information from deployments and heavy demands for communication with users and for user support.

***General-purpose technology developers should conduct joint demonstrations of their components*** used together to accomplish end-to-end user scenarios. This type of engagement yields two benefits. First, the demonstration gives infrastructure providers and HPC scientists a clear example of using the components to resolve problems familiar to them. Second, it provides an opportunity to expose more people to each of the partners’ components, as presumably each partner has different (though most likely overlapping) user bases.

General-purpose technology developers are interested in technology components that provide functionality complementary to their own components and can be used with their own components to build end-to-end capabilities for users. General-purpose technology developers look for modular components; and they want such components to be available with specific versions so that once they demonstrate an integrated solution, they can be sure that the integration with the original version will continue to be available to their users. ***Components should offer the ability to customize behavior, even at the expense of being complex,*** because the general-purpose technology developer will take advantage of the configurability and hide the complexity from users. Nevertheless, the default behavior should both work and be simple.

***Documentation and training for general-purpose technology developers should emphasize application programmer interfaces and service interfaces*** for these components rather than direct user interfaces. Examples of how the component has been integrated in scientific systems will help other developers formulate possible integration strategies.

The integration issues reported by general-purpose technology developers include dependency conflicts between components that require different versions of the same third-party code, language mismatches, and frequent changes in other components. General-purpose technology developers chafe at being reliant on others for responses to issues (e.g., bugs, feature requests, performance issues), so to retain their interest and trust, one must ***be responsive to their requests.*** General-purpose technology developers also want to intercept and handle faults from components

they are integrated with. Recovery strategies may vary widely from component to component, however, based on the requirements of the component itself. ***Component suppliers should work closely with other general-purpose developers to understand how they believe that faults should be handled*** and to provide ways to configure the components accordingly.

Diagnostic features that assist in distributed debugging are important for general-purpose technology developers. Since it is difficult to obtain good information from deployments, ***remote monitoring and error tracking should be made available***.

## 5.2 For Further Study

The interviews in Appendix D are a rich source of data, much of which we were not able to analyze fully because of lack of time and resources. This section summarizes four of the tantalizing avenues that we have left unexplored.

### 1. Interview more users

To further refine the user profiles, capture additional requirements, and perhaps identify new user types, we recommend interviewing additional users. We particularly recommend interviewing more representatives of the following categories: educators, scientists, developers, and project leads; operations staff at HPC facilities such as TeraGrid, caGrid, and OSG; and producers of distributions such as VDT and Rocks.

We also recommend interviewing people in a variety of roles who work on the same project. The idea is to build an unbroken chain of user perspectives on a per-project basis, starting from the lowest-level infrastructure providers to the highest-level end-users. Collecting this type of vertical data for several projects should enable an interesting examination of the interdependencies among users.

### 2. Expand the user profile sections to include methods data

The user profiles in Section 4 describe the goals and issues associated with each user type. They do not present work-related methods reported by the users, though the interviews contain a significant amount of that type of information. Given time constraints and the need to make choices about the content we summarized, we decided that user goals and issues were a higher priority than methods.

Adding user method information to the profiles could aid in identifying problem-solving techniques that are common within each user type. Combined with the issues data, the presence of common methods for solving problems would be a starting point for identifying opportunities to improve the day-to-day work of users in each category.

### 3. Use other lenses to analyze the data

Developing a classification system for tagging the data and developing useful slices of the data to look at once it was tagged were nontrivial tasks. We selected the “technology interaction” lens as an analysis method after trying several others and realizing they were either not structured clearly enough to generate reliable results or not generating results relevant to technology developers.

We are confident, however, that other categorization approaches could be used to examine the data and would reveal interesting patterns or trends that would have implications on strategies for technology development and outreach. We welcome others who are curious to try them out.

#### **4. Conduct a technology interaction census for science and engineering professionals**

This study identified types of users based on common sets of technology interactions. The four user categories were formulated after only 20 interviews, and we were gratified to find that none of the remaining 10 people interviewed appeared to signify a new category. However, this does not disprove the existence of individuals or even large populations with different mixes of technology interactions. We did not interview, for instance, any end-users of the scientific portals built by some of the domain-specific developers we interviewed. Nor did we attempt to identify subsets within each of the four user types – for example, “domain-specific developers in bioinformatics” or “HPC scientists in high-energy physics focusing on analysis of community datasets” – but this might be possible with a larger sample size.

Measuring the sizes of these populations and identifying additional technology interaction types will be a key step toward both maturity and sustainability for distributed computing products. The population size is vital to understanding the potential reach of our products into these groups and for setting specific goals for outreach. The relative sizes of the groups could also point to obvious prioritization strategies. For example, if the HPC domain-specific developers – the gateways to the scientists – are a relatively small group, then we could more easily reach *all* of them, which would in turn result in potential adoption by large portions of the end-user scientist population. More generally, whichever user type appears to be the smallest population is most likely a good target for aggressive outreach.

A representative survey of science and engineering professionals that focuses narrowly on eliciting technology interaction data could perhaps measure the populations of both the existing user types and new ones identified in the census. If additional data were collected focusing on, for example, discipline or high-level techniques used, one could further refine the categories into sectors narrow enough to be useful in developing marketing and outreach strategies.

## Acknowledgments

This work was supported by the National Science Foundation Office of Cyberinfrastructure, grant number 0534113 “Community Driven Improvement of Globus Software” and in part by the Office of Advanced Scientific Computing Research, Office of Science, U.S. Dept. of Energy, under Contract DE-AC02-06CH11357.

We are grateful to each of the people we interviewed for sharing their time and their thoughts and for granting permission to publish their words. We thank the members of the NSF-sponsored CDIGS project team for their suggestions, assistance, and critical reviews, particularly Charles Bacon, Martin Feller, Laura Pearlman, and Cristina Williams. Ann Chervenak and the members of the *dev.globus* Incubator Management Committee released us from committee responsibilities to work on this project. Karl Czajkowski, Mark Hereld, and Steven Tuecke also provided critical reviews and excellent suggestions. Ann Zimmerman at the University of Michigan provided general suggestions and advice and an informal review of our study methodology. Michael Wilde graciously served as a test interviewee and helped us improve the interview script. Gail Pieper edited our final drafts and made this report significantly more readable. We made extensive use of the TAMS Analyzer software (an open source software product available at SourceForge.net) in our qualitative analysis. The figures were drawn using the OmniGraffle software package.

## Appendix A Study Methodology

This study was designed to document the experiences of distributed computing technology users. The inquiry was not hypothesis-driven. We collected data by asking people general questions about their work, and we examined their responses, looking for commonalities and patterns that might warrant further investigation. We include the data in this report to enable scrutiny of the findings and facilitate further study.

We created a generic interview script that encouraged interviewees to describe their work on a single project of their own choosing. The script contained a mixture of closed and open-ended questions, generating both structured data (e.g., demographics) and unstructured data (e.g., descriptions of goals). Practice interviews were conducted to refine the script.

To capture a wide variety of viewpoints, we sought to interview people from multiple science disciplines, job types, and U.S. research projects. We drafted an initial candidate list through ad hoc means – brainstorming for known projects, scanning names in our mailboxes, searching Google, and the like. We categorized the potential interviewees according to several attributes: project, funding agency, science discipline, role in the project, type of project. We then looked for gaps in attribute values and expanded the list of candidates to fill them. While not formally representative of the community as a whole, the list represented a broad cross-section of people. An anonymized list of the people we interviewed can be found in Appendix B.

We sent individual emails to each candidate requesting an interview. Fifty-two invitations yielded thirty interviews. All interviews were conducted one-on-one over the phone and recorded with the interviewee's permission. A transcript was produced and sent to the interviewee for review. All edits requested by the interviewees were applied; all transcripts received either tacit or explicit approval. The interviews can be found in Appendix D.

### ***The Interview Script***

The design of the interview script was informed by guidelines outlined in *Observing the User Experience*<sup>143</sup>. The interview opened with general questions, moved to specific inquiries and ended with a general wrap-up. Interviewees were advised to skip questions that they did not wish to answer or that were not applicable. During questioning, prompts were often used to trigger ideas and elicit more detail. The script served as a loose guide; when a person's response diverged from the question at hand, the interviewer did not interrupt but eventually returned to the script.

#### **Setting context**

##### ***Question 1 "Please provide a one-minute overview of your project"***

- (Prompt: "What is the project's name?")
- (Prompt: "Which agency funds the project?")
- (Prompt: "What field does your project belong to?")
- (Prompt: "What is your job type?")
- (Prompt: "How long have you been a jobTypeX?")

#### **Learning about discipline-specific goals and approach**

##### ***Question 2 "What are the main goals of your project?"***

- (Prompt: "How will the success of your project be measured?")
- (Prompt: "What are the professional measures of success for you?")

##### ***Question 3 "What are you investigating?"***

<sup>143</sup> *Observing the User Experience: A Practitioner's Guide to User Research*, Mike Kuniavsky; Morgan Kaufmann 2003; ISBN: 1558609237



**Question 4 "What is your method for investigating XXXX?"**

(Prompt: "How do you work?")

(Prompt: "How do you keep track of interim results, if at all?")

(Prompt: "How do you test work-related hypotheses, if at all?")

(Prompt: "How do you document your results?")

**Question 5 "In what ways, if any, do you interact with simulations in your work?"**

(Prompt: "How, if at all, do you share simulations with others?")

(Prompt: "How, if at all, do you interact with the inputs to your simulations?")

(Prompt: "How, if at all, do you interact with the output of your simulations?")

(Prompt: "By what mechanisms, if any, is access to your simulations controlled?")

**Question 6 "Describe how, if at all, you interact with data in your work"**

(Prompt: "How do you share work-related data with others, if at all?")

(Prompt: "By what mechanisms is access to your work-related data controlled?")

**Question 7 "What resources, if any, do you use in your work today? And by "resources" I mean infrastructure such as compute cycles, data sensors, data storage, etc."**

(Prompt: "How, if at all, do you share work-related resources with others?")

(Prompt: "By what mechanisms, if any, is access to your work-related resources controlled?")

(Prompt: "How do you locate available resources for use in your work?")

(Prompt: "What types of information, if any, do you need to know about a resource in order to determine if it is suitable for your work?")

**Question 8 "What software do you currently use in support of your work?"**

(Prompt: "What scripting languages, if any, have you used in the past year?")

(Prompt: "What programming languages, if any, have you used in the past year?")

(Prompt: "What workflow tools, if any, do you use in your work?")

(Prompt: "What parallel computing tools, if any, do you use in your work?")

(Prompt: "If the need for new software-based functionality arises in your work, how do you acquire it?")

(Prompt: "How, if at all, do you share software with others?")

**Learning about the user's problems****Question 9 "What challenges do you face today in accomplishing your work-related goals?"**

(Prompt: "What types of information do you need in order to address the challenges you face?")

(Prompt: "What technology-related obstacles do you currently encounter, if any?")

(Prompt: "By contrast, can you provide examples of technologies you find very useful today?")

**Question 10 "Can you think of any work-related tasks that decrease your productivity?"**

(Prompt: "Describe, if any, repetitive tasks associated with your work")

(Prompt: "Describe, if any, time-consuming phases of your work")

**Learning about the Globus user experience****Question 11 "Which, if any, Globus data components do you directly interact with in your work today?"**

(Prompt: "Did you install the componentX client yourself?")

(Prompt: "Did you install the componentX server yourself?")

(Prompt: "How many people currently use the componentX server besides yourself? If you are uncertain of the number, please respond 'I don't know'.")

**Question 12 "Which, if any, Globus security components do you directly interact with in your work today?"**

(Prompt: "Did you install the componentX client yourself?")

(Prompt: "Did you install the componentX server yourself?")

(Prompt: "How many people currently use the componentX server besides yourself?")

**Question 13 "Which, if any, Globus execution components do you directly interact with in your work today?"**

(Prompt: "Did you install the componentX client yourself?")

(Prompt: "Did you install the componentX server yourself?")

(Prompt: "How many people currently use the componentX server besides yourself?")

**Question 14 "Which, if any, Globus information components do you directly interact with in your work today?"**

(Prompt: "Did you install the componentX client yourself?")

(Prompt: "Did you install the componentX server yourself?")

(Prompt: "How many people currently use the componentX server besides yourself?")

**Question 15 "Which, if any, Globus common runtime components do you directly interact with in your work today?"**

(Prompt: "Did you install the componentX client yourself?")

(Prompt: "Did you install the componentX server yourself?")

(Prompt: "How many people currently use the componentX server besides yourself?")

**Question 16 "Why do you use componentX instead of an alternative technology?"**

**Question 17 "What are the major challenges you face using componentX today?"**

### **Wrap-up**

**"Is there anything else you'd like to say to the people who build software systems for use by people like you?"**

### **Summarizing the Data**

In Section 3 and in each of the profile summaries in Section 4, our goal was to provide a concise summary of the ideas expressed by the interviewees without adding our own interpretation or further context. We used the following method to construct these summaries.

After we transcribed the interviews, we used a qualitative text analysis tool to tag excerpts of each transcript, noting any instances where the interviewee mentioned goals, issues, or points of satisfaction. After tagging the transcripts, we generated reports showing all of the excerpts related to each of these topics.

Within each topic, we collapsed duplicate statements and then identified one level of generalization. This formed the second and third tiers of Figures 4-6. We worked diligently to generalize without interpreting. That is, we grouped instances where interviewees expressed the same general idea using different words or details specific to the interviewee. We avoided drawing inferences about motivations, connecting to larger themes, or introducing other forms of context beyond the excerpt. We did not correct factual errors. The summaries reflect what the interviewees said.

The summary data resulting from this process appears in Appendix C. Illustrations of each summary appear in Section 3.

To create the profile summaries in Section 4, we grouped the interviewees into four categories based on the technology interactions they described in their interviews. After these groupings were made, we applied the same summarization process described above to the data contained in each group. These summaries are the basis for the four user profiles in Sections 4.2-4.5. Because the profile summaries are based on subsets of the data, the summarization process resulted in different intermediate categories (the second tier of the figures) for each profile.

## Appendix B The Interviewees

The observations in this report are based on the interviews appearing in Appendix E. This section summarizes the demographic data of the study participants. Rows are color-coded to indicate the user type identified in Section 4.1.

**Table 9: Users interviewed**

KEY	
	HPC Domain-specific Developers
	HPC Scientists
	General-purpose HPC Infrastructure Providers
	General-purpose HPC Technology Developers

ID	Project	Sponsors	Science Fields	Job Types	Time On Project	Featured Quote
1	CNARI	NIH	Neurology	Developer, Scientist, System Administrator	5 months	The scientists' happiness is my main measure of success.
2	LEAD	NSF	Computer Science, Atmospheric Science	Science and Technology Liaison, Portal Developer	4 years	Troubleshooting requires knowledge about software internals.
3	[withheld]	NIH	Neuroscience, Computer Science	Scientist, Developer	1.5 years	The Grid is a black box to me.
4	ENZO	DOE, NSF	Astronomy, Astrophysics	Developer, Scientist	6 years	The reason my tasks are so time-consuming is failure.
5	MEDICUS	National Cancer Institute, Private	Medicine	Project Lead	3 years	Performance improved from days to seconds.
6	Lattice QCD	DOE, NSF	Lattice QCD [Physics]	Scientist, Developer, Project Lead	30 years	The Grid idea is great, but there are barriers to making it work today.
7	TIGRE	State of Texas	Biology and Medicine, Air Quality Modeling, Geophysics	System Architect	2 years	If you add up all the tools you don't get a good user environment.
8	MILC	DOE, NSF	Physics, Elementary Particle Theory	Professor	22 years	I am trying to understand where Grid computing adds value.
9	LIGO	NSF	Gravitational Wave Physics, Physics, Astrophysics, Gravitational Wave Astronomy	System Architect, Scientist	5 years	Globus enables more science.
10	ALCF	DOE	Material Science, Biology, Nuclear Engineering, Astrophysics	Storage Engineer, Project Lead	1 year	Solving problems is easy once you have all the data.

ID	Project	Sponsors	Science Fields	Job Types	Time On Project	Featured Quote
11	Open Science Grid Engagement VO	NSF	Bioinformatics, Meteorology, Material Science	System Designer, Developer	6 months	The difficulty is not that things break, but detecting that something is broken.
12	CNARI	NIH	Medicine, Neurology, Neurobiology, Psychology, Speech Pathology, Computer Science	Scientist	30 years	The right approach is to be highly collaborative with domain specialists.
13	TIGRE	State of Texas	Education and Outreach	Scientist	2 years	We play a strong bridge role in connecting people with technology.
14	MPICH-G2	NSF	Computer Science, Meteorology, Cosmology, Hydrology, Biology	Professor, Principal Investigator	8 years	I start with benchmarks and follow up with real applications.
15	Access Grid	DOE, Microsoft, NSF	Collaboration Technologies, Computer Science, Engineering	Project Lead, Developer	2.5 years	We provide mechanisms for sharing video, audio and applications.
16	OptIPuter	NSF	Computer Science, Geoscience, Bioscience	Project Lead	5 years	We assume a world where lightpaths can be scheduled between computers.
17	FermiGrid	DOE	High Energy Physics, Astronomy, Non-Accelerator Physics	Computer Professional, Assistant Group Leader	10 months	GRAM2 is kept alive by the need to interoperate with European experiments.
18	UCLA Grid Portal	U. of California	Information Technology	Programmer Analyst	5 years	We provide an appliance for each cluster that acts as a parallel head node.
19	TeraDRE	The TeraGrid	Visualization	System Architect, Programmer	6 months	I would like sites to serve 100,000 or more users per week.
20	PUMA2	NIH, NSF	Bioinformatics, Genomics	Project Lead	7 years	We can provide our users with fresh data more frequently because of the Grid.
21	PASTA	NSF	Ecoinformatics	Lead Scientist, System Architect, System Administrator, Developer	3.5 years	We work to enable discovery, access and synthesis of distributed datasets.
22	GridShib	NSF, TeraGrid	Grid Security Middleware	Middleware Architect	4 years	Our goal is to bring attribute-based authorization to the Grid.
23	GNARE	NSF	Genomics, Bioinformatics	Project Lead, System Architect, Developer	3 years	It would take forever for a biologist to get this machinery working.
24	Network for Computational Nanotechnology	NSF, Purdue University	Nanotechnology	Associate Director for Technology	almost 4 years	The vast majority of people who could use supercomputing are excluded.

ID	Project	Sponsors	Science Fields	Job Types	Time On Project	Featured Quote
25	nanoHUB	NSF	Nanotechnology	Application Engineer	n/a	The diversity of the systems we run on is a problem.
26	CEDPS	DOE	Computer Science	Project Lead	1 year	Our goal is to make it easier to troubleshoot Grid applications.
27	Angle	NSF	Computer Science	System Administrator, Research Programmer	7 years	The end goal is to automatically detect network anomalies.
28	VO Services Project	US Atlas, US CMS, Open Science Grid, Fermilab	High Energy Physics, Non-High Energy Physics, Computer Science	Project Lead	1.5 years	The production worthiness of infrastructure is of the utmost importance.
29	Adaptive Cyber-infrastructure for Threat Management in Urban Water Distribution Systems	NSF	Civil Engineering	Project Lead	7 years	Our framework must adapt to changing conditions from the problem & the Grid.
30	Adaptive Cyber-infrastructure for Threat Management in Urban Water Distribution Systems	NSF	Environmental Sciences	System Administrator, Developer, Researcher	1.5 years	The scriptable interfaces at various sites are not consistent.

## Appendix C Integrated Summary Data (for Section 3)

These excerpts underlie the integrated summary figures in Section 3. They are excerpts taken from the interviews in Appendix D that have been edited for flow and readability.

### C.1 User Goals

#### C.1.1 Conduct and promote scientific research

##### *Extend scientific understanding*

- The main goal of the project is to detect gravitational waves and to conduct gravitational wave astronomy. In other words, to learn about our universe through the use of gravitational waves.
- This is all done in the context of research in the recovery from stroke, which is a disease of brain vessels that is very devastating.
- The researchers are interested in analyzing the response of the brain to various stimuli. They do this to study how patients who have suffered a stroke improve over time.
- When genomes are sequenced, they are represented as strings of letters, which signify different nucleotides. Once a genome is sequenced, you want to know what functions the genes perform and what physiological and metabolic processes the genes are involved in. So starting from just the alphabet soup of the sequence, by the end of the analysis you know
  - how many genes this organism has,
  - what they do,
  - how this organism lives (because we're reconstructing double helix properties),
  - does it have any pathogenic or nonpathogenic factors,
  - what does it transport in the cell, and
  - what does it produce.

So pretty much by the end of the analysis, not through experiments but by using pure bioinformatic methods, the biologists know quite a bit about the organism already.
- As far as the cosmological part of it goes, we're trying to understand the hierarchy of structure formation on various scales, from the larger scales and the universe, down to galactic scales. This is a tremendous number of orders of magnitude in physical scale - in space and time.
- We hope to be able to account for observations, to determine the cosmological parameters of the universe we're living in, to provide theoretical underpinnings for observations from things like the Hubble space telescope, or the James Webb space telescope, or any of those major projects.
- The main goals are to understand quantum chromodynamics (QCD), which is one component of the Standard Model of elementary particle physics. This includes
  - calculating the masses of particles that interact strongly; those particles are made out of quarks and gluons,
  - calculating their decay properties, and
  - ... studying QCD at very high temperatures, and possibly with nonzero chemical potential.
- The goals are to understand the interactions of quarks and gluons, and applying that understanding to the discovery of new, fundamental parameters of elementary particles.

##### *Apply science to practical problems*

- The end goal is to automatically detect network anomalies. As a first step toward that goal, we want to be able to interactively analyze the data to gain better understanding of it. This will allow us to devise better algorithms that will automatically do that for us. Such anomalies could include
  - a user transferring large amounts of data
  - the presence of a probe or
  - some kind of an attack.

Any kind of anomaly, but we're looking at the problem from a behavioral point of view, as opposed to mining actual content.

- The science goal is building a dynamically adaptive weather simulation system.
- The purpose is to identify a contaminant source in the system from the measurements that come out of sensors or water quality meters.
- We apply these algorithms to this problem of water security, in which we want to identify a contaminant source and its release history.
- The goals are to develop optimization algorithms, to develop the simulation part to work under these environments, and to apply them to this source characterization problem, and finally to locate these contaminant sources accurately under different scenarios.
- Our project focuses on the identification of contamination sources in water distribution environments. The problem scenario is one where you have a large water distribution network and in one area an intentional or accidental contamination occurs. A water distribution system is a network of pipes, junctions, tanks, reservoirs, etc. If a contamination is introduced you want to identify the location of the source as quickly as possible. In this project we are trying to identify contaminant sources through simulation-optimization. Based on limited information we try to identify the source location to better apply remediation measures.
- Comprehensively, we will be developing a prototype system that city authorities can take and apply to their own problems.
- We will also want demonstrate it to an urban water distribution system.

### ***Eliminate barriers to scientific investigation***

- Now that you have all these images collected, what are the cool questions? Some cool questions, if you have many, many deployments are, "Give me all the computer tomography images of the twenty-year old males who have lung cancer." These are totally relevant questions to ask in the medical domain, but we cannot do that today because the data are not aggregated. The critical thing here is that if you build a Grid which is connected to clinical data you can come up with very interesting epidemiological questions. "Give me all the cases of a specific disease in a specific area of the country" - and then you can see a big picture.
- Our users include over 5,000 people using one simulation or another, and they cover the spectrum from undergraduate, graduate, post-graduate, industrial users. Exactly what they're trying to accomplish with these tools is hard to tell.
- The focus of the project is to study what happens to distributed computing in an environment where there is no bandwidth limitation. It assumes a world in which people can schedule dedicated lightpaths between distributed computers, on demand, in the same way that they would schedule supercomputer clusters. We want to understand how those assumptions change application architectures, as well as change application users' perceptions of high-speed networking:
  - What are the middleware capabilities that need to be developed that are still missing?
  - What will the endpoints look like that connect to these high performance services? Will they be desktop computers? Will they be browsers?
- To take Grid technology deeper into the academic infrastructure than it has gone so far. Our particular project is targeted at the state, but we're also working with the Open Science Grid, SURA (the Southeastern Universities Research Association) and several regional organizations.
- So users don't need to learn about how to use a particular cluster or job manager.
- What we want to accomplish in the project is to show that there's technology available today, such as the very strong security infrastructure in the Globus Toolkit, which can be used in an intelligent way to build a security model for patient authorization and privacy for health data.
- In particular not to address computational sciences and their problems, but people with real problems to solve in laboratories and experiments.

- To share resources that would otherwise not be available to ordinary researchers. They don't have to go and buy their own equipment. They can get their research done even if they don't have lots of money.
- Moving forward, we want to build WSRF-compliant services on top of both Globus Toolkit 4.0 Java and C WS Core.

### ***Build a case for continued financial support***

- The main goals of the initial phase are to have an operational system, have an initial set of users (scientists, researchers and educators), and form collaborations to keep it going longer term
- The project was composed as a demonstration project in the sense that we have to demonstrate the capabilities in these areas. But we're just now transitioning into creating the conditions for a production-scale Grid.
- We're charged with bringing up Grid applications in three targeted application areas: biosciences and medicine, energy exploration, and air quality modeling. These areas were chosen as examples of the application of Grid technologies to economically useful and interesting topics.

## **C.1.2 Satisfy user requirements for systems used to do science**

### ***Establish provisioning/allocation mechanisms that efficiently satisfy varying demand***

- To automatically handle job submissions, accepting genomic sequences from the users and finding available resources to process them.
- By the end of the project we should have an application framework deployable on the Grid that can adaptively adjust the resource requirements to the problem requirements.

### ***Establish and employ security mechanisms that support dynamic, inter-organizational collaboration***

- To provide a unified authorization and authentication structure.
- To bring attribute-based authorization to Grids. In the past authorization has been somewhat of a weak point in Grid middleware. It was previously centered on the idea of identity-based authorization (i.e., the grid-map file). As we now know, that doesn't scale. So in order for Grids to grow, we need a new approach to authorization and we think that attribute-based authorization is one possibility.
- Enable virtual organization administrators to create a structure inside the virtual organization using concepts such as groups and roles. We then provide access authorization based on this organizational structure, such that users can present themselves with groups and roles. Different authorization privileges and execution environments are enabled, depending on the roles and groups that the users present.
- The main goals are to efficiently and compliantly communicate medical images in a secure fashion so that patient privacy is guaranteed.
- We also work on the security model to make sure the privacy protection is there. This will actually be very important to further explore the medical field. When you have a patient-doctor relationship, you as the patient sign consent that only the doctor who is treating you or the staff of that facility is allowed to see your medical charts. So now Grid enables us to communicate all this medical information wherever we want.
- So one motivating factor is to develop a security model that allows the same doctor-patient relationship being translated onto the Grid, allowing data access, but only if the patient authorizes it. So we are developing what we call a patient-centric authorization model as a way to approach this.



- We want to standardize the protocol used by resource gateways (policy enforcement points, in jargon) to communicate with policy decision points (PDPs). PDPs are servers that keep the policies for privileges to those resources. It is very important for us to make sure that these protocols are common so that developments of middleware in the U.S. can be immediately plugged into authorization infrastructures developed in Europe (and vice versa). ... The implementations are different, but if we achieve a common protocol, then we can achieve interoperability.
- Another thing that is very important to us is providing support to the storage groups in defining what is called the next-generation of storage authorization models. Access to storage is one of the big use cases that we are trying to make right in collaboration with our stakeholders.

### ***Establish and maintain system stability***

- The goal I'm really looking for is to have a mean time between failures of three months. That's what I'm targeting from the data replication side.
- Making the infrastructure all highly redundant so we don't have any one single point of failure.
- To provide stable computing facilities for people to do cutting-edge science. We are not about doing science; we are a resource/infrastructure provider. So our primary goal is to keep the equipment up and running, have as stable an environment (i.e., not changing, as few crashes as possible.)
- One of the motivators for this data collection effort came out of a weather event. A collection in Houston was down for several days due to a hurricane, because they turned the machines off and put them in a truck to move them somewhere else. So part of the goal is to be able to replicate data around the state so that kind of thing won't happen.

### ***Establish uniform diagnostic mechanisms that satisfy debugging needs in dynamic systems***

- The goal is to make it easier to troubleshoot Grid middleware and Grid applications, where troubleshooting doesn't just include failures but also includes performance-related issues. The nature of Grids makes it quite difficult to figure out the source of failures, and much of the underlying middleware lacks the right hooks to make it easy. So the goal of this research project is to figure out what is missing and try to get it added. We are trying to get many pieces of Grid middleware to use a common logging format and to log the right stuff. The hope is for OSG to report a noticeable drop in number of failed jobs, also to report a decrease in the amount of time it takes to track down problems.

### ***Provide compatibility with existing system components***

- There are very different modalities out there that we must take care of, and there's a lot of incompatibilities between the very different devices. Because of that, one current work focus is to make sure that the project can handle all these different vendor-specific imaging devices.

### ***Acknowledge the requirements of all users***

- Then there are the more political metrics related to how happy people are with the way we conduct our business. Do we have an open process to consider input from different stakeholders? Do we have large groups that are not considered in our requests for input?

## **C.1.3 Expand the resources available to a specific community**

### ***Provide computing systems to scientific users***

- We hope to evolve into a center-type project where we used distributed computing resources for this type of problem.

- To expose mass storage in a shared way.
- To have a unified systems support structure.
- We focus on other aspects too, such as providing access to computing resources. In this case success is measured on how production-ready the infrastructure is. This is a question of whether we can meet the baseline of our users job flows, of access to files, etc.

### ***Federate institutional computing resources***

- To create a structure within our institution so that six or seven interest groups with big pots of computing can share each other's resources. This is by far the bigger focus: to share the resources amongst ourselves, as opposed to sharing resources with people outside our institution.
- We want to share resources among the cluster users because there are so many clusters. We have 15 nodes here, 15 nodes there, 20 nodes over there, etc. Individual researchers own them and they are not used all the time. Cluster usage increases when people have an interesting project to do; after they are done they focus on something else, like analyze the data. During those times, lasting maybe days or weeks, the clusters are not used.

### ***Aggregate cross-institutional resources***

- I'd like to try to get production data analysis running on sites that are external to the DataGrid so federate into other Grids, such as Open Science Grid and TeraGrid.
- I'm also working on getting some of the data analysis pipelines or workflows to run in more sophisticated ways across multiple computing sites. Right now our users tend to pick a site and go run there. And while they use some Grid tools to do some of the data finding, they don't typically use Grid tools to leverage more resources than are available at a single site. So another of my metrics for the coming year is to try to enable production data analyses that run across multiple DataGrid sites in a continuous way.
- The goal is to get the scientific code that the researcher uses ported to the Grid, such that he can run it at larger scale and get much better speed up with what they do.
- I'm working to enable the user's local PC to participate.
- The main goals of the project are to find the contamination source as quickly as possible using available computational resources from as many sites as possible. So we try to accrue as much resources as we can to solve a time-sensitive problem. One important component of our project is find out which sites have the most resources available and try to offload our computations to that site. So applying computation to the science problem is one of our major goals.

### ***Run existing scientific applications/codes at higher resolutions***

- What we are trying to do is integrate computation systems with atmospheric science models and instruments. From the science perspective, we are trying to look for more ways of running the huge computational science models at high resolutions. That's been a big challenge. So running a weather forecast at storm scale and at tornado scale is something we've been trying to do.
- The big challenge for us is in three-dimensional radiative transfer, which is how light basically interacts with a fluid medium. Our code is being extended right now to incorporate these frontier pieces of physics, that up to now we haven't had enough computer power to include. So, there is a finite list of things we'd want to add. I wouldn't want to be too strict about this point, but we'll probably add most of the physical things we want to add in the next two to three years. And then it will be mainly a question of just how much computer power can we get our hands on. Our goal will be to be able to run problems that continue to match the most powerful resources available in the US, whether they're DOE or NSF.

### ***Enable scientific applications that require coordinated use of multiple systems***

- We are creating a tool called MPICH-G2, which is a Grid MPI, that can be used to solve computational problems that cannot otherwise be solved. Every time that we have solved a problem that no one was able to do before, we are very happy because MPICH-G2 is enabling technology. That's our goal. And we keep pushing that envelope farther and farther out - as far as we can. I also have a personal interest in trying to make this stuff work as a way of improving usability.

### **C.1.4 Expand the community that can use a specific resource**

#### ***Make scientific data accessible to more potential users***

- Establishing a framework where neuroscientists, especially people who are involved in brain imaging, can store, analyze and share their data in an effective manner.
- Our project is aimed at using database technology to store data that is electrophysiological or neuropsychological in nature. The data consist of hundreds or thousands of timepoints in a time series. Each timepoint consists of a large amount of data itself, such as a brain image or an electrophysiological recording from different sites on the brain. Using database technology to store the data is a way of allowing much more useful access to the data. Also through interactions with Grid computing experts at the University of Chicago, we found that it's a much better way to interface with distributed and high-powered computing devices in order to process the data.
- The main goals are to efficiently and compliantly communicate medical images in a secure fashion so that patient privacy is guaranteed, using existing security mechanisms and other standards-based technology provided by the Globus Toolkit. An example of the vision is that a patient comes into the hospital, and this patient's record is not only existent in the hospital, but also available on the Grid so that other healthcare providers can access the data and also add to it. So wherever you go as a patient Grid can basically follow you, aggregating all the information that exists for you at various health providers, and collecting them at the point of care. The challenge is to provide a technical solution that can scale to allow a large number of healthcare providers to interact and share images.
- Now Grid enables us to communicate all these medical information wherever we want.
- These are totally relevant questions to ask in the medical domain, but we cannot do that today because the data are not aggregated. The critical thing here is that if you build a Grid which is connected to clinical data you can come up with very interesting epidemiological questions - "Give me all the cases of a specific disease in a specific area of the country" - and then you can see a big picture.
- We plan to put up really nice dashboards that our collaborators can look at and see the current state of replication throughout our DataGrid.
- The primary project involves an architecture we call PASTA, which stands for Provenance Aware Synthesis Tracking Architecture. As part of the system there are 26 sites, which are spatially distributed across the continental United States, two in Antarctica, one in Tahiti and one in Puerto Rico. Each of the sites is collecting scientific data. The goal of the architecture is to pull data from them in a seamless way, based on both the metadata records and open access to the actual data file. These data are brought into a centralized data warehouse to enable data discovery, data access and synthesis of distributed datasets within the network.

#### ***Make scientific applications accessible to more potential users***

- Diving down a bit, my real project is the DataGrid. And the purpose of the DataGrid is to enable as much science as possible to be conducted using the project data. The data has to be analyzed. It's very computation and data intensive. And the main goal of the project is to build infrastructure

(tools, middleware, end-user tools, services and systems) that enable scientists to efficiently analyze the data and conduct their research.

- Establishing a framework where neuroscientists, especially people who are involved in brain imaging, can store, analyze and share their data in an effective manner.
- The main goal of the project is the analysis of large volumes of genomic data.
- To do the analysis we take the genomes from a national biology information repository, which is called the National Center for Biotechnological Information (NCBI). We then analyze a genomic sequence with an array of bioinformatics tools so it will be easier for the researchers to answer the questions that they usually have.
- We perform high-throughput analysis of sequence data to deliver results as fast as possible, making Grid resources available to the community. At the highest level, the goal is to create a gateway to the Grid for use by biologists.
- Once the data are actually centralized we develop interfaces to enable users to explore the data. I think one term is "exploratory". So we're developing web-based applications that include discovery interfaces, plotting routines, different types of data download mechanisms, and allowing end users to integrate different datasets so they can generate more or less synthetic products on the fly.
- Part of our goal is to make this next leap from science that takes place at the site to science that takes place at the network level. This would be a national, if not global scale. So the anticipation is once these datasets become available to end-users, that simulation and modeling will begin taking place. That's probably on the horizon within the next 1-5 years.
- As a first step toward that goal, we want to be able to interactively analyze the data to gain better understanding of it. This will allow us to devise better algorithms that will automatically do that for us.
- From the computer science part, we've been trying to make all these legacy Fortran codes and the legacy applications run in a service oriented architecture. And providing users direct access to advanced computational resources from a portal-based environment.
- The engineering goal is to build a service-oriented architecture in support of the science goal. So we are building the service-oriented architecture and Grid middleware ourselves while trying to leverage as much as possible the tools already available in the community.
- The main goal would be to enable researchers access to computing simulation codes to further nanoscience and nanotechnology.
- To provide online simulation services to a group of nanotechnologists around the globe.
- To change the expectations of experimentalists and educators regarding theory and modeling and simulation, and ultimately to change the way they do work - really seeing the concept of "simulate first, build later" be pursued in several areas of nanotechnology.
- To put the simulations in the hands of the people who need them and who wouldn't otherwise have access to them. We feel that simulation itself generally speaking is a very powerful tool to be used by people in the research or industry (or even undergraduate studies or whatever.) It's fundamentally useful across the board. But not everybody has access to everything, so we're trying to fill that niche as best we can.
- We want to move actually nanoscientists to nanotechnology, so we want to put simulation tools into the hands of people that normally wouldn't touch simulation with a ten-foot pole. The target audience is experimentalists that have work to do in the lab and they want to maybe design before they build. They're educators who want to train their students. They're students who want to study nanoscience and simulate structures. And they are potentially industry people as well as government persons.

- The main goal is to develop optimization algorithms that are less sensitive to the distributed, heterogeneous nature of the Grid resources and work reasonably well under different conditions.
- We have optimization algorithm development that will work in the Grid environments. We'll also enhance the simulation tool to work under these conditions, and then demonstrate the work.
- Software like Abacus and MATLAB has expensive licenses. If a resource is needed for two months then you don't want to buy that license for a year. If a cluster has a license then the cluster owners can let others use it. These things will be much easier in a Grid environment.
- To demonstrate applicability of Grid technologies to a wide variety of economically interesting and intellectually useful activities in the state.

### ***Make computation services accessible to more potential users***

- One goal is to provide a unified point of access for inbound Grid jobs, in order to share our institutional resources with the Open Science Grid.
- We are using the Globus Toolkit software to enable users to submit jobs into the multiprocessor system in a way they is both familiar to them and easy to use.
- The TeraDRE is a distributed rendering environment. What we mean by rendering is that we take models that are primarily generated from scientific data or from computer graphics and render the frames to make an animation. We want to make it easy for people who are not computer programming experts to render their jobs very quickly, and also to provide a certain level of flexibility to add more rendering technology.
- To bring new users on to the Open Science Grid.
- To bring in industrial partners and to help the business of the state as well, in terms of access to more resources.

### ***Make scientific colleagues accessible to more potential users***

- Our technology attempts to provide an environment where people can interact as naturally as when they're in the same room.
- We try to provide mechanisms for sharing not only visual and audio input, but also interactions with applications. We've also looked at allowing people to interact with remote instruments and computation.
- I would like to enable people to
  - share data and an application with others in a meeting,
  - grab the software and immediately set it up without requiring any depth of expertise, and
  - interact in a way that makes sense to the user (in terms of handing data files off to others, guiding people through a tour of visualization data, sharing equations, etc.)
- To bring in industrial partners and to help the business of the state as well, in terms of access to new knowledge and collaborations.

## C.2 User Issues

### C.2.1 Reliability

#### **Systemwide**

- As far as mitigating the effects of system failures for this frontline work where you're basically using an entire computer system at a site, one idea is to move away from the batch-queuing model. Move to a model that is closer to a physical experiment, as if you're using for instance, an astronomical telescope. In other words, it would be much more beneficial to us to be able to run for a long time, but to book that runtime at some point in the future and to have systems staff on call when the reserve timeslot begins, to fix anything that occurs.
- And at 60,000 processors (or whatever it's going to be) I suspect that computation at that scale will not be possible using the current approach to batch production. How on earth would you assure a user that when their timeslot came up that every single component of the system was functional? And how long would it stay in that state, given that the mean time between failures is proportional to the component count?
- Having jobs crash, often due to a node going down -- that's the most frequent reason that a job fails to complete: having to fix things up.
- I'm running in this 2,000-4,000 CPU range at the moment. And within the next year we expect that to go up to at least the 32,000 CPU range, if not a factor of two more than that. The unreliability that I see in filesystems, even in batch-process launching systems, disks, monitoring tools - you name it. Nothing really works reliably at the 2,000- or 4,000-processor level today. I am extremely doubtful about it working at a level ten times greater than that.
- I've heard from both OSG and from the LHC Grid and from a number of other Grid projects. They all give roughly the same answer: somewhere around 25 percent of their remote job submissions fail. This is a shockingly high number. In general they don't know why the jobs fail - they can only guess why. The top reasons cited for failure are basic authentication problems. You know - the user might not be in the right gridmap file. There are also disk-related issues such as running out of disk space during the act of staging in some input file, or they don't have the right permissions, etc. But then there are a whole lot of other failures that fall into the unknown category
- It seems like always somehow NFS is involved in some very sticky way when we're dealing with GRAM2 and it's not a pleasant experience. The biggest hurdle that we overcame to get to where we are now is by throwing hardware at the problem and getting a BlueArc NAS server, which is a very, very high capacity NFS appliance. Before that, NFS was crashing more than monthly, triggered by the kind of NFS activity that GRAM2 does. We still crash every once in a while - maybe once every other month or so - but nowhere near as bad. We have some idea [about what is triggering the failures]. In short, GRAM2 is doing hard links across NFS, and either the NFS client-of-the-day or the NFS server-of-the-day is not always reliable enough to implement that right.
- Stability for me, in the context of PVFS, means running without failures. Servers don't hang. User jobs run to completion. So right now PVFS hangs and jobs have to stop because they can't write data. We've got to get to the point that something figures that out, a backup comes into play, and the job can continue.
- The difficulty for us is not that things break; the difficulty is in detecting that something's broken. I may not even know who owns the site - it's just a black box to me - and something went wrong. Now I have to figure out what went wrong. So I do a little bit of probing, and then either tell the remote site what went wrong, or fix my stuff. Most of the time it is easy for me to figure out what is going wrong once it is detected.

- There's many ways a job can fail, obviously. One of the ways is that the job will be successfully submitted to a site, something will run, but not successfully. Let's say that a filesystem goes away while the code is running. To detect that type of failure is not that easy. This is because different error codes are returned, and some schedulers will say the job was successful while others will say that it was not successful.
- When everything is working, it's great. So a lot of what we need is more fault tolerance and improvements in the way errors and exceptions are handled. The errors can be due to hardware or middleware at any level.
- From a workflow point of view it's dealing with failures. If you can move 99 files with a batch script and they all got there safely, it takes you as much human time to deal with the one that didn't as moving the 99 that did. So human intervention to deal with the failures is the expensive time.
- It would be good to really understand what kind of failures Grids like OSG experience today. Most of what I know is somewhat anecdotal. Getting a picture of this is a hard problem. I don't know that they know. TeraGrid is the same way. It would be nice if there were somebody tracking and documenting failures in an organized way. It's much more common that you suffer a failure in the first few seconds, than it is the last few seconds. If any node, for example, can't see the parallel filesystem, that's fatal to a user job but it might be something that can be fixed quite quickly by a sysadmin. But if you're running in batch, you wait days (if not weeks) till your batch job starts, it fails instantly, and you have to go through the whole thing again. So the operation of these things needs to be made reliable in both physical and human terms. You've got to have systems support to overcome that kind of problem in real-time.
- We have so many things to keep track of we're losing the ability to keep track of it. We're building tools that need the information to function; they need the information to leverage the infrastructure. Tools need the information to help the users get the work done. But the systems that provide the information are breaking underneath the load. Then all these tools and infrastructure become unworkable and work stops.
- When we were first implementing the site selector we were relying on remote site information. We ran a daemon on every remote site that would in turn report back to the site selector. But that didn't work out so well because if the remote site went down, then the whole system went down.
- The Grid as it exists today as a computational resource provider is at the maturity level of the telephone system 80 years ago. What I mean by that is if you wanted to place a phone call to somebody, you would call the operator and say, "Tomorrow at noon I would like to place a long-distance call to so-and-so." And you pray that all the connections will work and you are able to make that phone call. The Grid is not yet a service that you can dial up, instantly connect to, and repeat again and again without hiccups. The Grid needs to work more like the telephone network. I just drove 99 miles and I'm almost certain I went through several service providers while I was talking on my cell phone, but I didn't have to think about it at all during our conversation. There are reliability issues with the Grid software that's out there. File systems fill up, certificates expire, and jobs fail. Maybe computational scientists are knowledgeable enough to put up with that, but not end users-not experimentalists.
- When a simulation fails it should be retried automatically, and if it fails again or can't be retried for some reason, it should be reported clearly to the user with something more than an obscure error code. We're quite far away from doing that with any sort of reliability.

## **Service Level**

- Not all the tools work as well as you would hope in terms of doing what they say they'll do, or having bugs, or "Oh, we didn't think of this yet". So a lot of maturity issues with some of the tools we try to use.
- GRAM2 has been more stable and reliable than GRAM4. That is the only reason I prefer GRAM2 over GRAM4. I need at least 70% success rate to consider a service stable. Ideally we want it to be much higher, but with GRAM4 we are seeing a much lower success rate. I certainly don't want to

blame everything on GRAM4. We've seen hardware failures on the cluster side. But I would say GRAM should improve the way it responds to hardware and network failures.

- I personally would like to have many of the basic services, like GridFTP and GRAM, be more reliable before I see more features coming out.
- It should be a stable service to our user community, a reliable deliverer of services. This probably means we shouldn't use the development versions of the software tools. We probably should use only the production versions. If the production versions are compatible, it will be much easier on the application developers.
- The ability to pull things together so easily now using a web interface. But at the same time this opens up problems, because it really allows anybody without a formal background in software development to develop these applications. The concern is similar to my pet peeve with Visual Basic: the resulting code seems very fragile, and you have to be very careful how you use it. Things break, or the maintenance of those types of applications become a nightmare at times.
- The goal I'm really looking for is to have a mean time between failures of three months.
- Give us the hooks to make it redundant if you don't do it yourself.
- Let's say you decide that within the software you're developing you want to rely on this particular open source software that seems really cool. You need some assurance that this piece of software will have longevity before jumping into it. Or you may decide that you are better off writing it from scratch, which you really want to avoid if at all possible.
- The most difficult thing was implementing the site selector. No matter what we did, we always ended up having jobs that would fail or just sit there not doing anything. And we always encountered a new set of problems that were not taken care of in the previous implementation. We had to keep changing, keep looking. Some sites would show us as running forever; the jobs would have the status as running, running, running ... and nothing was really happening.
- I guess the major challenge is still in the Grid aspect of this. It's still difficult to have a job run the first time you submit it anywhere and that shouldn't be the case. So our typical issue is in site selection. As I mentioned earlier, we pick a site randomly from a list that of sites we believe are operating. We believe the site is operational because it was operating an hour or two ago. But it may have stopped working in the interim. The problem is that our rate of submission failure is higher than we'd like to see under high loads.

## C.2.2 Diagnostics

### **Error Messages**

- When these problems are happening, for instance when hardware fails, the middleware we rely on gives cryptic error messages which we cannot read and parse automatically so that we can adapt to it. As I mentioned before the GridFTP "login incorrect" error does not provide us with sufficient information. In other contexts the source of login problems typically are on the client side, but in the GridFTP case it is often a server side problem. So what I would wish is when middleware cannot determine the particular error (and it's reasonable that it cannot determine everything), I would rather it propagate the original error message. Send it up the middleware layers of the architecture, instead of misinterpreting something and issuing a misleading error message.
- Another type of information that is lacking is documentation about errors. For example with GRAM, all we get is an error code, and there is not enough documentation explaining the error. We then have to google to find out how other users handled the problem. Sometimes we even need to go as far as to dig into the GRAM source code to determine under what conditions the error code is sent. There is some documentation, but not at a level enough that we can use it.
- Sometimes we get weird errors that don't really reflect what's going on. Other than that it's fine. Weird errors like, "Error code 17" that supposedly means one thing but is most commonly due to



something else. For example, it says, "could not create job description file", when the real reason is the user doesn't exist.

- We can't use a tool that produces a three-kilobyte file of error messages when something bad happens. Something happens and then all the processes start sending messages saying, "I can't do this. I can't do this, etc." We can't manage that. That's too much information, which is just as useless to us as no information.
- If we see certain errors right up front, we cannot directly take that and send it to, for instance, the TeraGrid helpdesk. I have to do at least an hour of digging. Because if I send directly an error message to the helpdesk, they will reply, "This is something to do with your client side. There is something wrong." So I dig deeper and deeper and go through my usual tests, and see, "Oh this service is down. Ok here is what's happening."
- Their error messages are very cryptic. The most common error we get is "login incorrect", but it has nothing to do with an incorrect login. It's something like a hardware problem, or there's a node goes down in a striped GridFTP server, or the allocation is out of the limit, or some scheduler is paused. For all these conditions we get the same "login incorrect" message.
- Error messages are the number one thing. With the gatekeeper that's what I usually have had issues with. Globus error messages are worse than Microsoft's -- worse in the sense that they really are not helpful at both the administration level and the client level. It doesn't have to be this way, if you look at the stack and the processes that actually do the execution. The functionality that GRAM and the gatekeepers are doing, we've been doing for years. Why is this so hard to do? This is a real challenge ... it's really frustrating for the end user and admin. Because it's not bad enough that you have a problem to solve from a users support standpoint, but it could be a heisenbug. Or it could be repeatable. Identifying it - that's one of the hardest things to determine in any system.
- There was this one site where all the resources were dual-CPU nodes. When we submitted BLAST [bioinformatics tool] jobs each job would be assigned to one CPU. If the next job were assigned to the second CPU on the same node it would crash due to memory problems. The problem was that neither Condor nor Globus would report this type of crash. So in this case half of the job would run, and then if another BLAST job came in, that other half would crash out. So you get an inconsistency, in the sense that incomplete output would come back and not be reported anywhere across all the software layers. Condor didn't catch it. Globus didn't catch it. Nobody caught the error. So we assumed that it was all completely done, and we loaded the results into our Oracle database. And then once the user looks at it, he sees bad results because of it. We could never figure out how to fix this problem. We had a long discussion with some of the OSG sites. Then they implemented a policy where they would not submit more than one of our jobs onto each node.
- It isn't particularly useful from our standpoint to put "Globus error 43" in front of the user. He will look at that and say, "I have no idea what that is." So in that sense not a whole lot of information is given back to the user beyond an indication that something failed. We do try to deduce the problem from the error report to a level like "the transfer of input files failed." And then maybe suggest that perhaps their file doesn't exist. If we can tell them that much, we will. But by and large the user doesn't see much error feedback.

## **Troubleshooting**

- How you go about troubleshooting when things don't work. Example: So I do a globus-url-copy from one center to another and I get an error message saying "End of file encountered". And the file at the other end is of zero length. Now what do I do? Right now, I send an email to the administrator asking, "What does this mean? Why didn't it work? It worked six months ago."
- Solving problems is easy once you have all the data in front of you. It's getting the data and knowing what data to get that's the hard part. Networks are notorious for this, right? They're black boxes. Very rarely are you lucky enough to have access to somebody who can actually find out operational status on routers and the like. So you have to infer what's happening by using things like Iperf, netperf, pipechar, etc.

- We're worried about things like firewalls and security policy getting in our way. I don't know if security policy is a technology obstacle or not, but it could be considered one. Part of the problem with logs is that there is potentially sensitive information in there, and if you strip out the potentially sensitive information, you often lose the ability to do troubleshooting.
- When anything goes wrong with your certificates, site certificates - anything like that - it's completely beyond the scope of anything a user can do. And usually it's beyond the scope of what the computer center personnel can deal with as well. It usually means that you're just crippled for a couple of days until the one guru at site X can actually figure out why what used to work no longer does.
- When new software tools become available it would be useful to know
  - what the new features are,
  - whether the features are compatible with previous versions, and
  - where the incompatibilities might impact the other parts of the system because in this case we'll know what to troubleshoot.
- When you write multithreaded code, bad things can happen: deadlock. Another condition that also appears in serial code, but is perhaps more pronounced in multithreaded code: accidental memory overwrites. A tool to handle that at large scale would be tremendously useful. By large scale I mean hundreds of distributed processes - even thousands.
- Finding out what to do next or troubleshooting is not something I am capable of doing - not at this point, without going through a learning process, which I didn't have time to do.
- The bigger missing documentation is in the troubleshooting area, where something is happening and we need to find out how to deal with it. Troubleshooting type of documentation is not only for me but for system administrators - they struggle without this. Because whenever something happens, we immediately post it to [help@teragrid.org](mailto:help@teragrid.org), and the system administrators try to figure things out. All the troubleshooting right now requires knowledge about the internals of the software. So only experienced people can troubleshoot right now. So if expertise is missing on the admin side, the issue keeps spinning for three or four days.
- Another problem that I find that is lacking is the ability to debug an application. It would be really handy to have a mechanism that would allow a developer to attach a remote debug utility to a Globus gatekeeper such that a deeper understanding of problems could be obtained. Though I suspect that this is quite possible to do if the gatekeeper is installed locally, it is not quite the same as being attached to a production gatekeeper. From the admin perspective, one of the issues is the ability to capture information when it happens. If a user were having difficulty submitting a job, it would be handy to have a trigger that would capture information when the user tries an action. It's often not the case that it's the middleware that may be having trouble - it could be the backend systems - but the ability to capture and repeat the user's actions would allow for quicker debugging. It would be helpful if this type of functionality was included as part of a web admin console. Such that both users and admin could see trace logs, etc.
- One of the things I ran into with the DRE was initially GSIFTP was configured like an xinetd server that's up and down all the time. So for every user call you are creating a process and killing the process. Unlike doing HTTP download where I could basically hit the web server constantly with new connections, GSIFTP really didn't like that, and it died. In fact when you hit any box running twenty of these against it, it basically came to its knees very quickly because of the overhead of starting processes. If I am user writing against this service, how do I find out how it's configured? I would like for a service to be able to allow me to connect to it, not really do anything but give me back some information about how it's configured so I can make a choice on how to use it. Am I starting up this process for every connection? Or is there a throttle placed upon me? How many other servers are running right now? Maybe I don't want to run right now - but I don't really have enough information to decide. It's a black box. When I print, I am able to connect to the printer spool, and it shows me the entire spool. You can inspect the queues. You can inspect the job queue if you had an alternative means, but I don't think you can see other people's jobs or how many jobs are running through GRAM. MDS is supposed to do that sort of thing, right?

- The issue when things go wrong on the Grid is trying to figure out what happened. It can be something on server side - some variable was set wrong. But you have to track it down and be able to replicate it and there's really not a way on either side to replicate that. As to the type of failure I'm describing: Java reported it as a failure, right? It may not be a GRAM failure. For example, we have issues with stage-in and stage-out sometimes (e.g. when a disk dies, auto mounter fails, it's full, or there is an open file handle still) and it's trying to write over somebody's files. From an application developer's perspective, the ability to access real time trace information would be helpful. It's not very helpful sometimes to just submit a ticket and wait for the gatekeeper admin to take a look at it. It is typically the case that the developer knows more about the gatekeeper software than the admin anyway. I guess what I'm looking for in terms of information is the ability to see the log files remotely, or some similar access through a web console or something.
- There are a lot of issues. For example "Globus Error 17", followed by some cryptic and non-meaningful words - when I tried to track the problem down, the trail leads to a log file - syslog, messages.log, and then it goes to another file. I need to track through all these things to find it because the gatekeepers don't remember. I can't query the service as an admin. There are no admin functionalities. There's no way to ask the service, "Hey, this job failed. Tell me where it went. Give me the attributes of that." It's not easy. It's not easy to debug things when things go wrong. And users really don't have a clue. They get back this thing that says "Error". The biggest problem is when somebody has an error and you need to track it down. I mean that's the hardest one. It would be great if a diagnostic tool or monitoring framework existed.
- I guess Condor-G is still a little bit of a black box to us, and that means we have to ask more questions. The other problem is when it doesn't work, we don't know whether it's the Condor layer, the Globus layer, or some other layer that's failing. It adds layers of complexity (maybe a little too strong a word) on top of the process you're trying to accomplish. Sometimes it works transparently. Other times it fails and you don't know why. The user sees a little bit of diagnostic information - not a lot. Condor and Globus log everything, so there's always a log file that has some kind of error report in it. But based on our experience, it doesn't tell the user how to fix the problem. Even if it did, he still wouldn't be able to fix things, because he's the user of the tool, not the developer.
- We need to get back to the intermediate files that were left by the application run. Each stage of the Grid process has a log file: for the file transfer there's a log file, for the execution section there's a log file. These are created when the job is run under the user's home directory system. We need access to that, so we can go back to where that particular job was run and dig into it a little bit.

## C.2.3 Communication

### ***Globus Developers***

- The other thing that is a little bit difficult: sometimes I think there's a reliance on the email lists for archiving information. And that's great, because sometimes the details really only exist in an email list and you want to be able to find them. But it can be hard. There are so many email lists I'm trying to monitor.
- It would be helpful if some of the campaign details again were in a more centralized place. And information that was exchanged through the email lists that's pertinent to the roadmap or the campaigns could end up in this other place.
- It would help us to know the assumptions that Globus developers are making on the various files. I'm referring to what are they doing with locking and where the state of the gatekeeper is living (for both Web services and pre-Web services.) I know the broad strokes, but we'll need to know a lot more detail when we do the redundancy work. We'll be emulating the service, and we need to know as much as much of nitty-gritty implementation details as we can. Right now we find these things out by trial and error. [prompt asking for more information on what it means to "emulate a service"] Take our work with our job forwarding as an example: What we did is we took a file

that lives down in a Globus library, condor.pm, and rewrote it to do what we wanted it to do. We're basically emulating the pre-Web services "2119/jobmanager-condor" interface [fragment of the conventional network address for remote Condor jobs]. If you send a job to that on our Grid, it's not really Condor underneath the covers; it is our own proprietary system.

- Sometimes when trying to work with various Globus developers, I get the feeling that there's nobody really in charge. Everybody seems to say, "Well, I don't know. Is that more important than this? Is that more important than that?" And the developers are very hesitant to commit to anything without talking to somebody else first. One week it'll sound like it will be a priority, and the next week the work will get bumped. From the outside perspective there doesn't seem to be a lot of cohesive direction and vision. It seems like a lot of firefighting and jumping around. All the Globus developers we've worked with have been great to work with. But every single time you ask, "Hey, can you add this?" they'll say, "Well, sure, but I've gotta find out if this is more important than that." I always get that response. I'm talking about tasks that take somewhere between a half-day and two days.
- The Globus team has gotten better at this, but there are still times where the team appears to be self-focused or focused inward. This doesn't apply across the whole team. But some folks seem to be focused on infrastructure for infrastructure's sake, as opposed to infrastructure for other people to build on. There are still some pockets of that occasionally. But that's certainly not the rule. As I think about it, the teams I've interacted closely with - the RLS and GridFTP teams - I can say that's the opposite. They tend to be very supportive in terms of reaching out, asking for use case scenarios and requirements, and being responsive to input.
- We're trying to be more deliberate about designing and thinking ahead, without going overboard, in terms of how the pieces fit together. I'd say that's going to be a larger component of what we do now. Going out and talking to the different middleware providers to understand what their roadmaps are, so we can try to get an insight into what things are going to look like one year, two years - even three years from now.

### **General Feedback Opportunities**

- The way the centers work all the information comes down and there's no feedback, this conversation notwithstanding, from the poor users at the end of this who are forced to use poorly designed and inadequately supported computers. And they suffer terribly in loss of scientific productivity dealing with the endless failures at every level of these systems.
- There is no feedback from the users to the center management or to the NSF, in terms of the cost in human resources in using these systems. The current round of the NSF program is a perfect example: this obsession with buying a petaflop computer for political reasons, presumably to brag about it internationally or something -- with
  - No input whatsoever from the userbase,
  - No clear understanding of how it possibly could be used, and
  - No input from the end-users as to its architecture, its characteristics, or what it will support.

### **Differences in Worldview**

- Another challenge is working with the domain science community and trying to understand their needs -- trying not only to advance their science but also to advance your own. Because one of the problems we face as computer scientists is that we are seen as the technicians for the domain scientists. The advancement of computer science is seen as secondary, as opposed to something that could be an equal partnership. Establishing this type of relationship takes a bit of education on both sides actually.
- Because it's old doesn't mean it should be ignored. Ninety percent of the science codes in a recent Oak Ridge survey were found to be written in Fortran, for example. No one in the computer science community can be bothered to help a Fortran programmer anymore. They probably don't even know Fortran. But in science it's still tremendously important, and C is the next one behind

that. We're not going to switch languages. I'm sorry to say that the DARPA HPCS language initiative is pie-in-the-sky.

- Generally I find it hard to imagine that the people who do these security services have ever done any large-scale computing project. For instance, I'm moving files from Pittsburgh to NCSA, so every time a job finishes (and these jobs run for over a year, one after another) I have to transfer a file. So the notion of typing in a password and doing that by hand is very annoying, compared to having it happen with some sort of automatic system.
- In some areas of science people use packages a lot and they're used to the idea of typing in some sort of GUI and hitting "go" and getting some sort of answer. We're about as far away from that as you can possibly get. It doesn't even make sense to me to consider such a thing. When the computation time is measured in months, any kind of traditional view of that just doesn't work. For example this one simulation that I'm running I started working on it in the last week of November and it's now June. It's not over yet. These things don't fit well with the Computer Science idea of running myriad little processes.
- Most of the services seem to do what I have been otherwise able to do for a decade or more (such as moving files with scp or ftp.) So I'm trying to understand where the value is added. Maybe UberFTP is able to move my file about twice as fast... I haven't yet tried it between Pittsburgh and NCSA.
- The people who are doing this ought to have some experience using these systems for large-scale projects. My feeling, perhaps out of ignorance, is that most of these tools that I've seen that are called Grid tools reproduce services we could do before. They seem to have complicated names and complicated protocols. And, regarding the security, the tools are not designed to run jobs that will run for months and months and months without too much user interaction.
- End-users can't distinguish between their domain science and your infrastructure: middleware, Grid logins, infrastructure, clusters, their particular science application. Maybe they've even been given a science application from someone they're working with. They can't distinguish among them - it's all "The Computer" to them.
- When I go to a new computer at the Pittsburgh Supercomputing Center I find they do an excellent job of organizing the information that I need to know in order to use that computer. In contrast, the Grid-related documentation for things like getting a certificate is organized into very small chunks. They weren't organized in the steps I wanted. As I mentioned, there are lots of different acronyms. The documentation seems to have lots of different paths because there are so many different ways of doing things. As a user I want one way that is going to work, and work easily.
- For many middleware developers, they often think that's the only thing left to do after you install their middleware. We have to gently point out that application porting is a mere starting point that occupies a small fraction of our time, because it is in fact so easy. It is by no means the end of the story. If we just have someone who has a computationally intensive application that needs a lot of CPU hours, then we hook him up with our Grid, TeraGrid, or OSG, and we're done. But very often there's a great deal of social interaction to be done. There's a great deal of organization building.
- One challenge is the mindset of "you users adjust to what we have, rather than we actually do something that you can use."
- Don't deploy a toy application of some fishes in a bowl, but demonstrate that you can really host a real application that is actually driven by users requirements. Such requirements might be true interaction with simulation tools, not batch processing - real interactive science. Then you'll experience what it takes to build a real application that serves not just one specialized user but a whole slew of different users. You'll find that these users don't have the ability or willingness to put in certificates left and right. They don't have the ability to rewrite front-end codes. Users are not as sophisticated as you think they might be.
- There is a distinction between tool developers and researchers: Given an arbitrary deadline, are you going to finish the paper for the conference that's due on Friday, or are you going to make

your system work reliably by Friday? I wish more people would answer, "I'm going to make my system work reliably"

- I think that I'm not a typical representative of the distributed computing community because I come from the HPC community. So I put a premium on performance, whereas the stuff that has been coming out of the distributed computing community has not focusing on performance. So that's one area where I would like to see some improvement. I would like to see those things perform well and with less bloat.

### **Intraproject**

- I'm a sysadmin so this may not be true of everyone, but I find meetings, collaborating with outsiders, getting everybody up to speed, exchanging docs to be very, very time consuming. So ideally I can have everything in my lab and have my students stop by my office to answer their questions, write something on the whiteboard, and have them start running it immediately. That saves me a lot of time. I'm comparing this to bringing different groups together, having weekly or biweekly meetings, exchanging our little limited views of each other's work, and trying to make sense of how we are going to put things together. That seems like a big, big time drain.
- I, like my a lot of my colleagues, spend entirely too much time in meetings, on telecoms, and answering email - and not nearly enough time being able to just sit down and solve the problems I need to solve. When you're in these collaborations there's just so many people you have to coordinate with and it just takes so much time; it can literally eat up a third of my time. It's not all bad, but there are days I wish I could just lock myself in my office and do nothing but write code, because I'd get a lot done.
- In a project as distributed as OSG is, you spend a lot of time in meetings and writing emails, just communicating issues back and forth. So I think that's my main problem. If you want to be a part of something, you have to invest a lot of time. But the big problem is how distributed the project is. There are so many sites, there are so many projects and experiments, and they all have slightly different agendas. So something that might be important to you, nobody else might care about (or the other way around). Or you're getting pushback from people on something that doesn't make sense to you.
- When you start coordinating with groups like OSG, it's pretty complicated, because they're such a big organization with so many different conference calls. It's hard to figure out. Also just trying to keep up with them in e-mail lists - some of these lists get hundreds of messages a week.
- There are many challenges. You know, one is maintaining an effective communication link between the project partners. That's a challenge because we have to coordinate the work and have regular meetings. The other challenge is getting the students to communicate effectively with one another. So it's primarily communication between the teams that's a major challenge.

## **C.2.4 Technology Adoption**

### **Support Burden**

- I solve the engineering problems, and I do everything else that's needed to support this lab. I'm very short on time. If I have a choice between using local CPUs to do work, or Globus CPUs controlled by somebody else to do the same work, I'll use my local CPUs. I know it'll take me an order of magnitude less time to set up something on my nodes to give the numbers that need to be computed than it would take for me to schedule the use of the Globus nodes on the TeraGrid (for instance). The setup work I'm referring to includes account setup for the students that need to access the resources, scheduling time to run our code on them, and installing prerequisite software. To use the remote resources I need to either
  - work with the admin of the remote sites to install things or
  - devise instructions for the students on how to compile the software in their home directories so they can be ran that way.
 Or I can just use my nodes and just get it over with much, much quicker.

- If we can get a lot of users going on our clusters with minimal interaction with user support staff, we should be able to do as well when getting them on the Grid.
- Monitoring our infrastructure decreases my productivity. We have a big academic course in the spring for example, where my research goals and development tasks have to be ignored because the burden of supporting the tools is great. What I mean by supporting the tools is trying to see what's happening on which resource.

### **Difficulty of Use**

- It's not like they can go to the website, download a Microsoft installer package, do a double click, and the software installs and you can start it. It's not working that way, and it will probably never be that way because you have to have credentials being created.
- The complexity and reliability of the tools we have to work with is a key problem for us. If you add up all the tools, you don't get a very good user environment out of them. It just still seems to be too hard for our users. So, for example, there's only been one person I've worked with so far that can really just figure out the stuff on his own. But he's really an exceptional kind of person this way.
- The management of credentials by users directly is too difficult. Our user base is not sophisticated enough to manage their credentials directly, so we are moving away from that approach. We'll be beginning to rely on MyProxy and similar types of credential repositories so users don't have to manage their PKI credentials themselves.
- The VDT is very helpful as far as getting things deployed much easier than in the past. But then after that, trying to get users working with those deployed tools is still a problem, and takes a lot of our time to help users. So ease of use is still a big problem.
- Globus is really cool in terms of being on the forefront. But sometimes it is a little harder for people to use. One issue is getting end users used to using certs. Why use certs if we can get a proxy cert from MyProxy with a username/password? The Globus security level that everyone uses is actually more secure than your bank - more secure than your credit card. Why? We're just making it harder. Why are science gateways so successful? Because they hide the complexity of the security. You can create an account and submit a job. Not any job ... ah, that's the key. I think issue of security should be posed in terms of levels thus the complexity of the security mechanism can match the needs. If we want a wider scope of users, then we are going to have to make it easier for them. Sometimes it makes sense to increase the security level on access depending on what the user is trying to do.
- It's one of my fundamental beliefs that tools are what make software. This is not always a popular belief because developing tools takes additional time and resources, but the benefits are that the developer/user base will increase if the tools make the underlying system easier to use. Tools can help by removing the need to manually develop core framework pieces and provide a way to architecturally institutionalize the software development process.
- One of the main reason I used Globus components is that I don't have other options on the Grid. I can't just install any other technology. But on the server side components that I control, I tend to use technologies that are natively installed and are easy for users. One thing that I've learned in my years in industry is that if it's not easy, then it won't get done. I think that holds true for the infrastructure that we are building.
- One thing that I should like to emphasize is that the production worthiness of the infrastructure is of utmost importance to us: Make sure the software not only has quality attributes like performance or maintainability but also has quality attributes such as usability and the ability to operate the infrastructure. This implies a need to provide all sorts of bells and whistles all around the software, such as the ability of doing monitoring and operational tools to manage administrative sides of the services. Not having this means we must provide such services around the software to make it usable by our users.

## Cost of Use

- I would not consider installing it myself. I don't like the overhead. When anything goes wrong with your certificates, site certificates - anything like that - it's completely beyond the scope of anything a user can do.
- GridFTP carries all the baggage of Globus with it, but it's the only component we're interested in. Really it's just an FTP program - why on earth do we have to bother with all the certificates and all the stuff that goes with it? All we want is point-to-point transfer to be fast and reliable.
- If I could be convinced that a non-GSI version of GridFTP was stable and secure, I'd use it. I hate GSI. It's very good at what it does, but it is a pain. When I was involved in GridFTP development, GridFTP didn't have problems - GSI had problems. Once I could get people past the GSI issues and get it all configured, GridFTP just runs.
- One time I had an expired Grid certificate, and at NCSA it was quite easy to generate a new Grid certificate, but only because I had taken a (paper) file folder from my office, which normally I would not have with me, that has my default password (because I'm traveling right now.) So if I hadn't decided at the last minute to take this file, I would not have been able to get a new certificate.
- Running a CA, deciding whom you trust - that's all a large pain - a very large pain. For example, you have to get a CA certified by TAGPMA and buy special hardware. And to be blunt: after all this is done, as a user we don't gain much of anything. No additional capabilities - you can access the same machines as you could before. You know, it's a big hassle for some potential benefits, like delegation, having your own agents out there to do things for you. There is some potential there, but it just hasn't come to pass that we've needed it.
- If we were to start using all your RPCs, your secure MPI, and the secure-wrapped standard libraries that have been modified for Globus use, if we were to use all that plus GridFTP, sharing a common authentication infrastructure, then the burden of all these additional layers and centralized key distribution would seem worthwhile. But if we have eight nodes, and we just want to launch our scripts remotely, that seems like a lot of work.
- The time burdens associated with setting up the application-specific environment on the remote machine is a big challenge. The University of Chicago Grid experts handled the Globus-specific setup, so I can't comment on challenges associated with that.
- There are times when I need to connect to the Grid elements directly. This is when I used the Globus components to either submit a job or transfer data. I consider Globus a pretty heavy-weight tool for the most part. It's not something that I would tend to use without a strong requirement to do so.
- I took a Globus Toolkit 3 workshop while I was working at NASA JPL and very much interested in learning about this technology. That workshop opened my eyes that I will not clutter up my application with all the requirements that Globus imposes on my application. I talked with a person that was teaching the course, asking, "Have you ever considered talking to end users of your framework?" The answer was, "No, we haven't done that." I think that was 2003. It was very eye opening. One person in that class actually said I should be rewriting my application in Java. I have a 200,000-line application written over many person-years. I'm not going to rewrite that in Java.
- Globus GSI-FTP and others need to do the authentication through using the certificates. They tend to take a longer time. I think people might say six seconds is not a big latency, but when you have many interactions, six seconds adds a lot to that. We don't say six seconds is not bad at the MPI level. I understand there are technology challenges, but I think there should be less cumbersome methods for authenticating the requests. I don't have high hopes for certificates. They may provide better security, but they bring the performance down a lot. So I go with the idea that maybe for eight or ten users, we can integrate the ssh keys into the tool itself. This is a focus of the Java CoG development work on our project: to provide the ability to import the user's public key and use this to launch the jobs.



## Infrastructure Difficulties

- I find it very difficult to figure out how to register these certificates at different sites, because I have a different user name at NCSA and at Pittsburgh and at SDSC. So first of all I found it difficult to find the right place to start looking for documentation about how to get my certificate registered at a new site. I found it easy to google and figure out how to get a certificate. But then to get the Distinguished Name registered and hooked up to each individual account took me a long time to find the right place to start looking for documentation to do that. And then once I found the documentation, some instructions said to use gx-map - other places said to use gx-request. In almost every case, neither one was on my path, and I had to hunt for probably thirty minutes before finding it so I could actually use it.
- I find that I often don't have the right commands by default in my path.
- Some of the large bioinformatics applications, like in CAMERA [Community Cyberinfrastructure for Advanced Marine Microbial Ecology Research and Analysis] projects have amounts of data that are 100 times larger than what we deal with, and they have no idea how they will deal with that. We were trying to download at least small chunks of their data, and it was taking hours and hours. Sometimes we can use GridFTP, and sometimes you can't - because in some cases we don't have a GridFTP client available to us.
- Once I get a transfer to Oak Ridge running right it's pretty much fixed, except for congestion on the network (like somebody kicking off a big job at the same time I did) - that is, until the next time they do a kernel upgrade and blow away all the modifications you made.
- One of the big things keeping GRAM2 alive is interoperability with European experiments. They are not going to GRAM4, as far as I know.
- Security is out of my control and requires bizarre and completely Byzantine communication between centers. For example, try using an NSF certificate at a DOE site: Dead on arrival.
- There are no obviously robust methods that we use to help us get around this. Now, I believe there's a thing called Reliable File Transfer (RFT) that might help. But again, we don't just use NSF TeraGrid, we use DOE centers as well, and it's not clear to me that they would implement anything like that.
- Typically it's difficult to go to a regular Globus site and just run our code because we rely on so many external libraries and tools that are not part of the Globus standard install. Sometimes it's possible to request these things to be installed at the site, sometimes not. If it's not, a lot of these tools can be compiled and installed in a home directory and run from there, as opposed to assuming the system has them.
- UberFTP: the major challenge I faced is the first time I tried to transfer a file, it only transferred one tenth of it and then it stopped. And in general it's not always clear who to ask for help because it's always a transfer between two different sites. So you have to get both sides involved, which can sometimes be difficult. Sometimes they don't communicate with each other - they'll only communicate with me. They may have different help systems.
- What I find annoying is there are so many authentication schemes. For instance in our DOE SciDAC-funded project we have computers at Jefferson Laboratory, Brookhaven National Laboratory and Fermilab. Everyone uses a different security system.
- When it comes to GridFTP and moving things in and out: We only control one end of the transfer. We can make sure the machines on our end are beefy enough and are configured correctly and tuned right. But if the guy is trying to transfer the other end off his laptop, we'll only go as fast as his laptop. Data transfer is an interesting problem in that respect. It's a two-ended problem. If you're trying to schedule the transfer, it requires co-scheduling - you must schedule resources at both ends. You don't have control over your own destiny: you can control your end and you can coach the other end. But if they don't have the hardware there's nothing you can do. And that actually gets quite frustrating. While I know that rationally they understand it, all the user knows is he's not getting what he wants.

- If you email them about GRAM2, they will be very responsive, but if you email them about GRAM4, their response is only best effort. I think it is this way right now because GRAM2 is the production version for OSG. If GRAM2 is failing for them, the site is considered to be failing. If GRAM4 is failing, it is not that big of a deal for most people.
- It's not been solvable in a general way. Metadata very quickly becomes very application specific. And most scientists have perhaps not as good a system as they would like, but they have a system of some kind that they already use for tracking metadata. So we don't provide a standard system service that could be called a metadata service.
- Keep GRAM2 around in addition to GRAM4, especially in the Open Science Grid at large. The Open Science Grid doesn't have an information system to reasonably send GRAM4 stuff around. Our whole information system right now is tied to GRAM2.
- The file transfer I was doing from NCSA to Fermilab is going to a special tape archive at Fermilab that's managed by dCache. And so to make this work I have to use an SRM-copy. And the SRM-copy was failing. And the reason it was failing is that Fermilab has to set up certain map files to make that work, and those are not being properly maintained. Maybe it's that not enough people are doing this kind of thing, so these things are not being maintained to the point that it makes it easy to rely on whether or not it's going to work. So then finally just fall back to the old FTP again, and that works, sort of. But in order to get the files onto the tape archive at Fermilab the scp has to go through two stages. You have to move files from a disk to a disk, then you have to move them from the disk to the tape. So that's a painful process and doubles the amount of work
- There are the hardware problems: dealing with the sluggishness on some of the networks like the ESNET, which we've had some problems with recently. The file transmission rates are painfully slow, errors occur, and then we have to retransmit.
- Users of my Grid have to negotiate separately with each site they want to use. So that means they need to contact the person at each site who has the power to authorize them. This is hard at a lot of the sites because they're set up to serve their local campus users. It's easier for TACC because being a TeraGrid site, TACC can say, "Sure we'll give you a little starter account, and go through TeraGrid to get more cycles."
- Sites seem to have GRAM2 installed, and it's working. I don't have a problem with that. And unless we use the Falkon component of Swift, we don't really need GRAM4. When we use the Falkon component of Swift, GRAM4 is required. And that restricts us to a subset of Grid sites. So I prefer GRAM2.
- The only problem that I've run into in my limited use of GRAM was the varying functionality on different deployments. I don't think that's really a criticism of GRAM, but the specific case was on the TeraGrid. You could specify that you want nodes of a certain type. The types of nodes I wanted were visualization nodes, and there was a way on the TeraGrid version of GRAM to specify those nodes. So I wanted to submit the job to the TeraGrid visualization nodes from the CoG Kit on the Windows desktop. But I was using the regular CoG Kit on Windows; it wasn't the TeraGrid-enhanced version. So it didn't understand those extensions. That was a frustration. I thought I had struck gold when I was able to submit a job from my Windows desktop, but then I hit that limitation and was disappointed. Maybe those extensions have since been rolled into Globus proper. But I was told at the time was that there are as many different flavors of Globus as there are flavors of Linux.
- The challenge is the different versions deployed at the Grid locations ... understanding what you should use. For example, I experienced problems between GT versions 4.0.1 and 4.0.3. It was in the job descriptor - it was a serialization bug. The symptom was, "I cannot deserialize this", basically. I immediately understood the problem. These things are compiled stubs and the other end wasn't recompiled to match. It was probably a bug that was fixed in one place but not the other. I didn't dig much deeper beyond saying, "Oh, A works, B doesn't, go with A." It's a problem. It's a cross-version compatibility problem, and that is an issue in the Grid world. That's why I just follow whatever is working on the gatekeeper right now instead of using the new

features. If we built roads in the same way we use infrastructure and build our current IT infrastructures, our roads would be very scary. We would have bridges where we would drive off.

- Constrained networks are a problem for us. If participants are on a low-bandwidth link in the middle of nowhere, it is problematic both for them and for their collaborators. Their networks would be incomparable, some of them on very fast networks and others on very low bandwidth networks.
- One of the things that I was thinking about is a BitTorrent because I'm interested in getting data back to the client. I don't care about moving it from one high-performance system to another. I want to move a terabyte of data back to the client - or have the choice to move it someplace else - but I want that choice. If I'm on my local cable modem connection and I'm downloading stuff, it's going to take a while. Not everybody has gigabit networks - that's the thing we have to realize. There are still universities out here that have 10 mbit connections, and these other technologies cropping up in the consumer world are handling that. Do you have a gigabit in line at home? Yet you can run Skype at home. I can Skype over wireless, which is kind of interesting. So yeah, it is more user focused.
- We haven't shared software with others because the site selector is really specific to our architecture. It is difficult to make it more generic because it depends on things like Condor queue and VDS for site lists, which is not in standard use by everyone. People don't use Condor all the time, and very few people use VDS. Let's say we had a daemon running on each of the remote sites, inspecting the queues. Further, let's say that a given daemon saw that a hundred nodes were free, and reported that back to GNARE. But this still didn't guarantee that the jobs will be run. The hundred nodes might be reserved for somebody else.
- Some sites had restrictions based on the VO; some sites would only allow N jobs at a time to be run by our VO. And some sites were dedicated to supporting a specific VO so they would restrict our jobs. So even when the daemon saw free nodes, our jobs might just sit in the queue forever.
- One difficult part was that we have a huge number of sites, both from OSG and the six or seven TeraGrid sites. We had access to so many sites, but then the challenge was to do the selection across such a large pool. If I have seventy different sites to choose from, then I need to understand
  - where to submit the job,
  - how to know which sites are available,
  - how to know that a site has failed,
  - what should happen if a site fails, and
  - how to know that a site will be going down.

When we started working on this we didn't have any information services built-in to these Grids. So there was no GridCat. There was a version of GridCat, but it wasn't up-to-date.

- There is a problem related to the diversity of systems that we run on, in particular on TeraGrid. We can build a tool on one site, and it'll only run on that site. It doesn't run anywhere else. And this is especially true of the parallel tools. Diversity is kind of a double-edged sword. You may find that an application runs really well on a particular type of architecture and not so well on others. In one sense diversity is a good thing, but the flip side is you have to be able to develop for all. So for us that means logging into all of them, re-porting and building code, and maintaining the application across all those platforms. That's one of those barriers that will stop the casual user.
- We sometimes encounter older software versions on TeraGrid. Unlike some TeraGrid users, we generally want the latest versions. But the latest updates are not always available, so sometimes we need to install the software ourselves in user directories. We've encountered this in the past with MPI2 and Java.
- Another problem is that sometimes we even have to investigate what is the best strategy for communication. We are using file-based communication and there are more than a few methods to transfer files between sites. We need to investigate which one works for us, because we tend to have smaller files, so latency rather than bandwidth is an issue for us. Even on TeraGrid we have tgc, gsi-ftp, scp, and a couple of other file transfer mechanisms. So we have to pick and choose which mechanism works best between the sites that are available to us. And I don't if there is an

easy way out from that. In some places scp will be the best way, and in other places globus-url-copy will be the best. GSI-FTP seems to work well between ANL's TeraGrid site and NCSA's TeraGrid site. Performance varies though, even at different times in the day. For some trials we are able to get better performance just using scp instead of GSI-FTP and other commands. It just depends on the load on the Globus service, I guess. So we had to model even that, and that's an extraordinary level of information that we don't need to deal with.

- Incompatibility between the sites is the biggest challenge. Whatever we do, we have to make sure that it works on every site. So we have to do the manual listing ourselves. So if there were a compatibility layer that ensures the resource allocation mechanisms work on all sites as expected, that would eliminate a lot of testing on our part. The MPI2 functionality, like connection sockets and dynamic process management and spawning new processes: this functionality would be useful to us. But right now it is not available on many TeraGrid sites.
- For security reasons, no compute node is allowed to communicate with outside machines as far as I know. The issue with security is that we are not able to directly communicate to a running a process due to security concerns. In order for us to send in the next set of parameters to the simulation component, we have to copy a file from the optimization framework through there. We can't connect directly through sockets. No compute node is allowed to communicate with outside nodes. So the bad performance I alluded to earlier is due to having to transmit data via files. We should be able to stream the information directly to the compute node rather than resorting to writing our stuff to a file, copying the file, and then redoing this over and over. And though they are small files, the latency itself is a killer for us. So it's more that the performance is bad because of the file-based communication, not having to do with the technology itself.
- The problems I experienced with the Globus job submission mechanism [GRAM4] happened around a year ago, so some of the information may be dated. But after that experience we are waiting for the word from the CoG Kit group to give us the go ahead to try Globus again. But at that time they told us they were not able to submit the jobs properly and were having certificate issues and so forth. So when we talked to the CoG Kit folks, they say that the TeraGrid deployment schedule is delayed so we need to wait another six months before that stuff becomes available on all sites. So they were acknowledging the incompatibilities at that point.
- We had to resort to this file-based communication because we can't directly communicate through a running job. We would like to stream in new input into a running process. My application really needs that. The simulation component produces results that feed into the optimization framework. We had to resort to using files to communicate between the two because we can't directly communicate with the running process. We are exploring ways to eliminate file-based communication wherever possible.
- The scriptable interfaces at various sites are not consistent, so our scripts to interact with the resources need to be site specific. If you want to get the backfill resources available at site A, you have to have a special script to talk with it. So one challenge is that the scriptable interfaces to the queuing systems at various sites are incompatible. And the Globus Toolkit functionality is supposed to hide that, but we've still encountered incompatibilities. This was six to eight months ago that we had to write our own custom scripts to do this task. At that time we tried using Globus job management services at various sites. But the word from one of our collaborators within Globus at that time was that the deployment scenario for certain Globus components on TeraGrid was not on track. So, at that point, we had to resort to writing custom scripts. The scripts went to particular schedulers and not the Globus gateways. And they would just state the input files and then launch the jobs using parameters that would minimize queue wait time.

### ***Lack of Knowledge***

- I think what really needs to happen is to give training courses to hospital IT- on what is Grid technology in general and what is a concrete implementation of it. That would really help because then we wouldn't have to repeat these discussions with every single institution when we do a deployment. But my gut feeling is that it may be a little too early because this whole field still has

to mature - not only in the Grid domain, but especially the interaction with the clinical hospital domain and the overall healthcare enterprise.

- I would also find more tutorial-like information quite useful. For example, I read the whole Globus Toolkit 4: Programming Java Services book, and I practiced a lot of examples in there. This is kind of helpful, and we would like more examples. Like when we are writing clients to a GRAM service, we look for more tutorials or even CoG help in some sense. So more tutorials would be helpful.
- It is difficult to figure out where to find the right documentation. Once I do find the documentation it's very hard to understand - it's full of acronyms and refers to unfamiliar and unnatural concepts.
- It seems to me that in order to get Grid solutions, you have to be pretty tech savvy. Getting the certificates, doing the job submission, doing the DAG of the workflows on Condor, managing the security: all of that seems to be an enormous barrier for actually getting jobs done on the Grid. It didn't seem to me like there is any mechanism on the TeraGrid or OSG to sit down and work out how to solve problems together.
- People write great developer guides. And that's great for somebody who is a developer. But what about the rest of us? What the hell does this thing do? Even with GridFTP we've tried, but people sometimes just don't get the big picture of GridFTP - concepts as simple as "What's a client? What's a server? What's a third party transfer?"
- The learning curve for some of the software is quite hard. It takes me five months to train people to use some of our software. That's a long time. But I'm not sure it's related to the software/technology. I think it is that some of the concepts are difficult.
- They say, "Oh good. You're here. You can do this now." And you have to say, "No! That's not what we're here for. We're happy to help, but you have to keep doing your science." The group is populated by a mix:
  - people who are just getting started
  - people who haven't figured it out yet
  - people who really don't know how to use the tools
  - the senior professor who is now going to ask you from this day on how to log on to his email (I kid you not!)

So more than half of the time when we try to work with people who ostensibly are researchers and scientists in their field, they say, "You're so much more qualified than I am to do this that it's hopeless for me to do more."
- We're working to convince the radiologists that this is a good paradigm that is beneficial to them. And the radiologists are not technically at a level where they understand what the underpinning is, but that's not really necessary. So the approach we've taken is kind of the soft approach of learning by doing. For the sites, if they see "Oh, this is working", or the security model is gaining trust, then you build confidence. After building confidence you can go to the next step.
- We've certainly hit challenges with the road to Web services: writing Java submission scripts that can be submitted through Web services. The XML - that's the place we sit down with the user. We'll just do it with them because many people look at XML and, well, it doesn't fit their worldview. But it's utterly trivial to sit and work with them, saying, "This is how you specify the batch queue."
- With Globus there's lots of information that you need to get the Grid certificates:
  - \* Knowing which one is the best one to get
  - \* Knowing how you use those certificates to authenticate
  - \* If you've gotten one from somewhere, how you get to another place and get authenticated there
- If someone were trying to do this who didn't know the area that well, it would be kind of tough. Since for example, compare it to something like Linux. You can take one of a number of distros and then kind of do all your shopping there. There's a bit of that going on in the Grid area, but I think not everything needed is covered in them yet. So for example metascheduling: we couldn't go to a distro like VDT and pluck it out of there. It's not in there yet. It's not mature enough. There

hasn't been a consensus yet on what is the best one. So this is where it's a good thing that a number of us on the project know the area. So we know the providers, contact them, and work with them directly.

- The lack of technologically trained neuroscientists. I've been trying to hire a postdoctoral fellow who's a neuroscientist to do computer modeling for two years. So one technologic obstacle is training more computationally sophisticated biologists.
- The technology challenge that I face is the learning curve involved in using different software.
- There are still not enough people in the world, as far as I'm concerned, that have real in-depth Globus knowledge. And certainly they're hard to find and hire. So we train people up here, to the best that we can. But it's still hard to say to someone: "I'm thinking about putting RFT into this service, but I need to understand where it's going to break. I need you to stand up an RFT, and throw larger and larger requests at it until it breaks." Or "I want you to throw larger and larger requests at it and tell me the load and memory footprint on the machine." I could do that with my staff, but I usually end up getting it kick-started and spending a little more management time than is optimal. That is not a comment on my staff because they're all good, hardworking, smart people. They just don't have some of the expertise, especially with of the Web services stuff now in Globus Toolkit 4.
- There was a case in which I was creating configurations at Pittsburgh and doing analysis at IU. This is a perfect example for distributed computing: where you run a job at one place, and you put the output somewhere else and run some subsequent step of the job. I could do all that just fine with ssh, using the network, queuing a job at IU. And nobody ever picked up the challenge of trying to do that with Grid commands. All I needed was a little background script running at Pittsburgh that I could understand quite well.
- I would say the general challenge is the fact that I'm not fully in control of the environment where I have to do the work. I don't have the tools or I don't have the knowledge to fully figure out what's going on when something goes wrong. Most of the time the showstopper for me is the information that needs to be discovered. I sometimes encounter problems that cannot be solved without finding additional information. I may have a vague idea, but somehow I need to figure out the best way to solve the problem. For example, I may not be fully familiar with the queuing systems (at least on the Grid side). Or I may not be familiar with some Grid environment variable settings that I should set.
- More documentation is needed. Most of the time the main pages and documentation are good but some applications lack it, so it's a general thing.
- The bandwidth speed of NSF reads and writes is still an issue. We don't have much experience with parallel filesystems. That is one area we will be experimenting in future to see if it solves our problem. For our new clusters we are going to experiment with PVFS and Lustre.
- Go to MATLAB's website, MathWorks, and look in their toolboxes. That would be an adequate level, where every aspect of their product has an example. MATLAB has all the API documentation just as Javadoc generates, but they also have examples. Globus gives you the API, but without the context of how to use them in more than just the trivial examples - it needs to be a little bit more than that. One of the other things I use a lot is Java Almanac. That moved to Example Depot [[www.exampledepot.com](http://www.exampledepot.com)]. Basically it is a repository of how to use various different APIs within Java. You could go in there and examine examples of a package like `javax.imageio`. Enough to get you over the hump of getting started - that's usually the problem. The examples are compilable. They're usually small, not significant.

## System Integration

- "Keep it simple" would be the only real advice I would have. You know - the KISS principle. Users these days have got an unbelievable amount of extra work to do compared to the supercomputing programs twenty years ago, when all you needed was a Fortran compiler and a Cray XMP and you were absolutely the best in the world. The complexity of it now is so great that

I see it breaking down. It isn't worth our time to consider adding more sophistication, because if we've got any spare time or any spare brain cells, adding sophistication within our code in terms of things that are domain-specific to us.

- An important consideration is which tools are extensible and allow me to build on top of them. This is in contrast to tools that try to provide a complete solution that force me to rip and replace stuff. I try to stay away from the rip-and-replace tools because we just can't do that. Such tools offer solutions but require me to give up other stuff that I'm already doing in order to use them.
- Educating my home institution about the Grid infrastructure itself. I spend a fair amount of my time making sure that things we need to implement to make Grids work aren't going to get tripped over by the folks who do security. This is, of course, a very common theme. We spend a lot of our time making sure there's adequate communication there so that nobody shuts down my Grid servers because they don't have username/passwords expiring every ninety days.
- I can look at a class of tools that do certain things - purporting to do something, for example "manage data transfers" or "manage workflows". I then try to decide if the tool offers bright shiny new functionality but will at the same time be unstable and I won't be able to rely on it.
- I have mixed feelings about Globus as it is. It forces the user to implement their code in some very specific ways. So there's a certain mindset that you have to work with. You cannot just take your code and just pop it in there if you really want to take advantage of Globus. Otherwise, you're just putting your own code on Globus using your own socket code, disregarding Globus security, and you're just using Globus to schedule things and gain access to machines. So unless you do it the Globus way, you're not really utilizing it.
- I look at is the interfaces. I'm going to usually have to put some glue in place to pull these pieces together in reasonable ways, and how easy is it going to be for me to do that gluing? Do I have an API that's only supported in one language? If it's not my first choice for the project, I'll need to extend outside of our area of expertise. Or is it something that's technology or API agnostic and I can easily just write whatever I want?
- If there will be some continuity between the releases that would be helpful. Ideally we would not need to rebuild everything in our system to accommodate the new changes. It would be really good to somehow lighten the burden of transitions to new releases.
- It's wonderful when new things are developed, but every time there is a new tool available, it means that in awhile we will need to rewrite the whole system to accommodate the new release. And this can be a problem. I understand that new technologies are being developed and that's why they are getting better and better. But it's a little hard on the application developers when new versions are not compatible with the other parts of the system.
- Most of our applications actually don't code against any APIs. They need to have the environment and security managed for them. So I can't go to a project data analyst and say, "I need you to link with this library so that your tools will interact properly with the security." They expect the infrastructure to operate at a level either above or below that, depending on how you characterize it. They just want to run their job, and they want everything to be handled for them.
- There are a lot of different bioinformatics tools and currently we are mostly installing them locally to run them. Sometimes we are submitting it on the network, but probably we just need somehow a more developed system in bioinformatics for Web services. So it should be probably Web-service based, the whole infrastructure. Note that we don't have any distributed algorithms. All of the data and all of the parallelization we are doing is embarrassingly parallel.
- To me, Globus is a set of daemons and infrastructure that
  - provides a unified security mechanism with cryptography, key exchange, and authentication on each service using a common set of keys,
  - provides a uniform remote procedure call interface,
  - provides some file transfer protocols using multiple underlying network protocols,
  - has some scheduling capabilities (I guess limited to per node scheduling), and
  - contains a set of standard libraries of tools that one can rely on being available on Globus nodes,

In order to start doing something outside of the Globus-provided services but staying within the Globus security network, one has to learn additional APIs and how to code things up. So it seems like there's a high startup cost to use Globus. To me it's cheaper to put 20 CPUs behind a firewall and a private network with no way in except through some gateway node that's well secured. I can then just run whatever I want behind that firewall using the most approachable, easiest to use toolkits with the least overhead and with least restrictions on how we code things up. I know there are many people who truly want to distribute their processing across multiple data centers. To them security will be more important. But as long as our project will fit within our local cluster that we can handle ourselves in the back of our lab, it's too much additional work to do it the Globus way.

- We use a lot of libraries and toolkits like R, which typically are not included in a standard Globus install. So we can't rely on them being on other Globus clusters. So we're shipping our own precompiled code that maybe has R installed in the home directory. We do things like that to get our code running, but in the process we're missing the point of Globus.
- What kind of logging does it have? Is this a tool I'll be able to drill down easily when I think there's something going wrong? Can I turn up the logging levels so I can really get a picture of what's going on?
- From the security realm, there are a number of solutions based on proprietary tools. Some people are interested in those because they seem to offer to the users a better experience. I use the term "seem" because I'm not convinced that they actually offer a better experience for the user. But the problem with these integrated solutions is then on the backend there's no choice about how to hook them in to other systems and services.
- The attempts from Globus-related teams (I don't think these are Globus Toolkit proper) to provide tools and infrastructure to help with metadata and provenance have not scaled. And especially in terms of provenance, they've required too many application-level changes. The approach was, "Just do everything this way then you'll get the provenance information." But there's no way to "just do it this way". That's not the way my users can be approached. They are going to do their science. The science is going to lead, and all the other stuff has to be tacked on.
- The things I need to do from a syntax perspective are completely different between GRAM2 and GRAM4 and require a rewrite of my stuff. And the RSL versus the XML is completely different. All that stuff is completely different - but functionally, no.
- Another problem has to do with software dependencies. When you leverage a technology, you need to look at its dependencies and compare it with your own to see if they clash. Java WS Core has a very large set of dependencies. This is not an issue for GridShib for GT, which is a plug-in for GT and sits on top of Java WS Core, because it was built from the ground up to work with Globus Toolkit. But one of our standalone components, called GridShib SAML Tools, has its own set of dependencies because it has its own standalone code base. At one point I was asked to investigate incorporating it into GT. The idea was to have it deploy into the GT codebase in the same way that GridShib for GT deploys into GT. This work is still not finished because I've not yet figured out how to reconcile the dependencies.
- There are language requirements, first of all. If a library only exists in C++ and you're developing in Java, that's a mismatch. And there are also compatibility issues in terms of what version of Java is required. That's always a question.
- There can be conflicting dependencies. When you look at somebody else's open-source software, they have a set of dependencies and you have a set of dependencies. The first question you have to ask is, "Are there any major conflicts in terms of those two sets of software dependencies?" Because if there are, you need to resolve those conflicts before you can even begin to leverage the open-source package.
- In my experience the most difficult part has been to connect the user interface to the component that generates jobs and submits them to the Grid resources. That was the most technically challenging part of the project. The "dynamic" aspect of the system is in the resource selection



logic, which dynamically selects resources from OSG or TeraGrid to run the analyses. This feature has been a big challenge to implement. We had problems in part because there was no existing resource reservation system we could use.

- For projects that are at the forefront of the technologies (like authorization interoperability) we have the challenge that some of the standards we're planning to adopt (such as XACML) do not have a stable and accepted implementation. So, currently for example, there are two implementations of the libraries, one is by OpenSAML community, and the other one is by the Globus team. And both are noncomplete; both try to address the same issues. There are different tweaks that the different groups do to the specifications in order to be able to implement things. And so there is always this question of what implementation should we use.
- From the TeraGrid side we find there is sometimes a need to write our own custom scripts. To my knowledge there's no TeraGrid resource query tool that provides us with sufficient information to build adaptive behavior into our framework. So we have to write our own scripts to do that.

### **Cultural Barriers**

- Sometimes there are terribly, terribly intrinsic issues to deal with. For example in the petroleum engineering field we find they have extremely powerful, well-developed expensive codes that the providers are happy to give you almost free academic licenses for. Really shocking how open they are with their code - you can download it almost like you would a piece of shareware. Extensively developed code. But then if you turn around to a particular researcher and say, "Ok, let's put this on the Grid." You find that they stop like a mule at a door because they won't let go of their data. It's the data that's important in that field. They are highly proprietary, having to do with detailed field measurements of oil-bearing strata. They are absolutely unwilling to let that part from what their perception of what a secure space is. So we have to spend a lot of our time working with them to assure them about data security and implementing tools to make sure that they always feel in control
- The same people who are probably logging on with cleartext passwords to POP email accounts react with great skepticism when you approach them with an absolutely locked down X.509-secured, strong cryptography solution for controlling access to their data.
- There's tremendous chaos in the identity management area. Everybody thinks they're in charge of identity management. Everybody! It's like when I first started teaching, I went home and told my wife, "Everyone thinks they're my boss: students, the dean, my funding agency." The problem is that none of them are wrong. Certainly your university thinks they're in control of all of the computer identities associated with you. The Virtual Organizations that you work with all want control. EDUCAUSE and Internet2 think they've got a good scheme. TeraGrid has its own thing and they want to be able to decide who in your university can log on to their resource and they're not interested in your opinion about it.
- Whenever I mention proxy certificates outside the Globus community I get strange looks from people. In fact I've gotten negative remarks. There are people in the Internet2 community who just do not subscribe to proxy certificates, even though they're well defined in an RFC. They just don't buy it.

### **C.2.5 Specific Technology Issues**

#### **C WS Core**

- C WS Core: The examples and documentation: I know they're working on that, and it will get better. But right now there's not as much documentation and not enough examples.

#### **GRAM**

- A lot of times a cluster user will modify their .profile [file holding unix environment settings] to set their environment for their jobs. They want those values to be used for the job via GRAM4, but

they aren't. In contrast, if they submit the job directly to the scheduler those values will be picked up.

- Both GRAM2 and GRAM4 are lacking in the same thing, and that's the ability to do co-scheduling. That's the biggest problem for us. Both GRAM2 and GRAM4 are great for saying I need 10 nodes on that machine, and I want you to run this application when you get them. And I don't want to worry about specific scheduler syntax. I don't care. I'm just going to specify the job in XML and let GRAM talk to the scheduler for me. GRAM is great at that, negotiating to put you in the queue, notifying you when the job is running, etc. That's perfect. But we don't run jobs like that. None of our MPICH-G2 jobs run like that, meaning on a single machine. All of our MPICH-G2 jobs necessarily run on two or more machines. It is imperative that the jobs are co-scheduled: that each job is launched is launched near the same time. It does us no good, in fact in some cases it does us harm, if one job actually gets through the queue executes on machine A, and then two hours later the second job gets through the queue and begins running on machine B. It doesn't do us any good. We need to make sure that they both get through the queue and hit the nodes at the same time.
- Documentation could always be easier to find. In the effort of deprecating GRAM2, the Globus documentation has been made very hard to find - at least it was the last time I looked a few months ago; I haven't even checked recently.
- GRAM4 is a huge resource hog. It takes 700 megabytes of memory to sit there and do nothing.
- The biggest concern for GRAM4, however, is the GRAM container goes into hibernation or stops for a while without any log messages. And it just comes back by itself after a few hours.
- The challenge that we have to solve eventually is try to figure out what the GRAM4 analogue of the GRAM2 forwarder will look like. How are we going to implement in GRAM4 what we've done for GRAM2 for our Grid:
  - Will we just put GRAM4 in front and keep GRAM2 in the back?
  - Will we try to do a GRAM4-to-GRAM4 thing?
- The other thing is Globus' nasty habit of (at least one time in three, and sometimes more) deleting the file you would like to see before you can get at it. This is with regard to debugging GRAM2.
- There are also issues that we have with GRAM, be it 2 or 4, with regard to job auditing. It always takes investigation into at least three log files to get a full picture of what has happened with a job. Not all of the authentication information is in the right place always, etc. There could be more information.
- We just went through an issue that turned out not to be a GRAM4 issue, but a Condor-G issue. It took us two or three weeks to debug that and identify the problem. It turned out that some authentications and authorizations didn't play nice together. Also, Condor-G was making calls when it ought not to (or not making calls when it should). So one GRAM4-related challenge would be working with the external callouts that are common in OSG.
- A lot of the RSL attributes that are defined, if you can find them on the webpage, are not implemented in the backend scripts. So, for example, we just added some memory support into ours here. Given that our nodes are multicore, we need to allow our users to say, "I need this many processes with this much memory per core." So that results in us putting one process per core per node, or perhaps one process per two cores per node, depending on the amount of memory they need. So that wasn't a big deal, but it was something we had to add in recently. The LSF.pm scripts did not include support for taking the min memory XML-based RSL attribute and turning it into the right LSF line in the submit script.
- Given our stakeholders it's unlikely that we'll be rid of GRAM2 any time in the LHC era. I expect we'll have to keep it going for at least five years, maybe more.
- There is an issue when the GRAM4 state thing is mounted on a shared file system. This could really put a crimp in what we were trying to do with our high availability.

- I really haven't found a clear document somewhere telling me everything that goes on from cradle to grave with job submission. I mean at a level that an admin needs: "When this breaks ... when you get this, go here." I wouldn't buy a software product without that because it's a requirement for what I consider to be enterprise-class software. Is Globus enterprise class? I expect it from Oracle, Weblogic, or an SAP. Either that or a phone number of a helpdesk or my service engineer that I can call. I guess I'm just used to running these things 24/7, 365 days a year, and living with a pager.
- The older gatekeeper software I think really had a problem with scaling. CMS proved that here. At one point in time our PBS had 4,000 jobs trying to submit against it that blew everything up. That's a scale issue; I think they both suffer scale issues. The gatekeeper becomes completely sluggish and loaded very quickly, and I can understand that. Maybe it's not so much the software, but perhaps the box isn't sized correctly, or is there a way to provide lateral scaling? Are there actually load statistics available? What's the limit? Another thing: I couldn't find anyplace was where I could set a hard limit. One that allows me define the point at which to say, "We're busy; go away." I would love to have that feature.
- The only other problem I have is things getting stuck. Sometimes the state files that are stored get out of syn and I can't get it back in sync. From a user standpoint there's no reference to what to do. I know the GRAM2 state files are out of sync because it says "stale state" (or something like that) in one of the error messages. I forget the precise details. It was weird - it went away eventually, but I don't know why. I think it had something to do with the way some state files are stored.
- It would be nice if I could define a job script saying I want to run a job using between 4 and 200 processors. So the job starts running with four processors. Now whenever new resources come along, I would ideally have a mechanism within the application that tells me when sixty additional processors become available. So it's not like you're specifying a fixed number of resources in your RSL script. That means you are stuck to those resources. Currently if you want to change resources, you have to submit a new RSL script. The main problem that I have with Globus is the lack of ability to change resources while things are running. Everything depends on RSL scripts. We have multiple servers running at the same time, and there is no way using the RSL scripts for us to change resources while things are running. So the way we are doing it, we are using Python scripts and files to communicate, rather than depending on those more efficient things because that is the most portable way we've found. So basically, we move files, start a new job, and then everything is independent. It's just we have a script that's monitoring the progress of different jobs that are running. Within a user's space, they don't allow you to - at least as far as I know - change resources while things are running. So pretty much once something starts running, that's it. And then if you want to start a new one, you have to submit a new RSL script with these multiple resources. So you pretty much have to submit new things every time, rather than some way of manipulating within the job that's running.

## **GridFTP**

- An engineering guide written for sysadmins (or people about to install a GridFTP server.) There should be a document that walks you through the thought process:
  - How big does the machine need to be?
  - How big do the drives need to be? how fast?
  - What should the network connectivity look like?
  - Should I run a striped server? should I not run a striped server?
  - Should I run GSI?
- One big problem we have is with the firewalls, with active/passive settings. We need to have different combinations of active/passive settings depending on the hosts. For instance, for some host-pairs we need to make the source active and the destination passive, but for others we need to make both active. So it's been really crazy and we've had to do all sorts of hacks to switch settings.
- I really don't want to make a big strong pitch for GUI-based tools, but certainly in the area of data transfer that would make our life easier. So if I could get a hold of the developers of CGFTP and say, "Make this real or make this go away," I'd do that.

- I tried to install the GridFTP client myself, but it failed on Solaris, and then I gave up because I could use it from Fermilab. I didn't try to track it down further, but when I was trying to install it, it looked like it was trying to pull half of the Internet onto my workstation. Part of the problem was, I think, that I ran out of disk space.
- The interface to GridFTP is a bit clunky - we would like something to be as simple as scp. So I gather that the TeraGrid project has done a fairly good job encapsulating some of the knowledge you need into tools such as tgcpc, but I get the impression that some of those things aren't really maintained so well.
- The lack of a GUI-based client for GridFTP is a barrier to some of our users. We've tried this CGFTP thing that's coming out of China Grid in some highly incomplete state. That satisfied a couple of our users. Some of our users like GUIs, and they don't like using the commandline to move things around.
- One thing we observe is that instead of using Globus url copy, sometimes using just scp is fast enough for us because Globus increases the latency, sometimes by a factor of two or three. So we are better off copying the files using scp instead of going through Globus for small files.

## Information Systems

- I essentially need to figure out:
  - Can I build my code here?
  - Will my problem fit on this machine?
  - What is the operating system?
  - What is the software that's already been installed?
  - Related but different: Is my prerequisite software installed?
  - The number of CPUs
  - The number of nodes
  - The amount of memory
  - The disk quotas
  - The scratch disk space
- It would be absolutely great if there were some information system - actually I guess it's probably not for Globus, because it's probably domain-specific knowledge. The information system would enable finding the services, finding the information, and somehow linking it in a simple way. In this case it could be distributed services and distributed data.
- MDS4 Index: It breaks, it's slow, and it's overly complex, in terms of the model. What I mean by that is
  - XML and XPath is more than is needed 98% of the time;
  - Java makes it quite heavyweight for small things;
  - The last I heard they were running in memory instead of out of a disk-based database, which hogs a lot of memory.
- The number of data products we are responsible for is growing quickly. Therefore the number of files is growing quickly. And for us the big issue is not so much the raw data size, because in a sense it's still a terabyte a day. But now instead of being divided over a couple thousand files a day, now it's over tens of thousands of files per day. And they're all different sizes. We have to track so much information about the data now. And so, as has been the case for the past couple of years, we're getting killed by the metadata.
- There are many monitoring tools out there such as INCA-based services, and the TeraGrid user portal has some of them and the WebMDS is supposed to have monitoring information. But based on our practical experience, we see that the frequency of those tools monitoring GRAM and GridFTP is not accurate. So we end up having our own tools to test in real time whether or not GRAM and GridFTP are up and running. If they are down, we immediately blacklist that host until someone manually goes and brings them back up. Until they are back up, we submit to a different resource. The reason we need to test in real-time is we have seen many examples where the monitoring tools aren't showing whether a service is really available. If you ping GRAM or

GridFTP, it works fine; but if you actually try to do some functional thing (like transfer a file or submit a job), it fails.

- If I can find all my bank history in a split second, why can't I find a machine that meets my requirements in less than ten seconds?
- I have looked at this, but the information WebMDS contains really doesn't support any of my needs. Most of the information in there, besides finding the box or queue name, I don't find useful. The reason is I think these information services are storing the wrong information. I believe we need to approach this problem from a different standpoint or path where we can describe the entire system end-to-end – a graphical way of going in and clicking on boxes and so forth and pulling up lists of software. Most people say, "Oh well. You gotta run static linked stuff." Well, try statically linking X Windows into your application CMS installs for their distribution when they install an OSG site. It's approximately 4 Gbytes for every version of their software. They take the parts of a Linux distribution and chop it down because they're trying to maintain consistency. I don't have ways of discovering this from our current information services.
- eBay was at the last TeraGrid conference. They gave a really good talk that hit a chord with me. It was about asset management. - not asset management in terms of what hardware you have, but what software, what services, what's this, what's that, how are these things connected? This is even more important. I envisioned an information system that's much more than we have now. That allows us to drill into these things and figure out how things are connected: this service talks to that service, to that box and that box has this amount of stuff in it - that sort of thing, to get to that level both from both an admin's perspective and from a user's perspective. Think of how things are dependent. For example, I click on something and it has a reference link saying, "I use Globus." Okay, I click that box, and it takes me down to Globus, and it says all right which part of Globus you are using, which service or whatever. And I click on that, and it says, "Oh you're using this set of software under these revisions," or I click on the node and it tells me the node has X Windows installed, this version and these libraries installed. It is the asset management. If you talk about asset management in business, there are two camps. One is the actual tagging of a box for tax purposes - that's the bean counters asset management view. If you're talking about the manufacturing engineer's asset management view, he'd want to know where that machine is connected, where the power goes, what the machine requires in order to function, how it plugs into the entire system, how many other processes depend on that piece of equipment being up, is this mission critical? So from a user standpoint, maybe I could start setting up more complicated requirements. What are my chances of finding something like this out there, finding it in detail? But especially in a Grid world and academia and so forth, we have such a turnover rate that these systems start becoming beneficial to the actual host environments as well. It gives them a management tool to manage how things are connected. But it is very interesting that eBay is looking at this, and they were looking at RDF [Resource Description Framework] as a technology that could do this for them.

### ***Install/Deployment***

- A lot of the more advanced configurations and uses of Globus seem to be not as well documented. So for me, that means I'm generating each key by hand for each user, and distributing the signatures to each node to allow the user to log in. It seems very painful and very complicated. And for what I was tasked to do (enable users to copy files and launch jobs remotely), it seems like a lot of work.
- A significant amount of user problems related to Globus or associated Grid stuff is simply that the client is configured incorrectly.
- Another technology-related obstacle I encounter is the issue of coherence of a given set of software. It is not possible to implement just one piece. Even all of the Globus Toolkit clearly is one piece. So the technology obstacles are ones of keeping the different components into a compatible state.

- In the two-and-a-half years of my project Grid there has not been a time when we've had the latest software versions installed on all of our machines. We are still not up to date.
- So then when we go to add pieces like our metascheduler, if that falls out of synchronization with some features of the Globus Toolkit, it can cause us problems. Nobody owns these problems. We have to solve them because they're our set of choices of what to include.
- The challenge of maintaining a very big and very complicated software stack on more than 3,000 machines is very difficult. The solutions we have in place now for managing this are not adequate. I send the instructions to the sysadmins, and they say, "What? This is crazy." And I say, "Yeah, I know, but it's all we've got right now." So getting a very complicated software stack distributed and running on all these machines is difficult. But this is mostly not a Globus problem.
- It is turning into a situation where you can't even use a distributed file system to get the software out there. There are more and more requirements, and more and more stuff has to be pushed out locally to every single compute node. Of course the Grid was sold in a totally different way when it first came out. It was supposed to just live on your batch system host, and you wouldn't have to worry about it. The nodes wouldn't have to know about the Grid. In practice on the OSG this is not the case. The OSG stack for every single worker node these days includes all the Globus clients, such as globus-url-copy and the Web service equivalent. It includes Grid security certificates and certificate of authority files, which are used for authenticated file transfers. And then there are many more things the OSG has on top of Globus, the latest of which is gLExec, which is used for pilot [technology from gLite] jobs. Several of the big virtual organizations have this technology. You might have one guy sitting at FermiLab sending out Condor Glideins all the way across the Grid, and pilot uses gLExec to determine the appropriate userid the jobs should run under. In the big picture gLExec pushes responsibility for authentication and authorization to every single compute node. The software stack to support this is very complicated.
- Version consistency, standardization - that's clearly the name of the game here. The pace of change of some of the software is dizzying.
- What I'm kind of looking for personally is like when I install my favorite linux distro (or cygwin on windows, or fink on the mac). I can go and pick out what I want, and it almost always works. You get a menu, you pick, it installs it and everything works.
- Something that I would like to avoid is installing various applications on the Grid. If the prerequisite applications and libraries (and anything else the scientists need) were already set up, that would be amazing. I am installing R on most of the Grid sites, and I am installing Octave and something else and something else. That takes a bit of time and is the part that I would prefer not to do so often.
- The biggest challenge with regard to GSI is probably the lack of platform support, given that Windows was our primary platform. Aside from that, on the platforms where it was supported it was always a challenge from a build and packaging standpoint. For example, we have some scripts for building Access Grid packages that included some fraction of Globus. But because of the way Globus is packaged we ended up shipping a lot more of Globus than we needed to. Personally we were okay with that because we wanted to support Grid computing through the AG. But that added a significantly long step to our build process. When we took that out, one of the comments from the Australians was that building an AG package went down from 50 minutes to 3 minutes. Our bundling of the Globus code was in June 2003, and we removed it in late 2005. We were using the GT2 C code from the GT3 distribution.
- Another idea that would get around trying to architect a modular jar system is to provide a service that would build a custom jar for you. One could imagine a web site that allows me to select the functions that I needed, which then assembles a jar or a set of jar files that I could download. This would be the à la carte model for deploying software for developers. In this way the interface layers and interdependencies could be better controlled. Globus advertises itself as being modular, but one of the things I'm finding is there is a lot of overlap in the packages. If I only want to use reliable file transfer, I don't want to have to use any other stuff. I want a nice stovepipe architecture, with respect to the packages. When the Autojar runs, I found a lot of crossover. It's

partly the reason why I run Autojar, to pick out the necessary ones instead of deploying all the JARs in the directory. Eliminating the overlap would really help the understanding. In my mind I see Legos. I see the client and server portions both being Beans in some sense. That's what they should be like - components that I can assemble. The Legos may not fit in all situations, so granularity level is a concern. It is hard trying to get that inter-package dependency down a little bit, there's going to be some. For example, the transport- it might be common amongst them all, and I need it. But it's helpful to identify that component, so I know I need this component for X. It's not helpful to have just a big directory of JAR files.

- In order to do a successful Grid in a particular Grid domain, some level of central management is needed. The admins could serve as experts in installing and debugging stuff, but also can quickly identify problems. It wouldn't take too many admins to do it. You could centralize them. They would be doing the application management level stuff, not OS level stuff.
- Perhaps start changing the Grid software stacks into more of a subscription service rather than having the local site admins install it. Basically the idea is that a site puts up a box, does a base container install, and registers with the Grid domain, and then all the software updates are installed and maintains by some central authority. This same model could also apply to development containers such that developers can keep up to date with the version of software that has been deployed.

## Java WS Core

- I think the Globus MySQL instructions and the way the database is connected should be revised. You have to install a specific version of the driver. Why is that? These kinds of things are a little bit nagging. And some of the drivers you can't even get anymore because they're outdated.
- The major problem with the Java container is that the database connectivity just sucks. In all the compilations (I started with 4.0.0 and then all the subsequent versions until the latest) the ODBC drivers always give me problems. Compiling this container can be a very simple thing, but it can also be a very painful, depending on whether or not you need these database drivers in there
- There's also the issue of the guaranteed delivery of notifications. Let's say we have a sensor that needs to aggregate some data. So every now and then it pops up and says, "Ok, here's my data for the past hour" and sends the data to a service. At the same time it would check for any pending updates from the service, so it would process notifications sent by the service while the sensor was offline.
- More dynamic IP address handling is needed. You know, how the container handles the network coming and going needs work. I think the container's notifications could be a good fit for us, but we need something that works better in that environment. So the use case I'm dealing with (in a different project) is where there's a sensor somewhere connected by a GPRS cell phone. The sensor gets different IP addresses every time it connects. It's just up for a few minutes and then goes down again. The current notification framework doesn't really work well in that dynamic scenario
- The common use of PostgreSQL in the toolkit should be revised. I think MySQL is more common than Postgres.
- We need more examples to help us figure out how things work. The existing tests are really good, but that's not enough. We need more examples. You can see what I mean by looking at like Mathwork's documentation page: how they introduce a concept. In the "Build A GT4 Service" tutorial, there was an example. But hearing the questions asked during that session, a lot of people didn't get it because it was like drinking from a fire hydrant. You would just uncomment some lines and redo the process again. They didn't understand what this was. Sometimes, I don't expect them to understand it. But the thing is it wasn't clear how these things merge together. It took me the longest time to figure out how the EJB technology worked with the JNDI lookup, with the get a home and get the interface. That took me awhile when I first started. It's like I had to wrap my head around it because normal C programming doesn't do these things.

- One of the things I see is that there are too few examples demonstrating the public interface layer of the Globus Java core technology. There is Javadoc that you can walk through, but I don't believe that there's really enough there. I think to myself: if Globus were a company, would I buy the product based on whether or not I could use it? I would tend to say no because I don't have that layer of documentation that I need to get started.
- I think the biggest challenge is that it is a moving target. We've had to recalibrate or recode at least two times (maybe more) because the GT authorization framework continues to evolve. It's evolving at a relatively rapid rate. Since we depend on it, we have to adapt to changes, and that's created some work for us. That's a challenge, though not insurmountable. We've been able to deal with it, especially thanks to one of the Globus developers who really is on top of things. I guess it hadn't been as bad as it could've been, but it's a moving target.
- Our software depends on Java WS Core. It depends on CoG jGlobus. And that's good, as far as it goes, until I find a problem. I've discovered a number of low-level bugs in jGlobus/Java WS Core. And these bugs don't tend to get fixed very fast. I don't know why. Even though I go through formal channels to report them (they're in Bugzilla), they don't get addressed. So that poses a problem. And so I end up duplicating code, which I hate to do. But to keep my project moving forward, that's been my approach.

### **MyProxy**

- I don't know where the documentation is, and if it doesn't work the first time, I have no idea what to do.
- The major challenge I had with MyProxy was debugging, figuring out problems with the trusted CA within NCSA's MyProxy, and not being able to use the new NCSA MyProxy client portion because of incompatibilities of the Bouncy Castle libraries. That was the hardest thing to debug. I had to generate code around it; partly that's also my own fault because I'm trying to do something other than the standard model. The standard model is to install all trusted certs on the box and go from there. I was trying to prove the point that you could do stuff without installing certs, and fetch them as needed. I really wanted to make this easy. And it's still easy - you don't know they're being installed, but they are being installed. I placed the burden on myself to keep them updated.

### **RFT**

- RFT is very good when you set all the optimal settings. On the default setting the performance is very bad compared to GridFTP. But if you tweak all the parameters, we get the optimizations. So we need to find out and learn some external tools to provide these optimization values. For example, we need to look more into MDS and see what are the optimal configurations to set between two hosts.

### **RLS**

- As it turns out, relational databases are not the best way to model our data. We don't really use the relational aspects of it. What we really want are fast index hashes. So what I've asked the RLS developers to think about is abstracting RLS so that it can support other plug-in backends, just like the GridFTP supports other data storage interfaces (DSIs). I would like RLS to support different DSIs. It should have the relational database as the default but also provide the option of using other methods of representing user data and its mappings between logical and physical filenames. Then what we would do would be to write our own backend based on a hash table approach. Because I really like the RLS API, and I like the model. I'm very happy with it as a service at that layer. What I want to get away from is the relational database backend because I don't think it's going to scale for us going forward five years from now.
- We tried to deploy RLS on a 64-bit machine, and during our critical production mode it did not scale beyond a certain limit - so very low scalability in 64-bit mode. We told the RLS developers, and they identified some problems in the C globus IO libraries. They gave us some fixes, and there



is still an open bug report about it. In general the scalability issue has haunted us a lot, and so we've had to find workarounds - on 64-bit machines. Things are fine on 32-bit machines.

- One of the problems we had - there are two components to the RLS program: one is RLI and there is another component. One of these RLS components didn't get immediately updated, and we couldn't figure out how to fix this. Whenever we listed a component in RLS, if we immediately queried it, we were not getting those components back immediately. So we had this problem, and nobody could figure out why it was happening.

## Security

- It seems that if I were to have 70 Globus nodes under my control, which I have not reached yet, but if I had 70, I can foresee difficulties associated with centralized account management. Key management for Globus seems very complicated. Until I go beyond 20 or 30 nodes it's been suggested to me just to keep the keys locally on all the machines and not try to centralize everything - it's much easier that way. Some heavy Globus users have suggested this to me.
- The standard Globus instructions basically lead the new user into an exercise of SimpleCA and building their own X.509 capabilities. These instructions are essentially useless in the context of any large-scale deployment where you actually have to trust each other and you need to build a foundation for trust.
- There was a problem in the security code in GSI that would cause connections to get hung up. There was no timeout so the connections just ended up forever hung. They never timed out, and our server would hang. It ended up happening under particular network conditions where the MTU size was too big ... I don't remember the details. It had to do with firewalls also. At the time support for GT2 had gone downhill. The sun was setting on the GT2 code so there was limited support for it. Either we had to fix the problem ourselves or migrate to GT3. We ended up patching the GT2 code for a while.
- It basically functions when I try to use it - most of the time, not all the time. I had a case (actually just a couple days ago) trying to connect to Pittsburgh Center with it, and it would refuse five times in a row and then the sixth time it would be ok, and I don't really know why. This was running GSI-OpenSSH. I never got an explanation as to why. And it's not the first time this has happened. The answer normally is "Wait a few minutes and try again." I don't really have any choice; it's not working. I either have to go somewhere else or try again. And if that's the site you need to get to, well, you wait and try again.
- The fact that the C implementation and the Java implementation don't do exactly the same thing is a problem. I mentioned policy signature files earlier. While the C implementation does consider them to define the namespaces that CAs are allowed to sign, but the Java version does not. So there are inconsistencies between the different releases. So you might have a version of the Globus Toolkit, and you expect the different versions to do the same things and sometimes they don't.

## Workflow

- Our Grid gateway uses VDL. We haven't transferred all of our domain-specific applications to VDL. Some time ago there was no recursion, but I think the issue may be addressed in Swift.
- VDL is good, but the problem with VDL was that it wasn't very stable. But then currently we are running pretty well with it. So I don't know what the future of it will be because I know that now it is called Swift
- Another problem we had was VDS kept changing all the time: from VDS, to Pegasus, and now to Swift. It's been a changing like every year.

## C.2.6 Other Social Issues

### **Allocations**

- One of the biggest challenges is getting a large enough amount of computer time to do the calculations that we would really like to be able to do but just can't accomplish right now.
- The usual funding issues where you're writing grants and waiting six months to find out if you did or didn't get it. Especially for solicitations that have less than a ten percent success rate - that is pretty counterproductive.
- Some of the problems produce a whole lot more data output than others. Those are best done at my home institution because we can manage the volume of output. So when we are finished writing that proposal, what we have is a list of projects and a list of machines, and for each machine we have an estimate of how much time we want on it. The proposal goes off to the committee, and the committee either approves or rejects it. Sometimes the committee is forced to provide an allocation on a different machine than we requested, but hopefully one with similar physical characteristics. This is actually tremendously inconvenient and is one of the ways that the NSF allocation structure hinders computational science. We are forced to request resources from all four of the major national centers because no one center could provide the resources that we need for a year. We could physically do every calculation we're interested in at my home institution, but we would require four or five million hours per year, which would be somewhere between thirty and fifty percent of the entire machine. So you understand that we can't get that. It wouldn't be good for the center to only have two applications.
- Another dimension to this is that some of the things users want may be very different from priorities we have. Because the AG tries to be a research project, there are also research-type priorities that we need to pursue. How do we execute that? We try to get money, decomposing the problems into workable subsets.
- A lot of middleware is developed as a research effort and papers are being written about it. I don't see many papers written about latex, for example, because it's actually infrastructure that works. So to me, building an infrastructure means creating and operating something that's useful for people even though there may be nothing novel to publish about the underpinnings. The funding agencies tend to tie the creation of infrastructure into research activities, but they need to fund it and evaluate it differently.
- I think the compute centers that we have, the HPC compute centers either on the TeraGrid or even the ones that exist by themselves, are serving an elite few. And though those people like to solve real science questions, I think the vast majority of people who could use the supercomputing are excluded.

### **Community Awareness and Collaboration**

- For us to do the research, we want to do we need to have Globus and (ideally) Condor and some other software using the new format logs. We also need to get OSG to deploy the central log file collection stuff. Some of these things are not super-high-priority items for everyone involved. So it can take a lot of phone calls and prodding. Everybody agrees it's a good idea. It's just not always on the top of everybody's priority stack.
- One challenge is that there is so much development happening now, in so many directions, by so many people, that it's becoming harder and harder to keep track of all the developments. Trying not to reinvent things and trying to leverage what's already there is a constant challenge.
- People have a tendency to write and rewrite monitoring applications as if forgetting all the enormous amount of work done on this topic by people before them.
- When you need changes or modifications or improvements, it's not under your control. You do not directly control the developer resources to get those changes done. And so you have to go back and ask them, and you don't really have much to offer in return other than the greater good of what

you're trying to do. We've been doing well operating under these constraints, but it hasn't been easy all of the time. There have been times when we've been told no, flat-out, "No, we aren't going to do it; I don't care how much you need it." And then other times when we've been told, "Yeah, but it will take a while." And in some cases, it takes a long while. Also there are other times when you get it right away. No one's out to get you, but they all have their own agendas. Your requests need to be fit into larger priorities, and sometimes the requests are not given as high a priority as you would like. It's not that they're trying to hurt you; it's just that they have other bells to answer. That's hard. There's no way around it; I don't know how else to put it. It can be a problem at times.

- It's open source work, so sometimes there are problems in other people's code that need to be dealt with in order to achieve your own objectives.
- We sign and keep all of the certificates ourselves, so we don't have any challenges within our portal. But if you have to go to somebody else's portal, then all the trust relationships get complicated because they have to trust you and you have to trust them. It is a challenge if you want to interact with another organization.
- One challenge is in instilling a mindset in the community to develop tools - not applications, tools. I don't think there's enough work with respect to that, and I don't know why. I think getting tool-level people engaged is important, because that says Globus is behind a standard. And there are people developing tools to that standard. This eases some of the load because you have a bigger market of tools. Tools are everything.
- The fact that the collaboration is very distributed. We have fractions of people working on the project. They are scattered around the U.S. (for the base part of the program) and around the world for the other collaborations (like authorization interoperability.) This is clearly a challenge for managing and controlling the development processes. With the GT2 framework in general, the fact that it's pretty much a frozen development can present a challenge. We have contacts with the Globus Toolkit developers, so for exceptional things we can have some features added to the infrastructure. But the challenge that we face is that it is in production everywhere for our stakeholders and it's not actively developed anymore, because the Globus Toolkit has moved to the Web service version. This is challenging to us. And this is also why there are groups that are investigating the new technologies. But before you convince yourself that they are really production quality ... well, it takes a long time.
- This situation affects us most with the GT2 gatekeeper. For example, in the VO services project we would like to pass more context back and forth between our authorization plug-in and the gatekeeper. This would require a change to the GT2 gatekeeper code and our code (to adapt to a different API). And it takes a lot of effort to try to bring this thing up again. We have now the agreement with the developers that they will work with us. But then on our side the people who were following up with that don't have so much effort anymore. So things got stopped. It's a frozen piece of development, so it's difficult to make it alive again if you need to change anything.

### **Lack of Time**

- Having to switch contexts often is a challenge. Working on one project, then something comes up in another project, and then another thing comes up. Segmenting the tasks too much decreases my productivity. The minimum fragmentation that I would take and still be productive would be a half-day. Spending half of one day on a task, then I can switch context. But less than that is not big enough.
- The main challenge is it's difficult to concentrate on one thing because I'm spread so thin. When we made the transition from being funded by NSF to being TeraGrid-funded, my involvement in the project went from full-time to half-time. That means there is less time that I can devote to that development.
- There was some discussion at some point about writing a site selector based on the MDS information services. Like we could use MDS to maintain a list of sites instead of using a VDL site catalogue. But we never found time to work on this idea. Once we got the portal up and running, our immediate focus was on building the user community for our infrastructure.

## C.3 Satisfaction Points

### C.3.1 GRAM

- GRAM is great. In one fell swoop it allows me to specify jobs in a single language, and hides all the details of every local job manager (which nobody wants to learn.) Plus, it has the entire security infrastructure built-in. That's a hard, hard problem to solve.
- GRAM4: because the staging support is better. It's a pretty big improvement over GRAM2 in that sense, because you can do smarter staging (like the whole filelist). That maps much better to our Condor description files. And in general the architecture is better.
- The Rendezvous Service was written to provide all-to-all exchange of information in the Web services context. We were able to whittle it down to a reasonable API. And it was even a great exercise because I showed up with a need that the GRAM developers didn't foresee. We talked with them and nailed down the requirements, and it was done. This was about a year or two ago. It was great.
- We do use GRAM4 - we've started to use GRAM4. We also use a new thing, which I can't tell you exactly where it sits. It is called the Rendezvous Service. It may be its own service; it may sit inside of GRAM4 - I don't really know. But it was critical for us as we moved from pre-Web services Globus to Web services Globus.
- The GRAM interface is really cool and provides us with one common interface for all the job managers. This is important for us since for our requirements we need to use multiple clusters and each has its different job managers. GRAM enables us to use the uniform mechanisms for both job submission and monitoring. It also supports the authentication mechanism that we like.
- We didn't see any particular need to support pre-Web services. And so as an experiment we tried just GRAM4. It seemed to have advantages in terms of being stateful and allowing us to interact more closely with our potential services. Certainly from the point of view of just job submission it was relatively trivial to adopt.
- We use GRAM4 for our job submission. That's where we started two years ago. We've had a little bit of struggle along the way - of course it's been improved. We really have never regretted that decision. We will support GRAM2 job submissions to our batch-oriented resources on request, but we've never had a request that couldn't be satisfied by teaching the person how to submit via GRAM4.
- GRAM2 is the best thing available right now. It just makes everything easy. It completely hides all the complexity. And one of the reasons is the interoperability it offered between OSG and TeraGrid. All we needed was a GRAM endpoint, and that's it. Globus actually makes different sites (like the many TeraGrid and OSG sites) appear like they are the same site. I mean Globus just completely shields us from the fact that they are different. In the end it looks like one type of site to us. As long as I have a GRAM pointer that I can submit jobs to, I don't have to worry about what scheduler is there. The only thing I need to worry about is the hardware - if my executables are able to run there or not. Like for example if it is a 64-bit machine I need to send the correct type of binary to the machine. Other than that the whole mechanism stays the same for OSG and TeraGrid.

### C.3.2 GridFTP

- GridFTP: Our major challenge has been the LSOF, but it looks to be addressed, and we're very excited about leveraging that functionality. Other than that - jeez it's completely reliable and moves tons of data. What else could I want?
- It is the way that exposes the highest rate of network transfer from filesystem to filesystem across a transcontinental distance. So for example if I have to move output from Pittsburgh to San Diego,

GridFTP buys me a factor of two or three over bbFTP, which I could look after myself. And that two to three is essential.

- We like GridFTP because it's both a protocol and implementation. The protocol has a specification that led to different implementations of the protocol. For example dCache, which is a storage element developed at Fermilab, has its own implementation of the specification. So the protocol has been demonstrated to be specified well enough to be implemented by different providers.

### **C.3.3 VM**

- For redundancy we're setting up two physical machines, each with four virtual servers. We're using Xen to do the virtual server. We talked to the leader of the Globus Virtual Workspaces service and received some advice in the early stages that helped us figure out what to do there. It was very helpful.

### **C.3.4 Install/Deployment**

- The Globus installation has improved a lot. I've been installing Globus from the first version to today. The installation is very good, the documentation - the problems we encounter during the installation - they've been documented very, very well.
- The challenges that are associated with using Globus (many people feel it's complex and hard to implement) were greatly lessened by our decision to adopt the VDT-based installs. We've worked very closely with the VDT team, including with security updates (a couple of which actually made it back to the Globus repository).

### **C.3.5 RLS**

- So we rely very, very, very heavily on RLS. I think it's true that we run the largest RLS network in the world. When RLS goes down, you better believe we know it. And we have to jump into action. Fortunately it very rarely goes down now. We're quite pleased.
- Our project testbed can be distributed across multiple locations because of the replicas. The replicas also allow us to choose the fastest available compute server for our computations

### **C.3.6 Security**

- The Grid security infrastructure is one of the things we completely rely on and are building our tools on. For example, all these remote job submissions and remote file transfers are completely based on GSI. So we absolutely love GSI-based authentication. This is something that has been really helpful and paved the way for the portal-based computation.
- We'd be lost without X.509 authentication. That just solves a whole raft of problems.
- With Grid technology with the security model, you can do quite a better job electronically: you can do an audit, verify certificates, verify attributes, etc. These mechanisms are way better than what clinical practice is right now, because many of the documents today are in writing, stored in physical files, reports end up in the trash, etc. So there are many places in the current system where private information is exposed to the outside. This is one way that I think Grid technology can help, because it has a very good security model.
- You just walk into a sysadmin's office these days and say, "We're using GSI - the Globus security infrastructure," and they get it. They know that it's been looked at by many eyes (maybe even their own). They trust it, so you don't have to convince them it is secure.
- If you have an IGTF-accredited CA, that's enough, because other large-scale projects throughout the world get these sets of trust anchors. So they know whether or not to trust the credentials of your CA, and on what basis. They know that you have, for example, been in-person identity-proofed by someone in the chain. They also know the CA is run in a method that does not allow a graduate student to walk in and issue their own certificates. So since these things are widely distributed and commonly accepted, it's very easy to start a virtual organization. We always make

the point that authentication is not authorization, but it's a starting point. They can then know the quality of the Grid credentials coming in and use that as a basis for signing membership to the Virtual Organization. It becomes barrier-lowering because I can accept a certificate issued in Czechoslovakia, for instance.

- The Globus GSI and everything that's built around that ecosystem now works really well for us. We can hook into it in so many different ways. We can set up services that manage the delegation for the users. The only thing the users have to do is enter the system once using something like MyProxy. Everything else is handled for them. That works really well.
- Because GridShib is a plug-in for GT, we need to support the GT authorization framework. I'm happy to say that it's a really nice framework. And it continues to be refined and enhanced. It works fine with us, and there's no point in considering an alternative, even if there were one. Of all the Globus security components, the authorization framework is nearest and dearest to my heart.
- I love Globus security. It addresses all of the use cases that interest us. It supports digital signature with capability of doing encryption, integrity checks; there is the ability of doing delegation, all the expected steps in the authentication processes, the ability of having control lists, signature policies ... it's a very complete suite that does what we need. There are people who are starting to use a different infrastructure, like OpenSSL for example. For the time being we are happy with the GSI infrastructure, which is, by the way, deployed everywhere by our stakeholders.

### C.3.7 I/O libraries

- IO was very good, but XIO is even better because it allows us to build protocol stacks. The protocol stack design will make it much, much easier for us to introduce new technologies in the future. So when it was all Globus IO-based we had to go through the exercise of pushing GridFTP into MPICH-G2. It didn't kill us, but it wasn't easy. We had to munge the code pretty heavily to shove that stuff in. And we had to go through the same exercise when we had to push UDT into MPICH-G2. With Globus XIO, all of this can be very neatly wrapped into an XIO module that either I or someone else writes. That's the recipe for rapid prototyping. I won't have to mess with the MPICH-G2 code at all anymore. I'll just write an XIO module, and one line of code to activate the module in MPICH-G2, and it's done. And you get it for free. That's a big step forward for us. That's a big help.
- The Globus data conversion library is indispensable. We need it. If it were to go away, we'd have to write it from scratch ourselves. MPICH-G2 is responsible for doing the data conversion between big endian and little endian machines, for example. The MPI application is not going to give a darn about that. I need to care about that. It's an ugly job that no application should have to write. One library should write it once and then provide it. We provide it in MPICH-G2 because the Globus developers wrote it.

### C.3.8 MDS

- The Index Service lets us broadcast whatever we want. It's easily configurable, and we are comfortable using that. Also, we don't need to install any additional software because it is part of the Globus Toolkit.

### C.3.9 Java WS Core

- In the past six months I've had the opportunity to dive into Java WS Core and understand that deeply. I really do appreciate the effort and expertise that went into building that code base; we've leveraged it significantly.

### C.3.10 General

- We went to a scientist who was highly computationally bound working on a small set of machines and created a portal environment to encapsulate his workflow. In the two-week demo the scientist had access to far more compute cycles than he had been able accumulate to date. He said he got

publishable work out of it. ... The scientist has since gone on to get research funding to buy clusters and then contribute those clusters to our project.

- Globus is certainly the dominant technology. It is compatible with a lot of the larger projects we want to interact with.
- Client-side backward compatibility is important. When a new version comes out, I should not have to rewrite my software. This has been a major concern in the past but has been much better lately. If the clients can talk to the services in the same way and get the same functionality, that would be good.
- Globus rules. It's gonna save the world. I'm not kidding. We rely quite heavily on Globus. Globus does all process management, the start-up, the security. In Globus we're using IO for all inter-machine communication. Globus is the one software that we use across all applications.
- I am attracted to the WSRF framework. I think that, especially for someone like myself who's not a computer science person, it allows me to quickly and easily leverage things like resource properties and the publication of resource properties: the lifetime management, the subscription, and notification - all the nice things. It allows me to leverage them quickly and much more easily than I could do if I had to do all that stuff by myself. After all, I am a physicist, and I'm dangerous when I'm writing code. The more code other people write for me, the better
- I certainly appreciate all their efforts. The Globus Toolkit has really succeeded in what I think is one of its primary missions: enabling more science. Without a doubt, Globus has made more science possible. Period.
- I really appreciate the overall effort. For Grid the whole paradigm can only thrive is if there's an open source and standards-based implementation, and the Globus Toolkit is delivering exactly that. One problem in the medical domain is that the internals of every equipment vendor, both software and hardware are proprietary. There are some standards on the interface side, but internally it's all proprietary. I think that the whole concept of service-oriented architecture presented in the Globus Toolkit and the Grid paradigm can have a major impact on how medicine is being addressed from a technology side.
- I think it's very good that this survey of users is being done. And I think it's something that should have been done five or more years ago.
- I want to say that I appreciate having people who build the software actually look at how people are using them. So this interview process is useful.
- I'd say within the last six to ten months we've been seeing more Globus developers showing up on some of the OSG calls and being there as a resource, if needed. I know where to find them if I need them.
- It's really not a lot of work to customize the Globus tools to fit their own users' use cases. I see TeraGrid doing that, and I think that's great.
- The Grid has made a qualitative difference in what we can do. For example, to analyze the data in preparation for a new release of our data products: If you want to do it on the cluster sometimes it's very difficult to get nodes. And on forty nodes the analysis will run for weeks. But on the Grid we can immediately do it, and it will take so much less time. We can provide our users with fresh data more frequently because of the Grid.
- Keep doing it. The work is incredibly valuable, and I just don't think we're at a point where we can stop. It's like we're building a car and we haven't put the engine in, so we can't start it yet.
- When the time is right, we're certainly going to start utilizing those services and those applications. Exactly how and when and other details have yet to be decided. But yeah, go for it.
- Thanks. But don't stop.
- Some people are perfectly ready for Grid technologies – for access to highly distributed computational resources. Those who say, "I've been waiting for you to come along," they have an

application that needs cycles or needs to move data or somehow needs a cluster. Dealing with these type of folks is so easy. We enjoy it so much. We give them credentials; we adapt their application. Ours is Web services only, so we wrap it in Web services submission script. Maybe we even build them a small service. We drop the tech in, and they're happy

- The formalization of Web services I think is an incredible technology to allow machines to communicate with machines in a very standard and structured way. Though there are still questions on what to use in the Web services area: Do I use SOAP? Do I use REST?
- Honestly those types of remote procedure calls have been around for a long time in the Unix environment using sockets, but they were unique to those platforms. I think the convergence now with the standards is really making things fly.
- There's no other technology out there that provides compute resources, data management resources, and security at the level that the Globus Toolkit does. Period.
- We really like Web services because it lets us build tools that are better suited to scientific workflows. They really are services - the steps that we need to accomplish. That could be better if we had user-pluggable services. We're not there yet.
- We're working in the field of radiation modeling for cancer therapy, and there's a proton accelerator here in the state that has been built by the M. D. Anderson Cancer Center. This large-scale \$120 million facility has huge modeling needs. We thought it would be a very hard problem to move the medical data around. But we found that there are tools in the caGrid software stack that are not only well-suited, they're actually explicitly written for the purpose of moving medical image data around with high security using Grid tools.
- That is an area where we thought we'd need to do a lot of development. Instead we found a complete working infrastructure that we just didn't know about
- We've recently become an incubator project for our netlogger work. There have been no hassles other than trying to convince our lawyers that Apache and free BSD licenses are effectively the same thing - which is our problem, not Globus's problem. I've been pretty impressed with the whole incubator process. I like the fact that you get a Wiki, a bug tracker, and a CVS repository. And if you need something configured, it seems to happen pretty quickly. The lead dev.globus infrastructure person seems really good. I was impressed with the whole incubator startup process and how smoothly it all went.
- When computer scientists actually try to use existing software to do the tasks they're writing new software for, they develop software that is highly domain-relevant. I don't know whether acquiring this domain knowledge is something that needs to be done by the team itself, or whether they need to have close collaborations. I think the collaboration that we have with the University of Chicago Grid experts may be very valuable in that respect. As neuroscientists we've had our frustrations, and those frustrations are being solved by some of these new approaches. Actually, that's not fair to say. They're not being solved, but we're working towards solutions. We'll see in five or eight or ten years whether we've really had a good effect. I agree with the University of Chicago Grid experts' approach, which is to be highly collaborative with the domain specialists. I applaud that and think it's the right way to go.
- With the Globus team I generally feel like when I ask questions, they're quite responsive.
- Based on our profession interactions with other large-scale centers, we can quite often implement a framework that would have been beyond the reach of a researcher left to his own devices. That is very satisfying.
- Leveraging large software projects like MPICH and Globus (and even to a lesser degree GridFTP or UDP or UDT) is great because it's a tremendous leg up. You leverage it. There's a bunch of code that's there, and you apply a little bit of work, and you get a tremendous benefit from it without having to do all of the work.



- Much of the basic functionality, such as data transport, is already included in the Globus Toolkit. And there was not a lot of effort required to make Grid technology work for the medical domain. And that was very neat because you don't have to reinvent the wheel.
- Thank you so much because somehow the use of the Grid and the use of the Globus really, really, really made such a huge difference in what we are doing. We can do so much more science after using the Grid. So just my deepest, deepest, deepest and sincere thank you.
- The Quickguide to installing the container is very good, very straightforward, and clear. Even for somebody who is a beginner, this is a straightforward document.
- The toolkit model fits me exceptionally well. I'm really looking for tools and infrastructure that allow me to build on top of them. So they have to be extensible; they have to have nice APIs and hooks that I can layer my own stuff on top.
- We chose it because it seemed to be the dominant technology for Grid services. And we were interested in going the Web services direction. We haven't found a reason to revisit that decision.
- With the Grid technology we deliver the images from any clinical trial center into the radiologist's own review workstation. That's capable now with Grid technology. And that's a big reward for us to see that really happening. And the radiologists are very pleased with this. This actually engages more radiologists in clinical trials than before. So we can actually improve the quality and quantity of research being done.
- Good job. The fact that people are actually looking for feedback and helping when issues come up - that's amazing. That's what open source and free software is all about. I'm happy with the situation as it is today.
- One technology we find very useful today is Globus and Grid. It has come a long way to help us.
- I've had very good luck working with Globus developers, and it's been a rewarding experience for me. I can't really think how that could be improved. It's working rather well, I think.
- I've looked at the codebase and understand it from a conceptual point of view. Part of the motivation was to see if we could learn something from a design or implementation perspective. We've employed techniques seen in CAS. So this is an interesting use case for the Globus code. As a developer, not only do I build components that depend on Globus code, but I also use it as an example. I've studied the Java WS Core and CoG jGlobus codebases extensively. They've been a tremendous help in developing some of my components.

## Appendix D The Interviews

*Important disclaimers:* The interviews represent snapshots in time. All answers are relative to the interview date shown in the left corner of each table heading; circumstances may have changed since the interview. Many interviewees are involved in multiple projects and serve in multiple roles, so the full extent of each person's work is likely not represented in the interview.

### D.1 The scientists' happiness is my main measure of success

Interview ID=1 31 May 2007	ANSWERS	ANNOTATIONS
<i>Pre-interview question: Do you interact directly with Globus software in your work today?</i>	Yes	
<b>Establishing context</b>		
<b>Q1.1 Please provide a one-minute overview of your project</b>	I am involved in a collaboration with a brain imaging research group in the medical center at the University of Chicago. They are interested in analyzing the response of the brain to various stimuli. And they do this to study how patients who have suffered a stroke improve over time. So they bring in patients and they run experiments where they record the patient's MRI brain images when the patient is being subjected to some stimuli. Once they have that data they come to us to help them process it. Essentially we are working with them in organizing this patient data and in helping them process the data.	
<b>Q1.2 What is the project's name?</b>	CNARI	
<b>Q1.3 Which agency funds the project?</b>	National Institute of Health	
<b>Q1.4 What field does your project belong to?</b>	Neurology	
<b>Q1.5 What is your job type?</b>	Developer, Scientist, System Administrator	
<b>Q1.6 How long have you been a &lt;job type&gt;?</b>	Five months	
<b>Learning about discipline-specific goals and approach</b>		
<b>Q2.1 What are the main goals of your project?</b>	The goal is to get the scientific code that the researcher uses ported to the Grid, such that he can run it at larger scale and get much better speed up with what they do.	
<b>Q2.2 How will the success of your project be measured?</b>	By the researchers' satisfaction in getting more work done. That would be measurable by the number of papers the scientists write after using our system and the quality of those papers. Also by how many resources they are able to use on the Grid to run their code.	
<b>Q2.3 What are the professional measures of success for you?</b>	The scientists' happiness is my main measure of success. Also the extent to which the scientists like the system we propose. That they use it and gain an advantage from using the system. Also if they become interested in using the Grid in their future research, whether or not we're involved.	
<b>Q3 What are you investigating?</b>	The actual work I do is setting up workflows for the scientists to use in addressing their scientific problem. That includes writing the workflow in the Swift workflow language, which you can think of as a Grid scripting language. I also set up the resources that the workflow will use. This includes installing prerequisite applications, such as the R package and AFNI, which is a visualization and brain image-processing package. My work also sometimes involves looking at the scientists' code. Understanding the code, working with the scientists to parallelize it. This involves figuring out where the loops are in the code and unrolling them to build a workflow that takes advantage of the parallelism. Then sending the subcomponents to different machines.	

Interview ID=1 31 May 2007	ANSWERS	ANNOTATIONS
<b>Q4.1 What is your method for investigating &lt;phenomena&gt;?</b>	<p>First I meet with scientist and we have a discussion. I try to figure out what the problem is they're working on. Most of the time I try to understand the actual scientific part of the problem because I like learning from different people.</p> <p>Then I ask them questions related to the execution of the problem. Most of the time they already have a solution and they just want to improve it by running on the Grid. I determine how fast their solution runs in a single-machine setting and what the components are (such as the external applications and libraries that are used.) And then I examine their existing application code and get a short explanation from them about it.</p> <p>Then I go away for a couple of days to try and understand what they did. I try to replicate it in my own setting and start unrolling the problem and splitting it into pieces. Finally I create the workflow, and thereby put the pieces back together. I give them back the workflow with everything in place, such that they can run it on the Grid and speed up their solution.</p>	
<b>Q4.3 How do you keep track of interim results, if at all?</b>	<p>We have wiki and we write down the main stages of the project, and which stage we're working on, the requirements and the goals. We document what the scientists want at the current stage, and what might be a showstopper problem or something that may need attention later.</p> <p>Essentially, these are not very complex applications so we use the wiki mostly for reporting purposes, so the project leader knows where we stand. Most of the current project status is in our heads because it's just a couple of people who are working on a project, and we all communicate extensively about the progress of the work.</p> <p>Typical project stages include:</p> <ul style="list-style-type: none"> <li>- understanding the problem</li> <li>- figuring out how to disconnect the problem into pieces</li> <li>- giving the workflow to the scientists</li> <li>- adding different codes to the workflow with the help of the scientists to improve the solution</li> <li>- installing the required applications on the Grid.</li> </ul> <p>These are main tasks involved in "Grid-ifying" projects.</p>	
<b>Q5.1 In what ways do you interact with simulations in your work?</b>	Unless the scientist has a code, which consists of simulating some phenomena, we do not do simulations per se. We mostly do the grunt work.	
<b>Q6.1 Describe how you interact with data in your work</b>	Well this project has a lot of data. Essentially the data consists of the brain scans collected from the patients. And that data has to be imported into a database. So a schema was designed to enable the scientists to ask relevant questions about the data. The schema design also facilitates Grid-enabled data processing. So essentially that's it. The scientists give us the data. They tell us what the data means. We sit down together and organize it. We come up with this schema, and then I do the importing into the database so that they can access it in an organized way.	
<b>Q6.2 How do you share work-related data with others?</b>	I don't. I'm the only person working on this project, so I don't have to share data with anyone. The scientists will probably share the data within their community. That's why the database was implemented, so that they can easily share the data.	
<b>Q6.3 By what mechanisms is access to your work-related data controlled?</b>	The scientists all have log-ins to the database, and they have access to various subsets of the data that they're working on.	
<b>Q7.1 What resources do you use in your work today?</b>	<p>The scientists have a big database server machine where we store the data, and we're using it as the backend for data access.</p> <p>I develop on my own laptop or on one of the workstations at my home institution. And then for testing I use some Grid sites.</p> <p>No big resources involved, nothing fancy.</p>	
<b>Q7.2 How do you share work-related resources with others?</b>	I check everything that is related to this project into the Subversion repository. This allows my colleagues to could look at it and tell me their opinions and suggest improvements. The workflow that pertains to this project is in a place where everyone from my team can access it and look at it.	
<b>Q7.4 How do you locate available resources for use in your work?</b>	<p>Resource acquisition is based on my own personal experience:</p> <p>Sometimes I prefer using TeraGrid because it's a bit easier to use. You have a login shell, so it gives me a lot of power. And some other times I talk to colleagues to see what they're using. Maybe someone has already installed some of the applications that I also need so I could reuse their effort.</p>	

Interview ID=1 31 May 2007	ANSWERS	ANNOTATIONS
<b>Q7.5 What types of information do you need to know about a resource in order to determine if it is suitable for your work?</b>	I need to know how loaded that resource is. In other words, how big the batch execution queue is. Also how friendly it is in terms of my ability to install applications, or providing plenty of CPU cycles and storage. For some sites I have had to come up with tricks to get around disk usage limits.	
<b>Q8.1 What software do you currently use in support of your work?</b>	The workflow is being built using Swift. And Swift uses the CoG kit. And that in turn uses the whole Globus stack like GRAM and GridFTP, etc. As far as the kind of software I use to develop or test my work: I use Eclipse, Subversion and the main UNIX tools.	
<b>Q8.2 What scripting languages have you used in the past year?</b>	I use bash and awk for the small problems. When it's a bigger problem I use Python or whatever seems to solve the problem fastest.	
<b>Q8.3 What programming languages have you used in the past year?</b>	On this project I haven't done too much programming in the past because I was mostly building the workflow, which is a high-level scripting language. I have been involved lately in MATLAB a bit and also with R, because these are the languages the scientists use to solve their problems.	
<b>Q8.4 What workflow tools do you use in your work?</b>	Swift mostly.	
<b>Q8.5 What parallel computing tools do you use in your work?</b>	None.	
<b>Q8.6 If the need for new software-based functionality arises in your work how do you acquire it?</b>	For this project we did not have to make any changes to the software to add any new software functionality because they already had it there. But in another project we disconnected the scientists' work from commercial software and replaced it with open source software. So essentially if some functionality needs to be there I usually look for open source solutions and try to stitch them together to solve the problem.	
<b>Q8.7 How do you share software with others?</b>	Everything that I do is in Subversion and the CVS. Other than that, we don't do a lot of software development. I have some projects on the side, but they are not at a stage where they can be shared with others.	
<b>Learning about the user's problems</b>		
<b>Q9.1 What challenges do you face today in accomplishing your work-related goals?</b>	I would say the general challenge is the fact that I'm not fully in control of the environment where I have to do the work. I don't have the tools or I don't have the knowledge to fully figure out what's going on when something goes wrong. Other challenges would be maybe the lack of resources, but usually this gets solved pretty quickly. That's usually not the showstopper. I can find replacements on anything that I need to keep going on. By "resources" I mean either applications or physical resources like CPU or storage.	
<b>Q9.2 What types of information do you need in order to address the challenges you face today?</b>	I usually need the domain knowledge about the problems that I'm trying to solve. For example I may not fully understand the application code that I'm getting from the researcher. But most of the time the showstopper for me is the information that needs to be discovered. I sometimes encounter problems that cannot be solved without finding additional information. I may have a vague idea, but somehow I need to figure out the best way to solve the problem. For example, I may not be fully familiar with the queuing systems (at least on the Grid side). Or I may not be familiar with some Grid environment variable settings that I should set.	
<b>Q9.3 What technology-related obstacles do you currently encounter?</b>	I don't have many technology obstacles to worry about. Essentially, I always find a way to solve the problem using some open source technology that is available. So there hasn't been any moment in time when I have to actually sit down and implement a full tool to solve my problem. I always find something that would help.	
<b>Q9.4 By contrast, can you provide examples of technologies you find very useful today?</b>	I'm using such a wide set of technologies in my work, that I can't state any preference for one. Everything that I use is very useful to me (obviously, otherwise I would not use it).	
<b>Q10.1 Can you think of any work-related tasks that decrease your productivity?</b>	Having to switch contexts often. So working on one project, then something comes up in another project, and then another thing comes up. Segmenting the tasks too much decreases my productivity. The minimum fragmentation that I would take and still be productive would be a half-day. Spending half of one day on a task, then I can switch context. But less than that is not big enough.	

Interview ID=1 31 May 2007	ANSWERS	ANNOTATIONS
<b>Q10.2 Describe repetitive tasks associated with your work</b>	At a high level I'm doing the same stuff. I get the problem from the scientists. Figure out how to "Grid-ify" it. Then I work with what they have and what I can get my hands on, eventually giving the solution back to them. So essentially, it's a repetition. But given the fact that all these problems are really different – they mostly belong to different domains – it's not repetitive. It's more like let's solve this problem and that problem. It's only repetition at a very high level, which doesn't bother me.	
<b>Q10.3 Describe time-consuming phases of your work</b>	Something that I would like to avoid is installing various applications on the Grid. If the prerequisite applications and libraries (and anything else the scientists need) were already set up, that would be amazing. I am installing R on most of the Grid sites, and I am installing Octave and something else and something else. That takes a bit of time and is the part that I would prefer not to do so often. But other time-consuming parts are quite interesting, like learning about the scientist's problem and trying to understand what's going on. That's the fun part. With regard to this specific project, having to process the huge amount of data was a bit time-consuming. They have huge files that needed to be processed. I had to write scripts to format the data and import it into the database. That took a couple of days or more. But that's part of the job and it's done now. I don't really have anything that's bothering me with respect to this project.	
<b>Learning about the Globus user experience</b>		
<b>Q11.1 Which Globus data components do you directly interact with in your work today?</b>	Directly, I don't use any of them. I only use Swift directly. But Swift uses CoG Kit, and CoG Kit uses GridFTP.	
<b>Q11.2 Did you install the &lt;component&gt; client yourself?</b>	I installed Swift myself. I just downloaded an archive, untarred it and it's ready to go. CoG is packaged with Swift, so I don't need to worry about installing that. And I use GridFTP servers that are already deployed; I didn't need to set up one.	
<b>Q11.4 How many people currently use your &lt;component&gt; server</b>	So besides me, the main scientists of the project that I'm involved with plan to use Swift. I give the scientists the results of my work when it is almost done. Then we enter a testing phase. During that time we are sharing the installation of Swift; they use my installation to focus on what's relevant to their work.	
<b>Q12.1 Which Globus security components do you directly interact with in your work today?</b>	None.	
<b>Q13.1 Which Globus execution components do you directly interact with in your work today?</b>	Swift uses (most of the time) GRAM2. It also uses GRAM4 in some of the cases. Sometimes when I'm debugging, I use GRAM2 just to verify that I have a valid proxy or that the job manager accepts my job. But that's seldom, whenever I need to debug something. A quick GRAM run allows me to validate my proxy or to validate that everything is in order on the remote side. I try doing a /bin/hostname or something like that just to see that it works.	
<b>Q15.1 Which Globus common runtime components do you directly interact with in your work today?</b>	When we use Swift through Falkon it uses the Java WS Core service container.	

Interview ID=1 31 May 2007	ANSWERS	ANNOTATIONS
<b>Q16 Why do you use &lt;component&gt; instead of an alternative technology?</b>	<p>CoG: Because that's how Swift was developed. Currently, they might choose to add a different provider, but at this point, GridFTP through CoG proves to be a good enough solution.</p> <p>GRAM2: The same answer. Because it's there. It generally allows me to send in the same jobs to any Grid site. That is the main advantage of using GRAM2. We do have some providers like a plain Condor provider and a PBS provider that would allow you to connect the local Condor or PBS pool, but that's not general enough. It's not cross-site. With GRAM you can connect to any Grid site. [prompt asking why GRAM2 vs GRAM4] Sites seem to have GRAM2 installed and it's working. I don't have a problem with that. And unless we use the Falkon component of Swift we don't really need GRAM4. When we use the Falkon component of Swift, GRAM4 is required. And that restricts us to a subset of Grid sites. So I prefer GRAM2.</p>	
<b>Wrapping-up</b>		
<b>Is there anything you'd like to say to the people who build software for use by people like you?</b>	<p>Good job. The fact that people are actually looking for feedback and helping when issues come up – that's amazing. That's what open source and free software is all about. I'm happy with the situation as it is today.</p>	

## D.2 Troubleshooting requires knowledge about software internals

Interview ID=2 31 May 2007	ANSWERS	ANNOTATIONS
<i>Pre-interview question: Do you interact directly with Globus software in your work today?</i>	yes	
<b>Establishing context</b>		
<b>Q1.1 Please provide a one-minute overview of your project</b>	I work in the LEAD project. The LEAD project, like most academic research projects, is one of the ITR projects. We are in our fourth year. What we are trying to do is integrate computation systems with atmospheric science models and instruments. The project is based on a workflow system called WOORDS (Workflow Orchestration for On-demand, Real-time, Dynamically-adaptive Systems.) The project is designed to run atmospheric models and use the output to steer instruments to get new data to serve as input to the models. So the project is based on a dynamic and adaptive system.	
<b>Q1.2 What is the project's name?</b>	LEAD	
<b>Q1.3 Which agency funds the project?</b>	The National Science Foundation	
<b>Q1.4 What field does your project belong to?</b>	Computer Science, Atmospheric Science	
<b>Q1.5 What is your job type?</b>	Science and Technology Liaison, Portal Developer	
<b>Q1.6 How long have you been a &lt;job type&gt;?</b>	Seven years in this area, four spent on this project	
<b>Learning about discipline-specific goals and approach</b>		
<b>Q2.1 What are the main goals of your project?</b>	The science goal is building a dynamically adaptive weather simulation system. So the engineering goal is to build a Service-Oriented Architecture in support of the science goal. So we are building the Service-Oriented Architecture and Grid middleware ourselves while trying to leverage as much as possible the tools already available in the community.	
<b>Q2.2 How will the success of your project be measured?</b>	There are multiple ways to answer this; it has been the subject of debate. So since this is a research project, success measures are similar to any other research project. We've also been having a lot of educational component to the project. We've been supporting student contests and educational activities. Many people are also using the system right now in production. We've been supporting some of their advanced research activities. For example, every spring the National Storm Prediction Center produces a spring forecast, and they're using the LEAD system for that. So a good measure of success is getting the users to use the system in a user-friendly way. And also to produce some good research and papers out of the project.	
<b>Q2.3 What are the professional measures of success for you?</b>	I am a liaison between meteorology and computer science. So I try to take the big use cases and transform them; I explain them in more detail to the meteorology PhD students and the guys working at IU. So one measure of success for supporting these activities is being able to successfully deploy them and run them and make sure they're stable. For the research component of the project, making sure the requirements and use cases are clearly understood by the computer science folks, and also to explain the computer functionality and features to the meteorology folks.	

Interview ID=2 31 May 2007	ANSWERS	ANNOTATIONS
<b>Q3 What are you investigating?</b>	<p>From the science perspective, we are trying to look for more ways of running the huge computational science models at high resolutions. That's been a big challenge. So running a weather forecast at storm scale and at tornado scale is something we've been trying to do.</p> <p>From the computer science part, we've been trying to make all these legacy Fortran codes and the legacy applications run in a Service Oriented Architecture. And providing users direct access to advanced computational resources from a portal-based environment. So users don't need to learn about how to use a particular cluster or job manager.</p> <p>So IU has been primarily the development center for the whole of the project. There are eleven universities in total – nine primary collaboration universities. Most of the non-IU groups are primarily application people, and data specialists and education folks. There are two big areas of work that we do at IU: one is the data side and the other is the workflow and orchestration side.</p>	
<b>Q4.2 How do you work?</b>	<p>At IU we have many students and staff working on these tools. So I take these tools and integrate them with the applications. I do integration and testing of all the different services and tools. So the CS people develop the tools and I apply them.</p> <p>For the CS people I am the user, and for the science people I act as a developer and answer their questions on the CS people's behalf.</p>	
<b>Q4.3 How do you keep track of interim results, if at all?</b>	I don't do much tracking of the development.	
<b>Q4.4 How do you test work-related hypotheses?</b>	<p>We have certain project goals. The science people come up with a use case. Then we drill down to figure out what components are needed to enable them to do their weather forecast.</p> <p>For example, we might determine that we need to search for X data, and then run Y computational model. So we take a data search tool and look at it to see if it needs to be modified for the use case. Then we apply it to a computation and run it on TeraGrid, for example, run it with Globus and see if we are getting the feedback.</p> <p>We then show it to the user and see if he's happy with the status and the monitoring and what he's getting back. If not, we'll go back and work with the developers to improve the user-friendliness of the tools.</p>	
<b>Q4.5 How do you document your results?</b>	<p>So once we create the tool and deploy it in the portal system, we create a screencast tutorial [<i>the project uses Wink software to create their screencasts, <a href="http://www.debugmode.com/wink/">http://www.debugmode.com/wink/</a></i>] to demonstrate how to interact with the tool. Then we put a link to the tutorial on the LEAD portal's help pages; by taking the tutorial the user can figure out the steps he needs to follow to use the new tool.</p> <p>Other than that, there is documentation that goes into the annual reports and into academic papers.</p> <p>We also track usage of the tools with monitoring software, and use bugzilla for tracking bugs and feature additions.</p>	
<b>Q5.1 In what ways do you interact with simulations in your work?</b>	So before applying the tools to the real models we apply them to simulated systems to test things like scalability.	
<b>Q5.2 How do you share simulations with others?</b>	<p>All the simulations are made available on the portal for people to use. The LEAD internal scientists need to go through an account approval process. Then we let them run large-scale simulations and real-time computations. We also grant limited anonymous access to the resources. There are limits because some of the simulations and production workflows can be very compute-intensive and time-intensive. So for example if they're doing a weather forecast, we only allow a prediction run covering the next twelve hours for a very small region or at a very coarse resolution.</p> <p>We know roughly how many resources are needed for the simulations. We leave allocations to the user on a semi-honorary basis, asking them to run once in a day.</p> <p>We get monitoring emails to understand what users are doing so if we need to we immediately disable their accounts. So far no one has abused the system so we've never had to disable an account.</p>	



Interview ID=2 31 May 2007	ANSWERS	ANNOTATIONS
<b>Q6.1 Describe how you interact with data in your work</b>	<p>The weather simulations we work with need various types of data. Some of the data are static and don't change, like the geographic information defining particular regions and also the survey information.</p> <p>We also have streaming data coming in from the remote real-time devices like the radar and many satellites and some other model data run by the NOAA centers. It comes from various sources.</p> <p>So we induce those data, then the user can query the data in a spatial and temporal way. An example query might be: I need the data for a particular region. So these data have spatial boundaries and are valid for only a certain time limit.</p> <p>So we take these temporal and spatial data searches and apply them to our workflows. So we</p> <ul style="list-style-type: none"> <li>- transfer those data</li> <li>- we address it to a certain location</li> <li>- we catalog it</li> <li>- we transform the data</li> <li>- after the computation is complete the data are transferred into permanent storage</li> <li>- and we catalog the location.</li> </ul>	
<b>Q6.2 How do you share work-related data with others?</b>	<p>We have different levels of data. Much of the input data we are integrating is community data, so anyone can search upon it and use it. Most of the data is available publicly via anonymous FTP.</p> <p>The data resulting from computations are catalogued in personal metadata catalogs. The user has the option to share it, but by default it is private. So the user must go explicitly to publish the data to the outside world, otherwise it is secure.</p> <p>The personal metadata catalog is built as part of the project. It is called myLEAD. Professor Beth Plale and her group develop the catalog at IU. myLEAD is one of the tools integrated as part of my work. I work with both data and orchestration groups, so we interact very closely.</p> <p>We also use some Unidata-based tools from the University Corporation for Atmospheric Research (UCAR) in Boulder Colorado. The UCAR folks are also a core data group in the LEAD collaboration.</p> <p>Basically Unidata produces a streaming tool that we use for radar data called Local Data Manager (LDM). Also we use a catalogue service from them called the THREDDS (Thematic Realtime Environmental Distributed Data Services) catalogue, which serves data via the OPeNDAP protocol. So there are some tools coming outside of IU too.</p> <p>The LEAD portal is the one stop place to access all of these tools.</p>	
<b>Q7.1 What resources do you use in your work today?</b>	<p>We have a LEAD testbed, which is distributed around five LEAD partner locations: Indiana University, University of Oklahoma, NCSA, Boulder Colorado and Huntsville Alabama.</p> <p>For LEAD computations we completely rely on TeraGrid.</p> <p>IU Data Services also provide some allocations for personal data storage.</p>	
<b>Q7.2 How do you share work-related resources with others?</b>	<p>All of the LEAD testbed resources are internal to the project.</p> <p>As part of the TeraGrid we have a LEAD Gateway Allocation. So our administrators control user access to the TeraGrid resources via the LEAD portal.</p> <p>The LEAD portal sits on top of Globus middleware, such as the CoG kit, GRAM and GridFTP. So:</p> <ul style="list-style-type: none"> <li>- the users authenticate with the LEAD portal (the portal uses the MyProxy-based PURSe credential repository)</li> <li>- then the portal as such authenticates with the Grid services using GSI-based authentication mechanisms.</li> </ul> <p>The LEAD portal is a one-stop place to access all the LEAD services. The portal is built using Service Oriented Architecture principles: we have application services and other persistent services, which in turn interact with the actual grid services on the TeraGrid. So as such, the LEAD portal is a lightweight portal framework.</p>	
<b>Q7.4 How do you locate available resources for use in your work?</b>	<p>We have a fixed set of resources from TeraGrid that we use, so we pick one among them. Many times based on the use cases, such as our UT activities, we reserve them (we do advanced reservation). And we pre-schedule that to run on those one or two resources.</p> <p>Otherwise we have an interaction tool from UCSB called Batch Queue Prediction (BQP). So we use that tool to determine the smallest queue wait time on TeraGrid and we submit jobs to that.</p>	

Interview ID=2 31 May 2007	ANSWERS	ANNOTATIONS
<b>Q7.5 What types of information do you need to know about a resource in order to determine if it is suitable for your work?</b>	<p>The first thing we need to make sure is that GridFTP and GRAM are available and up-and-running. There are many monitoring tools out there such as INCA-based services, and the TeraGrid user portal has some of them and the WebMDS is supposed to have monitoring information. But based on our practical experience we see that the frequency of those tools monitoring GRAM and GridFTP is not accurate. So we end up having our own tools to test in real-time whether or not GRAM and GridFTP are up and running. If they are down we immediately black-list that host until someone manually goes and brings them back up. Until they are back up we submit to a different resource.</p> <p>The reason we need to test in real-time is we have seen many examples where the monitoring tools aren't showing whether a service is really available. If you ping GRAM or GridFTP it works fine, but if you actually try to do some functional thing (like transfer a file or submit a job) it fails. And their error messages are very cryptic. The most common error we get is "login incorrect", but it has nothing to do with an incorrect login. It's something like a hardware problem, or there's a node goes down in a striped GridFTP server, or the allocation is out of the limit, or some scheduler is paused. For all these conditions we get the same "login incorrect" message. So these automated monitoring tools will not help. So what we do is prior to doing an actual real submission we do a test run. We try to remotely execute a /bin/date or transfer a file, and see if that worked. If so, then we transfer the actual file or do the actual job submission.</p>	
<b>Q8.1 What software do you currently use in support of your work?</b>	<p>We have a whole stack of service-oriented architecture built in the lab here:</p> <ul style="list-style-type: none"> <li>- a portal framework, based on GridSphere</li> <li>- a BPEL-based workflow engine built on a workflow orchestration language called Grid Process Execution Language (GPEL)</li> <li>- a web service library that takes any legacy application and converts it into a web service (the Generic Application Service Factory, a.k.a. GFac)</li> <li>- a workflow composition utility, which allows registered web services to be composed in a workflow and acts like a client to the workflow engine</li> <li>- and at the TeraGrid sites we have a whole stack of clients to GridFTP, and RFT and data registrations like RLS</li> </ul> <p>We develop Java portlets in JSF, JSP and Velocity and deploy them into the GridSphere portal framework.</p>	
<b>Q8.2 What scripting languages have you used in the past year?</b>	Perl, python, jython, shell scripts	
<b>Q8.3 What programming languages have you used in the past year?</b>	Java, Fortran	
<b>Q8.4 What workflow tools do you use in your work?</b>	Expire, GFac, GPEL	
<b>Q8.5 What parallel computing tools do you use in your work?</b>	Fortran, MPI	

Interview ID=2 31 May 2007	ANSWERS	ANNOTATIONS
<b>Q8.6 If the need for new software-based functionality arises in your work how do you acquire it?</b>	<p>First of all we look around to see what's available because we want to leverage what's going on. So we look through all the community tools and we completely evaluate three or four of them. We look to see how well it meets are needs.</p> <p>The LEAD project is highly demanding in terms of both science and computation. So sometimes we find we need to extend some of the tools and contribute them back to the community. If there's a good enough justification we build it from scratch.</p> <p>We evaluate software by comparing it to similar tools in the community. For instance if we were evaluating workflow tools: we'd take a look at Tavern, Kepler, Triana and other popular workflow tools. We would look at them with our requirements in mind: what do we want to achieve in the project? Continuing the workflow example we might decide that we need to enable long-running workflows as well as the ability to run the workflow as a background process. Perhaps too it should be dynamically composable, with support for both pausing and restarting the workflow. We started writing our current workflow tool four years ago because the tools of the day did not meet LEAD's requirements.</p> <p>Also whenever we write tools we try not to limit them to LEAD's goals, and to make sure that the tools are applicable to other domains.</p>	
<b>Q8.7 How do you share software with others?</b>	<p>The code and documentation for the software we're developing is completely available on the web.</p> <p>We participate in the TeraGrid Gateway and other calls, and tell other guys what's available.</p> <p>And if anyone is interested, or if anyone is reading our papers and coming to us, we completely help them use those software and toolkits. And if any modification is needed we put them in our bugzilla and consider them when we are developing and further enhancing the software.</p>	
<b>Learning about the user's problems</b>		
<b>Q9.1 What challenges do you face today in accomplishing your work-related goals?</b>	<p>The biggest problem we face is the reliability of middleware we depend on. Since the LEAD project has very broad goals and we had so much to achieve we completely relied on community middleware. We are building only what's absolutely needed. Unfortunately the software we depend on is not so reliable in the deployment sense. We are facing many hurdles and problems. I'll explain the deployment problems by using GridFTP and GRAM as an example. We've been working with them for a long, long time. But say when we deploy it and run with thirty concurrent users, it crashes badly and silently.</p> <p>So the deployment on the TeraGrid – the Grid middleware – is not working as well as we would like it to. We see the users of the TeraGrid resources are going on fine. People logging in and submitting batch jobs directly to the scheduler are happily using them. But in the metrics we can see that Grid middleware, for instance at NCSA, goes down for three days. It's the people relying on Grid middleware that are encountering these problems.</p> <p>For example sometimes GRAM submits the job fine, but it says the job never completes. But actually the job <i>has</i> executed and completed and exits fine. GRAM misses the status change, so it keeps telling us the job is active whereas the job actually returned from the backend.</p>	

Interview ID=2 31 May 2007	ANSWERS	ANNOTATIONS
<b>Q9.2 What types of information do you need in order to address the challenges you face today?</b>	<p>When everything is working, it's great. So a lot of what we need is more fault tolerance and improvements in the way errors and exceptions are handled. The errors can be due to hardware or middleware at any level. So when these problems are happening, for instance when hardware fails, the middleware we rely on gives cryptic error messages which we cannot read and parse automatically so that we can adapt to it. As I mentioned before the GridFTP "login incorrect" error does not provide us with sufficient information. In other contexts the source of login problems typically are on the client side, but in the GridFTP case it is often a server side problem. So what I would wish is when middleware cannot determine the particular error (and it's reasonable that it cannot determine everything), I would rather it propagate the original error message. Send it up the middleware layers of the architecture, instead of misinterpreting something and issuing a misleading error message.</p> <p>Another type of information that is lacking is documentation about errors. For example with GRAM, all we get is an error code and there is not enough documentation explaining the error. We then have to google to find out how other users handled the problem. Sometimes we even need to go as far as to dig into the GRAM source code to determine under what conditions the error code is sent. There is some documentation, but not at a level enough that we can use it.</p> <p>I would also find more tutorial-like information quite useful. For example I read the whole <i>Globus Toolkit 4: Programming Java Services</i> book and I practiced a lot of examples in there. This is kind of helpful and we would like more examples. Like when we are writing clients to a GRAM service we look for more tutorials or even CoG help in some sense. So more tutorials would be helpful.</p> <p>But the bigger missing documentation is in the troubleshooting area, where something is happening and we need to find out how to deal with it. Troubleshooting type of documentation is not only for me, but for system administrators – they struggle without this. Because whenever something happens we immediately post it to <a href="mailto:help@teragrid.org">help@teragrid.org</a>, and the system administrators try to figure things out.</p> <p>All the troubleshooting right now requires knowledge about the internals of the software. So only experienced people can troubleshoot right now. So if expertise is missing on the admin side, the issue keeps spinning for three or four days. I know a handful of users who can tell any problem happening and they can go fix it. But the new system administrators say, "We've been looking at the documentation and we're trying X, and we're trying Y, and we're trying Z, and none of them work. So troubleshooting for both users and the system administrators needs a lot of work.</p>	
<b>Q9.4 By contrast, can you provide examples of technologies you find very useful today?</b>	<p>I don't want to be completely on the negative side because we have been getting a very positive feedback in general.</p> <p>The Grid security infrastructure is one of the things we completely rely on and are building our tools on. For example all these remote job submissions and remote file transfers are completely based on GSI. So we absolutely love GSI-based authentication. This is something that has been really helpful and paved the way for the portal-based computation. Where you can run it right from the portal in a secure way on multiple resources.</p> <p>Even the GRAM and GridFTP are absolutely wonderful. When they work they give much better performance than other transfer protocols. The main problem is the reliability, but otherwise they're very good.</p> <p>The Globus installation has improved a lot. I've been installing Globus from the first version to today. The installation is very good, the documentation – the problems we encounter during the installation – they've been documented very, very well.</p> <p>As far as non-Globus tools, GridSphere has been very well-documented and active user forum. I like GridSphere's installation and documentation a lot.</p>	

Interview ID=2 31 May 2007	ANSWERS	ANNOTATIONS
<b>Q10.1 Can you think of any work-related tasks that decrease your productivity?</b>	Monitoring our infrastructure decreases my productivity. We have a big academic course in the spring for example, where my research goals and development tasks have to be ignored because the burden of supporting the tools is great. What I mean by supporting the tools is trying to see what's happening on which resource. The reason for this is if we see certain errors right up front we cannot directly take that and send it to, for instance, the TeraGrid helpdesk. I have to do at least an hour of digging. Because if I send directly an error message to the helpdesk they will reply, "This is something to do with your client side. There is something wrong." So I dig deeper and deeper and go through my usual tests, and see, "Oh this service is down. Ok here is what's happening." These kinds of things hamper my productivity a lot, I would say.	
<b>Q10.2 Describe repetitive tasks associated with your work</b>	Primarily what I do over and over again is deploy various workflows. Building different workflows for different categories of users. This is part of my role as a workflow developer. So I do that a lot – deployments and workflows. <i>[prompt asking for more information about "categories of users"]</i> One of the big challenges for us in the past few years has been trying to tackle a wide range of users: from high school students to advance research scientists. We have high school students for UT activities, we have undergraduate and graduate students for our spring national collegiate forecast, and the LEAD scientists who are doing their cutting-edge research and science.	
<b>Q10.3 Describe time-consuming phases of your work</b>	Composing workflows and making these legacy applications into services. I've been talking about the computational side, but there's a whole different world on the weather modeling side. For example the weather forecasting model called WAR. This work very time-consuming to compile and test. Deploying a weather model is my most time-consuming work.	
<b>Learning about the Globus user experience</b>		
<b>Q11.1 Which Globus data components do you directly interact with in your work today?</b>	GridFTP, DRS, RLS, RFT	
<b>Q11.2 Did you install the &lt;component&gt; client yourself?</b>	Yes	
<b>Q11.3 Did you install the &lt;component&gt; server yourself?</b>	No, for the Globus services many times we go with the GT4 Quickstart Guide lightweight installation for some of the services. For RLS and DRS we install it all ourselves	
<b>Q12.1 Which Globus security components do you directly interact with in your work today?</b>	MyProxy	
<b>Q13.1 Which Globus execution components do you directly interact with in your work today?</b>	GRAM2, GRAM4, MPICH-G2	
<b>Q13.2 Did you install the &lt;component&gt; client yourself?</b>	Yes	
<b>Q14.1 Which Globus information components do you directly interact with in your work today?</b>	Sometimes we use WebMDS and to get the hostnames of GRAM4 deployments on a one-off (as opposed to real-time) basis. Our collaborators at RENC1 are also beginning to use MDS4 for monitoring.	

Interview ID=2 31 May 2007	ANSWERS	ANNOTATIONS
<p><b>Q16 Why do you use &lt;component&gt; instead of an alternative technology?</b></p>	<p>GridFTP: primarily because it uses GSI authentication. If we are doing SSH and FTP we need to worry about authentication. Also, we use the APIs from Java CoG, and all of our services are Java based. Also, performance certainly matters, and GridFTP offers better performance than other transfer mechanisms.</p> <p>GRAM2: First of all the GRAM interface is really cool and provides us with one common interface for all the job managers. This is important for us since for our requirements we need to use multiple clusters and each has its different job managers. GRAM enables us to use the uniform mechanisms for both job submission and monitoring.</p> <p>It also supports the authentication mechanism that we like.</p> <p>GRAM4: Actually I prefer using GRAM2 over GRAM4. But certainly we are moving to a completely web service-based architecture, and we are inclined to use GRAM4 more. One of the reasons is it is web service-based. It has file-staging features, but we haven't used it extensively; we've tried it briefly.</p> <p>The other thing I would say is that we want to make sure we can interoperate with GRAM4. We have our own web service eventing system. We want to make sure instead of polling it that we are able to use GRAM4's push mechanisms. This is because when we are polling we miss state changes; so we are looking to GRAM4 down the road.</p> <p>RFT: we are certainly looking at RFT because we have a need for reliable transfer with the retries and exponential mechanisms. Its useful where we've done a high-costing compute job and need to transfer the results reliably. It also has a web service interface, which fits into the architecture we are moving towards.</p> <p>Also third party transfers are cleaner and better defined in RFT, as compared to third party transfers in GridFTP. RFT works better.</p> <p>RLS: We started using RLS because we are completing moving from using URLs to URI-based systems. We have this name resolution, and whenever we get a data product we register it so that it can have multiple replicas. The LEAD testbed can be distributed across multiple locations because of the replicas. The replicas also allow us to choose the fastest available compute server for our computations.</p> <p>DRS: We tried it but we are not using it in production. It supports distributed RLS, and it supports multiple locations, and it does the transfers with RFT. But we have different use cases and it wasn't a good fit and it wasn't quite ready at the time we were looking at it. The data movement capabilities weren't working when we looked at it, and so we had to move the data manually and register it with the RLS.</p> <p>Also there's a whole data research effort underway on the Unidata side and so we're using some of those tools too. I'm not a data person, so I'm not exactly sure why DRS is not currently used in production.</p> <p>MyProxy: MyProxy is really a very good technology. We have these scientific users, and it's really hard to convince them even to do a simple grid-proxy-init on the portal. So we've been permanently storing the Grid credentials for users and managing it ourselves in the portal. As a credential repository MyProxy has been great, and servicing our needs well.</p> <p>MPICH-G2: Because our application people asked for it.</p> <p>WebMDS: We use it to look at the gatekeeper version, whether the gatekeeper is running. As a high-level check.</p>	

Interview ID=2 31 May 2007	ANSWERS	ANNOTATIONS
<b>Q17 What are the major challenges you face using &lt;component&gt; today?</b>	<p>GridFTP: one big problem we have is with the firewalls, with active/passive settings. We need to have different combinations of active/passive settings depending on the hosts.</p> <p>For instance, for some host-pairs we need to make the source active and the destination passive, but for others we need to make both active. So it's been really crazy and we've had to do all sorts of hacks to switch settings.</p> <p>GRAM2: as I mentioned earlier, cryptic error messages and undocumented error codes are a problem.</p> <p>GRAM4: GRAM2 has been more stable and reliable than GRAM4. That is the only reason I prefer GRAM2 over GRAM4. I need at least 70% success rate to consider a service stable. Ideally we want it to be much higher, but with GRAM4 we are seeing a much lower success rate.</p> <p>I certainly don't want to blame everything on GRAM4. We've seen hardware failures on the cluster side. But I would say GRAM should improve the way it responds to hardware and network failures.</p> <p>The biggest concern for GRAM4, however, is the GRAM container goes into hibernation or stops for awhile without any log messages. And it just comes back by itself after a few hours. And that's been one of the things we've been following with the help desk.</p> <p>RFT: RFT is very good when you set all the optimal settings. On the default setting the performance is very bad compared to GridFTP. But if you tweak all the parameters we get the optimizations. So we need to find out and learn some external tools to provide these optimization values. For example we need to look more into MDS and see what are the optimal configurations to set between two hosts.</p> <p>Or software based on recommendation of the GridFTP developers called King Software that gives the bandwidth between two DNS servers. And also we've been looking at this NWS service from Santa Barbara where we can dynamically determine the striped bandwidth and specify the bandwidth size. So we need to do more work on our side, which we haven't yet done, to provide the parameters to RFT so right out of the box it is much slower than GridFTP.</p> <p>RLS: We tried to deploy RLS on a 64-bit machine, and during our critical production mode it did not scale beyond a certain limit. So very low scalability in 64-bit mode. We told the RLS developers and they identified some problems in the C globus IO libraries. They gave us some fixes and there is still an open bug report about it. In general the scalability issue has haunted us a lot and so we've had to find workarounds – on 64-bit machines. Things are fine on 32-bit machines.</p> <p>MyProxy: Can't think of any challenges.</p> <p>MPICH-G2: We've encountered deployment issues, such as:</p> <ul style="list-style-type: none"> <li>- the right intel compiler is not installed</li> <li>- the MPICH is not built in the CTSS version we get – we have to interact a lot with admins to get the right software installed</li> </ul> <p>For example on the Argonne cluster itself we had to wait for three months before get the right compiler and MPICH-G2 software were installed</p> <p>WebMDS: Not a heavy user of it, so really can't comment</p>	
<b>Wrapping-up</b>		
<b>Is there anything you'd like to say to the people who build software for use by people like you?</b>	<p>Many cool things are coming out of Globus development. I personally would like to have many of the basic services, like GridFTP and GRAM, be more reliable before I see more features coming out.</p> <p>Client side backward compatibility is important as well. When a new version comes out I should not have to rewrite my software. This has been a major concern in the past, but has been much better lately. If the clients can talk to the services in the same way and get the same functionality, that would be good.</p>	

### D.3 The Grid is a black box to me

Interview ID=3 1 June 2007	ANSWERS	ANNOTATIONS
<i>Pre-interview question: Do you interact directly with Globus software in your work today?</i>	no	
<b>Establishing context</b>		
<b>Q1.1 Please provide a one-minute overview of your project</b>	We have processing-intensive applications that involve mainly Monte Carlo simulations, permutations of datasets, and basically iterating over computer-intensive processes thousands of times. People do this daily, so we need a lot of computing cycles and computing nodes. That need is what started our working collaboration with the people of the Computational Institute at the University of Chicago.	
<b>Q1.2 What is the project's name?</b>	<i>Project name withheld at interviewee's request</i>	
<b>Q1.3 Which agency funds the project?</b>	NIDCD, which is part of NIH	
<b>Q1.4 What field does your project belong to?</b>	On the application side: Neuroscience On the development side: Computer Science	
<b>Q1.5 What is your job type?</b>	Scientist and Developer	
<b>Q1.6 How long have you been a &lt;job type&gt;?</b>	One year, seven months	
<b>Learning about discipline-specific goals and approach</b>		
<b>Q2.1 What are the main goals of your project?</b>	Establishing a framework where Neuroscientists, especially people who are involved in brain imaging, can store, analyze and share their data in an effective manner.	
<b>Q2.2 How will the success of your project be measured?</b>	Success in the short run will be measured by evaluating the achievement of certain milestones that we are funded for, as specified in the project grant submitted to the NIH. Success in the long run will be measured by speeding up analyses, assisting people to share information in the Neuroscience community, and by having people use the software we're developing.	
<b>Q2.3 What are the professional measures of success for you?</b>	Generating usable software product	
<b>Q3 What are you investigating?</b>	I study how language is understood in different brain regions	
<b>Q4.1 What is your method for investigating &lt;phenomena&gt;?</b>	We use one method which is called fMRI (functional magnetic resonance imaging) We also use a method called ERP	
<b>Q4.2 How do you work?</b>	We collect the data using machinery. We collect observations, for instance, on how people understand a given word. I say a word "dog, dog, cat, cat" and see how the brain reacts to them. The fMRI identifies reactions in the brain in a non-invasive way. We have many observations, so we have general scientific procedures that extract the signal from the noise. Our data tells us which parts of the brain react to words like "dog" and "cat". It's basically analyzing digital data stored in files. We take them, we average them, we run processes that help us filter noise and we try to get at the signal.	
<b>Q4.5 How do you document your results?</b>	We have either manual or digital lab notebooks where people document what they're doing. Or in the worse case, README files are created in the directories where results are stored. I use an electric lab manual and some README files. Many of the results are stored in databases. That's one of the things that started the project - when we began storing the data in databases and noticing that helps preserve the data and make it accessible later on.	



Interview ID=3 1 June 2007	ANSWERS	ANNOTATIONS
<b>Q5.1 In what ways do you interact with simulations in your work?</b>	<p>We use simulations like they're typically used in science. We use mainly Monte Carlo simulations – this actually might differ from how they're used in high-energy physics – because we don't use them to simulate the results of a certain model.</p> <p>So, as to what we do: We assess some properties of our data. We generate fake datasets where data are actually samples from a normal distribution. And then we try to see to what extent the data we have differ from random datasets.</p> <p>A concrete example: We collect data about the brain, and we have 100,000 units in which we measure the brain – tiny parcels of information. So we try to identify groups of parcels that are active together in the brain; such a grouping would be called an active region. So if you hear language, an active region would be some lobe in the brain.</p> <p>Because we sample so many units of resolution – let's say, in our example, 100,000 – even if you were just by chance sampling from such a enormous space, you would get some units that are active in close proximity to each other, that are in fact false positives.</p> <p>So we generate random datasets to understand what properties such false positives might have. That is, we generate false datasets to know what cluster sizes of active units one might get purely by chance. And once we learn about that from the simulations we go to our own data and see what's not likely to be due to chance, as compared with the simulation.</p>	
<b>Q5.2 How do you share simulations with others?</b>	The project currently includes me, two other developers and the PI on the Neuroscience side. I am the only Neuroscientist who works with the data within the project, so within the project I don't really need to share them.	
<b>Q5.3 How do you interact with inputs to your simulations?</b>	<p>We generate the inputs to the simulation in different ways. Sometimes we will just sample data randomly from a normal distribution, and do this many times in Monte Carlo simulations. So not much interaction there.</p> <p>In other cases, each simulation will be defined as some permutation of the original data, within a general statistical context of permutation-based testing. That is, we generate a permutation using some algorithm from the original data and analyze that permuted dataset (just like we do with our real data).</p> <p>We do that many, many times to see if our real data differs from permutations that one can generate out of it. So generating a permutation is a somewhat more complex matter than just generating a noise from distribution.</p>	
<b>Q5.4 How do you interact with the output of your simulations?</b>	On the broadest level we have an algorithm that looks at the output of the simulations for certain properties. We build a distribution that characterizes that output from all our simulations. So the output of 10,000 simulations will be a distribution of a certain parameter we're interested in. On the basis of this summary distribution that we construct, we make inferences from our data.	
<b>Q6.1 Describe how you interact with data in your work</b>	All data are digital. We use open source software that takes the brain images we collect at the fMRI scanner, and build files out of them that we manipulate using open source software. We then apply a series of transformations to the data until we get results that we can draw conclusions from. So we set up workflows with input and output running on unix machines.	
<b>Q6.2 How do you share work-related data with others?</b>	<p>A typical way, obviously, when a project is finished you publish the results in a scientific journal.</p> <p>Before the project is finished when you're in the process of understanding your data and you need to share data, there are two ways:</p> <ol style="list-style-type: none"> <li>1) You send people a file and maybe the workflow script that generated the file, or some hints about what you're doing to this datafile and how you're analyzing it. This is what we used to do about a year or two ago.</li> <li>2) The newer way that is tied to the project is basically storing the results of certain stages of the workflow in databases which are accessible over the internet. Then accessing the database with SQL queries.</li> </ol>	

Interview ID=3 1 June 2007	ANSWERS	ANNOTATIONS
<b>Q6.3 By what mechanisms is access to your work-related data controlled?</b>	In you send a file in the first scenario there is no security. You send it via email or put it on an ftp server, which is anonymously accessible. If it's in a database, then the person needs to know the username and password in order to do the query. That's the security layer on the database.	
<b>Q7.1 What resources do you use in your work today?</b>	Me personally, I use resources that have been made available to the lab in which I work by the Computational Institute. The lab devours enormous amounts of space. I think now we probably take between 3 and 4.5 terabytes, which we pay for. All of us use the local teraport resources, which is a 240-250 processor cluster hosted at the Computational Institute at the University of Chicago. Most recently we've started, as part of the project, to make the software that we're working on runnable over the general Grid (TeraGrid) resources, rather than just the teraport. These are three classes of resources that we use. We also have local unix servers in the lab that we do processing on, but the bulk is done on the teraport.	
<b>Q7.2 How do you share work-related resources with others?</b>	I don't have any local resource that's dedicated to me. There is a lab, it has machines. Each person works on their own computer. And there are some generally-available unix servers from the days we started using the teraport. If someone needs to run a process or some software that's on the teraport, they run it locally.	
<b>Q7.5 What types of information do you need to know about a resource in order to determine if it is suitable for your work?</b>	Because of how the system is being developed at the CI, all of that is kind of a black box to us. So access to grid resource, job submission, setting up the jobs, staging out the jobs and wrapping them up back to our machines is all something that we don't have to deal with. That is something that is basically mediated by an application developed at the CI, which is called the Swift framework. We use that as our go-between for everything between us and the Grid. We're kind of blind to it. They will locate the resources for us, and I think they will also do a lot of the authentication and the certificates – basically the overhead that's involved with Grid security.	
<b>Q8.1 What software do you currently use in support of your work?</b>	We use three main types of software: One is a general-purpose mathematical language, which is called "R" ( <a href="http://www.r-project.org/">http://www.r-project.org/</a> ). It is the open source analog of "S" and "S-Plus" software. Another one is software that is dedicated for analysis of brain images, which is called AFNI, developed by the NIH. And the third one is software that lets us manipulate the anatomy of brains, so to speak. It's called FreeSurfer, and it's also funded by NIH and developed by a research group at Harvard. As to the lab notebook: every person uses his own. I use a Mac application called "NoteBook" from a company called Circus Ponies ( <a href="http://www.circusponies.com/">http://www.circusponies.com/</a> ). It's basically general software that lets you take notes in a structured way. As to databases: this is actually something that started this project, and I'll take the blame for it. So basically I developed a system here that uses relational databases to store and make available fMRI data. This is a basic principle in the current project: - we store our data in relational databases - when we analyze the data, we don't read them from files - we read them by issuing SQL queries, retrieving the data that we need from whatever database Typically for each research project we construct the database that will contain as many tables as we need. So it's a central part of how we work during data analysis.	
<b>Q8.2 What scripting languages have you used in the past year?</b>	R, perl, shell script, awk, ced	
<b>Q8.3 What programming languages have you used in the past year?</b>	perl, R	
<b>Q8.4 What workflow tools do you use in your work?</b>	We're not using Swift on a daily basis. We're still in the process linking between Swift and applications we need to run	

Interview ID=3 1 June 2007	ANSWERS	ANNOTATIONS
<b>Q8.5 What parallel computing tools do you use in your work?</b>	We don't use MPI libraries or Atlas libraries or any parallel coding tools. The closest we can get is to submit jobs to multiple computers.	
<b>Q8.7 How do you share software with others?</b>	All the software we use is open source.	
<b>Learning about the user's problems</b>		
<b>Q9.1 What challenges do you face today in accomplishing your work-related goals?</b>	Conducting analyses rapidly is the major bottleneck. Getting simulations done quickly and finishing analyses quickly are the hurdles.	
<b>Q9.2 What types of information do you need in order to address the challenges you face today?</b>	I don't think we need more information. If I need anything, it's more compute cycles, in order to break up large jobs into smaller packets.	
<b>Q9.4 By contrast, can you provide examples of technologies you find very useful today?</b>	Grid computing, computer clusters tied in with relational databases are a good synergy today for scientific work.	
<b>Q10.3 Describe time-consuming phases of your work</b>	Analyses take time and simulations take time, and sometimes they take two days, and if we had two hundred processors they would take two hours. But we're not complaining. Apart from that there's nothing, at least personally, that holds me back from what I want to do.	
<b>Learning about the Globus user experience</b>		
<b>Wrapping-up</b>		
<b>Is there anything you'd like to say to the people who build software for use by people like you?</b>	<p>I participated in the Midwest Grid workshop about two months ago. I think there is a gap between people in science who might use cluster or Grid services and the people developing the software. I think my project is lucky to have intermediaries working on the project: Grid experts who are basically translators. They are translating our needs to the Grid or cluster-level and communicating back to us.</p> <p>I don't know how often this occurs. I do not know of a general mechanism that scientists can use to communicate need scenarios and discuss with experts developing the software on how the needs might be met.</p> <p>An example need scenario: I have 2,000 simulations, and I work in a small college where we have 20 pentiums. So, how do I get a grid solution in place for the student body to be able to run the simulations?</p> <p>Perhaps you could have a person who knows the Grid side of things, maybe with some scientific background, who talks on the phone with the person who needs to achieve the computation. Talks with them on the phone for an hour or so, sets up a general mechanism so that the non-expert understands what they need to do to get the Grid working for them to solve their problem.</p> <p>It seems to me that in order to get Grid solutions you have to be pretty tech savvy. Getting the certificates, doing the job submission, doing the DAG of the workflows on Condor, managing the security: all of that seems to be an enormous barrier for actually getting jobs done on the Grid. It didn't seem to me like there is any mechanism on the TeraGrid or OSG to sit down and work out how to solve problems together.</p>	

## D.4 ▼ The reason my tasks are so time-consuming is failure

Interview ID=4 1 June 2007	ANSWERS	ANNOTATIONS
<i>Pre-interview question: Do you interact directly with Globus software in your work today?</i>	yes	
<b>Establishing context</b>		
<b>Q1.1 Please provide a one-minute overview of your project</b>	<p>ENZO is a code that is used in computational astrophysics – theoretical work. It has components that can be used for general problems in astrophysical gas dynamics. Also it can be used for doing very large-scale calculations of cosmological structure formation. These calculations, particularly the latter, are currently running at the leading edge of what the NSF and DOE are capable of providing, in terms of computing centers.</p> <p>These are generally batch-oriented processes that have to be run in many chained, sequential jobs. The data output is enormous – typically on the order of a hundred terabytes. We archive almost everything that we output. At my home institution I personally own something in excess of four hundred terabytes of ENZO data in the archive at present.</p> <p>ENZO will, we hope, become a petascale application. And we can expect all of its computational and data demands to grow possibly by as much as three orders of magnitude over the next three to five years.</p>	
<b>Q1.2 What is the project's name?</b>	ENZO	
<b>Q1.3 Which agency funds the project?</b>	The National Science Foundation, the Department of Energy	
<b>Q1.4 What field does your project belong to?</b>	Astronomy and Astrophysics	
<b>Q1.5 What is your job type?</b>	Developer, Scientist	
<b>Q1.6 How long have you been a &lt;job type&gt;?</b>	Six years	
<b>Learning about discipline-specific goals and approach</b>		
<b>Q2.1 What are the main goals of your project?</b>	<p>As far as the cosmological part of it goes, we're trying to understand the hierarchy of structure formation on various scales, from the larger scales and the universe, down to galactic scales. This is a tremendous number of orders of magnitude in physical scale - in space and time. The code that we're using, ENZO, is a product of at least twenty years of evolution in those areas. It's the work of many people. We hope to be able to account for observations, to determine the cosmological parameters of the universe we're living in. To provide theoretical underpinnings for observations from things like the Hubble space telescope, or the James Webb space telescope, or any of those major projects. Bottom line is that cosmology is not actually useful to anybody, but it's fun.</p>	
<b>Q2.2 How will the success of your project be measured?</b>	<p>The primary measure for us is publishing in the peer-reviewed Astrophysical Journal, Physical Review, Nature – major publications like that.</p> <p>One version of ENZO is also a community code, so it is used worldwide by a number of people, usually not at the scale we are running. We're the developers here in San Diego, so we tend to run the pioneering calculations. In fact that's my job: I do the first run of the next largest scale that we manage to run. So I hope to be the point man in developing the petascale version.</p>	
<b>Q2.3 What are the professional measures of success for you?</b>	Peer reviewed publication. Getting some credit for the number of years I've put in to the technology of ENZO, and seeing it used. Continuing to increase its fidelity in solving these physical problems.	

Interview ID=4 1 June 2007	ANSWERS	ANNOTATIONS
<b>Q3 What are you investigating?</b>	<p>My focus is on the High Performance Computing aspect of ENZO. That is, at the moment we can run right at the largest scale that the machines available to us can handle. We're ready to move up to the next scale, which will be represented in the first instance by TACC Ranger, and the other NSF Track 2 systems that will follow on behind that.</p> <p>Our goal will be to be able to run problems that continue to match the most powerful resources available in the US, whether they're DOE or NSF. Because astrophysicists have the advantage of knowing our governing equations, unlike fields like biology where they're not really sure what they're looking for.</p> <p>So we can always adapt very quickly to the largest computer hardware available. And we don't see any end in sight yet, really, in terms of our computational requirements. If you could provide us a usable computer today that was a million times faster, we would just simply use it. A billion times? I'm not sure – I might have retired by then.</p> <p>In astrophysics, the physical scales are extreme. Also the classic problems are involved, ranging from nuclear physics (which is well understood) to gravity at the larger scale. In the middle of that range you've got difficult issues that come up in astrophysics everywhere: rotating magnetized fluids that are properly constrained under a gravitational field of some sort.</p> <p>The big challenge for us is in three-dimensional radiative transfer, which is how light basically interacts with a fluid medium. ENZO is being extended right now to incorporate these frontier pieces of physics, that up to now we haven't had enough computer power to include. So, there is a finite list of things we'd want to add. I wouldn't want to be too strict about this point, but we'll probably add most of the physical things we want to add in the next two to three years. And then it will be mainly a question of just how much computer power can we get our hands on.</p>	
<b>Q4.1 What is your method for investigating &lt;phenomena&gt;?</b>	<p>There's a group of us here at the Laboratory for Computational Astrophysics. Within the US there's an extended group of at least a dozen of us who work either with developing ENZO or using it in specific sub-domains of astrophysics and cosmology. We work together to publish papers and make academic progress.</p> <p>We share the results freely amongst ourselves; most of our results are archived here at my home institution, and they're freely available to anyone who wants them.</p>	
<b>Q4.2 How do you work?</b>	<p>ENZO is a very big code. It has a lot of components. The things that are holding us back fall into quite specific areas to do with the way the code has evolved over time. And we need to restructure what we're doing to remove those obstacles, and at the same time provide a clear structure for how to augment ENZO and some physics components that may be added at a later stage.</p> <p>So really, it comes down to being familiar with the workings of the entire code. Also the way the code is used at full scale, the target architectures, compilers and all their idiosyncrasies.</p> <p>Compared to a professional software developer, we probably don't use anything you would consider advanced computational science technology. Obviously there are components that we rely on, some of which were developed a long time ago, right? For instance: a Fourier transform is a Fourier transform. And we're using some solvers that come from the Department of Energy. So where things are available that will work, we do use them. It's overall management of the whole thing – it's a question of stepping up.</p> <p>Perhaps this will address the question better: when I started here, we were doing problems about 4,000 times smaller than what we can do now. And every factor of two or so in the scale of the computing system from where they were running (~64 CPUs) up to present (4,000 CPUs) has exposed some new algorithmic deficiency which I've had to fix.</p> <p>So we work by extending what we do. We work by growing up the scale of the problem, which is constrained by the scale of the computer we have access to (and funding, of course.)</p>	
<b>Q5.1 In what ways do you interact with simulations in your work?</b>	<p>We have no abstract model of ENZO used for testing the simulator itself. That might be something we choose to try to develop. We have occasionally tried to work with a group here to get profiles from ENZO. But that hasn't been terribly successful because ENZO is a very poor thing to benchmark.</p> <p>The only benchmark that makes any sense is "how long does it take to go from the beginning to the end of an entire simulation sequence". So if you were to ask, "How many teraflops does ENZO do on how many processors?" my response would be:</p> <ul style="list-style-type: none"> <li>• Do you mean at the beginning of the simulation?</li> <li>• At some physically interesting point, such as where the quasars light up and start reionizing the universe?</li> <li>• What kind of problem do you want to do? Do you want to do one with an adaptive mesh or one with a fixed mesh?</li> <li>• What physical things do you want to include? Do you want to include star formation and feedback?"</li> </ul> <p>Any of these choices would affect a benchmark. Changing any of these would invalidate the conclusion beyond a single test. So I try to discourage people from "benchmarking" ENZO.</p>	

Interview ID=4 1 June 2007	ANSWERS	ANNOTATIONS
<b>Q5.3 How do you interact with inputs to your simulations?</b>	<p>The way we start is a few of us will be involved in thinking about the next physics problem we want to attack. And that will involve a lot of discussions before we commit to doing something – to design what it is that we want to do from a physics point of view.</p> <p>It may seem strange, but the input to an ENZO simulation consists of possibly between one and two hundred parameters in an ascii file. Once that ascii file has been edited by one of us and checked by all of us, that's where it starts. It is very small – a kilobyte, ten kilobytes or something.</p>	
<b>Q5.4 How do you interact with the output of your simulations?</b>	<p>Ah, the output. It consists of a time series of three-dimensional physical fields and particle positions. So for example, the current simulation I'm running at NERSC consists of a <math>2000^3</math> non-adaptive grid with eight billion dark matter particles in it. We're tracking about ten or so physical three-dimensional fields (things like density, temperature) and also the positions of those eight billion particles and their velocities.</p> <p>So each dump at any instance in time is about 700 gigabytes (700G) of HDF5 files. We also use the same mechanism for checkpoint and restart. The rate at which those 700G files are produced depends on the phase of the calculation. On a 2,000-processor machine the average dump rate is approximately once every one or two wallclock hours.</p> <p>Then all that data has to be moved from where it is generated to where it can be analyzed and archived. Many centers, for example, have excellent MPP computers, but then fail to provide any suitable computers for doing the analysis.</p> <p>Case in point: we prefer to move data to SDSC where we have a number of very large shared memory systems (IBM POWER4 pSeries 690 Plus) that enable us to put those 3D fields into memory directly and manipulate them. It also has the advantage of being on the same federation switch as the main compute engine of DataStar, which is still pretty close to the state of the art for doing these simulations. It is also on the same system that manages our HPSS archival system. So things are very well integrated at SDSC.</p> <p>However, if we are forced to do the calculation elsewhere, the question becomes "How practical is it to get the bits back to SDSC in order to do something with them?"</p> <p>Now, we are planning to do more development of the analysis codes so that they can run concurrently with the simulation, because this model simply won't work at petascale. It will not be physically possible anymore to move the data. It may not even be physically possible to store the data, even temporarily for more than a few days given the rate at which it would be produced. And it will be economically unfeasible to archive it. Probably practically impossible to archive it as well.</p>	
<b>Q5.5 By what mechanisms is access to your simulations controlled?</b>	<p>We have accounts with allocations of service units coming from several different sources. For example, we won a NERSC INCITE award, and that is currently running on computers at NERSC and Berkeley [<i>Berkeley National Lab</i>]. We are in a collaborative project with Lawrence Livermore National Laboratory, and we run simulations on machines there. That system is not classified exactly, but it is restricted access.</p> <p>Then within NSF we have accounts on four or five or six machines that are spread across the centers: Texas, NCSA, Pittsburgh and San Diego.</p> <p>Generally speaking we simply log into to these machines using secure shell from our laptops or machines at our home institutions. We log in individually and manage the simulations. I hope I've already made it clear that any one simulation may have to be run as twenty or more sequential batch runs. So there's a lot of waiting. Which is why we each individually use different systems across the NSF that were chosen for political expediency, and also as in some cases for their unique computational characteristics.</p> <p>It's really quite a manual process. Workflows and so forth don't seem to really help in this arena. We have a workflow group here, but they're not interested in talking to us, as our work involves batch computation. We usually just want the entire machine to ourselves for as long as we can get it.</p>	
<b>Q6.1 Describe how you interact with data in your work</b>	<p>One of the things we do to make sure the simulation is progressing correctly is use our own tools to make three 3D volumes of specific fields, or projections through 3D objects. We then make graphical images usually using IDL software running on one of our p690 machines here.</p> <p>The offline analysis of the whole thing - extracting all the science – can take months to years after a simulation is complete. Specialized tools developed by individual researchers may be used in the analysis phase; these tools are beyond the scope of ENZO itself. We don't know the details of how people in Germany are examining the results an ENZO output. We have no idea.</p>	
<b>Q6.2 How do you share work-related data with others?</b>	<p>Certainly all the major simulations I've been involved in are all world readable in SDSC HPSS. On a particular machine we tend to be in the same unix group so we can all see the output.</p>	

Interview ID=4 1 June 2007	ANSWERS	ANNOTATIONS
<b>Q7.1 What resources do you use in your work today?</b>	<p>Our major DOE compute resources:</p> <ul style="list-style-type: none"> <li>- NERSC's Seaborg, which is an IBM POWER3 with about 6600 CPUs. This is quite an old system but it is extremely reliable. It has IBM GPFS as the parallel file system. HPSS is its archival system</li> <li>- We use NERSC's Bassi, which is a POWER5 with very similar characteristics for some smaller-scale stuff</li> <li>- I use an SGI Altix system at NERSC for retrieving data from their HPSS archive and transmitting it to SDSC</li> <li>- At Livermore [<i>Lawrence Livermore National Lab</i>] I use a system called Thunder, which is a 4,000 CPU Itanium 2 cluster running Linux Lustre parallel file system</li> </ul> <p>Our major NSF compute resources:</p> <ul style="list-style-type: none"> <li>- The primary system we use is DataStar, which is a 2,400 processor IBM POWER4</li> <li>- We use an SGI Altix at NCSA called Cobalt, which is a 3 terabyte shared memory system with 512 processors</li> <li>- We use the Cray XT<sub>3</sub> at Pittsburgh Supercomputing Center; that has 4,000 processors and Lustre file system</li> <li>- We use TACC Lonestar, which is the latest generation of Intel Xeon cluster; that has about 6,000 CPUs and also uses a Lustre filesystem</li> <li>- We have an IBM Blue Gene L here at SDSC, which I haven't used a great deal for production, but I did port ENZO to it and it's ready to use if we can find something that will fit in it. By "fitting" I am referring to memory.</li> </ul> <p>Memory is our largest constraint. IBM Blue Gene L has only 512 megabytes per node, and we don't really have any interesting physical problems that will fit into that node size. So at the moment we really haven't been able to use SDSC's Blue Gene for much, apart from verifying that we could use it if we ever had a problem that would fit.</p> <p>Additional resources include the predecessors of the above machines. I ported ENZO to the Cray X1E as well, using a system at Cray.</p> <p>So that's quite a lot of cycles we have access to. The total amount of time we have access to is somewhere between four and eight million CPU hours per year.</p> <p>Storage systems: we use local storage where it's available (local archival systems), but usually only as a safety measure. Most of our data actually comes back to SDSC. The reason for this in some cases, for example, Livermore is happy to give us cycles but will not accept the responsibility of storing our data. So we have to move it to SDSC – it was an explicit part of the project to do that.</p> <p>So we depend pretty heavily on long distance networks.</p> <p>We don't have any real-time data like sensors – this is just straight computation.</p>	
<b>Q7.2 How do you share work-related resources with others?</b>	<p>Theoretically, within the NSF anyway I believe, the PI would have a mechanism within the accounting system to restrict us individually. But actually we just work by a gentleman's agreement not to use all the time. We usually ask the PI if it's appropriate to use a certain chunk of time for a particular project.</p> <p>In some cases, like the NERSC INCITE run, the entire allocation was dedicated to a single run of ENZO on one problem. So we haven't had to worry about who's using what. I run that simulation. Likewise at Livermore I'm the only user so I don't have to worry about using someone else's time.</p> <p>Within the NSF specific graduate student and postdoc projects were spelled out in our NSF proposal, and also awarded specific time in each sub-section in that proposal. So we try to carry those projects out because in writing next year's NSF proposal it's good to be able to say we requested half a million SUs to do something and we did it and here are the results.</p> <p>So formally there's nothing to restrict us. Somebody could go crazy and try to use up all the cycles. But when you have an allocation of millions of CPU hours a year it's actually a tremendous problem to expend them quickly. For example on DataStar if I used the whole thing it's only 50,000 hours per day (if I could even get it for a day, which I never can.) So we're self-policing - we don't have to worry about this problem very much.</p>	

Interview ID=4 1 June 2007	ANSWERS	ANNOTATIONS
<b>Q7.4 How do you locate available resources for use in your work?</b>	<p>We applied for them: we wrote an NSF proposal to the Large Scale Resource Allocation Committee (LRAC). They meet twice per year; we're on the cycle where the awards are made in March. We write a ten- to fifteen-page proposal, which is a renewal for us so it might be a little bit less. We spell out precisely which calculations we wish to do in the coming calendar year with an estimate of what those will cost.</p> <p>And those proposals are peer-reviewed by the LRAC/XRAC/MRAC committees. I used to be a member for many years, so I'm very familiar with that process. Those proposals are peer-reviewed: they are chewed over, hacked apart and severely criticized by a large committee. The committee typically meets at one of the centers except the September meeting always happens at NSF's headquarters in Washington DC. And that committee recommends what allocations are made for the proposals, and also has the power to decide how some of the resources should be tied to sub-projects within a proposal.</p> <p>So the process for ENZO begins in January or so when we sit around and decide what it is that we want to do next year (in terms of physics.) Then we try to work out exactly what that will cost in terms of CPU hours and so on, and which computers can be used. Some problems are generic enough that they can be run anywhere. Some projects require a specific thing. For example the SGI Altix machine, called Cobalt at NCSA, is a large shared memory system that is highly advantageous for certain physical problems that we're interested in. Whereas other problems could run on a sufficiently large cluster-type machine anywhere.</p> <p>Some of the problems produce a whole lot more data output than others. Those are best done at SDSC because we can manage the volume of output. So when we are finished writing that proposal, what we have is a list of projects and a list of machines, and for each machine we have an estimate of how much time we want on it. The proposal goes off to the committee and the committee either approves or rejects it. Sometimes the committee is forced to provide an allocation on a different machine than we requested, but hopefully one with similar physical characteristics.</p> <p>This is actually tremendously inconvenient, and is one of the ways that the NSF allocation structure hinders computational science. We are forced to request resources from all four of the major national centers because no one center could provide the resources that we need for a year. We could physically do every calculation we're interested in on SDSC DataStar, but we would require four or five million hours per year, which would be somewhere between thirty and fifty percent of the entire machine. So you understand that we can't get that. It wouldn't be good for the center to only have two applications.</p> <p>Of course, this is a lot of the underpinning for TeraGrid. Because we're forced to use computers all the way across the US that aren't necessarily suitable for what we're doing – it's a political solution to a political problem – that politically justifies TeraGrid.</p> <p>Nirvana for us would be a single national supercomputer center that had a suitably configured machine of awesome capability where a few percent of that would be enough for us to get everything done at one place. Because pushing these bits across the country with the existing networks is terribly painful.</p>	



Interview ID=4 1 June 2007	ANSWERS	ANNOTATIONS
<b>Q7.4 How do you locate available resources for use in your work?</b> [continued]	<p>The things I'm wishing for don't really exist. On the other hand, it would be splendid if there were just one really large distributed memory system that was sensibly configured. Sensibly configured in this context means:</p> <ol style="list-style-type: none"> <li>1) the aggregate memory is large enough</li> <li>2) the per-node memory is large enough</li> <li>3) the interconnect speed is not a constraint, relative to the computational ability of the machine</li> <li>4) the parallel IO systems are sufficient to get the data out of the machine</li> <li>5) the archival-type storage is sufficient to maintain the usability of the whole system (e.g., if one of our runs uses up 90% of the available disk space, it will need to be archived somewhere so someone else can use the machine)</li> <li>6) and also truly prodigious long-distance networking and local receiving centers, in order to do some local work. I realize this is a tremendously difficult technical thing, because to drive sufficiently broadband network requires a large amount of computers on the receiving end, so that's not something that scales very easily.</li> </ol> <p>If you go back to the beginning of the centers program in 1985-86, every center pretty much got a state-of-the-art supercomputer. And they all had their own lesser machines for doing the kind of data manipulation things you want to do. And they all had good visualization centers. In those days it wasn't uncommon to get in a plane and travel to the center.</p> <p>We're almost returning the same kind of model because with the development of the planned Track 1 petascale system from NSF, the output of a suitably scalable application on that system will be so huge no one will be able to move it. You're going to have to travel to where your result is again. This is a real issue for us. It's inefficient.</p> <p>To return to the issue of locating resources today: the allocations are made annually. It's quite difficult to change them. It would certainly be extremely difficult to get more. All of the machines are allocated 100%.</p> <p>We do have good personal relations with some of the people at some of the centers, where we can discuss w/ them what we might be able to get. Suppose we wish to use the Cray XT at Pittsburgh. If we wish to use more than a few percent, then it would be a sensible to talk with them and see if we could get five percent of all the cycles. So there's quite a bit of personal maneuvering that goes on, but once the proposal's written, that's pretty much it.</p> <p>Then the allocations committee makes the allocations on particular machines at particular sites, and those sites are more or less obliged to provide them. So every site has an allocations coordinator who tells the allocations committee how many millions of hours are available on each machine at their site. Part of the LRAC process is to make the request fit within the global pool of cycles available to the NSF.</p>	
<b>Q8.1 What software do you currently use in support of your work?</b>	Fortran90, C, C++, HDF5, IDL, a few graphics packages that come out of the DOE. We will be using the hyper solver package from Lawrence Livermore Lab	
<b>Q8.2 What scripting languages have you used in the past year?</b>	csh, bourne shell	
<b>Q8.3 What programming languages have you used in the past year?</b>	Fortran90, C, C++	
<b>Q8.4 What workflow tools do you use in your work?</b>	none	
<b>Q8.5 What parallel computing tools do you use in your work?</b>	Obviously MPI We'll be looking at doing some hybrids work – in the past I have used Open MP. We do use Open MP in a couple of minor applications that support ENZO. But as of today, we just rely on MPI.	
<b>Q8.6 If the need for new software-based functionality arises in your work how do you acquire it?</b>	<p>That's a very difficult question to answer. If it's something like a solver package, like for example <i>hypre</i>, we get it from colleagues at say, Livermore.</p> <p>We wouldn't ever bind ENZO to anything that required money. We try not to bind it to anything that has any restrictions on its use at all. So my philosophy has been to not attach it to anything that is license-able in any way, beyond such things as compilers (which you would expect every site to have.) The code is completely generic and clean and has as few site-specific hooks as possible.</p> <p>Looking to the future, one thing that I'm very interested in adding and using in petascale ENZO would be Unified Parallel C (UPC). But this is so intimately related to compilers, that it's really the supercomputer vendors that must provide it – integrate it with their compiling system.</p>	

Interview ID=4 1 June 2007	ANSWERS	ANNOTATIONS
<b>Q8.7 How do you share software with others?</b>	Within the group we use SVN for source control. And those with which we wish to share have the password.	
<b>Learning about the user's problems</b>		
<b>Q9.1 What challenges do you face today in accomplishing your work-related goals?</b>	<p>Mmm that's a juicy question. Our immediate challenge is we need access to a machine that has more aggregate memory, more faster processors, a higher performance parallel filesystem that actually works – just more, more, more of everything. But I do want to add a caveat - because I mentioned earlier that Blue Gene is not a usable system for us. If the machines are all going to be multi-core, multi-node, then per-node memory matters to us to a considerable degree in addition to the aggregate memory of the system. That's something I find that people overlook to a great degree, but it's probably the single most important thing to us.</p> <p>The other constraint that I feel terribly is the reliability aspect of these large-scale systems. I'm running in this 2,000-4,000 CPU range at the moment. And within the next year we expect that to go up to at least the 32,000 CPU range, if not a factor of two more than that. The unreliability that I see in filesystems, even in batch-process launching systems, disks, monitoring tools – you name it. Nothing really works reliably at the 2,000- or 4,000-processor level today. I am extremely doubtful about it working at a level ten times greater than that.</p>	
<b>Q9.2 What types of information do you need in order to address the challenges you face today?</b>	<p>The way the centers work all the information comes down and there's no feedback, this conversation notwithstanding, from the poor users at the end of this who are forced to use poorly designed and inadequately supported computers. And they suffer terribly in loss of scientific productivity dealing with the endless failures at every level of these systems.</p> <p>There is no feedback from the users to the center management or to the NSF, in terms of the cost in human resources in using these systems. The current round of the NSF program is a perfect example: this obsession with buying a petaflop computer for political reasons, presumably to brag about it internationally or something. With:</p> <ul style="list-style-type: none"> <li>- No input whatsoever from the userbase</li> <li>- No clear understanding of how it possibly could be used</li> <li>- No input from the end-users as to its architecture, its characteristics, or what it will support.</li> </ul> <p>It's just – they throw a bucket of pigswill over you and you're supposed to do something with it. It's just awful.</p> <p>As far as mitigating the effects of system failures for this frontline work where you're basically using an entire computer system at a site: one idea is to move away from the batch-queueing model. Move to a model that is closer to a physical experiment, as if you're using for instance, an astronomical telescope. In other words, it would be much more beneficial to us to be able to run for a long time, but to book that runtime at some point in the future. And to have systems staff on call when the reserve timeslot begins, to fix anything that occurs.</p> <p>For example, it's much more common that you suffer a failure in the first few seconds, than it is the last few seconds. If any node, for example can't see the parallel filesystem, that's fatal to a user job but it might be something that can be fixed quite quickly by a sysadmin. But if you're running in batch, you wait days (if not weeks) till your batch job starts, it fails instantly, and you have to go through the whole thing again.</p> <p>So the operation of these things needs to be made reliable in both physical and human terms. You've got to have systems support to overcome that kind of problem in real-time. I imagine that's not too much of a problem inside the weapons labs [<i>US National labs backed by military funds are colloquially known as "weapons labs"</i>].</p> <p>But in the NSF, where you've got thousands of users chipping away and endless small jobs, while at the same time trying to accommodate people who want to use an entire center for a day or week at a time, it's a terrible human-scheduling problem in addition to the typical throughput-scheduling problem.</p>	
<b>Q9.3 What technology-related obstacles do you currently encounter?</b>	<p>Of course the unreliability you experience is directly proportional to the number of components you're using. The vast majority of NSF users are still stuck in the 100 CPU category. So they probably don't experience anywhere near the frustration that I experience because, all things being equal, they see only 1/20<sup>th</sup> or 1/40<sup>th</sup> of the failures that I do.</p> <p>And at 60,000 processors (or whatever it's going to be) I suspect that computation at that scale will not be possible using the current approach to batch production. How on earth would you assure a user that when their timeslot came up that every single component of the system was functional? And how long would it stay in that state, given that the mean time between failures is proportional to the component count?</p>	
<b>Q10.1 Can you think of any work-related tasks that decrease your productivity?</b>	<p>Networking: having to move the data from one computer to another.</p> <p>Archiving: archival systems are one to two orders of magnitude too slow.</p> <p>Capacity of disk systems: there are systems where the rate at which I can work is throttled by the actual amount of disk space at runtime. Even though a system may have hundreds of terabytes of disk it's no use if it's ninety-five percent full.</p> <p>Center policy: that is, what is the maximum length of a batch job that any one center will let you run. That varies from as little as six hours to some centers that have seen the light and it's almost unlimited. So really it's how do centers deal with the competition for resources at different scales.</p>	

Interview ID=4 1 June 2007	ANSWERS	ANNOTATIONS
<b>Q10.2 Describe repetitive tasks associated with your work</b>	<p>It comes down to a couple of things and they're both on the output side of the work. Let's say your batch job ran successfully. And now you have several terabytes of output sitting on the high-performance disk, and you've got to do something with it, which in this case means save it to a permanent storage system.</p> <p>Now, is the permanent storage system across the country, or is it local? That adds another wrinkle. However you do it, you first have to aggregate the results into large chunks. This is important either for transferring it effectively across a long-haul network, or it's also mandatory if you're going to archive it on anything that involves a tape medium. I learned that lesson early on, very expensively: if you ever intend to get anything out of an archive there had better be just a few very large chunks rather than a lot of very small chunks.</p> <p>The reason this is not automatic and so time-consuming is failure. These things fail. There are no obviously robust methods that we use to help us get around this. Now, I believe there's a thing called Reliable File Transfer (RFT) that might help. But again, we don't just use NSF TeraGrid, we use DOE centers as well, and it's not clear to me that they would implement anything like that.</p> <p>The same goes for GridFTP, by the way: the lingua franca that we're often forced to use is bbFTP (<a href="http://doc.in2p3.fr/bbftp/">http://doc.in2p3.fr/bbftp/</a>) because we can build it ourselves. GridFTP carries all the baggage of Globus with it, but it's the only component we're interested in. Really it's just an FTP program – why on earth do we have to bother with all the certificates and all the stuff that goes with it? All we want is point-to-point transfer to be fast and reliable.</p> <p>So from a workflow point of view it's dealing with failures. If you can move 99 files with a batch script and they all got there safely, it takes you as much human time to deal w/the one that didn't as moving the 99 that did. So human intervention to deal with the failures is the expensive time.</p>	
<b>Learning about the Globus user experience</b>		
<b>Q11.1 Which Globus data components do you directly interact with in your work today?</b>	GridFTP	
<b>Q11.2 Did you install the &lt;component&gt; client yourself?</b>	No. I deal with installing all kinds of software, but I wouldn't consider touching Globus software. It either exists because the center provides, or it doesn't.	
<b>Q12.1 Which Globus security components do you directly interact with in your work today?</b>	<p>If it's there I don't know it is. I don't want to know. Apart from using ssh to log in to something and GridFTP, I don't want to know any more than that.</p> <p>I would not consider installing it myself. I don't like the overhead. When anything goes wrong with your certificates, site certificates – anything like that: it's completely beyond the scope of anything a user can do. And usually it's beyond the scope of what the computer center personnel can deal with as well. It usually means that you're just crippled for a couple of days until the one guru at site X can actually figure out why what used to work no longer does.</p> <p>It's too fragile, it's too cumbersome, it's too complex, it's too difficult.</p>	
<b>Q16 Why do you use &lt;component&gt; instead of an alternative technology?</b>	<p>Within the TeraGrid, all of the partner sites are obliged to provide GridFTP and the major systems that I use have dedicated GridFTP server hardware. It is the way that exposes the highest rate of network transfer from filesystem to filesystem across a transcontinental distance. So for example if I have to move output from Pittsburgh to San Diego, GridFTP buys me a factor of two or three over bbFTP, which I could look after myself. And that two - three is essential. If I could get more – more is better.</p> <p>I know that users always say "oh the network's slow", but I know better than that. It's not the network wire provided by a telecom, it's</p> <ul style="list-style-type: none"> <li>- how fast is reading and writing from the parallel filesystem on each end</li> <li>- to what degree can we use striped GridFTP to get the maximum number of GridFTP servers reading and writing at each end and transmitting it down the middle</li> <li>- do the centers expend enough money in providing that service</li> <li>- is there enough hardware on the sending and receiving ends to match the actual physical network bandwidth</li> </ul> <p>Now at SDSC we're proponents of GPFS-WAN. We export a global parallel filesystem across the TeraGrid to other sites that will mount it. In theory that produces a wonderfully simple logical thing. You can see the filesystem you want to write to locally. But in actuality that isn't available on any of the computers that I use. So again, it's rather a miss for us.</p>	

Interview ID=4 1 June 2007	ANSWERS	ANNOTATIONS
<p><b>Q17 What are the major challenges you face using &lt;component&gt; today?</b></p>	<p>[GridFTP challenges] The interface to GridFTP is a bit clunky – we would like something to be as simple as scp. So I gather that the TeraGrid project has done a fairly good job encapsulating some of the knowledge you need into tools such as tgcp, but I get the impression that some of those things aren't really maintained so well.</p> <p>I write my own wrappers because I know the source and endpoints I'm going to use all the time. So I have a shorthand to make it easier to work with. Because it's a very clumsy-looking thing: too many parameters, too much knowledge – you don't have time to go in and tune these things. You find something that works and you stick with it.</p> <p>[Security challenges] Security is out of your control and require bizarre and completely Byzantine communication between centers. For example, try using an NSF certificate at a DOE site. Dead on arrival. So wherever those are, they're a nuisance. All we want to do is send bits. We don't care about certificates and so forth. It's irrelevant. Good old FTP would have been fine if it was striped and so forth. Also when you have this problem with them expiring. It just doesn't help.</p> <p>[Prompt asking for more detail about problems] The worst thing that happens to users... I don't even know how to express this because I don't know what I'm talking about, right? I'm a scientist, I have no need to know what this is; I'm not interested in knowing.</p> <p>However, every six months to a year or something, you'll find that what you used yesterday no longer works because some certificate somewhere has been changed or expired or whatever. The error codes you get are so arcane – you can't even tell that's what happened. Globus error messages are the pits from what I've seen – completely inexpressive in terms of what a user would understand them to mean.</p> <p>So when I encounter an error I pick up the phone and call some person who's job it is here to deal with such things and I say, "This doesn't work anymore. I have no idea why. Please fix it." And then I'm dependent on whether they take me seriously and how long it takes them to figure out what is broken. I'm nearly always right, but hey, it doesn't help because I can't fix it.</p> <p>I have often thought that I would simply just install bbFTP at all sites that I use, forget GridFTP and just suffer the loss in performance. Because if anything goes wrong – I'm one of these people that works twenty-four hours a day – and if something breaks it's always on a Saturday evening when you're in the middle of something and there's absolutely no one you can call. You can file a ticket, the ticket won't even be seen possibly until sometime on Monday. And it may not be acted upon until that person feels morally obliged to do it.</p> <p>So I would always, always, always vote for extreme simplicity in these critical functions. And to me critical functions mean "being about to move data when I need to move it." I'm usually under terrible time pressure to move stuff because of purge policy, or whatever.</p> <p>I actually issued a challenge to the management here, and I think it's still a good challenge. And that is: anyone who has never had to move just a single terabyte from one place to another should try it. Moving a hundred terabytes is monumentally painful. Like I said earlier, I envisage that moving ten petabytes will simply be impossible, so we have to figure out how not to do that anymore.</p> <p>I don't want to beat up on Globus, because I don't use 99.5 percent of it. I find that the professional computer science community has a very weird view of what computational scientists such as myself actually do. For us all the complexity lies in writing our own Fortran or C++, or whatever it is, to solve systems of equations that describe whatever it is we're trying to do. That's extremely difficult. And then on top of that, you add the requirements for distributed memory, which is a rather artificial requirement if you think about it. And then on top of that we have to worry about massively parallel IO. And then we have to start worrying about data transmission strategies.</p> <p>And now we have to start thinking about "How on earth could we possibly rewrite a working code to do error recovery due to processor failures that are going to occur several times an hour in a 50,000 CPU run?" It's asking too much. I see many straws approaching the camel's back that will put an end to this because the complexity is out of control.</p>	

Interview ID=4 1 June 2007	ANSWERS	ANNOTATIONS
<p><b>Q17 What are the major challenges you face using &lt;component&gt; today?</b> [continued]</p>	<p>Most users in the physical sciences still write their own stuff. It's all quite small scale - well it depends on your perspective, whether you consider a hundred thousand line physics code to be big or small. It's certainly the integration of many, many man years of effort. At the million line level it's man decades.</p> <p>We can't rewrite that software. We don't have the grant money, we don't have the people, we don't have the time. And so simplicity is what I strive for all the time. Things that can remain under our control. I think I've already mentioned that I'm extremely loath to link in any 3<sup>rd</sup> party software that I can't control.</p> <p>HDF5 is a wonderful exception, actually, in that I can compile HDF5 myself for each of the machines I use. I don't usually have to rely on some favor from some systems person to build and install it at a site, or to keep it up-to-date. Actually I go straight to the horse's mouth and deal with the HDF5 folks myself. And they are wonderful in terms of being responsive to suggestions and helping out. But that's a product that I can compile and manage myself. But there are components beyond that that I simply couldn't.</p> <p>Excessive complexity is not our friend. It's getting too difficult already. So if there's additional complexity in software, it may theoretically add some splendid functionality. For example workflow software: in theory you could turn an entire ENZO run into just a workflow program. And you could hit "go" and after a few months you could come back and collect your petabyte of output and go on your merry way. But it wouldn't last a day before it required human intervention. So it's not worth trying. You need humans on the job - they're faster and more responsive anyway. Having the least number of people you have to rely on actually increases your efficiency.</p> <p>[As an aside]</p> <p>I was in the applications working group and the user services working group when the DTF/TeraGrid project began. And it was an unholy mess. It still is an unholy mess. Because that project is like any project - it exists for its own reasons. Despite whatever noises people make about it being for the users, it isn't.</p> <p>There's a TeraGrid conference in Madison on Monday, right? One of the organizers of that asked me the other day, "There's 300 computer center and software people going to this event - but there are almost no users. Why is that? Why can't we get users to come?" That's the basic issue, isn't it? If they were discussing anything that actually was of use to us, maybe we would come. But most of it, frankly, is peripheral at best.</p> <p>I suspect that attitude has got a lot to do with which discipline in science you're in. In some areas of science people use packages a lot and they're used to the idea of typing in some sort of GUI and hitting "go" and getting some sort of answer. We're about as far away from that as you can possibly get. It doesn't even make sense to me to consider such a thing. When the computation time is measured in months, any kind of traditional view of that just doesn't work. For example this one simulation that I'm running I started working on it in the last week of November and it's now June. It's not over yet. These things don't fit well with the Computer Science idea of running myriad little processes, or something.</p> <p>[Is there a center that has many good characteristics?]</p> <p>I'm pretty high on NERSC. I think they do a very professional job. And I'm probably being a little traitorous here, but I'm sticking with that. NERSC does a better job than SDSC, or any of the NSF centers:</p> <ul style="list-style-type: none"> <li>- The consistency of their support of the third party stuff across their multiple different architectures is exemplary</li> <li>- Their documentation is first rate</li> <li>- Their pedagogical examples: the centers used to do such things, now we don't bother - we just point people NERSC's webpages</li> <li>- They have a very focused attitude to supporting very large computational projects</li> <li>- They're capable (obviously because of the political organization) of supporting very large, single computational experiments in a way that the NSF really hasn't caught up to. So for example the INCITE program is an excellent, excellent example. DOE, Oak Ridge and NERSC manage those things, and they make available explicitly enormous chunks of time to do things that are not possible to do elsewhere. And they do a very good job of it.</li> <li>- They turn on a dime. You know, some of the things I was complaining about such as not getting enough support? They do things, for example, in order to recognize my NSF certificates simply because I was doing this ultra-large calculation, and they made other systems and other network pathways available to me</li> </ul>	
<p><b>Q17 What are the major challenges you face using &lt;component&gt; today?</b> [continued]</p>	<p>Another bunch of people that deserve very high praise are the HDF people, who even make design changes in HDF to accommodate things that we need for ENZO. So there are spectacular examples of people really going out of their way to help us.</p> <p>Then there are others who have a business as usual approach, where if I can't solve my own problems then I am stuck. And I work at one of these places, so this is a very schizophrenic comment. I have to support myself.</p>	
<b>Wrapping-up</b>		

Interview ID=4 1 June 2007	ANSWERS	ANNOTATIONS
<p><b>Is there anything you'd like to say to the people who build software for use by people like you?</b></p>	<p>I don't wish to be too critical. I'd like them to understand that a lot of what may sound a bit niggly from me is the result of using these large systems pretty much twenty-four by seven for the last several years. I do recognize that ENZO is an absolute cutting edge thing. What we're trying to do, both with the code and given the machines, so I cut everybody a lot of slack even though I sound frustrated. It is frustrating experience. I wish it wasn't.</p> <p>"Keep it simple" would be the only real advice I would have. You know? The KISS principle. Users these days have got an unbelievable amount of extra work to do compared to the supercomputing programs twenty years ago, when all you needed was a Fortran compiler and a Cray XMP and you were absolutely the best in the world. The complexity of it now is so great that I see it breaking down.</p> <p>It isn't worth our time to consider adding more sophistication, because if we have any spare time or any spare brain cells, we want to add sophistication within our code in terms of things that are domain-specific to us. For example, if we can figure out better ways to do adaptive mesh refinement than somebody else, then we get some sort of competitive advantage. But it doesn't do us any good if somebody writes an AMR package that is completely incompatible with our code, which is the result of thirty man-years of work. No one is ever going to have the time or money to make those things work together. Besides which then we'd be connected to something that isn't supported.</p> <p>You probably noticed from my list of things we do is that it includes things that would be nonsense to even begin without. Obviously we're dependent on compilers. The vendors seem to be gradually giving up on the service, in terms of compiler support, that we've received over the last several years.</p> <p>Because it's old doesn't mean it should be ignored. Ninety percent of the science codes in a recent Oak Ridge survey were found to be written in Fortran, for example. No one in the computer science community can be bothered to help a Fortran programmer anymore. They probably don't even know Fortran. But in science it's still tremendously important, and C is the next one behind that. We're not going to switch languages. I'm sorry to say that the DARPA HPCS language initiative is pie-in-the-sky. As somebody who remembers Ada – this is even less likely to work.</p> <p>Simplicity helps. Because if we can absorb something else it has got to be usable in a standalone way.</p>	

## D.5 Performance improved from days to seconds

Interview ID=5 4 June 2007	ANSWERS	ANNOTATIONS
<i>Pre-interview question: Do you interact directly with Globus software in your work today?</i>	yes	
<b>Establishing context</b>		
<b>Q1.1 Please provide a one-minute overview of your project</b>	<p>MEDICUS is a Globus Incubator project, which is part of the development process within Globus. MEDICUS is a user community-based process in terms of development and objectives. There was a need in the Children's Oncology Group and the Children's Neuroblastoma Cancer Foundation to communicate large-scale three- and four-dimensional medical images between various sites to do quality assurance and central review in clinical trials. That project started in 2003, and in 2004 we started to work with the Globus team at the Information Sciences Institute of the University of Southern California.</p> <p>After some discussions we came to the conclusion that it would be possible for Globus to provide the necessary infrastructure to perform central review and quality assurance functions over the Grid. So that developed as a neat project because we found out that all we really need is to communicate medical images into the Grid, and all the other components like security, data management, data transport are already part of the Globus Toolkit. So the MEDICUS team basically only needed to work on the domain-specific components of the Digital Imaging and Communication in Medicine (DICOM) standard.</p> <p>So this has been basically implemented and being used within the two communities. There are 27 children's hospitals from the Children's Oncology Group and there are 13 hospitals from the Neuroblastoma Cancer Foundation using MEDICUS and obviously Globus as the underpinning infrastructure to communicate clinical trials imagery.</p>	
<b>Q1.2 What is the project's name?</b>	MEDICUS	
<b>Q1.3 Which agency funds the project?</b>	The project is funded from a mix of sources, including federal funds from the National Cancer Institute, and private funds from fundraising efforts (such as the CureSearch National Childhood Cancer Foundation.)	
<b>Q1.4 What field does your project belong to?</b>	<p>MEDICUS is relevant to medicine in general because it provides image communication, which has applications for both clinical and research purposes.</p> <p>Image communication can be used for clinical use. For example, when you go to your doctor and your doctor creates some images from you, such as an xray, this xray stays with the doctor. Now, if you're admitted to the hospital later on, these images are not available at the hospital. So using the MEDICUS interface the images can be communicated to the point of care, wherever you go.</p> <p>The second use is to communicate these images for research purposes. While technically there is no difference in handling these images for clinical purposes rather than research use. The major difference is that when you are in a patient-doctor relationship the doctor is authorized to know your identity. But in a research setting, when you collect data for clinical trials, for instance, this patient information, which is called Protected Health Information (PHI), cannot be exposed to any of the researchers. Therefore the mechanism for identity protection and security management is very critical in the framework of clinical use.</p> <p>This is an active field of research, where very different systems have been proposed to communicate clinical images for healthcare. Now that MEDICUS has successfully implemented a version for clinical trials, we want to provide the same solution for secure communication for clinical patient care.</p> <p>So we've basically worked on what we call the Patient-Centric Authorization Model, which uses X.509 certificates and the attributes-based SAML assertion technique to create a two-layer security model that allows PHI to be shared on Grids. This Patient-Centric Authorization Model is not specific to images. The model could be used to secure any kind of medical records, such as text information from your primary care physician, reports from a radiologist or pathologist, lab reports – any information that can be rendered as Health Level Seven data format (HL7).</p>	
<b>Q1.5 What is your job type?</b>	Project Lead	
<b>Q1.6 How long have you been a &lt;job type&gt;?</b>	Three years	
<b>Learning about discipline-specific goals and approach</b>		

Interview ID=5 4 June 2007	ANSWERS	ANNOTATIONS
<b>Q2.1 What are the main goals of your project?</b>	<p>The main goals are to efficiently and compliantly communicate medical images in a secure fashion so that patient privacy is guaranteed, using existing security mechanisms and other standards-based technology provided by the Globus Toolkit.</p> <p>So an example of the vision is that a patient comes into the hospital, and this patient's record is not only existent in the hospital, but also available on the Grid so that other healthcare providers can access the data and also add to it. So that wherever you go as a patient Grid can basically follow you, aggregating all the information that exists for you at various health providers, and collecting them at the point of care.</p>	
<b>Q2.2 How will the success of your project be measured?</b>	<p>Well this is a big endeavor. It is not so much a technology problem as it is a sociology problem. The background here is that the health providers are not really willing to share information. However there is more and more demand from the government and insurance companies to do exactly that. Now the challenge is to provide a technical solution that can scale to allow a large number of healthcare providers to interact and share images. I think Grid technology is the only viable solution out there. That really is the motivating factor – to show a use case and a technical implementation that demonstrates the technology is ready to address that problem.</p>	
<b>Q2.3 What are the professional measures of success for you?</b>	<p>There are two levels here. We pretty much have had a good success with the clinical trials domain that we've worked on the past three years. We have two running Grids and the physicians are very happy because they are able to launch a review of images without having to travel to the data centers where the images are acquired.</p> <p>So the physicians are finding they can read the images in their own environment. Reading really large images requires specialized equipment: DICOM display workstations that have enough grayscale display capabilities to read images diagnostically. And there's very different software out there and every radiologist has his own preferred environment for reading. And with the Grid technology we deliver the images from any clinical trial center into the radiologist's own review workstation. That's capable now with Grid technology. And that's a big reward for us to see that really happening. And the radiologists are very pleased with this. This actually engages more radiologists in clinical trials than before. So we can actually improve the quality and quantity of research being done within the Children's Oncology Group at the Neuroblastoma Cancer Foundation.</p> <p>The other reward that we see is that this technology is so well received and accepted within the radiology domain that we plan to explore further if this can be used in clinical settings as well. But this is very new and we don't yet have much experience with this.</p>	
<b>Q3 What are you investigating?</b>	<p>MEDICUS is an engineering project. It's a software development and I think specific goals within that project are to ensure that the radiology imaging workflow is flowing. That means that there are very different modalities out there that we must take care of, and there's a lot of incompatibilities between the very different devices. Because of that, one current work focus is to make sure that the MEDICUS project can handle all these different vendor-specific imaging devices.</p> <p>We also work on the security model to make sure the privacy protection is there. This will actually be very important to further explore the medical field. When you have a patient-doctor relationship, you as the patient sign consent that only the doctor who is treating you or the staff of that facility is allowed to see your medical charts. So now Grid enables us to communicate all these medical information wherever we want. And that brings a lot of opportunities.</p> <p>For example, consider the implications for medical images: if images are acquired of a patient, and later on the patient moves on to another hospital, the prior images are usually not available. And so you see a repeat of the same imaging being done just to follow-up with the patient, so you have a redundancy of imaging. This is not good for the patient and not good for healthcare at all because this creates additional, unnecessary cost.</p> <p>So one motivating factor is to develop a security model that allows the same doctor-patient relationship being translated onto the Grid, allowing data access, but only if the patient authorizes it. So we are developing what we call a Patient-Centric Authorization Model as a way to approach this. In the Patient-Centric Authorization Model, the patient carries his own private key with him, goes to his health provider, creates a patient assertion, finds that patient assertion with his private key and allows the health provider at that point to act on his behalf. That health provider then queries the Grid, discovers previous information about the patient (lab reports, diagnostic reports, images, etc.) and then uses them to treat the patient. So in that sense the Grid is being used as a discovery pool of medical records. The second advantage of the model is that the healthcare provider currently caring for the patient can publish new entities to the patient's health record. So a global health record is created for the patient.</p> <p>Obviously this concept has a sociological component to it, and it may not be feasible in the short term. But I think what we want to accomplish in MEDICUS is to show that there's technology available today, such as the very strong security infrastructure in the Globus Toolkit, which can be used in an intelligent way to build a security model for patient authorization and privacy for health data.</p>	



Interview ID=5 4 June 2007	ANSWERS	ANNOTATIONS
<b>Q4.1 What is your method for investigating &lt;phenomena&gt;?</b>	<p>For the engineering side of MEDICUS development we work using a top-down approach. We assembled an inter-disciplinary group in a room (Grid experts, medical informatics experts, software engineering people) and drafted a set of requirements. We identified the need for data transport, security, DICOM protocol translation into Grid protocols, and so on.</p> <p>After we had our requirements, the second step was to identify what was already available. And that was a really interesting step for me, because I learned that much of the basic functionality, such as data transport, is already included in the Globus Toolkit. And there was not a lot of effort required to make Grid technology work for the medical domain. And that was very neat because you don't have to reinvent the wheel.</p> <p>And that's how we vertically integrated all the components, like GridFTP for data transfer, and X.509 certificate methodology for security and so forth.</p>	
<b>Q4.2 How do you work?</b>	<p>With respect to the engineering aspects of our work, a fundamental principle of the MEDICUS implementation is that we try to keep every module as focused and specific as possible. So if there's any redundancy with something that already exists we try to avoid that. So for instance, our DICOM Grid Interface Service is not doing anything other than receiving DICOM images, compressing and caching the series, and then sending it out with GridFTP.</p> <p>At that point we don't want to know what's in GridFTP – we just use the public interface. And the same way with the security model. We try to use the vanilla service and not add anything fancy. Sometimes this happens in the Grid community, where people say “we need a specialized version of GridFTP.” We don't do that.</p> <p>In terms of sociological aspects of our work, there are two components to it. First of all, in deploying the two grids for the Children's Oncology Group and Neuroblastoma Cancer Foundation we found that hospital staff are, on the whole, completely unfamiliar with Grid technology. One way to deal with this is to provide a pre-configured Grid deployment for them. So what we do in the MEDICUS project is we provide them with a gateway machine with all Grid components pre-installed by us, send it to them, and they only have to network it.</p> <p>And this concept works very well, because it is very difficult to communicate in words what the Grid can do for them: some people are interested, some are not. At the end of the day, to convince them you need to have something that shows them the benefits. And this points to a weakness perhaps of the Globus Toolkit. Because the software itself today requires quite a bit of knowledge to compile, install and to maintain it. So that knowledge is not there at our site. Perhaps the other sites don't have that problem. For instance in the physics community it is probably easier to find people who are tech-savvy with Linux and compilers, etc. But in the medical domain that's not the case. So it is best to provide them with a plug-and-play solution.</p> <p>The other sociological component here is that all of a sudden you bring a totally new paradigm into a domain that is unused to sharing information. And the Grid paradigm is the exact opposite: it is all about sharing information, federating resources, aggregating information and data mining on large-scale datasets.</p> <p>This sociological issue is best tackled in small steps. So the MEDICUS project in this regard is starting with radiology and just making the medical images available. We're working to convince the radiologists that this is a good paradigm that is beneficial to them. And the radiologists are not technically at a level where they understand what the underpinning is, but that's not really necessary. So the approach we've taken is kind of the soft approach of learning by doing. For the sites, if they see “oh, this is working”, or the security model is gaining trust, then you build confidence. After building confidence you can go to the next step.</p>	
<b>Q4.5 How do you document your results?</b>	<p>There are two levels. There's the technical documentation: all the installation manuals, and documentation within the code, and so forth.</p> <p>Because it is a federally funded project, we obviously have to write progress reports, and this is the second type of documentation. It includes coverage of what the application is, the successes of the project, how many images were acquired, the satisfaction level of the radiologists, and so forth. So we do file progress reports with the NCI.</p>	

Interview ID=5 4 June 2007	ANSWERS	ANNOTATIONS
<b>Q5.1 In what ways do you interact with simulations in your work?</b>	<p>Not yet, but that's an interesting question. Because now that you have all these images collected, what are the cool questions? Some cool questions, if you have many, many deployments are, "Give me all the computer tomography images of the twenty-year old males who have lung cancer." These are totally relevant questions to ask in the medical domain, but we cannot do that today because the data are not aggregated.</p> <p>If you look at the National Institute of Health, which is a major medical research sponsor in this country: if they accumulate 10,000 cases per year in their archives, that would be a lot for them. If you go to the radiology department of a small neighborhood hospital, they probably do around 60,000-100,000 cases per year. Then you have, for instance, probably 80-300 hospitals in Los Angeles county. You get the idea that in the clinical domain on a daily basis you can get way more data than you can ever get in research.</p> <p>The critical thing here is that if you build a Grid which is connected to clinical data you can come up with very interesting epidemiological questions. "Give me all the cases of a specific disease in a specific area of the country" – and then you can see a big picture.</p> <p>The problem is that you have to overcome the sociological barriers to sharing data within the medical domain; you also have to overcome the issues with the law because patient privacy must be protected at all times. And if you can solve these two issues, you can aggregate information that is very, very relevant. And can outpace efforts at the National Institute of Health. Because many of these clinical trials accumulate only a very small numbers of cases.</p> <p>A lot of cases, if you look at them on an individual level... let's take cancer as an example: there are specific cancers that you may get one or two cases per year at one facility. More cases are needed in order to really understand the disease, so data must be drawn from multiple centers. That's what we're doing right now with MEDICUS. So you end with maybe 60 or 100 cases per year because data are collected from 20 or 30 hospitals.</p> <p>But the really interesting question is "How can we get the 20% of the population that has this disease?" This is why I think Grid technology is relevant within the medical domain. Because it provides the technical underpinning for asking these kind of questions.</p>	
<b>Q6.2 How do you share work-related data with others?</b>	<p>There is a law enacted in 1996 that is called HIPAA. The law basically states that the identity of the patient can only be exposed to people that the patient has authorized. If you go to a hospital and you get admitted you must sign a waiver of consent so they are allowed to know your identity. The law says that no one must be able to access your medical data without your consent. Your data can be exposed, but not your identity. So you may have medical records (lab reports, diagnostic reports, etc.) but you must de-identify that information if you want to use it for research.</p> <p>On the issue of securing your data that includes your identity information, it may be impossible to have 100% security here. Looking at a clinical practice right now, if you go to the hospital (not being a patient, just go there) eventually you would be able to steal some information. This is because there are so many points at which personal data is being exposed. There's only a best practice where hospital staff try not to have these things exposed.</p> <p>With Grid technology with the security model, you can do quite a better job electronically: you can do an audit, verify certificates, verify attributes, etc. These mechanisms are way better than what clinical practice is right now, because many of the documents today are in writing, stored in physical files, reports end up in the trash, etc. So there are many places in the current system where private information is exposed to the outside. This is one way that I think Grid technology can help, because it has a very good security model.</p> <p>But MEDICUS is not 100% perfect. We are a research project. We do not have the mother of all solutions, but we need to start somewhere, and Grid technology provides us with a wonderful foundation to start.</p>	
<b>Q7.1 What resources do you use in your work today?</b>	<p>We use a couple of SUSE Linux workstations; we have servers for storage of the medical data. We have what we call the Grid Book, which is our Grid deployment installation. It is a pre-configured Linux and Globus Toolkit laptop, which we send out to each of the participating institutions. And that serves as the imaging gateway there. Some centers that have a larger imaging workflow will have what we call the Grid Gateway, which is a rack server which does the image transport and communication.</p> <p>We get our images from the hospitals' Picture Archiving Communications System (PACS). This is a standard data storage system within the radiology domain. There are different vendor implementations for that, and all these PACS systems communicate over the DICOM protocol. We get these medical images from these PACS databases into our Gateway and from the Gateway we push the images into the Grid. And then on the Grid we can do replication with the replication management services. And to be able to find the images we also index them using the OGSA-DAI service.</p>	
<b>Q7.2 How do you share work-related resources with others?</b>	<p>We share the Grid Book, the data storage and the metadata catalog resources. In this sense the whole Grid deployment is shared with all the members of the virtual organization.</p>	

Interview ID=5 4 June 2007	ANSWERS	ANNOTATIONS
<b>Q7.3 By what mechanisms is access to your work-related resources controlled?</b>	We use certificate authorization technology. Because the data are already de-identified on our Grid you don't have to worry about patient privacy protection and patient authorization is unnecessary. This kind of policy we don't have to enforce because at the data acquisition site the patient signed a consent authorizing its use for research. Then we strip out all the identity information according to HIPAA compliance and send these datasets into the Grid. So we basically accumulate a huge pool of distributed data that is free of any private information.	
<b>Q7.4 How do you locate available resources for use in your work?</b>	<p>There is no way that a hospital would expose a PACS system to the Grid. Some people believe they can, but in reality it is not possible. The thing is if you want to communicate identity-laden images out of the hospital it must be patient-initiated.</p> <p>For the research case where the patient is part of a clinical trial, the participating health provider is required to send <i>all</i> data relating to the trial to the trial authority, which in our scenario resides on the Grid. So what happens is that the Grid Book is registered to the PACS system, and the PACS system pushes the images to the Grid Book.</p> <p>The Grid Book cannot query the PACS system. Such a query would be illegal in terms of HIPAA compliance because <i>all</i> of the hospital's clinical cases are in the PACS system, not just the trial participant's case.</p> <p>So what happens is the PACS system pushes the images to the Grid, and then the images become available. What we can do then is store the data on the Grid: you can replicate the data there, you can make the data available to any other instance in the Grid. But you must obtain the images first. There's no way legally to interfere with the clinical PACS.</p> <p>That's another thing: if you were to expose a PACS system as a resource on the Grid, you would interfere with image workflow at the radiology department. Let's say hypothetically that you expose the PACS server as a resource and all of a sudden there were 1,000 hits on the server: that would interfere with performance of the radiology department. That would cut back on the workflow, that would cut back on the revenue. So this is a total no no to expose PACS as a resource. And then there's also some legal issues associated with that, because you cannot tamper with medical equipment. There are some very fine lines that you cannot cross.</p> <p>So what is needed in order to make the Grid viable for the medical community is to present alternative or add-on functionality. For instance: PACS systems are required to have a replica because the health provider is legally bound to save their information for a certain number of years (the exact number varies by country.) So in order to do that, backup and offsite storage functions are mission-critical for the radiology department.</p> <p>So what has happened these past few years is that these PACS systems had offsite storage attached to them for disaster recovery and fault tolerance. So a good way to provide additional functionality to the PACS domain using Grid technology would be to provide offsite storage built on the data movement and replica mechanisms included in the Globus Toolkit.</p> <p>And that's exactly how MEDICUS was designed. Not designed to be exposing the PACS as a resource, but to be an additional component that adds to the PACS system by providing offsite storage of the images. Once you have the images offsite you can use them actively. So you decouple the local hospital workflows from the Grid-based workflows. Obviously the thing that you have is the redundancy of the images: one instance of the image exists in the PACS system and the other exists outside on the Grid. That's how you want it, because if the PACS system dies you can recover it from the Grid.</p> <p>So this is a good buy-in concept, and is the only way to bring in radiology imaging, as well as all medical data in general. This is because of the fact that Grid operations cannot interfere with the local hospital activities. And that's very different from the use cases that you have so far with Grid technology in the physics domain where data is being handled on a different sensitivity level.</p> <p>As far as finding resources local to the Grid, you have the images stored on a resource and there's reference to them in a metacatalog. And the metacatalog holds part of the medical information of different types: patient level, study level, series level, and image level. For instance, there's a unique serial number for every image that we use for finding specific images on the Grid. (That means we don't need to create unique identifiers because they come with the data.) Patient-level data includes non-PHI data like age of the patient (not birthdate, which is PHI.)</p> <p>In order to find these resources, you use a metacatalog, and the metacatalog carries the identifier for these images, and then you go to the replica location service and identify the physical representation of the data. And then you find one or more urls where the data is located and then you can get them using GridFTP as the data transport.</p>	
<b>Q8.3 What programming languages have you used in the past year?</b>	Everything is implemented in Java	

Interview ID=5 4 June 2007	ANSWERS	ANNOTATIONS
<b>Q8.4 What workflow tools do you use in your work?</b>	<p>No workflow tools. The workflow of discovery and publication of the images is being driven by the physician. So the physician sits in front of his console and says, "I need to have patient X. Give it to me." It's very straightforward; it's not a query of a complex nature.</p> <p>But in terms of workflow it would be actually interesting to add a component there in the future to say, "I'm a reviewer of clinical trials and I want to automatically get all the images on my desktop display workstation as they come in." So, kind of monitoring who is the reviewer of which dataset, and when they become available push the datasets to the workstation so the radiologist can review them. But that's more of a convenience tool; it's not mission-critical.</p>	
<b>Q8.6 If the need for new software-based functionality arises in your work how do you acquire it?</b>	<p>Like we did before: we try to find the expertise in the field and then try to engage with that group or person and see if they want to participate in the open-source contribution process. Actually there are some software efforts underway that are in the very early stages of extending MEDICUS. However the collaborators are taking the lead to drive their own efforts in their domain of expertise, so I leave it to them to provide specific information when they feel it is appropriate.</p> <p>But this raises an important point: the interest of the MEDICUS project is not to take over the whole domain, but rather to serve as a seed. Demonstrating what can be done with radiology data will hopefully show what can be done for pathology data and for neurology data, and other entities. We don't have expertise in those fields, but would like people to engage with MEDICUS and bring their expertise to the table.</p> <p>And that's how we look at the MEDICUS project in general. This is not a typical medical technology development model, but we would like to have contributors engaging in this open-source community effort, bringing their expertise to further build out MEDICUS.</p>	
<b>Learning about the user's problems</b>		
<b>Q9.1 What challenges do you face today in accomplishing your work-related goals?</b>	<p>The first problem is to install and deploy Grid technology within the medical domain. So bringing in Grid technology to a specific hospital is a major issue because they are unused to sharing images, they are unused to sharing data. They are very restricted on the detail level. For instance, opening ports on the firewall is a very major issue for these hospitals. It's not that they lack the technical expertise to control the firewalls. It's really the general concept in running hospital IT: you close everything and then you're happy. The problem is that with Grid technology you want to accomplish the opposite- you want to make everything available. So you have to find a balance and ways to accommodate the hospital's view.</p> <p>The way to do that is to convince them that these are standards-based security methods, there are large deployments, there are government, industry and academia examples of using the technology. And then finally you convince people to give it a try, and see if it's possible to open ports, to enable that communication. That's one level of the accommodation issue.</p> <p>The second level of the accommodation issue is that you cannot tell anybody in radiology, for instance, or even at the hospital IT, "Here's the link to the Globus Toolkit. Please install the Globus Toolkit. And here's the MEDICUS project. Install it. And here are the contribution examples detailing what you need to get. And then once you have done that we'll create a certificate for you, and you install the certificate in your container. And then that's it." This won't work because they don't have the expertise to do so. So that's the major problem with the user community is that they are not aware of how to install and deploy Grid technology.</p> <p>And I think that's the main problem with the Globus Toolkit. It's not like they can go to the website, download a Microsoft installer package, do a double click and the software installs and you can start it. It's not working that way, and it will probably never be that way because you have to have credentials being created. There's a lot of infrastructure that has to be in place and has to be leveraged.</p>	
<b>Q9.2 What types of information do you need in order to address the challenges you face today?</b>	<p>Well it's an educational problem. I think Globus already addresses this by having these training courses available. I think what really needs to happen is to give training courses to hospital IT- on what is Grid technology in general and what is a concrete implementation of it. That would really help because then we wouldn't have to repeat these discussions with every single institution when we do a deployment. But my gut feeling is that it may be a little too early because this whole field still has to mature. Not only in the Grid domain, but especially the interaction with the clinical hospital domain and the overall healthcare enterprise.</p> <p>The healthcare domain is very conservative. Technology adoption is very slow into that domain. So I don't expect that to be accelerated for Grid technology. But the way to do that is to have training courses and lectures on what Grid technology is and what it can do for them. In this respect I think MEDICUS is doing a great job because it advocates Grid technology within that domain on a specific use case.</p>	

Interview ID=5 4 June 2007	ANSWERS	ANNOTATIONS
<b>Q9.3 What technology-related obstacles do you currently encounter?</b>	<p>There are no technology issues. It's always to find a way through the specifics of the components that are there. In general you have these two ways to do it. Either you take what's there and try to put these things together to integrate them, or you say, "Ok, I don't like what's available. I will re-invent the wheel with the flavor I need." And depending on which paths you go, you have different obstacles. Obviously MEDICUS chose the integration approach, and so any problems we have would be related to the existing software.</p> <p>But in the case of MEDICUS we did pretty well, because we only need to have some very basic functions and the Globus Toolkit provides that. So on a technical level I really don't see any major obstacles. There is nothing I can point to that is a roadblock. But this is very specific to our project maybe. Other projects may have some specifics.</p>	
<b>Q9.4 By contrast, can you provide examples of technologies you find very useful today?</b>	<p>Any kind of mobile technology is very interesting. Because for instance, radiology is moving towards a scenario where the radiologist is not anymore on staff at the hospital, but interacts more like an outside consultant. The radiology may serve in that function for multiple hospitals, multiple healthcare providers- perhaps even owning his own practice. And so the hospital, or healthcare provider, is basically an imaging service. And then these images become available in some form, perhaps via virtual private networks, or Grids or whatever. But they become increasingly available over the internet. This is one direction I think the whole field will move, and if you have a mobile device like a PDA that is Grid-enabled and can control your workflow, that's very relevant for the medical domain.</p> <p>So imagine extracting relevant information out of a pool that resides on the Grid, to get a very quick overview of a patient. Let's say you're in an emergency room, and you have your PDA, and you can very quickly query the Grid, "What are the most important things I need to know about this patient?" in order to make a decision about emergency treatment. That is very relevant and has impact to the field. So I think any kind of mobile technology will be important. That's true for the Grid domain or any kind of internet-enabled methodology.</p>	
<b>Q10.1 Can you think of any work-related tasks that decrease your productivity?</b>	<p>Requirements specification and implementation can be painful, but also rewarding.</p> <p>The only specific thing that I can think of which might be done better is a Globus-specific issue. There is some inconsistency in how the manuals for the components, in this case GridFTP and RLS, are laid out and formatted. One specific critique is that it would be good to have one standard way of formatting the manuals, like manpage style under Unix. It would help to bring in the information in a quick and ordered way that people are familiar with. This was from some time ago, so the situation might be better today.</p> <p>Tutorials are very helpful. That's what I like about the <i>Globus Toolkit 4: Programming Java Services</i> book. It goes with a red line through a specific example. That's exactly what you need to do. Even if you're an experienced Java programmer and you have no idea about Globus, it provides a way to very quickly get your hands dirty in Grid technology. And then you will make significant progress on your development. So that's the right approach.</p>	
<b>Learning about the Globus user experience</b>		
<b>Q11.1 Which Globus data components do you directly interact with in your work today?</b>	<p>We looked into Reliable File Transfer, but we skipped that because of some semantic and performance considerations in our use case. That doesn't mean that RFT is not useful, it's just that we found it's not useful in the work we're doing right now.</p> <p>The other components are RLS, GridFTP</p>	
<b>Q11.2 Did you install the &lt;component&gt; client yourself?</b>	yes	
<b>Q11.3 Did you install the &lt;component&gt; server yourself?</b>	Yes; I have compiled the Globus Toolkit many, many times. I can tell you that it got better and better every time.	

Interview ID=5 4 June 2007	ANSWERS	ANNOTATIONS
<b>Q11.4 How many people currently use your &lt;component&gt; server</b>	<p>There are two end users interacting on each site: the CRA and the PI of the site, so with forty sites we currently have eighty end users.</p> <p>Then if you count the hospital IT people who participate in the installation, that would be another two people (the PACS administrator and someone from the network team) per hospital. This staff allocation is short-lived, because after the deployment the MEDICUS team is responsible for maintaining the system.</p> <p>Then there's the MEDICUS development team, which includes me, three people from the Radiology Department at University of Chicago, five people from the Information Sciences Institute at USC, and some Radiology advisors. So all-in-all the development team consists of twelve people, many working part-time on the effort.</p> <p>I should mention that the end users don't realize they are using the Grid. The feature provided by MEDICUS, as I pointed out before, is that when the radiologists issue a query they do so from their own workstation. They don't even see Grid. They don't even do single sign-on because the security is built in to the DICOM Grid Interface Service. So the service behaves like the hospital interface, in terms of query trees and storing images.</p> <p>So they don't even see the Grid, and that's one very critical requirement. They don't want to have yet another workstation to query from. They want to use their own workstation, which they're accustomed to, which will read the images all the day.</p>	
<b>Q12.1 Which Globus security components do you directly interact with in your work today?</b>	MyProxy, GSI certificates, GridShib	
<b>Q12.2 Did you install the &lt;component&gt; client yourself?</b>	yes	
<b>Q12.3 Did you install the &lt;component&gt; server yourself?</b>	yes	
<b>Q12.4 How many people currently use your &lt;component&gt; server</b>	<i>See 11.4</i>	
<b>Q14.1 Which Globus information components do you directly interact with in your work today?</b>	MDS	
<b>Q14.2 Did you install the &lt;component&gt; client yourself?</b>	yes	
<b>Q14.3 Did you install the &lt;component&gt; server yourself?</b>	yes	
<b>Q14.4 How many people currently use your &lt;component&gt; server</b>	<i>See 11.4</i>	
<b>Q15.1 Which Globus common runtime components do you directly interact with in your work today?</b>	jGlobus from the Java CoG kit, Java WS Core	

Interview ID=5 4 June 2007	ANSWERS	ANNOTATIONS
Q15.2 Did you install the <component> client yourself?	yes	
Q15.3 Did you install the <component> server yourself?	yes	
Q15.4 How many people currently use your <component> server	See 11.4	
Q16 Why do you use <component> instead of an alternative technology?	<p>GridFTP: Because we can use the X.509 certificates with GridFTP; it supports our methods for authenticating users. We also use it as a storage service, and what I mean by that is we use it for transferring the images from the hospital, as well as sending the images to the end users.</p> <p>RLS: Because it's integrated within the toolkit and it has all the components that we need.</p> <p>MyProxy: Because it's part of the toolkit, and it provides the functionality we that we need to have for these delegated credentials</p> <p>MDS: We are using it to expose the resources in our testbed. We have not deployed it yet in production because we're just beginning to explore it. At this time most of the sites have a specific storage path and they don't need to find alternative storage resources. But the more this Grid grows we want to have more replica sites which are directly controlled from the endpoint. So that, say, a hospital in Montreal is storing to some Canadian service provider. And then from there you have some intelligent replication management where you copy all the new files and serve them to another replica service provider in the US. These kinds of things eventually will come, but we haven't deployed it yet. The more services you get, you need an index. You need a yellow pages, you know?</p> <p>GridShib: Because it frontends to Shibboleth and Shibboleth is probably becoming the standard for entity-providing services at a federal level. There was a Gardner research report last December which I found very interesting where they announced that Shibboleth should be regarded by government agencies as <i>the</i> entity provider; I think that was with respect to university domains because they have this edu-person mapping profile in there. And again since the MEDICUS project does not want to reinvent the wheel, we integrate what's out there. So it is my understanding that for attribute management and identification based on attributes this combination of GridShib and Shibboleth is the standard.</p> <p>GSI: Because that's the private key infrastructure Globus is built on.</p> <p>CoG/jGlobus: We use the security libraries</p> <p>Java WS Core: We want to write our services in Java because we like the ease of building services in the Java language and we have a dependency on a DICOM library that is implemented in Java. One interesting sidenote with regard to potential performance issues: this is not something we're worried about at this point in time. You have to understand that in clinical trials it is still common practice to not communicate these images electronically. So what happens is that the institution burns a DVD and sends the image to the reviewer site. So a hardcopy mailing takes place to transfer the data. Or, the reviewers must go to a central review place, and so need to travel.</p> <p>But with our system they can all of a sudden query from their desktop. They are happy to see their queries popping up in realtime, compared to what they are accustomed to. So we are not at the point where we need to do a lot of performance optimization yet, because we've improved performance already from a matter of days to seconds. Eventually performance optimization will be interesting as the deployment scales. Then things like file transport and quick turnaround then become more important.</p>	

Interview ID=5 4 June 2007	ANSWERS	ANNOTATIONS
<b>Q17 What are the major challenges you face using &lt;component&gt; today?</b>	<p>GridFTP: None. This is probably the most mature software you have in the Globus Toolkit.</p> <p>RLS: Installing the database driver for the toolkit.</p> <p>MyProxy: None.</p> <p>MDS: None. Seems to be very mature.</p> <p>GridShib: This has nothing to do with GridShib actually, but the problem we had is as follows: The problem is to identify which of the workflows to enable: do you want to do active verification from the Grid service provider with that entity provider. The difficulty is more in figuring out how to map our problem to the use of GridShib-Shibboleth. We are still working on this and are just beginning to deploy and test an approach, so haven't heavily used it yet. So I don't really yet have enough experience to comment on GridShib.</p> <p>GSI: None. This is also very mature.</p> <p>Java WS Core: The major problem with the Java container is that the database connectivity just sucks. In all the compilations (I started with 4.0.0 and then all the subsequent versions until the latest) the ODBC drivers always give me problems.</p> <p>Compiling this container can be a very simple thing, but it can also be a very painful, depending on whether or not you need these database drivers in there, and the dependency on specific versions of Java, and drivers for MySQL. That could be improved.</p> <p>I think the common use of PostgreSQL in the toolkit should be revised. I think MySQL is more common than Postgres. And in that respect I think the Globus MySQL instructions and the way the database is connected should be revised. You have to install a specific version of the driver. Why is that? These kinds of things are a little bit nagging. And some of the drivers you can't even get anymore because they're outdated.</p> <p>Other than that, compiling the container and installing certificates is very straightforward. And I think that the Quickguide to installing the container is very good, very straightforward and clear. Even for somebody who is a beginner, this is a straightforward document.</p>	
<b>Wrapping-up</b>		
<b>Is there anything you'd like to say to the people who build software for use by people like you?</b>	<p>I really appreciate the overall effort. For Grid the whole paradigm can only thrive if there's an open source and standards-based implementation, and the Globus Toolkit is delivering exactly that. See one problem in the medical domain is that the internals of every equipment vendor, both software and hardware are proprietary. There are some standards on the interface side, but internally it's all proprietary.</p> <p>I think that the whole concept of service-oriented architecture presented in the Globus Toolkit and the Grid paradigm can have a major impact on how medicine is being addressed from a technology side. And I think MEDICUS is just a very small brick in this whole puzzle in showing that it is possible to use SOA and to communicate data and to make it available. And this couldn't be done without having the underpinning Grid infrastructure being implemented by people in the Globus team. So we're very appreciative for what has been done and I think Globus is a great project.</p> <p>Having Globus available makes a big difference, which we already see apparent in use cases like the Children's Oncology Group that they can all of a sudden communicate images, which they couldn't do before.</p> <p>Yes, you could choose other technology to do the same thing, but we choose Grid technology because it has more potential to add on new services. And we even haven't talked about that aspect. For instance when you have images, the images alone are not enough. You have to do image processing, which is very time consuming, needs expertise and so on. Some things can be automated and in a large scale when you get thousands of new images on a minute-by-minute basis, you need to automate that and integrate cluster technology. And all this is already present in the Globus toolkit and the Grid paradigm. There's no other technology out there that provides compute resources, data management resources and security at the level that the Globus Toolkit does. Period.</p>	



## D.6 ▼ The Grid idea is great, but there are barriers to making it work today

Interview ID=6 07 June 2007	ANSWERS	ANNOTATIONS
<i>Pre-interview question: Do you interact directly with Globus software in your work today?</i>	no	
<b>Establishing context</b>		
<b>Q1.1 Please provide a one-minute overview of your project</b>	We are doing computations in lattice gauge theory at six or seven national centers. These computations involve, among other things, reasonably large files (the files range from hundreds of megabytes to ten gigabytes.) We archive these files, we then pull them out of the archive and use them to study various physics questions. We call those "lattice files". Some of the analysis that we do also generates large files, some of which we like to store, some of which we throw away as soon as we generate them. So we operate between many different sites, and also between those sites and our home institutions. So I think as far as Globus is concerned that gives you an idea of what we do.	
<b>Q1.2 What is the project's name?</b>	Lattice QCD	
<b>Q1.3 Which agency funds the project?</b>	The National Science Foundation and the Department of Energy	
<b>Q1.4 What field does your project belong to?</b>	Lattice QCD	
<b>Q1.5 What is your job type?</b>	Scientist, Developer and Project Lead	
<b>Q1.6 How long have you been a &lt;job type&gt;?</b>	30 years	
<b>Learning about discipline-specific goals and approach</b>		
<b>Q2.1 What are the main goals of your project?</b>	The goals are to understand the interactions of quarks and gluons, and applying that understanding to the discovery of new, fundamental parameters of elementary particles.	
<b>Q2.2 How will the success of your project be measured?</b>	By the accuracy to which we can predict things, and the degree to which our predictions are confirmed by experiment.	
<b>Q2.3 What are the professional measures of success for you?</b>	By publishing our results. We talk to experimentalists who make various measurements, and we see whether or not the results of the measurement agree with what we have calculated.	
<b>Q3 What are you investigating?</b>	We have several projects. We are studying the properties of matter at very, very high temperatures, which would simulate the conditions of the early universe moments after the big bang. And these are also conditions that are produced artificially in a laboratory, in heavy ion collisions. We are studying the strong interaction effects – the quark and gluon effects – in the decays of heavy particles into lighter particles, where the decays are governed by the so-called weak interactions (radiative decays.) So the results of our computations are essential for extracting the information about the weak decays from experiment. We're also studying properties of the various elementary particles. Trying to understand why the particles have the masses that they do. And just fundamental issues about the interactions of quarks and gluons, and the theory of Quantum Chromodynamics.	

Interview ID=6 07 June 2007	ANSWERS	ANNOTATIONS
<b>Q4.1 What is your method for investigating &lt;phenomena&gt;?</b>	<p>The method involves solving a quantum field theory using numerical simulation. And the quantum field theory we're solving is the well-accepted theory of how quarks and gluons interact with each other. In order to solve it we have to do massive computations because we're simulating a theory in three dimensions of space and one dimension of time.</p> <p>We represent the space and time with grid points, which we call a lattice. We then solve for the interactions of quarks and gluons described as fields on the lattice. So the computations become more and more refined the closer we can put those lattice points together. We refine the mesh size, and we get better and better approximation of what we hope is reality.</p> <p>The computations are quite demanding because the interactions are rather complicated. So we manage to saturate any large machine at this moment in achieving the degree of accuracy that we want.</p>	
<b>Q4.3 How do you keep track of interim results, if at all?</b>	<p>I mentioned that we generate lattice files, and those are archived. We also generate a stream of log files that report results of simple measurements on these lattices. Those are quite small and easy to archive and move to our home workstations.</p>	
<b>Q4.4 How do you test work-related hypotheses?</b>	<p>There are a couple of different levels of testing:</p> <p>One is validating the code to ensure that it is computing what we think it's computing. We have ways of setting parameters so that the codes calculate things that we can calculate by hand. We have alternative approaches that we can compare, where we calculate with one algorithm and then calculate with another.</p> <p>There are a whole variety of different techniques that we use to make sure that we're calculating what we think we're calculating. We adjust parameters to what the effects are of some of the approximations that we make, to make sure that things are properly converged. So there are quite a variety of ways to check that the code is doing what it is supposed to.</p> <p>The other level of testing is validating the theory itself, and the approximations we've made, by comparing our results with experimental results. As far as validating the theory and the approximations inherent in the approach (such as setting the coarseness of the mesh on we're calculating), we calculate quantities that can be measured experimentally.</p> <p>Sometimes the experimental results have already been obtained. Lately the results have yet to be obtained, but will be very soon. So we're actually making predictions that are being confirmed. So far we've been doing pretty well.</p>	
<b>Q4.5 How do you document your results?</b>	<p>There are a couple of different levels of documentation. Of course we archive our critical results. Such results include the lattice files - any observables we generate with expensive computation. These files are then stored in mass storage somewhere on tape. So if someone needs to verify or check a result that we've obtained, we can go back to the tape, look at it and repeat some of the calculations. So that's documentation of what we've done.</p> <p>Then of course we publish our results, which is the dissemination of the results of our calculations. So we publish papers that are a distillation of our results, without the intermediate steps (which are archived.)</p>	
<b>Q5.1 In what ways do you interact with simulations in your work?</b>	<p>We have campaigns that run for months on end. But each of the jobs within a campaign may run for a few hours. We examine the results. We can also do measurements on the lattice that gets produced.</p> <p>But we're not interacting in real-time with any of our calculations, except the most primitive way, which is running a tail on the logfile to see where things are. We check that the job is still producing sensible results and not in some peculiar state because something happened to the machine.</p>	
<b>Q5.2 How do you share simulations with others?</b>	<p>Within the collaboration: we all have immediate access to the logfiles whenever we're working together on a project.</p> <p>With the community: through conference reports and publications.</p>	
<b>Q5.3 How do you interact with inputs to your simulations?</b>	<p>The first kind of input is a very small file that sets the parameters of the simulation. We interact with this file using a text editor.</p> <p>The other part is these lattice files that I mentioned. And those we just need to feed in to the hopper (local disk). We don't interact with the detailed contents of those files because they're many gigabytes.</p> <p>We also have job scripts- only a couple dozen lines long.</p>	

Interview ID=6 07 June 2007	ANSWERS	ANNOTATIONS
<b>Q5.4 How do you interact with the output of your simulations?</b>	<p>We have the log files that I mentioned [monitor w/tail]. We have tools for producing histories of these simulations. So we produce simple graphs that are easy to view on an X Window display. We use these tools for monitoring the progress of the simulations.</p> <p>Once the simulation is finished we go through an analysis process. Normally we bring the log files to our home workstation. This is the primary analysis. We run scripts on it: curve fitting and whatever else we have to do. But that's all done on a single processor.</p> <p>The other result would be a set of lattice files. And those become input into perhaps a subsequent calculation, which is examining and measuring some physically interesting quantity on those lattices.</p> <p>So then it's another campaign with a series of lattice files, and the result always in the end is a bunch of log files containing the quantities that we wanted to measure. Those require statistical analysis and interpretation. But all that is done on a home workstation.</p> <p>[Description of a campaign] These lattice files are producing a statistical picture of what we call the QCD vacuum. A QCD vacuum is a description of the field configuration of the ground state of Quantum Chromodynamics for a given set of parameters.</p> <p>A statistical sampling normally consists of several hundred of these lattice files. We need that many to do a statistical analysis and reduce the errors.</p> <p>Generating those files requires an enormous amount of computing. And I guess it's just the human factor: we normally design our parameters so we get the best results we can in a matter of several months of running.</p> <p>So this means submitting several hundred jobs over several months to produce these several hundred files. We call that a campaign.</p>	
<b>Q5.5 By what mechanisms is access to your simulations controlled?</b>	<p>When it's in progress, our jobs will normally run for several hours, depending on the job policy of the center. We don't tend to interact ever with a job that is running, except possibly for machine diagnostic purposes. But each of the jobs ends with some new increment, and then we can look at the result – look at the log file for that increment – and make adjustments in the parameters if we need to.</p> <p>So it's at that level of interaction, which is quite minimal. We don't need to go through some massive visualization process in order to steer the calculation. Usually just adjusting a parameter or two, based on a few parameters.</p> <p>Initiation of the simulations is controlled by collaborative agreement. We're a pretty close-knit collaboration, so we don't need to have artificial methods for authorizing access to the simulations. Normally the machine policy is that the files are user read&amp;write, so the person who's doing the project would be the one who has immediate control of the files, but everyone else can read them.</p> <p>Whenever we're working together on one of our projects we all have accounts on the machines that perform the computations at the centers. The home workstations – we tend to have accounts on each other's workstations too. So we have read permissions to look at log files and results if we need it, or we just email the results back and forth to discuss them.</p> <p>[Response to direct question:] Yes, we are a multi-institutional collaboration.</p>	
<b>Q6.2 How do you share work-related data with others?</b>	<p>By having access to the same machine or emailing it back and forth, depending on the quantity of the data.</p> <p>Regarding the multi-gigabyte data and log files: we all have access to the archives where those files are kept.</p>	
<b>Q7.1 What resources do you use in your work today?</b>	<p>For compute cycles we have allocations through the NSF's LRAC process for the centers at Pittsburgh, Illinois, San Diego, Texas, Michigan.</p> <p>And for through the DOE: Oak Ridge, NERSC, FermiLab and soon through Argonne also.</p> <p>We also have a specialized center at Brookhaven.</p> <p>We tend to get allocations on the order of tens of millions of processor hours per year.</p> <p>With regard to data storage:</p> <p>We archive files at San Diego on HPSS, at NCSA on their mass storage system and in collaboration with folks at FermiLab.</p> <p>I would guess on the whole that we use on the order of hundreds of terabytes of storage.</p>	

Interview ID=6 07 June 2007	ANSWERS	ANNOTATIONS
<b>Q7.4 How do you locate available resources for use in your work?</b>	We apply for computer time at the centers, and the storage resources are made available to people who have accounts there. Based on the time that was granted to the project, on a yearly basis the centers allocate some percentage of their available compute resources to us. To use the allocation one typically submits a job, and it is put into a queue and scheduled according to the local policies of the center. It will run whenever the resource is available.	
<b>Q7.5 What types of information do you need to know about a resource in order to determine if it is suitable for your work?</b>	We look at the capabilities of the processors and the switch, or the network that connects them. Typically we run benchmarks before we start applying for time so we know how our code performs.	
<b>Q8.1 What software do you currently use in support of your work?</b>	Mostly software we've written ourselves; it is called the MILC code. It is known in the community. It is disseminated, and a lot of people use it. The graphing tools mentioned earlier are quite low level; we're not using any high-level visualization tools in our calculations. There's also a piece of software that someone in our collaboration worked on many years ago called Axis for producing two-dimensional plots. For analysis we use a wide variety of software on our workstations. Tools like fast-Fourier transforms. Also another package that we're working on called SciDAC Software Suite for Lattice Gauge Theory that's being produced under the DOE SciDAC program.	
<b>Q8.2 What scripting languages have you used in the past year?</b>	perl, shell scripts	
<b>Q8.3 What programming languages have you used in the past year?</b>	C, C++	
<b>Q8.4 What workflow tools do you use in your work?</b>	none	
<b>Q8.5 What parallel computing tools do you use in your work?</b>	MPI, the message passing system included in the SciDAC Software Suite for Lattice Gauge Theory Occasionally I use TotalView for debugging	
<b>Q8.6 If the need for new software-based functionality arises in your work how do you acquire it?</b>	We usually write it, if we can't find it somewhere else (and usually we can't.)	
<b>Q8.7 How do you share software with others?</b>	We publish it on the web.	
<b>Learning about the user's problems</b>		
<b>Q9.1 What challenges do you face today in accomplishing your work-related goals?</b>	The first challenge is getting the computer resources we need in order to do our calculations. We've been reasonably successful there – as long as the funding agencies have an interest in providing them. We also have to do a lot of file wrangling – moving big files around the country. And that presents a reasonably large challenge as these files have grown in size. So it's challenging for software, authentication and network capabilities. Another challenge is to finding the time and human resources to produce the codes in order to do the physics that we want.	

Interview ID=6 07 June 2007	ANSWERS	ANNOTATIONS
<b>Q9.2 What types of information do you need in order to address the challenges you face today?</b>	<p>I don't think it's a question of needing information, more a question of needing resources.</p> <p>The tools that we use to move files typically are the standard unix tools included with ssh. And for that we don't need more information it's just painful. Painful in the sense of having to manage the transfers by hand, restarting transfers when they fail - all of this is done by hand.</p> <p>Then of course there are the hardware problems: dealing with the sluggishness on some of the networks like the ESNET, which we've had some problems with recently. The file transmission rates are painfully slow, errors occur and then we have to retransmit. So that's the painful part.</p> <p>Then of course we have to write the scripts in order to manage transfers of quantities of files.</p> <p>With Globus there's lots of information that you need to have:</p> <ol style="list-style-type: none"> <li>1) Getting the Grid certificates <ul style="list-style-type: none"> <li>* knowing which one is the best one to get</li> <li>* knowing how you use those certificates to authenticate</li> <li>* if you've gotten one from somewhere, how you get to another place and get authenticated there</li> </ul> </li> <li>2) How you install these tools on some systems where installation may not be quite so smooth</li> <li>3) How you go about troubleshooting when things don't work. Example: So I do a globus-url-copy from one center to another and I get an error message saying "End of file encountered". And the file at the other end is of zero length. Now what do I do? Right now, I send an email to the administrator asking, "What does this mean? Why didn't it work? It worked six months ago."</li> </ol>	
<b>Q9.4 By contrast, can you provide examples of technologies you find very useful today?</b>	<p>The new switches that connect the machines, and also the faster and cheaper processors. The progress of Moore's law really makes a difference in our work. So the hardware capabilities are very useful.</p> <p>And improvements in national networks also make a big difference in our ability to get our work done.</p>	
<b>Q10.2 Describe repetitive tasks associated with your work</b>	<p>There are certainly repetitive tasks. I try to reduce them by writing scripts when I can. But there comes a point when it's harder to maintain the scripts than it is just to roll up your sleeves and do it.</p> <p>Moving files around: if I have a list of a few hundred files, and I want to get from point A to point B, and I start the process and something happens (it breaks or dies in the middle) I have to retransmit. But I don't want to retransmit all of it. The process of sorting through which one succeeded and which ones did not, and restarting. It's a time-consuming and annoying process and is something that slows down work.</p> <p>Much of what we do is repetitive, and it's not something that we would expect some broad-based tool to help with. It just requires us to produce our own tools.</p> <p>Workflow management: to some extent I think we could benefit from some of that. We have not been using any of these tools. So for the computational campaigns I described earlier - the job processes that go on for months - some of them are simple enough that there's no need for a terribly complicated tool. But we would typically submit a few jobs and have a long, long list of lattices that need to be processed. When the job finishes successfully, we have to mark that one as done, and take that off the list. So there's a process, but it is fairly simple. Not a huge complicated dependency chain that's involved, but it's an arduous process to manage all of that - sort of a daily chore.</p> <p>Managing software: we have a big codebase with hundreds of thousands of lines of code. It is continually evolving: we improve it, we add more capabilities. Whatever we do often requires making changes in the rest of the code. Sometimes it means making global changes. Finding where those changes need to be made and implementing them is a task that involves a lot of repetition.</p> <p>Testing the code: to make sure that after upgrading it still produces the same results on the same problems; to confirm that we haven't broken anything.</p> <p>There's some amount of repetition there because we test it on many different platforms.</p>	
<b>Q10.3 Describe time-consuming phases of your work</b>	<p>The time-consuming phases of the work are generally where there is a certain amount of intellectual creativity that is involved. And it's enjoyable.</p> <p>I guess I would describe the repetitive tasks as being more onerous and time-consuming things.</p>	
<b>Learning about the Globus user experience</b>		

Interview ID=6 07 June 2007	ANSWERS	ANNOTATIONS
<b>Q11.1 Which Globus data components do you directly interact with in your work today?</b>	I should begin by saying that I am only a very casual user of Globus, though I am interested in having these tools work better for me. So I've not used GridFTP very much at all. We've used simple tools – I've mentioned globus-url-copy. As part of the process of moving files around we've been experimenting with that.	
<b>Q11.2 Did you install the &lt;component&gt; client yourself?</b>	I tried to install the GridFTP client myself but it failed on Solaris, and then I gave up because I could use it from Fermilab. I didn't try to track it down further, but when I was trying to install it, it looked like it was trying to pull half of the internet onto my workstation. Part of the problem was, I think, that I ran out of disk space.	
<b>Q11.3 Did you install the &lt;component&gt; server yourself?</b>	No. At Fermilab, this is done by people there.	
<b>Q12.1 Which Globus security components do you directly interact with in your work today?</b>	I'm not very knowledgeable about which ones I'm using, but I've got a DOE Science Grid certificate. I'm not actually involved in delegating authority or anything, I'm just a user. So I've got a certificate and authenticate through grid-proxy-init.	
<b>Q12.2 Did you install the &lt;component&gt; client yourself?</b>	I've had to copy the certificate to Fermilab to use it there. I've had to run tools that translate the certificates into things that grid-proxy-init understands, but that's not very much.	
<b>Q13.1 Which Globus execution components do you directly interact with in your work today?</b>	We've experimented with using globus-job-run to launch jobs at remote locations. In this case I was trying to automate scripts so I could move files from NCSA to Fermilab for analysis and processing there. And so at Fermilab I was launching a job at NCSA to push the files back to Fermilab.	

Interview ID=6 07 June 2007	ANSWERS	ANNOTATIONS
<b>Q17 What are the major challenges you face using &lt;component&gt; today?</b>	<p>The biggest problem was getting the certificates in the first place. As a novice user, I needed to find out what was the best certificate authority to work with. I had a number of choices: I could go with the DOE Science Grid, I could go with NCSA, I could go with a couple of other places; even here at Utah somebody was trying to set something up to do that.</p> <p>I found out that some people don't trust each other. I'm trying to remember the conversation – I think it was to the effect that the people at Fermilab were not going to be trusting my NCSA certificate. So I finally went with the DOE Science Grid because it turned out that NCSA would also accept those.</p> <p>So, knowing who accepts what is a very hard thing to find out. It seems that you need to do a lot of sleuthing in order to find out. So the first hurdle is just knowing what is the best certificate authority for the projects that you're working on. It might just turn out that you have to get more than one, if you have two people that mutually distrust each other.</p> <p>Once the certificate was in place the next part is just managing the Globus process. I'm sure there wouldn't be any problem for me if I ever needed to learn more about a Globus command, finding that on the web, because that's easy.</p> <p>But as I mentioned earlier, when I ran into this snag at Fermilab and tried to do a globus-url-copy and ended up with a file of zero-length at the other end: finding out what to do next or troubleshooting is not something I am capable of doing. Not at this point, without going through a learning process, which I didn't have time to do. So knowing where to get answers to those questions – maybe if I had done a little google research I might have found an FAQ somewhere that has that. But I haven't done that yet.</p> <p>Now, so that's the Globus part of the experience. The other part of the problem is a little more complicated. Because the file transfer I was doing from NCSA to Fermilab is going to a special tape archive at Fermilab that's managed by dCache. And so to make this work I have to use an SRM-copy. And the SRM-copy was failing. And the reason it was failing is that Fermilab has to set up certain map files to make that work, and those are not being properly maintained.</p> <p>Maybe it's that not enough people are doing this kind of thing, so these things are not being maintained to the point that it makes it easy to rely on whether or not it's going to work. So then finally just fall back to the old FTP again, and that works, sort of. But in order to get the files onto the tape archive at Fermilab the scp has to go through two stages. You have to move files from a disk to a disk, then you have to move them from the disk to the tape. So that's a painful process and doubles the amount of work.</p> <p>So, the Globus and Grid idea is great, but the problem is there are still a lot of barriers to making it work, and we'd love to see it working better than it is. What I'd love to be able to do at NCSA is to get files straight from tape at NCSA to tape at Fermilab. There would be a file conversion that goes on in the process, as part of the reason we're doing this. But at this point I'm not knowledgeable about how to get them straight out of tape at Fermilab. It would be nice to be able to run a job at Fermilab that just says "move these files from the NFS system at NCSA to the tape system at Fermilab, rather than having to stage them first to disk, which is what I'm doing right now. So maybe I'm not knowledgeable enough yet, and maybe there's a way it can be done with an SRM copy, but I'm not sure if they have an SRM broker at NCSA that goes straight from their tape system.</p>	
<b>Wrapping-up</b>		
<b>Is there anything you'd like to say to the people who build software for use by people like you?</b>	I want to say that I appreciate having people who build the software actually look at how people are using them. So this interview process is useful.	

## D.7 If you add up all the tools you don't get a good user environment

Interview ID=7 12 June 2007	ANSWERS	ANNOTATIONS
<i>Pre-interview question: Do you interact directly with Globus software in your work today?</i>	yes	
<b>Establishing context</b>		
<b>Q1.1 Please provide a one-minute overview of your project</b>	We are trying to deploy a Grid, a distributed system, across the state initially in academic institutions. Our goal is to support science in research and education around the state. The addition will be also trying to bring in industrial partners and to help the business of the state as well, in terms of access to more resources and access to new knowledge and collaborations. So as part of that, our Grid of clusters is using the Globus software as kind of its base software.	
<b>Q1.2 What is the project's name?</b>	TIGRE	
<b>Q1.3 Which agency funds the project?</b>	The State of Texas	
<b>Q1.4 What field does your project belong to?</b>	Crosscutting, with foci on: Biology and Medicine, Air Quality Modeling and Geophysics	
<b>Q1.5 What is your job type?</b>	System Architect	
<b>Q1.6 How long have you been a &lt;job type&gt;?</b>	Two years	
<b>Learning about discipline-specific goals and approach</b>		
<b>Q2.1 What are the main goals of your project?</b>	The main goals of the initial phase are to <ul style="list-style-type: none"> <li>- have an operational system</li> <li>- have an initial set of users – scientists, researchers and educators – using it</li> <li>- and to begin to form the collaborations and relationships to keep it going longer term</li> </ul>	
<b>Q2.2 How will the success of your project be measured?</b>	The easy answer is we have a set of milestones from the state, and we either do or not do them. The milestones are in terms of having capabilities deployed, having enough members of TIGRE and showing usage. That's the contractual answer. The more actual answer: From the results side so far we're focusing on researchers, so as always more science, more papers coming out of it. As far as the people building it: having firm collaborations and working relationships, processes, policies and being able to operate the Grid well. And at least some start of trying to bring additional institutions into TIGRE, both academic and other. So as an aside there were five original universities as part of it. So we have a sixth participating quite a bit now, and others edging on as well. This initial phase we're focusing solely on integrating Texas universities. The types of capabilities we're deploying include a user portal, a metascheduler, a customer service system. These are bits of functionality that were called out as needed for the Grid to work.	
<b>Q2.3 What are the professional measures of success for you?</b>	The first one, as someone trying to in part manage the project: hitting our milestones and satisfying the state. It would be great if we could contribute back to the community a bit, kind of like we're doing now saying, "This is what worked. This is what didn't. Hey can you guys improve this in this way for us." And try in general to improve the software we have floating around to build this kind of Grid. Those are probably the main things I can think of, aside from the more general stuff, the fluffier success of interest, users, expansion of the Grid itself.	



Interview ID=7 12 June 2007	ANSWERS	ANNOTATIONS
<b>Q3 What are you investigating?</b>	This is where if I were wearing my computer scientist hat there would be a different answer. For TIGRE, I am mainly investigating <ul style="list-style-type: none"> <li>- the Grid itself</li> <li>- the status of the software that's available</li> <li>- what needs to be improved to provide a good environment for our users</li> </ul>	
<b>Q4.1 What is your method for investigating &lt;phenomena&gt;?</b>	Essentially it's been what you would expect: we try to find potential software systems that will satisfy our needs, we try to deploy them and see how they work, and then we work with the people who write the software to try to improve it in ways that we need. Our search for potential software begins with first thinking about what we need. Our needs are determined <ul style="list-style-type: none"> <li>- from our project milestones</li> <li>- what we learn as we start building things</li> <li>- by helping our users get done what they want to get done</li> </ul> So we'll decide that we need something to do X. And then a lot of times, since a number of us already know the community, we'll have a couple candidate packages we already know. Then usually some quick googling or asking a couple contacts would find any other ones we might not know about. So say we're looking for metascheduling. We'll know of a couple software packages that do that. We'll poke around a bit more, we find a couple others, and then go from there. I must say that if someone were trying to do this who didn't know the area that well, it would be kind of tough. Since for example, compare it to something like Linux. You can take one of a number of distros and then kind of do all your shopping there. There's a bit of that going on in the Grid area, but I think not everything needed is covered in them yet. So for example metascheduling: we couldn't go to a distro like VDT and pluck it out of there. It's not in there yet. It's not mature enough. There hasn't been a consensus yet on what is the best one. So this is where it's a good thing that a number of us on the project know the area. So we know the providers, contact them and work with them directly.	
<b>Q4.3 How do you keep track of interim results, if at all?</b>	Our main vehicle there is our quarterly reports that we do for the state, where we wrap things up and document them well. Aside from that we have developer communications: mail list, wiki; that is more dynamic capture of the stuff.	
<b>Q4.4 How do you test work-related hypotheses?</b>	I'll interpret that to mean how do we determine if a chunk of software is doing what we need it to do. In our case we just deploy it and try it out. And by "try it out" I mean that the developers will try it out themselves to see if it actually works, the documentation accurately represents the way the software works, and then we usually have a user or two give it a shot as well. Our testing at this point is ad hoc – we don't develop a big test plan, or unit tests or anything. We just kind of essentially work through the different things we want to use it for.	
<b>Q4.5 How do you document your results?</b>	Test results are documented for internal use on our wiki. Then there are the quarterly reports for the state. I'm sure we'll have some papers come out of this. We create user documentation, but also refer our users to other documentation. So there's no need to write a user guide for a tool if the one provided by the author is good. So we will introduce the tool, describe it a bit and refer off to the remote docs.	
<b>Q5.1 In what ways do you interact with simulations in your work?</b>	We interact mainly as providers of the infrastructure to people running simulations. We are trying to enable them to run them on a number of machines in TIGRE, one at a time at this point. We act as both deployers of the Grid software and then working with these user groups to get their codes running in different machines and to help them use this Grid software to get their stuff done. So part of our work is providing systems support. The other part is a higher-level user support. This higher-level support involves working with the users to help them translate their applications to the Grid context.	

Interview ID=7 12 June 2007	ANSWERS	ANNOTATIONS
<b>Q5.3 How do you interact with inputs to your simulations?</b>	<p>It's relatively simple. A classic case is helping the user to move from a cluster to the Grid. So in that case it is usually best using some Grid tools to move their data onto the machine where they end up running, and then output data off of the machine.</p> <p>And then the other thing we're starting to do is trying to help users host data collections. This work is still pretty early, but we're trying to allow users to set up a data collection, add data to it, pull data out of it, all on demand.</p> <p>The data storage resources are owned by a TIGRE member. And we are helping TIGRE users turn that storage system into more of a data collection that can be queried remotely and that can organize their data a bit. So once again, kind of a translation from keeping a dataset in a parallel filesystem for use by a small group, to making it available to more people.</p>	
<b>Q5.5 By what mechanisms is access to your simulations controlled?</b>	<p>At this point the groups performing the simulation control access. To some degree it's via unix account permissions, but we have one group that is doing simulations for a broader community. So they're taking in requests through a web browser interface and then running the simulations in a portal account.</p> <p>It varies a bit, but essentially for each project they decide how they want to do it, as long as it fits within the policies of the resource owners.</p>	
<b>Q6.1 Describe how you interact with data in your work</b>	<p>As I mentioned earlier, we push data to and fro or pull it to and fro. And help our users do that. And then we're trying to work on data collections, which are organizing a bit more in terms of location and metadata describing what the data are.</p> <p>So it's still early, but we're trying to work with one user group to make air quality and weather data available. How do you organize it in such a way that someone else can make sense of it? The "metadata" is for that. That could include a simple directory structure, it could include tags already in the data – I'm not totally clear on that one at this point. It's the usual stuff. We have no TIGRE-wide back up system, although one of the motivators for this data collection effort came out of a weather event. A collection in Houston was down for several days due to a hurricane, because they turned the machines off and put them in a truck to move them somewhere else. So part of the goal is to be able to replicate data around the state so that kind of thing won't happen.</p>	
<b>Q6.2 How do you share work-related data with others?</b>	We mainly use systems where we have accounts on that system. Like we will say, "It's on system Y. Log on there and find it at this location."	
<b>Q7.1 What resources do you use in your work today?</b>	Compute cycles for parallel and serial jobs. Data storage systems. Networks for moving data around. Some of the air quality work is beginning to tie into sensors, but it's not totally tied in at this point.	
<b>Q7.2 How do you share work-related resources with others?</b>	<p>Most of the machines are clusters or similar. So the mechanisms consist of the usual for those: queues, policies for scheduling. In terms of networking, site specific network policies are used.</p> <p>TIGRE does not own the compute clusters, which is an issue. So users of TIGRE have to negotiate separately with each site they want to use. So that means they need to contact the person at each site who has the power to authorize them. This is hard at a lot of the TIGRE sites because they're set up to serve their local campus users. It's easier for TACC because being a TeraGrid site, TACC can say, "Sure we'll give you a little starter account, and go through TeraGrid to get more cycles."</p>	
<b>Q7.4 How do you locate available resources for use in your work?</b>	<p>There are a couple of levels to getting access to something. One is knowing the resource is available. So we have a web portal that shows machines you might be able to access. And the user must go to the machine owners and try to negotiate access, seeing how much they can get in terms of cycles or storage space or whatever.</p> <p>Then after you actually have access to the resource, the process becomes more dynamic. We're starting to try to use MDS for that, but up until now it's been ad hoc scripts and things. The scripts would poke the machine to see how busy it was.</p>	

Interview ID=7 12 June 2007	ANSWERS	ANNOTATIONS
<b>Q7.5 What types of information do you need to know about a resource in order to determine if it is suitable for your work?</b>	<p>We need to know lots of things. I essentially need to figure out:</p> <ul style="list-style-type: none"> <li>- Can I build my code here?</li> <li>- Will my problem fit on this machine?</li> <li>- What is the operating system?</li> <li>- What is the software that's already been installed?</li> <li>- Related but different: Is my prerequisite software installed?</li> <li>- The number of CPUs</li> <li>- The number of nodes</li> <li>- The amount of memory</li> <li>- The disk quotas</li> <li>- The scratch disk space</li> </ul> <p>The usual things that a scientific user wants to know about a cluster before deciding if it will work for them</p>	
<b>Q8.1 What software do you currently use in support of your work?</b>	<p>Right now we are mainly using the VDT software, and we're pulling bits out of it. The bits we're using, say from Globus, include the GSI, the WS GRAM, the GridFTP. We're starting to try to use MDS.</p> <p>Related things: the UberFTP, GSI-ssh, we have Condor-G for some job management. Recently we have put up GRMS and GridWay for metascheduling.</p> <p>We also have some user portal GridSphere, GridPort.</p>	
<b>Q8.2 What scripting languages have you used in the past year?</b>	perl, python, bash	
<b>Q8.3 What programming languages have you used in the past year?</b>	Java, C, C++	
<b>Q8.4 What workflow tools do you use in your work?</b>	Both GridWay and GMRS	
<b>Q8.5 What parallel computing tools do you use in your work?</b>	MPI, TotalView parallel debugger	
<b>Q8.6 If the need for new software-based functionality arises in your work how do you acquire it?</b>	<p>We are going open source. So it comes down to locating existing software if we can, and modifying if needed. And building if we are forced to.</p> <p>TIGRE is a deployment project, not R&amp;D. So we only do development on this project as a last resort.</p>	
<b>Q8.7 How do you share software with others?</b>	Though we pick up software from other places, we provide a download from our website. We use PacMan to package our technology.	
<b>Learning about the user's problems</b>		
<b>Q9.1 What challenges do you face today in accomplishing your work-related goals?</b>	<p>The complexity and reliability of the tools we have to work with is a key problem for us. If you add up all the tools you don't get a very good user environment out of them. It just still seems to be too hard for our users.</p> <p>So for example, there's only been one person I've worked with so far that can really just figure out the stuff on his own. But he's really an exceptional kind of person this way. I'm talking about tools that have already been deployed for the users.</p> <p>So the VDT is very helpful as far as getting things deployed much easier than in the past. But then after that, trying to get users working with those deployed tools is still a problem, and takes a lot of our time to help users. So ease-of-use is still a big problem.</p> <p>As far as other challenges:</p> <p>As I mentioned, complexity and reliability are fused together. Also some pieces are missing it seems. And not all the tools work as well as you would hope in terms of doing what they say they'll do, or having bugs, or "oh we didn't think of this yet". So a lot of maturity issues with some of the tools we try to use.</p> <p>A general issue that I seem to hit: if it all works - even with say VDT - great. But if things don't work you really need an expert to fix them. The stuff that breaks can be odd. And it's not something that a sysadmin would have a chance of figuring out.</p>	

Interview ID=7 12 June 2007	ANSWERS	ANNOTATIONS
<b>Q9.2 What types of information do you need in order to address the challenges you face today?</b>	<p>So if we take the ease-of-use challenge, you can always point at documentation.</p> <p>Another thing to point related to that would be better descriptions when things go wrong. So a lot of times the messages you get about why things went wrong aren't useful to someone who's not an expert, and a lot of times aren't useful to someone who is an expert as well. You have to go around and poke around a try a number of things before you can track the issue down.</p> <p>Another information thing could be – so VDT is kind of doing this – of saying, “Here's a set of packages that should all work together for you.” That's quite good. It's still not quite everything someone might need in that set of packages, but it's a good start.</p> <p>I guess what I'm kind of looking for personally, is like when I install my favorite linux distro (or cygwin on windows, or fink on the mac). I can go and pick out what I want and it almost always works. You get a menu, you pick, it installs it and everything works.</p>	
<b>Q9.4 By contrast, can you provide examples of technologies you find very useful today?</b>	<p>VDT helps us a lot, and the linux, cygwin and similar packaging tools are all very good at that level.</p> <p>Perhaps something else to look at is what Sun and Amazon are doing with their user-accessible machine rooms. Upload your data, upload your executable and it kind of just seems to work. That would be a great goal for the academic community. It would be a little harder problem – with different kinds of machines, and parallel applications, different owners... I'd also point off to a lot of the commercial web stuff. You know – you go to amazon or your bank. How easy to use that is, and how responsive. Like if I can find all my bank history in a split second, why can't I find a machine that meets my requirements in less than ten seconds.</p>	
<b>Q10.1 Can you think of any work-related tasks that decrease your productivity?</b>	<p>Debugging and spending more time than I think I should need to working with users on this stuff. By debugging I mean mainly figuring out what's going wrong – that's probably the main thing. It really shouldn't be that hard for them. If we can get a lot of users going on our clusters with minimal interaction with user support staff, we should be able to do as well when getting them on the Grid.</p> <p>There are also delays once you find the problem, either you fixing it, or someone else fixing it: getting that fix back in to the people who wrote the software. But some delay is expected; it takes a little time. So mainly just figuring out what's going wrong and why.</p>	
<b>Learning about the Globus user experience</b>		
<b>Q11.1 Which Globus data components do you directly interact with in your work today?</b>	GridFTP	
<b>Q11.2 Did you install the &lt;component&gt; client yourself?</b>	Yes	
<b>Q11.3 Did you install the &lt;component&gt; server yourself?</b>	Yes. For TIGRE we have the servers running on server machines. We also provide a client tools package that just includes the globus-url-copy and uberFTP clients.	
<b>Q11.4 How many people currently use your &lt;component&gt; server</b>	Everyone on the project uses it; I am not sure the exact number of users there are on the project.	
<b>Q12.1 Which Globus security components do you directly interact with in your work today?</b>	GSI-OpenSSH, grid-proxy-init, MyProxy	
<b>Q12.2 Did you install the &lt;component&gt; client yourself?</b>	Yes	
<b>Q12.3 Did you install the &lt;component&gt; server yourself?</b>	Yes	

Interview ID=7 12 June 2007	ANSWERS	ANNOTATIONS
<b>Q12.4 How many people currently use your &lt;component&gt; server</b>	Everyone on the project uses GSI-OpenSSH and grid-proxy-init (and again, I am not sure the exact number of users there are on the project.) For MyProxy, this is not the case, because right now everyone is getting long-term certs. But we are probably going to move toward giving everyone short-term certs through MyProxy at some point. This is strictly for a usability thing. So we're probably going to switch from people doing grid-proxy-init to people doing MyProxy login and not having to manage their own long-term user cert.	
<b>Q13.1 Which Globus execution components do you directly interact with in your work today?</b>	GRAM4 For TIGRE we do not support GRAM2, but some machines have it so we can do a little sanity check sometimes. But GRAM4 is the service we offer. We have GridWay a little bit up – we're having problems with it. We're using GRMS instead. Some folks have used GSI-OpenSSH to submit jobs to remote machines.	
<b>Q13.2 Did you install the &lt;component&gt; client yourself?</b>	Yes for GRAM4 There is no GridWay client	
<b>Q13.3 Did you install the &lt;component&gt; server yourself?</b>	Yes for GRAM4 I did not install GridWay myself	
<b>Q13.4 How many people currently use your &lt;component&gt; server</b>	Everyone on the project uses GRAM4 Because GridWay is not working, only one or two users for that	
<b>Q14.1 Which Globus information components do you directly interact with in your work today?</b>	We are trying to use the MDS4 Index service. I am not sure about the Trigger service.	
<b>Q14.2 Did you install the &lt;component&gt; client yourself?</b>	Yes.	
<b>Q14.3 Did you install the &lt;component&gt; server yourself?</b>	Yes as part of the default Globus container, I installed the Index service. But I think we're working on additional ones that I did not install.	
<b>Q14.4 How many people currently use your &lt;component&gt; server</b>	Directly usage includes three or so people who are trying to get the Index service to work. Indirectly we're trying to use it via GRMS, which we're playing with but not everyone is using yet because it still has problems getting information from MDS.	
<b>Q15.1 Which Globus common runtime components do you directly interact with in your work today?</b>	For TIGRE directly, none. I've personally (not TIGRE related) used the Java WS Core. But just so you know- not the WSRF parts. I used it as a WS container that supported GSI security.	
<b>Q15.3 Did you install the &lt;component&gt; server yourself?</b>	Yes	
<b>Q15.4 How many people currently use your &lt;component&gt; server</b>	Just me, because I don't have everything running yet.	

Interview ID=7 12 June 2007	ANSWERS	ANNOTATIONS
<b>Q16 Why do you use &lt;component&gt; instead of an alternative technology?</b>	<p>GridFTP: Because it support GSI security. Performance could be a reason, but it has not yet been a reason for TIGRE.</p> <p>GRAM4: We picked GRAM4 over GRAM2 because it is being developed, where GRAM2 is not. We picked GRAM4 over another technology because we needed to leverage a lot of stuff for the project to work, given our current funding level. VDT was a good thing to leverage, and it had GRAM4 in it.</p> <p>MDS4 Index: honestly we would like an alternative [see Q17].</p> <p>GridWay: we are not yet using GridWay; we're using GMRS instead because it's working better for us. We've not given up on GridWay yet. They're both under active development, so we're watching them. We are interested in GridWay because of its capability to automatically pick where to run jobs given multiple machines, as well as its support for simple workflows or sets of jobs.</p> <p>GSI certificates: Mainly because it is the default in the community for this type of Grid. By "Grid type" I mean TIGRE is a Grid of clusters, in contrast to a high-throughput Grids like Condor or BOINC.</p> <p>Java WS Core: I was doing some web service stuff, and I wanted to use Java and GSI. I was also using Axis as well.</p>	

Interview ID=7 12 June 2007	ANSWERS	ANNOTATIONS
<p><b>Q17 What are the major challenges you face using &lt;component&gt; today?</b></p>	<p>MDS4 Index: It breaks, it's slow, and it's overly complex, in terms of the model. What I mean by that is:</p> <ul style="list-style-type: none"> <li>- XML and XPath is more than is needed 98% of the time</li> <li>- Java makes it quite heavyweight for small things</li> <li>- The last I heard they were running in memory instead of out of a disk-based database, which hogs a lot of memory.</li> </ul> <p>GridWay: It's not a client server model, which is our preferred operating model. They had a problem with proxy propagation, which meant GridWay wouldn't run with a delegated credential. You had to go on to the GridWay machine, do a grid-proxy-init there, and then start doing GridWay stuff. This might change soon, not sure. I never quite understood it, but that's what the people were telling me who were evaluating it. There were also maturity issues with the software, with some bugs that need to be shaken out.</p> <p>GSI certificates: There are a couple of challenges. One is getting users to understand them. It's new to them. They don't get it – don't understand it. This is why I personally like the MyProxy with the built-in CA. You can tell the user, "Do a MyProxy logon." They give it a username and password, and they're good.</p> <p>Which I like because it is kind of like the Kerberos model. It seems at that level better for the user – quite similar to the MyProxy stuff. Essentially, "Run this command with a username and password, and don't worry about it." "If you want to check how long you can do stuff, here's another command." So that's the user side of the GSI challenges.</p> <p>The other side of the GSI challenge is the CA. Running a CA, deciding whom you trust – that's all a large pain. A very large pain. For example, you have to get a CA certified by TAGPMA and buy special hardware.</p> <p>And to be blunt: after all this is done, as a user we don't gain much of anything.</p> <p>No additional capabilities – you can access the same machines as you could before. You know, it's a big hassle for some potential benefits, like delegation, having your own agents out there to do things for you. There is some potential there, but it just hasn't come to pass that we've needed it.</p> <p>Java WS Core: I did not do any WSRF stuff, so I did not find any challenges in using Java WS Core that were not there for using SOAP or WSDL in general.</p> <p>GridFTP: I'm not sure there are any. It kinda just works and that's great.</p> <p>GRAM4: The challenges there relate to getting the backend scripts to work with the local scheduler. Also, one thing we've never liked is the default environment you get is very bare, which is not typical when you run a job on a cluster. Such as paths on setup.</p> <p>A lot of times a cluster user will modify their .profile [<i>file holding unix environment settings</i>] to set their environment for their jobs. They want those values to be used for the job via GRAM4, but they aren't. In contrast, if they submit the job directly to the scheduler those values will be picked up.</p> <p>The current GRAM4 default is not what you'd expect, so we actually change that here; we modify the GRAM scripts. Also I notice a lot of the RSL attributes that are defined, if you can find them on the webpage, are not implemented in the backend scripts.</p> <p>So, for example, we just added some memory support into ours here.</p> <p>Given that our nodes are multi-core, we need to allow our users to say, "I need this many processes with this much memory per core." So that results in us putting one process per core per node, or perhaps one process per two cores per node, depending on the amount of memory they need. So that wasn't a big deal, but it was something we had to add in recently.</p> <p>The LSF.pm scripts did not include support for taking the min memory XML-based RSL attribute and turning it into the right LSF line in the submit script.</p>	
<p style="text-align: center;"><b>Wrapping-up</b></p>		
<p><b>Is there anything you'd like to say to the people who build software for use by people like you?</b></p>	<p>Two high-level things:</p> <p>Make it easier to use.</p> <p>It needs to do more. We really need the next level of middleware (or whatever it is called) to make all this stuff at all usable. That could be another level of middleware, or it could be right at the top: a user environment type of thing.</p>	

## D.8 I am trying to understand where Grid Computing adds value

Interview ID=8 13 June 2007	ANSWERS	ANNOTATIONS
<i>Pre-interview question: Do you interact directly with Globus software in your work today?</i>	Yes	
<b>Establishing context</b>		
<b>Q1.1 Please provide a one-minute overview of your project</b>	I work in the area of Lattice QCD, which is a numerical approach to quantum field theory. It requires us to run many large-scale calculations. By large-scale, I mean currently projects that take on the order of a teraflop-year, and in the future may take tens or hundreds of teraflop-years of computation.	
<b>Q1.2 What is the project's name?</b>	We have a collaboration called the MILC (MIMD Lattice Computation) Collaboration	
<b>Q1.3 Which agency funds the project?</b>	The Department of Energy and the National Science Foundation	
<b>Q1.4 What field does your project belong to?</b>	Physics, Elementary Particle Theory	
<b>Q1.5 What is your job type?</b>	Professor	
<b>Q1.6 How long have you been a &lt;job type&gt;?</b>	22 years	
<b>Learning about discipline-specific goals and approach</b>		
<b>Q2.1 What are the main goals of your project?</b>	The main goals are to understand Quantum Chromodynamics (QCD), which is one component of the Standard Model of elementary particle physics. This includes: - calculating the masses of particles that interact strongly; those particles are made out of quarks and gluons, - calculating their decay properties, - and we're interested in studying QCD at very high temperatures, and possibly with non-zero chemical potential.	
<b>Q2.2 How will the success of your project be measured?</b>	By correct calculations of quantities that can be measured in experiments at places such as the Relativistic Heavy Ion Collider (RHIC), the Fermilab Tevatron, the SLAC B Factory, KEK B Factory, etc. So if we can calculate quantities and have them verified by experiment – if we're lucky enough to make predictions, not postdictions – that will make us happy.	
<b>Q2.3 What are the professional measures of success for you?</b>	Getting the Nobel prize. Well, I'm a full professor now, so things like getting prizes from the APS, becoming a distinguished professor, would be other steps. But more realistically: publishing papers that are highly cited, getting more grants, getting more computer time, etc. <i>[prompt asking if getting computer time is tied to getting grants]</i> We apply to the DOE and NSF for grant money to do our regular research. We apply to the LRAC for NSF computer time, as well as to NERSC for computer time. So they're separate proposals, and not necessarily directly linked. In our last application to LRAC one of the reviewers seemed particularly misguided and didn't seem to read very carefully. He commented that our monetary funding seemed small. We pointed out that different fields require different amounts of money. In our case, most of the support we need is for postdocs and graduate students. The other important support is the computer time itself, which is not measured in money.	



Interview ID=8 13 June 2007	ANSWERS	ANNOTATIONS
<b>Q3 What are you investigating?</b>	<p>We are investigating this quantum field theory, which is called Quantum Chromodynamics. And what we are trying to do at the moment is calculate the masses of particles made out of up, down and strange quarks; those are the three light quarks in nature.</p> <p>We are also interested in calculating the masses of heavier states that involve the charm quark and the b-quark. In fact there was an announcement just today regarding the discovery of a new baryon, called the “cascade b”, seen by the DZero collaboration at Fermilab.</p> <p>My graduate student has been trying to calculate the mass of this particle. There are statistical and systematic errors associated with the calculations, and the student’s errors are still bigger than the experimental error. He was hoping to get a prediction, but is not ready. But I have confidence that when he examines more of the configurations that we have saved (as a result of our previous investigations) and does further analysis, he’ll be able to reduce his errors. How much is still up in the air.</p> <p>So we are trying to calculate the masses of these particles and compare them with experimental masses that have been observed. Also we try to calculate some masses that have not yet been observed (which before today would have included this cascade b mass.)</p> <p>Then we are trying to calculate the decay properties of various particles. The ones we’re most interested in are particles made out of a quark and an antiquark. They’re called a meson. Most of the time we’re looking at the decays of spin zero mesons, which are called pseudo-scalars. We look at the lightest meson in nature (the pion) and the next lightest meson (the kaon), and we can quite accurately calculate their decays.</p> <p>We’re also very interested in mesons made with one heavy quark: a charm quark, or a b-quark. The charm quark meson decays were only measured about a year ago in experiment. We successfully predicted those decay properties to ten percent accuracy, which was about equal to the experimental accuracy. And we’re working hard to increase our precision because we know the experimentalists will collect additional events, and by having additional events they’ll be able to reduce their errors.</p> <p>We’re also interested in the properties of Quantum Chromodynamics at very high temperatures. There is an accelerator called the Relativistic Heavy Ion Collider at Brookhaven National Lab, where they’re producing something called the quark-gluon plasma. And we are interested in the properties of that state of matter.</p> <p>In particular, we’re trying to calculate what temperature you need to produce it; this is called the transition temperature. And we’re trying to calculate what’s called the “equation of state of that system”, which is the relationship between the pressure or energy and the temperature of the system.</p> <p>We’re also trying to calculate the masses of the quarks themselves – the up, down and strange masses. We have predictions of those.</p> <p>And then there are all the other things... low-energy constants of the chiral Lagrangian, etc. But I think this is probably enough to mention here.</p>	
<b>Q4.1 What is your method for investigating &lt;phenomena&gt;?</b>	<p>Using this theory Quantum Chromodynamics, we take the continuous space-time of nature and approximate it as a grid of points in space and time. And then we have our variables either live at the grid points or, in the case of a very important variable called a gauge field, on the links joining the grid points. We can then formulate this theory as a finite theory with a finite number of grid points and links. This gives us a finite number of variables, so we can put it on a computer system.</p> <p>Then we examine the system using various numerical techniques, the most important of which is something called “hybrid molecular dynamics.” So we use an approach like that to create typical snapshots of the fields that describe the system. The different snapshots correspond to the quantum fluctuations in nature.</p> <p>Then we average over the snapshots that we save, in order to calculate different physical processes. We have to do a lot of sparse matrix inversions. And this molecular dynamics involves frequent inversions. Thus it takes a lot of time to do the calculations.</p>	

Interview ID=8 13 June 2007	ANSWERS	ANNOTATIONS
<b>Q4.3 How do you keep track of interim results, if at all?</b>	<p>We take snapshots of the system as we evolve it. We have this grid with variables defined on it. And we archive snapshots of the variables in our system. So typically I might run, for three hours, say, at the Pittsburgh Computing Center and create the next configuration. I would store it on the disk. (Actually, at Pittsburgh the run would be one hour and twenty minutes.) I have five of these evolutions in a job. And at the end of the job I have one 4.3 Gigabyte file that I will archive on the tape system. And so we save hundreds of these configurations. Not all are that size – many are smaller. This represents our latest calculations, which generally get bigger with time.</p> <p>These archived files aren't results yet. With this snapshot of the variables we then look at different physical processes using a different computer program. But once the files are archived, and we share these files with the rest of the lattice gauge theorists in the world, people are able to look at different physical quantities on that grid of variables.</p> <p>Then what we do is average over an ensemble of snapshots; in our case an ensemble would be made up of five- or six-hundred snapshots. Once the averaging is done, we have an average that is much smaller than the totality of the data. And then we do fitting of this, for instance to get masses of particles.</p> <p>So generally things go in two stages. One stage is generating the configurations. The second stage, which can be done anytime later, is writing and running a program to measure some physical quantity. There are many possible physical quantities one could look at.</p>	
<b>Q4.5 How do you document your results?</b>	<p>We have been doing this for quite a while. So people know we have certain ensembles of configurations available. A number of them are made available through something called the Gauge Connection [<a href="http://qcd.nersc.gov">http://qcd.nersc.gov</a>], which is a service of NERSC. So we will deposit configurations that we are ready to share with the world at this NERSC repository.</p> <p>There is also a newer system called the International Lattice Data Grid (ILDG), which is supposed to be a searchable database of the configurations people have made available to the international community. And I believe we have some ensembles there. But since I'm the producer of the configurations, I spend less time thinking about what's available that way than how we store them.</p> <p>So I archive some of them, and then we archive files at NCSA and at SDSC. They're organized in a hierarchical fashion. So you should be able to see what's available with a little bit of searching around, if you have access to the mass storage system. So internally you can do that; externally you'd have to rely on the Gauge Connection or NERSC. Or email: "Hey! Do you have this available?"</p> <p>Each file comes usually with a little info file that tells when the data was produced and what parameters were used to produce it. And then as far as the intermediate files go, generally individuals are creating those and hopefully backing them up. I know I tend to back things up at the Pittsburgh Supercomputing Center.</p> <p>Then we all have lots of data on our own workstations. We could be more systematic about backing that up centrally. Because what can often happen is you have an idea of who might have run a given physics project, and then you have to email and say, "Where are the output files?"</p> <p>As far as documenting the measurements, we have one or several files for each of the snapshots, and then someone will average that up and do some kind of fitting procedure to extract a physically useful result. So you either make a table of what the fitted values are (and there may be different cases.) You may have to do some other fits, at a second level of fitting. At that point we're talking to each other and producing graphs that we hand around for people to look at. So generally a few people are involved in the actual fitting. Then the people who aren't doing the actual fits will examine the results of the fit and the confidence levels, and the plots, etc. And eventually it goes into a paper.</p>	

Interview ID=8 13 June 2007	ANSWERS	ANNOTATIONS
<b>Q5.2 How do you share simulations with others?</b>	<p>The configuration files are huge and people don't tend to collect them. So a collection of 500 four gigabyte configuration files is not something you want to bring to your computer because it's only useful to have it where the analysis is done. So these tend to stay on these central, large computers where we have backup/tape systems, large disks and multiple CPUs for performing analyses on the snapshots. The output from the snapshots tends to be on the order of one megabyte, roughly speaking. It's much smaller than the configuration files. So people (hopefully) archive them in case their workstation crashes. They will definitely bring them back to their own computer for subsequent analysis.</p> <p>Our code has been open for a long time; it's called the MILC code. And so we have an ensemble of applications that people can use to analyze our configurations, or to create configurations on their own. It has been freely available for years. And people will develop a new application, and there may be a time lag between when the code for that application is produced and it goes into the suite of applications that we make freely available to the world.</p>	
<b>Q6.1 Describe how you interact with data in your work</b>	That is an interesting way of putting things. I interact with data by creating it, archiving it, moving it around the country, and analyzing it. I spend a lot of time making sure files got moved.	
<b>Q6.2 How do you share work-related data with others?</b>	We make data available through the gauge connection, through the ILDG, and often by working together on the same machine, or by collaborators requesting a certain set of output data. And telling them where they can find it, or moving it to some machine where we have a commonly readable directory. I've also put things up on the web.	
<b>Q6.3 By what mechanisms is access to your work-related data controlled?</b>	By unix file permissions in most cases.	
<b>Q7.1 What resources do you use in your work today?</b>	<p>-The Pittsburgh Supercomputing Center, including the tape system FAR [<a href="http://www.psc.edu/general/filesys/far/far.html">http://www.psc.edu/general/filesys/far/far.html</a>]</p> <p>- NCSA</p> <p>- A computer at Indiana University that's part of the TeraGrid</p> <p>- A cluster dedicated to Lattice QCD at Fermilab</p> <p>- My desktop workstation at school</p> <p>- My laptop</p> <p>I have jobs in the queue at Pittsburgh, Indiana, Fermilab and I'm running benchmarks at NCSA</p> <p>We use the archival system at Pittsburgh, NCSA, SDSC, Fermilab and Indiana to some extent</p> <p>I also built myself a .75 terabyte raid system that's sitting in my office, but there's no tape backup; there's also only a 10 mbit network connection into my office, which is annoying</p>	
<b>Q7.2 How do you share work-related resources with others?</b>	<p>Generally a project that we agree we want to accomplish is assigned to a center. And someone is assigned to do the running, and that person tries to use up our allocation. We have discussions if something's not going quickly enough. In these cases we might move it to another center, or ask for a dedicated queue, or tell someone to let another person get more time.</p> <p>This level of coordination happens within the MILC collaboration, as well as other collaborations I'm involved in. For example we're also running with a different group of people called Hot QCD at the Lawrence Livermore Lab. Those runs involve classified Blue Genes computers, so we don't actually do anything there but see some of the output. A strange way of doing a computational project, but that's another place where, in principle, we have lots of resources.</p>	
<b>Q7.4 How do you locate available resources for use in your work?</b>	<p>We apply for time and then we have a certain allocation at each place, which we try and pay attention to. We apply:</p> <ul style="list-style-type: none"> <li>- to the NSF through the LRAC,</li> <li>- to the DOE; at NERSC we have an allocation,</li> <li>- to the Fermilab US QCD computing cluster</li> <li>- and then this year we'll be applying to the DOE's INCITE program</li> </ul> <p>We also try and get friendly user time when we can.</p>	
<b>Q7.5 What types of information do you need to know about a resource in order to determine if it is suitable for your work?</b>	That's a good question. We need to know how big it is. Generally what kind of CPUs it has, and if we anticipate it will be fast enough to be of value to us. We're usually less concerned about the per-node memory or the disk because we tend not to stress that as much as some other fields.	

Interview ID=8 13 June 2007	ANSWERS	ANNOTATIONS
<b>Q8.1 What software do you currently use in support of your work?</b>	<p>We write our own codes in C and sometimes in assembler. So we need good compilers.</p> <p>We need an excellent implementation of MPI.</p> <p>We tend to do some of the debugging on our own workstations. We could probably be more adept at using performance tools and debugging tools at the centers where the parallel computers are located.</p> <p>We also use various scripting languages, like csh, bash and perl.</p>	
<b>Q8.4 What workflow tools do you use in your work?</b>	<p>That's a very interesting question. Generally we tend to do it on our own. So we write scripts and babysitters that will monitor whether or not things are moving along the way we like, and automating certain tasks. We actually have very little experience with anything you would call a workflow tool provided by a third party. However, as part of the US QCD project, there is an effort to start using workflow tools at Fermilab. I'm not the person working on that, but that is an issue for our collaboration. To see if we can make things smoother.</p>	
<b>Q8.5 What parallel computing tools do you use in your work?</b>	MPI	
<b>Q8.6 If the need for new software-based functionality arises in your work how do you acquire it?</b>	By writing it.	
<b>Q8.7 How do you share software with others?</b>	<p>We have a freely available code database called the MILC code served up via the web.</p> <p>There's also other software infrastructure provided by the US Lattice QCD Infrastructure project, which we usually call the SciDAC project because that's who funds it and we know we're always talking about Lattice Gauge Theory.</p> <p>So if you were to go to <a href="http://www.lqcd.org/">http://www.lqcd.org/</a> you would probably find that type of software infrastructure involved. There are a bunch of libraries which people can use as the basis of code.</p> <p>So MILC is kind of a higher-level application, and there's a lower-level of code that goes by the name of QLA, QMP, QDP, and a few other things all of which start with "Q".</p> <p>So there is a big effort to provide community software within the field of Lattice Gauge Theory. And there are at least two other groups outside of MILC who are providing application-level code.</p>	
<b>Learning about the user's problems</b>		
<b>Q9.1 What challenges do you face today in accomplishing your work-related goals?</b>	<p>One of the biggest challenges is getting a large enough amount of computer time to do the calculations that we would really like to be able to do but just can't accomplish right now.</p> <p>And then, if we had those allocations, the challenge would be making sure the jobs all get run. Then hopefully the fun would be in analyzing them.</p>	
<b>Q9.2 What types of information do you need in order to address the challenges you face today?</b>	<p>Which funding agency is making compute cycles available, how big those computers will be, how to run on them, how to apply for time on them, how to compile jobs, where the scratch space is on the system, how to get to the archival storage system... things like that.</p>	

Interview ID=8 13 June 2007	ANSWERS	ANNOTATIONS
<b>Q9.3 What technology-related obstacles do you currently encounter?</b>	<p>Technology can throw obstacles my way when things don't work, and that can be frequent in the world of supercomputing. But for us I think the main thing is we'd like to have more computers and more powerful computers available to us.</p> <p>We typically run on up to 2,000 cores, and in the future I think we'd be happy running on larger numbers of cores. So we like very big computers and we like to be able to get very large chunks of time on them.</p> <p>We also need to have sufficient disk space. Also, access to archival storage and networking in order to move configurations or the snapshots off to another center for further analysis.</p> <p>So those are the main things for us.</p> <p><i>[prompt asking if interviewee is often told there are no more cycles available, or no more disk space]</i></p> <p>Well, having no more cycles is more likely than having no more disk space, because as I mentioned there are other users who may stress disk and memory a lot more than we do. Now I won't say I've never run out of disk space or had some file I wanted wiped off the machine because of some policy. But generally at the NSF centers we're not seeing that the disks are 99% full and jobs are failing because the disks are full.</p> <p>Now at Fermilab sometimes when I'm trying to transfer a file some raid array may get full and I'll just have to copy it again. But I don't generally tend to lose stuff. Occasionally, but not too much.</p>	
<b>Q9.4 By contrast, can you provide examples of technologies you find very useful today?</b>	MPI, scp, fast networking with low latency, fast memory access, high speed chip technology with good floating point performance	
<b>Q10.1 Can you think of any work-related tasks that decrease your productivity?</b>	<p>Fighting with security issues. Having people tell me to change my password all the time and then I can't remember it.</p> <p>What I find annoying is there are so many authentication schemes. For instance in our DOE SciDAC-funded project we have computers at Jefferson Laboratory, Brookhaven National Laboratory and Fermilab. Everyone uses a different security system, and I find it annoying. Generally I find it hard to imagine that the people who do these security services have ever done any large-scale computing project.</p> <p>For instance, I'm moving files from Pittsburgh to NCSA, so every time a job finishes (and these jobs run for over a year, one after another) I have to transfer a file. So the notion of typing in a password and doing that by hand is very annoying, compared to having it happen with some sort of automatic system.</p> <p>Having jobs crash, often due to a node going down. That's the most frequent reason that a job fails to complete. Having to fix things up.</p>	
<b>Q10.2 Describe repetitive tasks associated with your work</b>	<p>Running thousands of jobs to create the configurations, and running many hundreds of jobs on each ensemble of configurations to do the measurement routines.</p> <p>And then to move stuff around from one center to another.</p> <p>That kind of stuff gets very repetitive. And as I say, we do our best within our tools to automate that. For instance, I have several jobs in the queue, and as soon as one finishes it submits the next one, but I have to check to see whether anything crashed and whether anything actually ran.</p> <p>Similarly I have a babysitting script that can see whether any new file that needs to be archived has been produced, and then move the new file. This process would work automatically in a system where a password isn't required to be typed.</p> <p>The jobs I'm currently executing run so intermittently that I don't just put that in the background. But in times past a similar script might wake up every hour and a half and see if a file has been produced; it might transfer several files a day. The script that I'm thinking of probably now only gets to run every two days, so I just start it up when I need it.</p>	
<b>Q10.3 Describe time-consuming phases of your work</b>	<p>The repetitive tasks are somewhat time-consuming.</p> <p>And then the parts that are more fun are doing the fitting later, and trying to understand out what's going on. Just running the jobs is not all that exciting, but it needs to be done. I tend to call it blue-collar physics.</p> <p>Then we also have to write new codes, and that can be time-consuming. Creating the codes and doing the debugging.</p> <p>Preparing proposals is also time-consuming.</p> <p>Participating in conference calls because of all these big projects that require a certain amount of coordination.</p>	
<b>Learning about the Globus user experience</b>		

Interview ID=8 13 June 2007	ANSWERS	ANNOTATIONS
<b>Q11.1 Which Globus data components do you directly interact with in your work today?</b>	I don't think I've used any of those. Several of them I've never heard of. I did use UberFTP this week for the first time. I used UberFTP after complaining to the consultants at Indiana University and NCSA that my transfers from NCSA to IU were very slow. A system engineer in the NCSA consulting group suggested I try to use it. And after much struggle and failure I was finally able to use it.	
<b>Q12.1 Which Globus security components do you directly interact with in your work today?</b>	I used MyProxy this week and last week for the first time. I use ssh all the time without using a Grid certificate. But I often use a kerberized [Kerberos-based] ssh.	
<b>Q12.3 Did you install the &lt;component&gt; server yourself?</b>	No, I was using one at NCSA.	
<b>Q14.1 Which Globus information components do you directly interact with in your work today?</b>	I haven't heard of any of them.	
<b>Q16 Why do you use &lt;component&gt; instead of an alternative technology?</b>	UberFTP: I tried it because someone told me it might be faster than scp. MyProxy: because the NCSA engineer said, "Give this MyProxy command and then use UberFTP.	

Interview ID=8 13 June 2007	ANSWERS	ANNOTATIONS
<p><b>Q17 What are the major challenges you face using &lt;component&gt; today?</b></p>	<p>UberFTP: the major challenge I faced is the first time I tried to transfer a file, it only transferred one tenth of it and then it stopped. And in general it's not always clear who to ask for help because it's always a transfer between two different sites. So you have to get both sides involved, which can sometimes be difficult. Sometimes they don't communicate with each other – they'll only communicate with me. They may have different help systems.</p> <p>MyProxy: I don't know where the documentation is, and if it doesn't work the first time I have no idea what to do. One time I had an expired Grid certificate, and at NCSA it was quite easy to generate a new Grid certificate, but only because I had taken a (paper) file folder from my office, which normally I would not have with me, that has my default password (because I'm traveling right now.) So if I hadn't decided at the last minute to take this file I would not have been able to get a new certificate.</p> <p>And then I find it very difficult to figure out how to register these certificates at different sites, because I have a different user name at NCSA and at Pittsburgh and at SDSC. So first of all I found it difficult to find the right place to start looking for documentation about how to get my certificate registered at a new site. I found it easy to google and figure out how to get a certificate. But then to get the Distinguished Name registered and hooked up to each individual account took me a long time to find the right place to start looking for documentation to do that. And then once I found the documentation, some instructions said to use gx-map – other places said to use gx-request. In almost every case, neither one was on my path, and I had to hunt for probably thirty minutes before finding it so I could actually use it.</p> <p>General: It is difficult to figure out where to find the right documentation. Once I do find the documentation it's very hard to understand – it's full of acronyms, and refers to unfamiliar and unnatural concepts.</p> <p>I find that I often don't have the right commands by default in my path. Most of the services seem to do what I have been otherwise able to do for a decade or more (such as moving files with scp or ftp.) So I'm trying to understand where the value is added. Maybe UberFTP is able to move my file about twice as fast... I haven't yet tried it between Pittsburgh and NCSA. I should, because I often transfer files along that path.</p> <p><i>[prompt asking for example of good documentation]</i></p> <p>When I go to a new computer at the Pittsburgh Supercomputing Center I find they do an excellent job of organizing the information that I need to know in order to use that computer.</p> <p>In contrast, the Grid-related documentation for things like getting a certificate is organized into very small chunks. They weren't organized in the steps I wanted. As I mentioned, there are lots of different acronyms. The documentation seems to have lots of different paths because there are so many different ways of doing things. As a user I want one way that is going to work, and work easily.</p> <p>Most of the capabilities I read about I can do on my own. People have been talking to me about the Grid for years. And I'd say, "Tell you what: here's one of my job scripts – here's what I do. You convert it to Grid commands and we'll try it." And I never hear anything back from them. So:</p> <ul style="list-style-type: none"> <li>- we have to get files out of archival storage</li> <li>- we have to run a job</li> <li>- we have to put files back into archival storage</li> </ul> <p>In fact there was a case in which I was creating configurations at Pittsburgh and doing analysis at IU. This is a perfect example for distributed computing: where you run a job at one place and you put the output somewhere else and run some subsequent step of the job. I could do all that just fine with ssh, using the network, queuing a job at IU. And nobody ever picked up the challenge of trying to do that with Grid commands. All I needed was a little background script running at Pittsburgh that I could understand quite well.</p>	
<p><b>Is there anything you'd like to say to the people who build software for use by people like you?</b></p>	<p style="text-align: center;"><b>Wrapping-up</b></p> <p>The people who are doing this ought to have some experience using these systems for large-scale projects. My feeling, perhaps out of ignorance, is that most of these tools that I've seen that are called Grid tools reproduce services we could do before. They seem to have complicated names and complicated protocols. And, regarding the security, the tools are not designed to run jobs that will run for months and months and months without too much user interaction.</p> <p>So I think it's very good that this survey of users is being done. And I think it's something that should have been done five or more years ago.</p>	

## D.9 Globus enables more science

Interview ID=9 2 July 2007	ANSWERS	ANNOTATIONS
<i>Pre-interview question: Do you interact directly with Globus software in your work today?</i>	yes	
<b>Establishing context</b>		
<b>Q1.1 Please provide a one-minute overview of your project</b>	LIGO is a project funded by the National Science Foundation to detect gravitational waves from astrophysical sources. The LIGO laboratory runs and maintains three different interferometer instruments at two different sites: one being in Hanford, Washington and the other in Livingston, Louisiana. In addition the LIGO laboratory maintains large computing centers at the sites, as well as at Caltech and at MIT. Additionally, the larger LIGO scientific collaboration has a number of large compute resources available at the University of Wisconsin-Milwaukee and Penn State University. The collaboration also has some data analysis capabilities and clusters in Europe. My primary focus is on the LIGO DataGrid. The focus of the DataGrid is first of all to make data available after it is saved – comes off the instruments – make it available to all the analysis sites. And then secondly to enable scientists to efficiently use the data sites around the world to analyze LIGO data.	
<b>Q1.2 What is the project's name?</b>	LIGO	
<b>Q1.3 Which agency funds the project?</b>	National Science Foundation	
<b>Q1.4 What field does your project belong to?</b>	Gravitational Wave Physics, Physics, Astrophysics, Gravitational Wave Astronomy	
<b>Q1.5 What is your job type?</b>	System Architect, Scientist	
<b>Q1.6 How long have you been a &lt;job type&gt;?</b>	Five years	
<b>Learning about discipline-specific goals and approach</b>		
<b>Q2.1 What are the main goals of your project?</b>	The main goal of the LIGO project is to detect gravitational waves and to conduct gravitational wave astronomy. In other words, to learn about our universe through the use of gravitational waves. Diving down a bit, my real project is the LIGO DataGrid. And the purpose of the LIGO DataGrid is to enable as much science as possible to be conducted using the LIGO data. The data has to be analyzed. It's very computation and data intensive. And the main goal of the project is to build infrastructure (tools, middleware, end-user tools, services and systems) that enable scientists to efficiently analyze the data and conduct their research.	
<b>Q2.2 How will the success of your project be measured?</b>	So formally, the LIGO DataGrid is funded by a National Science Foundation grant, and we will be measured using some simple metrics about the types of services that we can stand up. We've committed in our project plan to accomplishing certain activities, making certain pieces of the infrastructure available and evolving them over time for use by the scientists. These include both user tools and services that support the user tools. More broadly though, the appropriate measure is really how much science is being accomplished by LIGO scientists using the LIGO DataGrid. This can be measured in terms of: - how quickly the scientists are able to run their data through the analysis pipelines - the utilization of the computing clusters - how much data (both raw and LIGO data) is transferred per unit time - how much user data (post-processed data) is generated and published per unit time And I guess at the very highest levels we would measure our success on the number of peer-reviewed scientific papers produced with LIGO DataGrid resources.	



Interview ID=9 2 July 2007	ANSWERS	ANNOTATIONS
<b>Q2.3 What are the professional measures of success for you?</b>	<p>My personal metrics for success are twofold.</p> <p>One of my most important specific projects is building, maintaining and evolving the infrastructure that replicates LIGO data from the interferometer sites where it's produced to all of the computing sites where it is consumed. And while we've been able to build infrastructure that is fairly robust and reliable, it has not been easy to maintain, monitor or manage.</p> <p>And so one of my personal metrics is going to be how much time is needed by administrators around the LIGO DataGrid to configure, maintain and manage the data replication infrastructure. To the extent that we can beat that down so that it really is a service they just stand up. If they don't need to baby sit it, and can really just let it do its thing with very little intervention, we will have succeeded.</p> <p>The goal I'm really looking for is to have a mean time between failures of three months. That's what I'm targeting from the data replication side.</p> <p>I'm also working on getting some of the data analysis pipelines or workflows to run in more sophisticated ways across multiple computing sites. Right now our users tend to pick a site and go run there. And while they use some Grid tools to do some of the data finding, they don't typically use Grid tools to leverage more resources than are available at a single site.</p> <p>So another of my metrics for the coming year is to try to enable production LIGO data analyses that run across multiple LIGO DataGrid sites in a continuous way. As a component to this – I don't know if we'll accomplish this by the end of the year – I'd like to try to get LIGO production data analysis running on sites that are external to the DataGrid. So federate into other Grids, such as Open Science Grid and TeraGrid.</p> <p>So my overall continuous metric that helps to judge how successful I'm being is the user adoption of the tools. The more users that actually use the tools day-to-day in a production way, the more I consider myself successful.</p>	

Interview ID=9 2 July 2007	ANSWERS	ANNOTATIONS
<p><b>Q3 What are you investigating?</b></p>	<p>I have two main foci, which can be grouped into two sets of projects.</p> <p>In terms of data replication: we built some infrastructure in the past for doing this on top of the Globus Toolkit, as well as other components. It's primarily based on pre-web services components of the Globus Toolkit. Moving forward, we want to build WSRF-compliant services on top of both Globus Toolkit 4.0 Java and C WS Core.</p> <p>We want to use Java where it makes sense. We can build Java services more quickly. They are easier to build because there's a lot of tooling. But they don't perform as well. We do have a few services that we think will need to be built on top of the C WS Core because they tend to involve a very high number of interactions and require faster exchange rates.</p> <p>So we're investigating and beginning to plan for that move. We'll be rolling it out in stages. Since our data replication tools have a number of services we can pick and choose, and we can move those services one at a time.</p> <p>Once we've accomplished that, then we can begin to really leverage some of the niceness that's in the Globus Toolkit. In particular, we want to be able to harvest a lot of the information that these customized services will be exposing as resource properties using MDS4. Not only to expose at each site the local state of their data replication, but also to aggregate that information. So we plan to put up really nice dashboards that our collaborators can look at and see the current state of replication throughout our DataGrid.</p> <p>Part of this work will be leveraging the MDS Index service. We also plan to use the WebMDS in a very large way, since we plan to generate a number of monitoring pages to monitor these services in a fair amount of detail. I've been getting beat up for years now that our tools are very difficult to monitor. And I'm tired of getting beat up for that. We're diving into monitoring in a big way, and I plan to do it on the back of MDS.</p> <p>I should also mention we're also going to use the Trigger Service in a fair amount of detail. Since we'll be doing all this monitoring, we want to be able to trigger on significant events. So when the service is not behaving the way we expect it to or when there are changes to the environment that the system needs to respond to, we can trigger on that and do the right thing.</p> <p>The second focus of my work is on the data analysis side, where I'm trying to get some of these production workflows to run across the LIGO DataGrid. We're investigating a couple of different technologies that will impinge on Globus, although I don't know that we'll be using Globus directly. We're building on top of higher-level services. We definitely want to use version 4.0.x of GRAM as the gateway.</p> <p>We're looking at using two different tools for helping to manage these workflows across multiple sites. The first is Condor-C. That will be one way to help us manage workflows across multiple sites. And that won't involve GRAM because I believe Condor-C will handle that between sites, keeping that in-house underneath the covers of Condor.</p> <p>At the same time, to go outside of our Condor pools, we'll be looking to use Pegasus, which is a project coming out of ISI. We've been working with that group to use Pegasus for some of the workflow planning that is executed and managed by Condor DAGman. It uses Condor-G to bridge out to non-Condor sites. And that will use GRAM4, which I'm excited about because of the scalability improvements and the monitoring hooks that we'll get with the newer version of GRAM.</p> <p>And I'd also like to set up some WSRF customized services that would address very specific LIGO workflows, so that the users are actually talking to this customized service. Then we can use the service to hide from the user some of the details about these other tools. So we can have a LIGO "face" that we present to the users without exposing them to the more generic tools like Pegasus, Condor-C, GRAM, etc. So we'll be building an infrastructure layer on top of the middleware so we can tie things up nicely for our users.</p>	

Interview ID=9 2 July 2007	ANSWERS	ANNOTATIONS
<p><b>Q4.1 What is your method for investigating &lt;phenomena&gt;?</b></p>	<p>I try to spend a fair amount of time actually deploying and using the tools. I can create quick easy deployments and make my own little testbeds. Because I personally find that I can't understand how a tool works in sufficient detail to actually design it into some infrastructure until I've really sat down and used it. I can read about it, read all the research papers, and go to talks and see slides. That can give me some ideas. But it's when I actually use the tool and try to understand what it does and how it works that I can find the sweet spot. As part of understanding them, I tend to need to stand them up myself.</p> <p>It's not very productive for me to try to use someone else's installation. People try to be helpful like that. For example the Pegasus team will say, "Well, you can just log in here and run the latest version of Pegasus to see what it looks like." That's a nice gesture, but at the same time I prefer just to get the code, compile it myself, deploy it myself so that I can really see what's involved. So there's a lot of my time that goes into those efforts.</p> <p>In the past I have been guilty of not doing enough design and not spending enough time trying to think clearly about what the usage scenarios are, and how the infrastructure should really work. We were so pressed to have anything working that we just threw a lot of stuff out there. We would try to iterate quickly but we'd end up usually tying ourselves in knots and going in directions that we couldn't support.</p> <p>So we're trying to be more deliberate about designing and thinking ahead, without going overboard, in terms of how the pieces fit together. I'd say that's going to be a larger component of what we do now. Going out and talking to the different middleware providers to understand what their roadmaps are, so we can try to get an insight into what things are going to look like one year, two years – even three years from now. Although I perfectly understand that everything changes so fast that it's hard to predict how a tool is going to evolve in three years. Three years is forever.</p> <p>But if I can at least get a sense of where RLS is going and RFT – things like that. What the priorities are. Then I can try to be able to plan a bit more than we used to for how we're going to leverage these tools.</p> <p>In the past we were allowed more flexibility because everything was so new and shiny. People understood that we were solving problems that no one had solved before. So we had a little extra slack and were given extra forgiveness for building things that weren't maybe as stable as one would like. But we've sort of outgrown the toddler stage now, and we're looked at as adolescents. We had better get our act together and start building production quality software that can be stood up for months at a time, scales well, has lots of documentation and is neatly wrapped and easy to install. So basically everyone wants everything. So we do have to step it up a bit.</p> <p><i>[Prompt asking for more information on what it means for "pieces to fit together"]</i></p> <p>How easy is it going to be to pull the pieces together and build a higher-level functionality or service on top that really meets the specific needs of the users I'm supposed to support? So I can look at a class of tools that do certain things – purporting to do something, for example "manage data transfers" or "manage workflows". I then try to decide if the tool offers bright shiny new functionality, but will at the same time be unstable and I won't be able to rely on it.</p> <p>Another thing I look at is the interfaces. I'm going to usually have to put some glue in place to pull these pieces together in reasonable ways, and how easy is it going to be for me to do that gluing? Do I have an API that's only supported in one language? If it's not my first choice for the project, I'll need to extend outside of our area of expertise. Or is it something that's technology or API agnostic and I can easily just write whatever I want?</p> <p>What kind of logging does it have? Is this a tool I'll be able to drill down easily when I think there's something going wrong? Can I turn up the logging levels so I can really get a picture of what's going on?</p>	

Interview ID=9 2 July 2007	ANSWERS	ANNOTATIONS
<p><b>Q4.1 What is your method for investigating &lt;phenomena&gt;?</b> [continued]</p>	<p>What type of documentation does it have?</p> <p>The type of access do I have to the developers is also an important component. With the Globus team I generally feel like when I ask questions they're quite responsive. And if they're not I feel I can always get in the car and drive down and stand in their office if I have to. So generally I'd say it's good. But that's not always the case for other tools.</p> <p>Licensing is much less important for us. We're basically the end users and we're not going to try to sell anything. So as long as it has a reasonable license we'll use it (as long as it's some flavor of open source license.) Although I should add that we'll only use at this point tools that are cost-free. We haven't paid for anything for a very long time, and I don't think we will because the last time we did it was a disaster.</p> <p>So those are some of the things I look at when I'm examining all the choices out there.</p> <p>A good example right now – I haven't talked about this so far from a project standpoint – but one of the projects I have on the backburner is upgrading our security infrastructure. We're thinking about supporting different use case scenarios, and we have to provide better security management services for our users.</p> <p>So there are a huge number of tools out there that we're looking at. There's Shibboleth and GridShib to hook in the Grid pieces. There's MyProxy and all the things it can do. There's the GAARDS suite of services from the caGrid project that's now an incubator. There are all the Kerberos-based tools, and this new OpenID [<a href="http://openid.net/">http://openid.net/</a>] thing.</p> <p>There's just an explosion of tools available. I'm just beginning to sit down and look at these and think about which tools we can build on, and which ones we can't. An important consideration is which tools are extensible and allow me to build on top of them. This is in contrast to tools that try to provide a complete solution that force me to rip and replace stuff.</p> <p>I try to stay away from the rip-and-replace tools because we just can't do that. Such tools offer solutions, but require me to give up other stuff that I'm already doing in order to use them. So the toolkit model fits me exceptionally well.</p> <p>I'm really looking for tools and infrastructure that allow me to build on top of them. So they have to be extensible, they have to have nice APIs and hooks that I can layer my own stuff on top.</p> <p>To provide a contrasting example from the security realm: there are a number of solutions based on proprietary tools. Some people are interested in those because they seem to offer to the users a better experience. I use the term "seem" because I'm not convinced that they actually offer a better experience for the user.</p> <p>But the problem with these integrated solutions is then on the backend there's no choice about how to hook them in to other systems and services. It means going to all of my application developers and saying,</p> <p style="padding-left: 40px;">"Well, we can't really rely on just using GSI or Globus as an option because there's only this one API now. You have to code against this API if you want your application to work in the security infrastructure."</p> <p>Most of our applications actually don't code against any APIs. They need to have the environment and security managed for them. So I can't go to a LIGO data analyst and say, "I need you to link with this library so that your tools will interact properly with the security." They expect the infrastructure to operate at a level either above or below that, depending on how you characterize it. They just want to run their job, and they want everything to be handled for them.</p> <p>So the Globus GSI and everything that's built around that ecosystem now works really well for us. We can hook into it in so many different ways. We can set up services that manage the delegation for the users. The only thing the users have to do is enter the system once using something like MyProxy. Everything else is handled for them. That works really well.</p>	

Interview ID=9 2 July 2007	ANSWERS	ANNOTATIONS
<b>Q4.1 What is your method for investigating &lt;phenomena&gt;?</b> [continued]	<p>I guess another example would be something like Condor. We do use Condor a lot and I'm a big Condor fan as a batch system, because it's reliable and it allows our users to get a lot of things done. But Condor really is a soup-to-nuts approach. They want you to give them your job and they'll handle everything for you. And I think that's great when it works. When it doesn't work it's frustrating because I can't hook in to the backend as easily as I can if I'm pulling components from a toolkit.</p> <p>So I really like the toolkit approach. I think it serves us well. I've always been fairly disappointed and shocked at how little other communities have used Globus as a toolkit. They tend to look to the toolkit for the end-user tools. The perfect example is globus-url-copy, which has been a great success that has evolved over time. But I think it really shouldn't be necessary.</p> <p>I would like to see other Grid projects code against the GridFTP or the RFT APIs themselves. They could be doing a better job of servicing their users if they didn't try to provide them a generic tool like globus-url-copy. It's really not a lot of work to customize the Globus tools to fit their own users' use cases. I see TeraGrid doing that, and I think that's great. But other Grid projects should be doing more of it, in my humble opinion.</p>	
<b>Q4.3 How do you keep track of interim results, if at all?</b>	<p>They're kept track of in my head. There should be a better system, but I guess it's both a good thing and a bad thing. It's a bad thing because if I get hit by a bus tomorrow, there would be a fair amount of knowledge about how LIGO could leverage these tools that would go with me.</p> <p>To be quite honest, on the positive side of it, it gives me a skill set that other people don't have. I'm not trying to hoard, but it does set me apart. I think it has enabled me to get in here and get paid to do what I do. Because I understand the science – my background is in gravitational wave physics – and I can talk to the physicists and I can understand what they're trying to accomplish. But at the same time I have a better knowledge of all the tools and what they have to offer. So then I'm in a good position to draw the lines between the two sides, and decide how best to apply the glue that brings them together.</p> <p>So I guess that would be my defense of why I don't do anything better. I should probably do some better things, in terms of tracking. And I guess if I was going to turn this to something the Globus Toolkit could do for me: To the extent that the toolkit developers could make the roadmaps even more transparent, I think that would be helpful.</p> <p>So what tends to happen is I tend to hear about some ideas or plans or great visions for how a tool might evolve. I usually hear about it to some extent through the grapevine, and it can be hard to track the progress of the concept. I know that a lot of the campaigns are documented in the bugs [<i>many short-term Globus work plans are documented as "CAMPAIGN" and "ROADMAP" entries in the Globus issue tracking system at <a href="http://bugzilla.globus.org">http://bugzilla.globus.org</a></i>]. And that is certainly better than other systems I can imagine. It's not particularly well suited for it though, because there can be a lot of cruft in there that you have to sort through. By "cruft" I mean things like comments from other developers, or users, that really aren't germane to what's going on. That's not always the case, but occasionally it is. So I wouldn't complain if there were another system that was just for tracking campaigns. But it's certainly better than nothing.</p> <p>The other thing that is a little bit difficult: sometimes I think there's a reliance on the email lists for archiving information. And that's great, because sometimes the details really only exist in an email list and you want to be able to find them. But it can be hard. There are so many email lists I'm trying to monitor, and I'm not even on all of them that I'd like to be, because I haven't had time to talk to majordomo to get on more lists.</p> <p>So in some sense it would be helpful if some of the campaign details again were in a more centralized place. And information that was exchanged through the email lists that's pertinent to the roadmap or the campaigns could end up in this other place. I realize that's extra work – they've already composed an email to tell the community what's going on, and then someone has to harvest that and put it into some other bucket. But it might be helpful.</p>	

Interview ID=9 2 July 2007	ANSWERS	ANNOTATIONS
<b>Q4.4 How do you test work-related hypotheses?</b>	<p>I prototype – lot’s of prototyping. So the good news is that we have an abundance of hardware. Hardware all over the place, because hardware is cheap. So it’s very easy for me to scrounge up boxes, create a prototype Grid, throw stuff down and do stuff. And now with virtualization technologies it’s even easier. I can get as much hardware resources as I need to do all kinds of testing. So that’s really not a problem.</p> <p>And that’s what we do – tend to prototype things. I might want to look at the details of RFT, for example, and see if it really has the queuing structures or behavior that I need. Or if I am worried about scaling, I can easily deploy an RFT and I can slam it, and see where it falls down for me under the types of transfers I expect it to manage. So that’s the easy part.</p> <p>The harder part is just having the people to do it, or the time to do it myself. So if I don’t have the time to do it then I need to ask someone to do it. And there are still not enough people in the world, as far as I’m concerned, that have real in-depth Globus knowledge. And certainly they’re hard to find and hire. So we train people up here, to the best that we can. But it’s still hard to say to someone:</p> <p>“I’m thinking about putting RFT into this service, but I need to understand where it’s going to break. I need you to stand up an RFT, and throw larger and larger requests at it until it breaks.” Or “I want you to throw larger and larger requests at it and tell me the load and memory footprint on the machine.”</p> <p>I could do that with my staff, but I usually end up getting it kick-started and spending a little more management time than is optimal. That is not a comment on my staff because they’re all good, hardworking, smart people. They just don’t have some of the expertise, especially with of the Web services stuff now in Globus Toolkit 4. It’s kind of foreign to them, so I have to spend some extra time getting them up to speed.</p>	
<b>Q5.1 In what ways do you interact with simulations in your work?</b>	<p>I don’t really interact in any significant way with simulations. But I wish I did. If I had time and I had someone to do it, I would love to have a Grid testbed that would truly simulate what we see in terms of networking. And – more than just networking – the response of certain of our resources, particularly the data replication.</p> <p>It would be really neat to have a Grid testbed where I could mock up the latencies that we see going across the Atlantic or even across the country here. Some of the network weather that we see. And also the behavior of some of the resources.</p> <p>Particularly at Caltech, where they’re using a special type of tape backend. Our GridFTP server there overlays on this tape backend. To the GridFTP server it looks like a POSIX file system, and GridFTP with the file DSI right out of the box can talk to it. It’s not like HPSS or some of the older style tape systems. At the same time it is a tape system, and so sometimes requests can take a long time – sometimes hours – to be serviced. That really changes things. And boy if I had someone who could set up a simulator so I could experiment with different scheduling algorithms and ways of handling those delays, it would be terribly exciting. But I just don’t have anyone that can do that for me.</p>	
<b>Q6.1 Describe how you interact with data in your work</b>	<p>The rough numbers we always use are a terabyte per day that we have to replicate to nominally seven sites around the world. We still use those numbers, but it’s actually much more complicated. It used to be just the raw data products, or the interferometer data, but now as we’ve given users the ability to replicate and manipulate the data there are all kinds of boutique, customized datasets being generated.</p> <p>So instead of having data coming out of three interferometers, in some sense we have data coming out of twenty-five instruments. It’s just that some of those instruments are computer codes doing different types of things.</p> <p>So the number of data products we are responsible for is growing quickly. Therefore the number of files is growing quickly. And for us the big issue is not so much the raw data size, because in a sense it’s still a terabyte a day. But now instead of being divided over a couple thousand files a day, now it’s over tens of thousands of files per day. And they’re all different sizes.</p> <p>We have to track so much information about the data now. And so, as has been the case for the past couple of years, we’re getting killed by the metadata. I really hope the Globus team can help us address this problem.</p>	

Interview ID=9 2 July 2007	ANSWERS	ANNOTATIONS
<b>Q6.2 How do you share work-related data with others?</b>	<p>There's three parts to this answer.</p> <p>First of all there's metadata about the data. And we have a customized metadata service, and that's the service that I'm going to move to WSRF in the next two months. So it's a completely customized metadata service. Right now it uses MySQL on the backend, but I hope to change that. It has the ability for clients to obtain metadata information from the service. Generally each site has a copy of the service and it obtains metadata from other sites.</p> <p>We have this notion of an authority site. So if a particular set of files is generated at some site, it has to be published in the metadata catalogues. Wherever that happens, we consider them to be the authority site. And other sites, if they need that metadata, they can get that from that authority site. So that's one way we share the metadata about the data.</p> <p>Then there's the data itself: how do users and applications find out where the data is within our DataGrid. So we rely very, very, very heavily on RLS. I think it's true that we run the largest RLS network in the world. When RLS goes down, you better believe we know it. And we have to jump into action. Fortunately it very rarely goes down now. We're quite pleased.</p> <p>And then for actually pushing data around the Grid, GridFTP is the technology we use. As far as I'm concerned we'll continue to use GridFTP, especially now with the small file support [<i>LOSF feature</i>]. Unfortunately the feature wasn't included in the 4.0.5 release, though I understand why the decision was made not to include it. As soon as we can leverage the <i>LOSF</i> functionality, we are going to do that in a big way. Because that really impacts our throughput. So we're looking forward to doing that in the late fall.</p> <p>So you can really say that our data problems are solved with three components: GridFTP, RLS and our customized metadata service. And that's why I'd like to get out of the metadata business and swap in a Globus component. Then I can do other things with my life.</p>	
<b>Q7.1 What resources do you use in your work today?</b>	<p>LIGO DataGrid has a number of fairly large Linux clusters. Caltech has roughly one thousand dual core nodes, which means roughly two thousand CPUs. Then each of the sites has around three to four hundred dual core nodes, totaling approximately two thousand cores. UWM has about six hundred dual core nodes. Penn State has about three hundred single core nodes.</p> <p>In Europe, there's a Linux cluster at Cardiff that has about one hundred fifty nodes, there's a cluster at University of Birmingham that's a similar size. There's a four-hundred node cluster at Albert Einstein University in Germany. There's a very large cluster planned at University of Hannover that will measure in the many thousands of cores. So we have a fair number of compute resources, which are almost all Condor-based.</p> <p>Our data storage resources are all over the map. Caltech's tape archive is something like fifteen petabytes; they also have fifteen terabytes of disk that acts as a cache for the tape storage. That's where all the LIGO data ends up – archived to tape. All the other sites have different types of spinning storage. So at UWM we have twenty-five or thirty boxes, each one with ten or twenty terabytes, for storing data. And all the other sites have storage that scales down from there. I think the small sites in Britain only have three or four terabytes of storage; most have thirty to one hundred terabytes.</p>	
<b>Q7.4 How do you locate available resources for use in your work?</b>	<p>On the data side it's RLS and GridFTP and our metadata service.</p> <p>For running workflows we don't, and that's a big problem and is what I'm trying to jumpstart. We don't have any infrastructure deployed and configured that enables us to understand in a useful way what the loads are at the sites in terms of computing. This is needed to make more intelligent decisions, either by humans or by infrastructure. We need to enable this functionality in the next year.</p> <p>I would like to see a reasonable MDS architecture that can harvest and aggregate information that's necessary so the pipelines (the scientists use the term "pipeline" in place of "workflow") can do more sophisticated cross-site scheduling.</p>	

Interview ID=9 2 July 2007	ANSWERS	ANNOTATIONS
<b>Q7.5 What types of information do you need to know about a resource in order to determine if it is suitable for your work?</b>	<p>Part of that is a research question that I am trying to ascertain. The biggest one will be data location. So understanding what data is at a site and what data is not at a site – that’s by far the most important question. Because the amount of data involved in analysis is so large that you really want to optimize on scheduling around the data.</p> <p>When we talk about running on sites that are outside of LIGO, we’ll have to move data. That’s definitely doable – if you do it right you can still win. You can move the data, run jobs, clean up after yourself, while still achieving more aggregate throughput than you would if you did not have access to the external resources.</p> <p>But certainly within the LIGO DataGrid it’s all about the data. We’re replicating the data as much as possible to all the sites. Since the replication happens outside of the workflows, then you really want the workflows to schedule around data location.</p> <p>After data location then you need information related to the details of the site:</p> <ul style="list-style-type: none"> <li>- where the storage is located</li> <li>- type of storage, such as whether or not it is an NFS-shared disk</li> <li>- local directories that should be used for certain workflows</li> </ul> <p>Then there are application-level details at each site. Our applications tend to rely on certain types of environment variables being set. It would be nice to understand whether or not they’re all set at all the sites. If we had a way to monitor that, trigger on it and be available for use by the infrastructure’s decision-making processes – that would be helpful.</p> <p>Those are more second-order effects. Basically everything else can be accounted for after the data location; that’s the big one.</p>	
<b>Learning about the user’s problems</b>		
<b>Q9.1 What challenges do you face today in accomplishing your work-related goals?</b>	<p>The major problems are the proliferation of metadata, first of all about our data. So the number of data products is really exploding. And there’s so much to track about all the data. And that’s even outside the realm of the science. I’m not even a scientist analyzing the data, and I still have problems tracking it. So that’s a major pain point for us.</p> <p>And there’s not only metadata about the data, there’s the metadata about the workflows, and by “workflows” I mean “analysis pipelines”. That’s a huge issue. We now have about one hundred and twenty users who do analysis. So these are the people who actually type commands and run codes. They generate a huge amount of data themselves. Post-processing the data, keeping track of it all, and being able to understand where a result came from (the issues of provenance) are all tremendous issues for us.</p> <p>And to be frank, the attempts from Globus-related teams (I don’t think these are Globus Toolkit proper) to provide tools and infrastructure to help with metadata and provenance have not scaled. And especially in terms of provenance, they’ve required too many application level changes. The approach was,</p> <p style="padding-left: 40px;">“Just do everything this way then you’ll get the provenance information.”</p> <p>But there’s no way to “just do it this way”. That’s not the way my users can be approached. They are going to do their science. The science is going to lead, and all the other stuff has to be tacked on. So that’s a major problem.</p> <p>So as to metadata: we have so many things to keep track of we’re losing the ability to keep track of it. We’re building tools that need the information to function; they need the information to leverage the infrastructure. Tools need the information to help the users get the work done. But the systems that provide the information are breaking underneath the load. Then all these tools and infrastructure become unworkable and work stops, and it’s just bad.</p>	
<b>Q10.1 Can you think of any work-related tasks that decrease your productivity?</b>	<p>I, like my a lot of my colleagues, spend entirely too much time in meetings, on telecons and answering email. And not nearly enough time being able to just sit down and solve the problems I need to solve. When you’re in these collaborations there’s just so many people you have to coordinate with and it just takes so much time; it can literally eat up a third of my time. It’s not all bad, but there are days I wish I could just lock myself in my office and do nothing but write code, because I’d get a lot done.</p>	
<b>Learning about the Globus user experience</b>		



Interview ID=9 2 July 2007	ANSWERS	ANNOTATIONS
<b>Q11.1 Which Globus data components do you directly interact with in your work today?</b>	<p>Definitely we use RLS and GridFTP.</p> <p>We do not use OGSA-DAI. OGSA-DAI is far too much overhead for getting into the details of the metadata and the data we need. We have so much metadata we can't take the performance hit of going through the layers of OGSA-DAI. We've evaluated it a couple of times and it has never been able to meet our performance requirements. Performance in terms of the ability to query for information and get results back. The pure query rate doesn't scale the way we need it to.</p> <p>It's a hard problem - it's nothing to take away from those guys. It's a really hard problem, and the advantage we have is we have a very specific schema, so we can make our custom tool work faster.</p> <p>We have not yet used RFT. Initially we didn't use RFT because in the first release it didn't support some of the features we needed, such as connection reuse. We ended up building our own customized client and tuning it a fair amount. And we continue to use that today.</p> <p>However our custom client is built on top of Python Globus, and as much as I like Python Globus as a tool, I can't build production software on it anymore. In our experience we find too many bugs because the user base is smaller. It can also take longer to develop production software using PyGlobus because the documentation is sparse and we don't have good code examples. If I want to build a customized metadata service I can go look at GRAM and RFT to understand how certain WSRF things are being handled in Java. And I can't do that with Python Globus; I don't have a production-quality piece of software that I can look to as a learning example. So we're going to be moving away from using Python Globus for production, but we'll certainly use it for prototyping.</p> <p>So I think we will be able to start using RFT now, and it's on my roadmap to try to fit it in. Actually it's on my roadmap to come down and bug the RFT developers because I need some things changed. So they'll get a visit from me by the end of the summer.</p>	
<b>Q11.4 How many people currently use your &lt;component&gt; server</b>	It depends on what is meant by "use". I'm the architect, and I design the system. I have two developers who write code around it for me. I have between seven and ten administrators who administer RLS across our data replication network. And then we have one hundred fifty users that interact with the system to figure out where data is.	
<b>Q12.1 Which Globus security components do you directly interact with in your work today?</b>	Today GSI-OpenSSH is widely used. GSI pre-Web service core is heavily used. I'll add SimpleCA - not that we use it in production - but I use it a lot when I'm prototyping. We do not use CAS at this point. In the future, which means very soon I hope, we will be leveraging the Delegation service and all the Web service versions of GSI. So hopefully that will happen in the next couple of months.	
<b>Q12.4 How many people currently use your &lt;component&gt; server</b>	All of our users use GSI-OpenSSH. That's how they get into the clusters. So there are around one hundred fifty users.	
<b>Q13.1 Which Globus execution components do you directly interact with in your work today?</b>	Today it would be none, which is disappointing for me. In the future I very much hope it to be GRAM4. I do not want to use GRAM2 because it does not scale and it does not meet our performance requirements. I do not see a way on our roadmap to use GridWay.	
<b>Q14.1 Which Globus information components do you directly interact with in your work today?</b>	Today we do not use MDS4, but I will be using the MDS Index, WebMDS and the Trigger service in a big way in the future.	
<b>Q15.1 Which Globus common runtime components do you directly interact with in your work today?</b>	Today I use the Java WS Core and the C WS Core because I'm writing services for our future use, so not yet in production.	

Interview ID=9 2 July 2007	ANSWERS	ANNOTATIONS
<b>Q16 Why do you use &lt;component&gt; instead of an alternative technology?</b>	<p>RLS: I'm not aware of any alternative technology that offers the same sensible API and scales and performs as well as RLS.</p> <p>GridFTP: At the time we started using GridFTP it was the only game in town. There are now other tools that will transfer data at least as fast (and some say faster). But I don't prefer them because they don't fit into the overall framework in the same way as easily as GridFTP does.</p> <p>I'm actually getting some pressure from Caltech to look at a Java-based data transfer tool coming out of Harvey Newman's group and being used by some of the high energy physicists. You look at their raw numbers and, sure enough, it will transfer data like mad. Very fast – and that's great. But it doesn't have the same security hooks in it. I can't use GSI and I don't really want to shoehorn the security in as an afterthought. I like the fact that GridFTP and all the Globus components think about security upfront, and it's part of one coherent system.</p> <p>So I'm very hopeful that the GridFTP LSOF feature will help with our throughput and I won't have to think about using one of these other superfast data transfer tools that move the bytes fast but are not part of a larger toolkit approach.</p> <p>GSI: I am not aware of an alternative technology that supports delegation of authentication and authorization decisions in the same way as GSI.</p> <p>Java and C WS Cores: I am attracted to the WSRF framework. I think that, especially for someone like myself who's not a computer science person, it allows me to quickly and easily leverage things like resource properties and the publication of resource properties. The lifetime management, the subscription and notification – all the nice things. It allows me to leverage them quickly and much more easily than I could do if I had to do all that stuff by myself. After all I am a physicist and I'm dangerous when I'm writing code. The more code other people write for me, the better.</p>	
<b>Q17 What are the major challenges you face using &lt;component&gt; today?</b>	<p>RLS: It turns out that a relational database backend doesn't make the most sense for LIGO's use of RLS. And the reason is because the things we track in RLS are data files, both coming off the interferometers and user-generated. Because of the kind of experiment LIGO is, all of our files are time-oriented. Every single file has a beginning timestamp and an ending timestamp. And timestamps index very, very, very well. And they hash very, very, very well. So whenever anyone asks a question, "I need a piece of data," what they're really asking is, "I need a piece of data beginning at X time and ending at Y time." And once you know the time range, then all the other metadata is secondary. If we've got five years of data, that's a lot of data files and time. But if you're selecting around an hour, you've already picked out a small subset of the data. Then drilling down with the rest of the metadata attributes is easy. And as it turns out, relational databases are not the best way to model our data. We don't really use the relational aspects of it. What we really want are fast index hashes. So what I've asked the RLS developers to think about is abstracting RLS so that it can support other plug-in backends, just like the GridFTP supports other data storage interfaces (DSIs).</p> <p>I would like RLS to support different DSIs. It should have the relational database as the default, but also provide the option of using other methods of representing user data and its mappings between logical and physical filenames. Then what LIGO would do would be to write our own backend based on a hash table approach. Because I really like the RLS API and I like the model. I'm very happy with it as a service at that layer. What I want to get away from is the relational database backend because I don't think it's going to scale for us going forward five years from now.</p> <p>GridFTP: Our major challenge has been the LSOF, but it looks to be addressed and we're very excited about leveraging that functionality. Other than that – jeez it's completely reliable and moves tons of data. What else could I want?</p> <p>GSI: The management of credentials by users directly is too difficult. Our user base is not sophisticated enough to manage their credentials directly, so we are moving away from that approach. We'll be beginning to rely on MyProxy and similar types of credential repositories so users don't have to manage their PKI credentials themselves.</p> <p>Java WS Core: The major challenges are personal. I'm not a Java guy, and so am just coming up-to-speed on Java. Again, I'm a physicist, so Java was not something I learned growing up.</p> <p>C WS Core: The examples and documentation. I know they're working on that and it will get better. But right now there's not as much documentation and not enough examples.</p>	
<b>Wrapping-up</b>		

Interview ID=9 2 July 2007	ANSWERS	ANNOTATIONS
<p><b>Is there anything you'd like to say to the people who build software for use by people like you?</b></p>	<p>I certainly appreciate all their efforts. The Globus Toolkit has really succeeded in what I think is one of its primary missions: enabling more science. Without a doubt, Globus has made more science possible. Period.</p> <p>I think that's a huge win. Or at least it's an important metric for me, and I assume it's an important metric for the Globus team. So that's fantastic. Three cheers and keep up the good work!</p> <p>On a different point: the Globus team has gotten better at this, but there are still times where the team appears to be self-focused or focused inward. This doesn't apply across the whole team. But some folks seem to be focused on infrastructure for infrastructure's sake, as opposed to infrastructure for other people to build on. There are still some pockets of that occasionally. But that's certainly not the rule. As I think about it, the teams I've interacted closely with – the RLS and GridFTP teams – I can say that's the opposite. They tend to be very supportive in terms of reaching out, asking for use case scenarios and requirements, and being responsive to input. And I think that's great.</p>	

## D.10 Solving problems is easy once you have all the data

Interview ID=10 2 July 2007	ANSWERS	ANNOTATIONS
<i>Pre-interview question: Do you interact directly with Globus software in your work today?</i>	Not yet, but I will	
<b>Establishing context</b>		
<b>Q1.1 Please provide a one-minute overview of your project</b>	The ALCF as a whole is a new Leadership-class computing facility. “Leadership-class” means a small number of users doing very high-end, cutting edge science. I am specifically involved in the storage and IO aspects of that. The Grid aspects of it, as planned currently, involve using GridFTP to move data in and out of the facility.	
<b>Q1.2 What is the project’s name?</b>	ALCF: Argonne Leadership-Class Facility	
<b>Q1.3 Which agency funds the project?</b>	Department of Energy	
<b>Q1.4 What field does your project belong to?</b>	As a facility we have users from a broad range of disciplines. Some of the primary initial ones are Material Science, Biology, Nuclear Engineering and Astrophysics.	
<b>Q1.5 What is your job type?</b>	My title is Storage Engineer, but I am basically the lead on all storage and IO issues.	
<b>Q1.6 How long have you been a &lt;job type&gt;?</b>	Not quite a year yet	
<b>Learning about discipline-specific goals and approach</b>		
<b>Q2.1 What are the main goals of your project?</b>	To provide stable computing facilities for people to do cutting-edge science. We are not about doing science; we are a resource/infrastructure provider. So our primary goal is to keep the equipment up-and-running, have as stable an environment (i.e., not changing, as few crashes as possible.) So in the spectrum between <ul style="list-style-type: none"> <li>- complete research software which compiles once and never runs again, and</li> <li>- bullet-proof production software that’s commercially available,</li> </ul> we’re closer to the production end of the spectrum. As a facility we have a broad range of functions. We have User Services (help desk, etc.), we have Performance Engineers to help people tune their code, we have sysadmins, facilities people, and so on and so forth. So again, our primary goal is to provide services to scientists so that they can do their science, run their codes, handle the data, store the data, and so on and so forth.	
<b>Q2.2 How will the success of your project be measured?</b>	A number of the projects that we will have running on our system run on other systems currently. But they are limited in the scale they can run: the size of the mesh, the number of molecules they can simulate, and so on and so forth. So our success will be measured in how much additional science they can address: bigger problems, more problems. It will also be measured in the amount of papers that they publish. That sort of thing.	
<b>Q2.3 What are the professional measures of success for you?</b>	Basically, “Do the storage and IO systems meet the needs of the user?” There are two pieces to this question. One is the objective aspect: does it meet specs? We have specs on bandwidth, storage sizes, uptimes, etc. And obviously I have to ensure that the systems meet all that – my team and I. The other half is the more subjective side of things, which is, “Are the users happy with it?” So, do they think we have sufficient bandwidth, sufficient storage, is the balance right, etc.	
<b>Q3 What are you investigating?</b>	I am investigating data storage and IO issues. This is a very large machine. We literally have over a thousand ports of 10 Gigabit Ethernet. So issues include: <ul style="list-style-type: none"> <li>- parallel IO to the storage</li> <li>- how should the storage be handled in terms of splits between online, near-line and archival storage</li> </ul> One of the big issues, assuming we can’t be a universe unto ourselves: <ul style="list-style-type: none"> <li>- how are we going to move data in and out of this place at a reasonable rate?</li> </ul>	

Interview ID=10 2 July 2007	ANSWERS	ANNOTATIONS
<b>Q4.1 What is your method for investigating &lt;phenomena&gt;?</b>	<p>We have some things that we do on our own. We have test suites that are under development. Some of them exist in a simple form, but we are improving those. These are test suites that basically help us ensure that as we do things we don't take a step backwards.</p> <p>For instance if we decide to make a step up on a Linux kernel, sometimes the Linux kernel TCP performance is not as good as the previous kernel. And these test suites help us identify that we took a step backward, and whether we either need to do something about that or accept the fact that other benefits outweigh this loss. So performing test suites and standard experiments is one of our methods, but it is actually a smaller part of our work.</p> <p>Much of our time is spent focusing on how our users work. Sometimes they come to us and say, "This isn't good enough." Then working with them to see what we can do to improve it. A lot of their problems are code-specific, because optimization depends on access patterns:</p> <ul style="list-style-type: none"> <li>- large contiguous reads</li> <li>- small reads that are regularly strided</li> <li>- completely random IO</li> </ul> <p>We optimize those differently.</p> <p>Some examples of things we run inside our test suite:</p> <ul style="list-style-type: none"> <li>- Iperfs for network</li> <li>- MPP test suite from MPICH, which is a parallel IO test suite</li> <li>- IO Bench, for standard disk benchmarking</li> <li>- the John May MPI-IO test suite, which is another parallel IO test suite</li> <li>- we run some standard filesystem correctness test suites that are out there in the open source world (such as FX-Linux)</li> </ul>	
<b>Q4.3 How do you keep track of interim results, if at all?</b>	<p>We aren't running as a facility yet. Our hardware doesn't get installed until October. So we are borrowing a lot from what's currently running at the Mathematics and Computer Science Division at ANL and then expanding beyond that.</p> <p>Right now test results are stored in text files. We desire to get to the point where they are stored in databases so we can run ad hoc queries and trends automatically. Right now we do them manually, unfortunately. For instance, we look at a given IO disk benchmark to see if it is trending up, or trending down or staying stable. So right now it's in text format and we do it by hand. Eventually we'd like to get to the point where it has alarms. Where the tests are statistically analyzed and a message or alarm is triggered by statistically significant variations. Those are easy.</p> <p>The hard problems to identify are trends. For instance: "Individually this one isn't out of whack [<i>colloquialism for "not working"</i>], but if you look at the last five they've all been trending down, which is bad."</p> <p>So that is what we do:</p> <ul style="list-style-type: none"> <li>- we run tests</li> <li>- we store the results somehow</li> <li>- we do statistical analysis on trends to answer such questions as, "Are we running faster? Is more storage being taken up? etc."</li> </ul> <p>As far as what we do with the results: they're largely internal because they're operational. We do try to work with the scientists using the system to contribute to their papers. Occasionally we'll write a paper for submission to a sysadmin-type conference, such as LISA.</p> <p>We run the tests any time we make a change. If we go down for maintenance, or if we upgrade the kernel, we always run the test suite after that. It's kind of our equivalent to release candidate testing. That's our way of making sure we haven't broken things. In our case "broke" not only means not running, but also seriously degraded performance. There's also the flipside: we made some change intending to improve performance, and testing tells us whether we succeeded or not.</p> <p>We don't analyze trends on a regular schedule because they tend to be very intense. We will periodically kick them off if we think there's a problem, or if it's just been awhile. There is some discussion just getting started amongst the large sites (ANL, Berkeley, San Diego, etc.) about standardizing tests and identifying best practices.</p> <p>If we run the entire test suite end-to-end it can take days. So what we actually tend to do is run pieces of it periodically, which can take anywhere from a matter of minutes to a few hours.</p> <p>As far as tracking interim results of user-based tests, right now we use the Trac [<a href="http://trac.edgewall.org/">http://trac.edgewall.org/</a>] system. But ALCF has an entire User Services organization that plans to move to a more "touchy-feely" system before we're up-and-operational.</p>	

Interview ID=10 2 July 2007	ANSWERS	ANNOTATIONS
<b>Q4.5 How do you document your results?</b>	We plan to put results in a database and produce graphs and reports that would be available at the touch of a button to provide administrators with a view of the current health of the system.	
<b>Q5.1 In what ways do you interact with simulations in your work?</b>	<p>The majority of the jobs that are run on our machine are indeed simulations. I do not interact with the simulations directly. I interact with the scientists who are running the simulations.</p> <p>Because we're still building the system, the interactions I've had so far with users have had to do with gathering requirements:</p> <ul style="list-style-type: none"> <li>- What do you plan on having in terms of IO patterns? Large contiguous reads and writes? Strided or random?</li> <li>- How long do you want checkpoints to take?</li> <li>- What are your disk space requirements? How big is your total dataset? How is it distributed? Are there many large files? How big? How many? Are there many small files? Are they random? Is there a distribution?</li> <li>- Do you read and write files? Or only read? Only write?</li> <li>- How long do you expect to keep them on disk?</li> <li>- Once you're done with them how long do you want to keep them around?</li> <li>- If they're scheduled to move to tape, how long do you want to keep them around?</li> <li>- How critical is it if we lose your data in a fire? What kind of disaster recovery do we need to have in place? Can you just rerun the simulation or is this irreplaceable data?</li> </ul> <p>Essentially: how big? How fast? What's the distribution? What's the retention policy? What are the disaster recovery requirements?</p> <p>These interactions were initiated because I had to design and plan for the storage and IO system and the goal was to get the best possible estimates I could of what user needs were.</p>	
<b>Q6.1 Describe how you interact with data in your work</b>	<p>I don't do data analysis. I'm not worried about the contents of the data. That's what the scientists do. My interactions with data are as a service provider. So I'm worried about moving it, accessing it.</p> <p>To a lesser extent I'm worried about metadata and finding user data. The reason I'm less concerned with that is to date it's not been solvable in a general way. Metadata very quickly becomes very application specific. And most scientists have perhaps not as a good a system as they would like, but they have a system of some kind that they already use for tracking metadata.</p> <p>So we don't provide a standard system service that could be called a metadata service. Whereas we do provide services for IO, data movement, archiving, etc.</p>	

Interview ID=10 2 July 2007	ANSWERS	ANNOTATIONS
<b>Q6.3 By what mechanisms is access to your work-related data controlled?</b>	<p>There's the operational data that we take, which is the things like the results of the test suites, and uptimes, etc. We secure this type of data within our standard security system. Most of it is accessible over our admin network, and that is one-time password only. It is the only accessible method, and only includes staff here at the ALCF. There are firewalls in place, intrusion detection systems in place, etc.</p> <p>Occasionally we write papers on our results. Also there are various workshops going on sponsored by DOE trying to develop best practices for large computing installations, and we participate in those.</p> <p>Then there's the actual data of the scientists themselves (the output of the simulations, etc.) We don't make our users use one-time passwords yet – there's some discussion that we might. We require them to use standard ssh keys. The normal DOE rules: no sharing of accounts, no sharing of passwords, keys only, user permissions, etc.</p> <p>One thing that we don't have yet but there's discussion that we might have, is UCNE: Unclassified Nuclear Engineering data. Even though it is unclassified they worry it about it; it has higher security requirements. We're working on a plan where it would have its own internal VLAN – its own partition in the storage system. That would probably be only one-time password protected. One of the interesting problems we're going to have:</p> <p>Though we will largely not be doing Grid work – operations will mostly be confined to our own facility – we will be using GridFTP to move things in and out. We are going to need to worry about X.509 credentials and what we're going to do about that.</p> <p>A lot of our users probably won't have X.509 certificates. And there's no guarantee that the other end is going to have them. It depends. If there is an existing GridFTP installation it most likely will have them, and we'll just have to worry about the users on our end.</p> <p>If it's a new project that hasn't done any work on the Grid in the past, they may be setting up GridFTP purely so they can move things in and out of ALCF.</p> <p>We're not sure they're going to want to deal with all the overhead of managing a GSI certificate authority. They could use DOE's [<i>DOE hosts its own certificate authority</i>] but still there are problems getting it installed.</p> <p>So we have to address that. I know the GridFTP team has done some work on SSH-based authentication, so we're going to take a look at that. We'll also have to look at what it will take to handle GSI certificates. So that's how we protect access to the user data.</p> <p>How user data are shared is up to the scientists. We don't control that, of course. They do it through all their normal mechanisms – through their collaborations, through papers being published, some of it goes into publicly-available databases (such as the gene sequences), etc.</p>	
<b>Q7.1 What resources do you use in your work today?</b>	<p>There's today vs. what's planned (which is what I really work on.)</p> <p>Today we have Jazz [<i>350-node computing cluster</i>] and the Blue Gene/L [<i>ANL's first teraflop-scale computer</i>].</p> <p>As to what I'm working on:</p> <p>I can tell you that we are bringing up a 500-teraflop compute facility with 5 petabytes of disk and 100 petabytes of tape. [<i>An IBM Blue Gene/P series machine</i>]</p> <p>Each node in the machine is a quad core with 2 Gigabytes of memory. A rack has 1024 nodes. So that's 4096 cores and 2 Terabytes of RAM per rack. We will have a one-rack system for testing and development, which will be named Surveyor. We will have an eight-rack, 100 Teraflop system, which will be called Endeavor. We also have a thirty-two-rack, 500 Teraflop system named Intrepid, providing 128,000 cores and 64 Terabytes of RAM.</p> <p>For networking: everything is 10 Gigabit-connected; we have well over 1,000 ports. Internally, we have more bandwidth than you can shake a stick at. We have a switching complex that it is capable of switching at line speed 2048 ports of 10-Gigabit. So that means inside the facility we can move 2 terabytes of data per second.</p> <p>We will have 768 10-Gigabit Ethernet ports going in and out of the 500-teraflop machine. Within the facility the IO system will have eighty-two 10-Gigabit connections going in and out of it. As far as connections to the outside world, we are currently bringing up one 10-Gigabit link, and are scheduled to bring up two more.</p> <p>There are 21 SANs, each of which has 4 file servers in front of it using PVFS and a GridFTP server installed on it.</p> <p>We also have graphics machines for doing post-processing visualizations.</p>	

Interview ID=10 2 July 2007	ANSWERS	ANNOTATIONS
<b>Q7.2 How do you share work-related resources with others?</b>	We participate in the DOE INCITE program, and that's the way people get time on the machine. So scientists write proposals and they are peer reviewed. There's no cost. Commercial companies can use this program as well.	
<b>Q7.4 How do you locate available resources for use in your work?</b>	We use the Cobalt scheduler.	
<b>Q8.1 What software do you currently use in support of your work?</b>	<p>We are a Linux-based shop – SLES 10 specifically. Cobalt for the scheduler. There's Blue Gene-specific software required to operate the machine. Our filesystem is PVFS. We're going to use HPSS for access to the tape. GridFTP for transfer of data in and out of the system.</p> <p>The graphics software is controlled by the experiments. They use what they use.</p> <p>Our testing framework is homegrown with scripts. For our acceptance tests we are testing using Karajan, the CoG kit workflow tool. And if that works out well for the acceptance tests we will use that as a harness to run all of our test suites. "Acceptance tests" are part of our standard test suite; they are used to determine whether or not to pay vendors for new equipment.</p>	
<b>Q8.2 What scripting languages have you used in the past year?</b>	bash and python. A number of the tools we use are written in python (Cobalt, bcfg2, etc.) I personally write shell scripts.	
<b>Q8.3 What programming languages have you used in the past year?</b>	C, Java	
<b>Q8.4 What workflow tools do you use in your work?</b>	Karajan	
<b>Q8.5 What parallel computing tools do you use in your work?</b>	I don't use them. MPICH and MPI-IO are the primary tools used in our facility, and are included in the standard software suite on the machine.	
<b>Q8.6 If the need for new software-based functionality arises in your work how do you acquire it?</b>	Pretty standard – what you would expect anyone to do. We go out and look to see if it exists someplace. If we find something that meets our needs we incorporate it. If that doesn't work we then look to see if there's something close that we can modify. And if that doesn't work, we write it.	
<b>Q8.7 How do you share software with others?</b>	<p>Everything we do is open source. All of our operational tools are open source, so they're freely available. The tools themselves are available (Karajan, bcfg2, etc.)</p> <p>But our explicit workflows are facility-specific enough that it doesn't really make sense to share. So we would share the fact that we're using Karajan and our experiences that we have with it (good, bad, indifferent, yes we'd suggest you look at it, no it didn't work for X reason, etc.) But we don't give out our actual workflows because they don't make sense any place except on our machine.</p> <p>Now, that of course is different than the scientific codes. Those are under the control of the different projects; we don't have any control over that.</p>	
<b>Learning about the user's problems</b>		



Interview ID=10 2 July 2007	ANSWERS	ANNOTATIONS
<b>Q9.1 What challenges do you face today in accomplishing your work-related goals?</b>	<p>Since we're not operational yet I will talk about our experience with the Blue Gene/L [<i>an older machine</i>].</p> <p>Probably the biggest problem that we are having right now is with PVFS. Performance-wise, bandwidth-wise it does a very nice job for us. It's not as stable as we would like it to be. And so one of the things that is a primary project for us is to work with the PVFS team to improve stability and robustness. We're going to work on adding heartbeat monitors and failover mechanisms, etc.</p> <p>Stability for me, in the context of PVFS, means running without failures. Servers don't hang. User jobs run to completion. So right now PVFS hangs and jobs have to stop because they can't write data. We've got to get to the point that something figures that out, a backup comes into play, and the job can continue. That's probably the biggest real-world problem we have.</p> <p>Another issue is not a big problem yet – but I know it is going to be. When it comes to GridFTP and moving things in and out? We only control one end of the transfer. We can make sure the machines on our end are beefy enough and are configured correctly and tuned right. But if the guy is trying to transfer the other end off his laptop, we'll only go as fast as his laptop.</p> <p>Data transfer is an interesting problem in that respect. It's a two-ended problem. If you're trying to schedule the transfer, it requires co-scheduling – you must schedule resources at both ends. You don't have control over your own destiny: you can control your end and you can coach the other end. But if they don't have the hardware there's nothing you can do. And that actually gets quite frustrating. While I know that rationally they understand it, all the user knows is he's not getting what he wants.</p>	

Interview ID=10 2 July 2007	ANSWERS	ANNOTATIONS
<p><b>Q9.2 What types of information do you need in order to address the challenges you face today?</b></p>	<p>To some extent there's nothing we can do about this problem. We can't buy hardware for them. We can't buy enough hardware such that we park our hardware everywhere. And, like I say, if the fastest thing he's got is a laptop, that's all you can do.</p> <p>But let's say hypothetically they have the hardware and the other end is just not tuned right. They're certainly not going to just open their doors and let me go in and start tweaking their system, as rightly they should not.</p> <p>So I guess we could define best practices. For instance for GridFTP: identifying best practices in terms of tuning operating systems, spec-ing machines for hosting GridFTP, etc. A lot of that information is there, but it's scattered and disorganized.</p> <p><i>[prompt asking how the causes of transfer problems are determined]</i></p> <p>Good old troubleshooting. It's hard to describe because the specifics depend on the problem.</p> <p>Sometimes the type of problem is obvious. If they're trying to move their data in then we know at least it is a GridFTP problem. And then I have a list of pretty standard things I can start doing:</p> <ul style="list-style-type: none"> <li>- run Iperfs between sites to see if it is the network</li> <li>- run all the test suites on our end to make sure we haven't hosed something</li> </ul> <p>If I can get on the other end (sometimes I can, sometimes I can't):</p> <ul style="list-style-type: none"> <li>- run disk benchmarks to find out if the remote storage system is a problem</li> </ul> <p>You just start picking the pieces apart. Is the problem in the link between the ends? No. Is our end working right? Yes. Then you start poking at the other end to see what we can find out.</p> <p>Sometimes you find that something was misconfigured, or just congestion on the network, or sometimes you find new flakiness. New software interactions, new Linux operating systems, etc. To use a realworld example: the base SLES 10 kernel is 2.6.16; this kernel has wicked bad TCP problems. But if you can get above that (we're running 2.6.20) they've fixed them. So you stumble across things like that. But so much depends on where the problem lies, and sometimes isolating the problem.</p> <p>Solving problems is easy once you have all the data in front of you. It's getting the data and knowing what data to get that's the hard part. Networks are notorious for this, right? They're black boxes. Very rarely are you lucky enough to have access to somebody who can actually find out operational status on routers and the like. So you have to infer what's happening by using things like Iperf, netperf, pipechar, etc.</p> <p>I would love to see the network become not a black box. Just like on your scheduler. You have a way of poking a scheduler to see what its status is, right? How heavily loaded are you? What's my expected queue time? Blah, blah, blah...</p> <p>It would be nice if the network had such a thing. Now that is a non-trivial problem – I'm not suggesting it isn't. But if we could get a network that allowed that, then you could begin to write software that could do optimizations and decide which routes to take.</p> <p>There are technologies out there that are aimed at the problem: things like GMPLS so you can set up virtual circuits. But right now it's still extremely human-intensive. You have to call somebody and negotiate, and it takes months to set it up. Then it's going to stay there for a long, long time. If you need it to change, that's bad because it's going to take months again.</p> <p>Whereas it would be nice if you could say, "Hey I'm going to run a transfer between ANL and ORNL. I need to move this much data and I need to move it this fast. Give me a path." That would be sweet.</p>	

Interview ID=10 2 July 2007	ANSWERS	ANNOTATIONS
<b>Q9.2 What types of information do you need in order to address the challenges you face today?</b> [continued]	<p>[prompt asking if the transfer problems are on a per-transfer basis]</p> <p>No. A lot of this stuff you can fix point-to-point. Once I get a transfer to Oak Ridge running right it's pretty much fixed, except for congestion on the network (like somebody kicking off a big job at the same time I did). That is, until the next time they do a kernel upgrade and blow away all the modifications you made.</p> <p>This type of thing happens, even at ANL. We participate in something called PingER, which is a project that sends pings to monitor bandwidths. And we got an email saying, "Hey. Your bandwidth got cut in half. What happened?" The problem was that the default machine builds for the division had changed, and all the machines were rebuilt, and all of our modifications in sysctl.conf that sets all the TCP parameters got nuked. So we lost half our bandwidth.</p> <p>[prompt asking if any storage-related information is needed]</p> <p>It would be good to make storage more of a first class citizen. Being able to poke the storage and find out how busy it is, what its topology is, how fast it can go, etc. That would be an interesting capability.</p> <p>For instance, "Is the storage dedicated to the machine, such that if I have this machine I also own the storage? Or is this like on TeraGrid where it's SAN, and while I may have the machine, there could literally be a thousand other machines beating against the same storage resource. So even though in theory it has this amazing amount of bandwidth it can give me, I'm not getting it because I'm sharing it with a thousand other people.</p>	
<b>Q9.4 By contrast, can you provide examples of technologies you find very useful today?</b>	<p>GridFTP is useful. I may have a bias here, but I would use whatever the best solution for the problem is, and I don't know of a better solution for moving data around.</p> <p>Iperf is an incredibly useful tool for doing network troubleshooting.</p> <p>bcfg2 for handling cluster configurations is pretty cool, pretty effective. The really cool thing about it is it can probe machines and generate reports about their configurations: "Yes, all the filesystems look exactly as they should. They have the right packages loaded, the right modules running, the right services running." Or it can flag changes and report them or automatically launch a rebuild.</p> <p>It might be interesting if somebody got a bcfg2 setup to do Globus – it might be a trivial way to push Globus out.</p>	
<b>Q10.1 Can you think of any work-related tasks that decrease your productivity?</b>	Do I prefer sitting in meetings, or sitting at my desk working on technical stuff? Obviously I prefer sitting and working on tech. But meetings are not counter-productive; they're just productive in a different way. You've got to exchange information; you've got to coordinate things on a project.	
<b>Q10.2 Describe repetitive tasks associated with your work</b>	The repetitive stuff we are writing tools for. Some of it is repetitive today, but we're working on that. For instance I'm new here, so Iperf isn't commonly used yet. Currently I'm running all the Iperfs by hand, but I'm also writing the Iperf modules to plug into our test harness.	
<b>Q10.3 Describe time-consuming phases of your work</b>	Much of the time-consuming stuff right now is because we're in the design phase. It is one time only.	
<b>Learning about the Globus user experience</b>		
<b>Q11.1 Which Globus data components do you directly interact with in your work today?</b>	GridFTP	
<b>Q11.2 Did you install the &lt;component&gt; client yourself?</b>	yes	
<b>Q11.3 Did you install the &lt;component&gt; server yourself?</b>	Yes, on the testbed. My team will be installing on the production facility.	

Interview ID=10 2 July 2007	ANSWERS	ANNOTATIONS
<b>Q11.4 How many people currently use your &lt;component&gt; server</b>	<p>On the Blue Gene/L [<i>the old machine</i>] I'm not sure of the number of users. The number of projects using GridFTP is seven or eight.</p> <p>That's one of the things about being a leadership-class facility. We are going to have a few really big users – not the other end of the spectrum. And when we are up-and-running on the big machine, we're expected to have on the order of twenty projects with one- or at the most two-hundred users. So by big facility scales, not many users.</p> <p>We won't have to deal with a lot of the problems that a place like NERSC does. They probably have to reset passwords on a daily basis because they have thousands of users. We'll probably need to do that on a weekly basis – it won't be a big deal. They probably have to help people get their ssh keys installed every day. We won't have those kinds of problems.</p> <p>On the other hand, we will have people who are trying to move around petabytes of data, or who want teraflops of performance – running at scales they've never run at. We're going to encounter all sorts of problems.</p> <p>We can guess where some of the problems might be, we just don't know which one is going to reach up and bite us. And we've designed around them as best we can, but there's always a bottleneck someplace, right? Otherwise we'd run infinitely fast. Start the lottery now as to which one is going to get us.</p>	
<b>Q12.1 Which Globus security components do you directly interact with in your work today?</b>	<p>So far we're not. For the testbed we've been running GridFTP without security.</p> <p>The testbed is internal to the facility; it's behind firewalls so we can get away without running security on that. Once we start actually opening the firewall to the outside world, we have to be secure one way or another. The alternative to GridFTP is scp, which is secure but dog slow, so performance is unacceptable. And therefore we're forced to provide an alternative to scp.</p> <p>If the GridFTP team can make the ssh-based transfers work, we might go with that. That's new, it's not as well tested so it's less stable, and it can't do delegation, so there is a concern about third party transfers. There are security issues with the ssh option.</p> <p>So what we may end up telling our users is, "There are two ways you can get data in and out of this place:</p> <ol style="list-style-type: none"> <li>1) You can scp it. Then you don't need to do anything. You can log on right now and scp.</li> <li>2) If you want better performance than that, GridFTP is the tool we have. And in order to remain secure, you must have a GSI certificate to use it."</li> </ol> <p>If I could be convinced that a non-GSI version of GridFTP was stable and secure, I'd use it. I hate GSI. It's very good at what it does, but it is a pain. When I was involved in GridFTP development, GridFTP didn't have problems - GSI had problems. Once I could get people past the GSI issues and get it all configured, GridFTP just runs.</p> <p>I need to worry about security because I have to let people move data in and out of the facility. Users want performance that is better than scp, and GridFTP is the only solution that I'm aware of that can meet that performance need.</p> <p>Therefore I'm locked into whatever security mechanism GridFTP uses. If you could provide me with an alternative mechanism that has all the same benefits and stability of GSI, yet was easier to install, I'd jump on it in a heartbeat.</p> <p style="text-align: right;"><i>[answer continued on next page]</i></p>	

Interview ID=10 2 July 2007	ANSWERS	ANNOTATIONS
<b>Q12.1 Which Globus security components do you directly interact with in your work today?</b> [continued]	<p>The big thing that GSI gives you that no other solution does is delegation. GridFTP particularly needs delegation because it does third party transfers. The way third party data transfer works is:</p> <ul style="list-style-type: none"> <li>- a client talks to two servers</li> <li>- the client tells one of those servers: hey start listening on a port</li> <li>- then somebody – you have no idea who in the hell it is – is going to connect to the client and start sending commands. You can see the danger, right?</li> </ul> <p>And that’s why in plain FTP servers many administrators disable third party transfers.</p> <p>Third party GSI-enabled GridFTP transfers work as follows:</p> <ul style="list-style-type: none"> <li>- the client connects to each server and delegates a self-signed credential</li> <li>- so a proxy is created on each server with the client’s user credentials</li> <li>- further communications during the transfer require the client and server credentials to be identical, or error conditions will be raised</li> </ul> <p>[prompt asking about scp]</p> <p>In my experience, for most data movement scenarios you want security only on the transfer commands. You want to make sure the right person is telling the right thing to move. Most of the scientific community does not care if the data being transferred is encrypted. But scp encrypts everything. That is one of the reasons scp has such horrible performance. It just opens up an ssh connection and ships everything over the connection after encrypting it.</p> <p>With GridFTP you’re encrypting the control channel and integrity-protecting it to make sure nobody is tampering with it, but by default the data channel is only authenticated. The data channel is not encrypted or integrity-protected by default. You can turn those options on if you want to – for instance medical people do – but the vast majority of scientific users don’t need to.</p> <p>scp also does not support third party transfers. Third party transfers are necessary because they allow me to move data at tremendous rates from a client on my laptop. For instance, say I have a very powerful GridFTP server at the data source and a very powerful server at the target destination. Without third party transfers I would have to route it all through my laptop (or whatever the client has to be running on.)</p> <p>Now, the counter argument is, “Yeah, but I could just log in over there.” Well, what if you don’t have an account over there? As an admin, there may be a GridFTP server that I am willing to give you permission to run on, but I may not be willing to let you log in to my machine. So the third party model allows you to invoke work remotely.</p>	
<b>Q13.1 Which Globus execution components do you directly interact with in your work today?</b>	Karajan We use Cobalt because it has Blue Gene-specific stuff in it. I think a Cobalt-GRAM adapter has been written but we do not use it. We’re not against running GRAM, but the reason to run it would be if it is requested by a user. Like our user also runs jobs someplace else using a different scheduler, so their scripts assume GRAM. We’ll install and run GRAM the day a user shows up and says, “I use GRAM. That’s the interface I’ve written all my code around.” But now because all of our users are Blue Gene users, they’re all running Cobalt or Loadleveler.	
<b>Q13.2 Did you install the &lt;component&gt; client yourself?</b>	yes	
<b>Q13.3 Did you install the &lt;component&gt; server yourself?</b>	Yes, on the testbed. My team will be installing it on the production facility.	
<b>Q16 Why do you use &lt;component&gt; instead of an alternative technology?</b>	Karajan: The big reason we looked at Karajan (or just workflow in general) instead of just writing scripts is that it has built-in semantics for parallelism. A lot of the tests, for example the disk benchmarks on the file servers, all run in parallel. And Karajan makes it trivial for me. I can define a list of resources (filesystems) and then I can say, “parallel for filesystem in list”. It’ll just in parallel invoke the same command on all of them simultaneously. <p>And then it does automatic barriers. Karajan waits until all the commands in the block are done before it goes to the next instruction. That would be a lot of work to write in something like python. I’d have to create all those constructs myself. So those are the big things: automatic barriers and parallel semantics.</p>	

Interview ID=10 2 July 2007	ANSWERS	ANNOTATIONS
<b>Q17 What are the major challenges you face using &lt;component&gt; today?</b>	<p>Karajan: The documentation could be better. In particular because they have two languages: they have XML and .k, and they kind of mix and match those in the doc. So sometimes I happen to be trying to use the .k, and it'll be frustrating because I can only find an XML segment. And, at least to me, converting from one to another is not trivial.</p> <p>And karajan is newer, so I've stumbled across some bugs. One was particularly insidious:</p> <p>I have this habit when I write my code I use C style comments. And so I'll do <code>"/**"</code> and then several <code>"**"</code>s and ending with <code>"*/"</code>. I build a box around my comments. Well there was a bug in the parser that if the number of <code>"**"</code>s was even (or odd? can't remember which) it didn't close the comment. And so I kept going, "Why the hell can't I get this to work?"</p> <p>On the flipside, the Karajan developer I've been working with has been outstanding at responding to my questions. In the problem described above I sent him the code, saying, "I'm at a loss. I don't get this." And he showed me a command that will dump the intermediate code for inspection. He looked at the intermediate code and figured it out. He's been really good answering questions. And he cranked out a Cobalt provider for me. So he's been very responsive and that's been nice.</p>	
<b>Wrapping-up</b>		
<b>Is there anything you'd like to say to the people who build software for use by people like you?</b>	<p>To people who are not experts and didn't write the technology, documentation is worth its weight in gold. And I don't know of a project out there that wouldn't benefit from having better documentation.</p> <p>Examples of documentation that GridFTP needs that it doesn't currently have:</p> <p>An engineering guide written for sysadmins (or people about to install a GridFTP server.) There should be a document that walks you through the thought process:</p> <ul style="list-style-type: none"> <li>- how big does the machine need to be?</li> <li>- how big do the drives need to be? how fast?</li> <li>- what should the network connectivity look like?</li> <li>- should I run a striped server? should I not run a striped server?</li> <li>- should I run GSI?</li> </ul> <p>An engineering guide to help you <i>plan</i> your installation prior to starting the installer. So you have information ready that you may need during the install. Looking at the bigger picture:</p> <p>People write great developer guides. And that's great for somebody who is a developer. But what about the rest of us? What the hell does this thing do? Even with GridFTP we've tried, but people sometimes just don't get the big picture of GridFTP. Concepts as simple as, "What's a client? What's a server? What's a third party transfer?"</p> <p>One thing that Karajan has is an "almanac". Basically it is a list of every single command – every single language construct in Karajan, with an accompanying example of how to do that language construct. Just a little five or ten line snippet of code. It shows you what the syntax looks like, and that's come in very handy. So if I know I need to use <i>parallel</i> for I can find a snippet of code that actually uses it. So I've found that to be very useful.</p>	

## D.11 ● The difficulty is not that things break, but detecting that something is broken

Interview ID=11 23 July 2007	ANSWERS	ANNOTATIONS
<i>Pre-interview question: Do you interact directly with Globus software in your work today?</i>	yes	
<b>Establishing context</b>		
<b>Q1.1 Please provide a one-minute overview of your project</b>	I work on the OSG Engage VO project. We help new users get on to OSG – especially users that are not in high-energy physics. The idea is to demonstrate OSG infrastructure.	
<b>Q1.2 What is the project's name?</b>	OSG Engagement VO	
<b>Q1.3 Which agency funds the project?</b>	NSF	
<b>Q1.4 What field does your project belong to?</b>	Bioinformatics, Meteorology, Material Science	
<b>Q1.5 What is your job type?</b>	System Designer/Developer	
<b>Q1.6 How long have you been a &lt;job type&gt;?</b>	Six months	
<b>Learning about discipline-specific goals and approach</b>		
<b>Q2.1 What are the main goals of your project?</b>	To bring new users on to the Open Science Grid.	
<b>Q2.3 What are the professional measures of success for you?</b>	For me it is about how easy it is for these users to work on the infrastructure. If they find it very useful that's what counts – and if they keep using after I give an introduction. The whole idea is that we approach users who have smaller projects and smaller teams. They usually have very little IT or computer science knowledge. They have knowledge enough to write the model, or to implement their idea, but that's pretty much where things end. So when we start working with these users we have to make things self-explanatory. They can't be exposed to a lot of errors. Things have to be pretty much self-contained and running. After a couple of weeks, the goal is for them is to be able to do their own runs without emailing us questions.	
<b>Q3 What are you investigating?</b>	A big part of the work is spent working with OSG, other VOs and resource owners. We verify sites, and as issues come up we interact with the OSG ticketing system. We act as a middleman between OSG and our users to make sure that they don't have to deal with all the day-to-day issues that come up.	
<b>Q4.1 What is your method for investigating &lt;phenomena&gt;?</b>	My approach is to do a lot of testing: site verification and probing things to make sure I run into problems before my users do. When a user submits a job and a failure occurs, things are managed in a way that the job will be resubmitted to another site. I then go to the logs at the failing site and look to see what's going on. So my job involves a lot of detective work, I would say. Finding out details about issues. I try to diagnose problems before the users even know that they have them. For example, if a user submits a big job at a site where I see a lot of other jobs failing, I can look at the logs and say, "Ok, that site has a problem with its filesystem." Then I can reduce the number of jobs dispatched to that site, and the user's jobs will not be submitted there anymore. If an error is detected then the site will fall off the list of available resources by itself. The workflow will be successful either way because the job will be resubmitted somewhere else. But time will be wasted submitting to the broken site – I try to minimize that. Sometimes a site looks ok for a while before it begins failing; that's what we have to watch for. So a big part of my work is writing tests and scripts to make sure I can detect problems early.	

Interview ID=11 23 July 2007	ANSWERS	ANNOTATIONS
<b>Q4.2 How do you work?</b>	<p>Each resource on OSG reports into a central advertising component, called the Resource Selection Service (ReSS). The advertisements include which Virtual Organizations (VOs) are supported. So we poll the ReSS (every ten minutes or so) and build a list of all the sites advertising support for our VO. Then we submit a couple of different jobs every six hours to each site, to probe and test them.</p> <p>So that's the first line of defense. If a site has an authentication problem for example, they will obviously fail our test and won't get advertised in our local metascheduler. That type of testing takes care of the sites that are obviously broken.</p> <p>Other times we break sites, for example, by submitting too many jobs. And those problems are a little bit harder to find. Those are the ones that we see, "Ok, we have a hundred jobs running, and now all the new jobs are failing for some reason." Maybe we filled up a filesystem or something like that. So sometimes we won't find out about this type of problem until six hours later when the next verification run comes around.</p> <p><i>[prompt asking how interviewee verifies a site]</i></p> <p>There's one job going to the fork job manager and one going to the scheduler. The jobs probe certain things that we know might be an issue. For example they make sure that the data and application directories exist and are write-able by us, and that they're not full. We also test for prerequisite software, and whether or not it has outbound network connectivity.</p> <p>Not all of the error conditions raised by the tests are fatal. If we can't write to the data directory, that's a fatal error. But for instance if we can write to the data directory but we don't have outbound access, that's still fine for some jobs. We pass information like that back, and use it later when matching jobs against that site. A user job can express a requirement for outbound network connectivity, so a site lacking that would not be considered as a potential resource for that run.</p> <p><i>[prompt asking if the interviewee personally writes the tests]</i></p> <p>Yes, I am the only person doing the testing/site verification work for the VO. But there is a distinction between the work I do for this VO and the global OSG infrastructure testing effort. What I'm doing is testing from our users point of view, which is different than the larger effort.</p> <p>For example, during the OSG operations center testing, they run under a certain user account. Now maybe the OSG test user encounters no authentication problems, but when running as a user in my VO, authentication problems <i>do</i> appear. And the same thing with filesystems permissions and all that stuff. So we call what we do "testing from the Engagement VO's point of view."</p>	



Interview ID=11 23 July 2007	ANSWERS	ANNOTATIONS
<b>Q4.2 How do you work?</b> [continued]	<p>[prompt asking if the interviewee's users all belong to the same VO]</p> <p>Yes, but by definition our VO is widely distributed. Our VO is a collection of sixteen users that may form new VOs by themselves at some point. Most of the VOs in place today are either a project-based VO, or an experiment-based VO. But the users in our VO may not even belong to the same field. There have been discussions about what to do as they grow and become really big, or if particular users really use a lot of CPU time. Should those users form their own VOs? Maybe, I don't know yet. I really haven't been involved in those discussions. The point is that my users are all over the place.</p> <p>But they are aware that they are members of the VO. A few of them know each other. Some of the users work in the same group or project. Others have been referred to us from one group or another. So some of them know each other, while others have no idea about each other.</p> <p>The Engage VO is not really a tech support organization, though we are helping users to get going. The idea is to provide an infrastructure that is easy enough that they don't need much tech support. We provide:</p> <ul style="list-style-type: none"> <li>- a matchmaking system</li> <li>- a submit host</li> <li>- example scripts and OSG wrapper scripts that they can adapt for their needs</li> </ul> <p>But once they're running the idea is that they should be self-sustained and we should not need to help them much.</p> <p>[prompt asking if the project has federated resources behind the submit host]</p> <p>No. I guess we're special in that way. Most other VOs are either based on an experiment or they are a resource provider, and we are neither. We don't own any resources on OSG, we don't have any single project or experiment. Resource usage within the Engage VO is opportunistic. We are getting leftover cycles on OSG that big experiments haven't used up.</p> <p>Our users know this as well. We don't make any promises on how many CPUs our users will be able to get or anything like that. Resource availability can vary a lot, but our users are just happy to get anything.</p>	
<b>Q4.3 How do you keep track of interim results, if at all?</b>	<p>Over time I do not keep anything. I just work within a six-hour (or maybe a day-long) time period, where it's just a moving window.</p> <p>As I mentioned earlier, when the tests are successful they spit out some extra classads – some extra little pieces of information. I take the real classad from ReSS, I add my little extra things to it, and then insert it into my local Condor process. That's how the resources are advertised within the Engage VO. When the next verification test comes around in six hours, previous advertisements are just thrown away.</p> <p>So I just keep a current view.</p>	
<b>Q4.4 How do you test work-related hypotheses?</b>	<p>[prompt asking how interviewee decides which tests and probes need to be written]</p> <p>That's just by knowing what things have been failing in the past. We have a pretty solid set of tests now. They're not very complicated. We may be testing like twenty or thirty little things. There's nothing that's like a huge test; most of them are very simple, like, "Can I touch a file in this directory?" Things like that.</p> <p>Development of the tests is ongoing. If I have a site that always fails X test, I will go back and see why that is. I need to determine if my test is broken, or if it's something I should report to the OSG operations center.</p> <p>Another thing that can happen:</p> <p>If I start seeing that jobs failing at a certain site, I'll say, "If the job is correct yet it is failing, then obviously we're not doing our testing well enough." And we add a new test to make sure that doesn't happen again.</p>	
<b>Q5.1 In what ways do you interact with simulations in your work?</b>	<p>Most of the models that are running in the VO are simulations, but it's not that often that I get to work on the simulation code itself. Most of the time to me it just looks like an executable with a set of inputs and outputs. I work with it on the level of a process that I need to support, although I need to understand the model enough to know what the inputs are, and what the outputs should look like. Very seldom am I involved in the coding on the model itself.</p>	

Interview ID=11 23 July 2007	ANSWERS	ANNOTATIONS
<b>Q5.3 How do you interact with inputs to your simulations?</b>	<p>This depends on different models.</p> <p>One user I worked with recently, for example, had five hundred thousand input files that needed processing. So a lot of the work there involved setting things up and making sure the jobs were picking the right inputs.</p> <p>In other cases you have pre-processing of inputs, such as with our weather models. Dealing with that is a little messier, where you actually have to understand a little bit more about the model. Making sure the inputs are correct by, for example verifying that they fall into the correct time period.</p> <p>Most of the inputs we see are file-based, though a couple of them are using databases.</p>	
<b>Q5.4 How do you interact with the output of your simulations?</b>	<p>For most of the runs we're doing, if outputs are created then I'm done. I just give them to the user and they tell me if it was a good run or not. There are very few instances where we do any visualizations or anything like that.</p> <p>The outputs are most commonly file-based.</p>	
<b>Q5.5 By what mechanisms is access to your simulations controlled?</b>	<p>The simulations are open. We tell our users to tell us that if they are worried about other users in our VO having access to their results. So far we haven't had anyone come back to us with that worry yet.</p> <p>But if that were the case we would probably set up a new user for our VO – VOMS will allow us to do that pretty easily – and just use Unix file permissions to separate that user from everyone else.</p> <p>As far as permissions to initiate the simulations, they are just executables owned by a user. We make sure the user has a wrapper script that will build the right commandline for them and copy the inputs to the right place.</p> <p>Most of the code we're working with is open source, so nobody has been worrying about protecting it yet. And if they were, like for example if there is a commercial code, then we would again set up a separate user and protect things that way. Then maybe use our advertising mechanism to identify the sites where it's installed.</p> <p>Our advertising mechanism is based on Condor classads and ReSS, as described in Q4.3. The end result is that the VO's classads describe available OSG resources from the point of view of the Engage VO.</p>	
<b>Q6.1 Describe how you interact with data in your work</b>	<p>There's a lot of data, and most of it is filesystem-based. I can open them up and look at them, but most of the time their contents don't make any sense to me. Like proteins and things – I have no idea – I have no background in understanding that. So most of the time I have to email the user asking, "Is this right or wrong?" or, "Should I use this input or that one?" and they will have to tell me.</p> <p>There's also output from the jobs, but I'm not doing anything smart with that data either.</p> <p>Other types of data I interact with include test results. There's some automatic parsing of that. There's many ways a job can fail, obviously. One of the ways is that the job will be successfully submitted to a site, something will run, but not successfully. Let's say that a filesystem goes away while the code is running. To detect that type of failure is not that easy. This is because different error codes are returned, and some schedulers will say the job was successful while others will say that it was not successful.</p> <p>So one technique we use for dealing with this is we write markers to the job output. The job wrapper will run a code and if it is not successful (if it finished prematurely or something like that) the marker will never be written to the output. After then run the job output is parsed on our side to assess whether or not it succeeded. If not, the job will be resubmitted. That's the best example of parsing the output logs.</p>	
<b>Q7.1 What resources do you use in your work today?</b>	<p>The only resource owned by the Engagement VO is a single machine called the <i>submit host</i>. It has GridFTP, Condor, etc. on it. Somebody else owns the remote compute resources, so we have no control over them. So our submit host acts like a gateway to OSG resources.</p> <p>Our user jobs are mainly submitted via Condor-G on our submit host to the GRAM2 servers hosted by the various OSG resource owners. As far as which schedulers are used by the remote resource owners, I don't know and I don't really care. That varies from site-to-site, and I don't really care what's happening behind there.</p> <p>Though recently we have tried to do more MPI jobs, and we're getting into a little bit more remote site configuration issues. This is because MPI is pickier about underlying hardware and setup. But when we run serial jobs we don't really care as long as we can get the CPU.</p>	

Interview ID=11 23 July 2007	ANSWERS	ANNOTATIONS
<b>Q7.2 How do you share work-related resources with others?</b>	If you're an Engage VO user you get login access on the submit host. We have one user who wants to run their own submit host, but it's pretty much a copy of ours.	
<b>Q7.3 By what mechanisms is access to your work-related resources controlled?</b>	All our users get their own DOE certificate. They use that to authenticate with the remote resources; that's how access to the compute resources is controlled. The only thing they need to do is that they join our VO. This means that their certificate subject goes into our VOMS server.	
<b>Q7.4 How do you locate available resources for use in your work?</b>	Via the ReSS plus our own additions to that. Then they're advertised and our local Condor does matchmaking between the users jobs and the remote resources. Our additions include: 1) site verification from our point of view, and 2) the additional site requirements users have for their jobs (what software is installed, what does the network connectivity look like, etc.) ReSS shows a fairly generic view of resources, so we add a little bit more detail to it.	
<b>Q7.5 What types of information do you need to know about a resource in order to determine if it is suitable for your work?</b>	All the information we really need is from ReSS, plus passing our verification tests. As a side note: twenty percent of the sites advertised as available in ReSS fail our verification tests.	
<b>Q8.1 What software do you currently use in support of your work?</b>	Condor for matchmaking, Globus for submitting jobs, and a lot of perl and shell scripting	
<b>Q8.2 What scripting languages have you used in the past year?</b>	Perl, python, shell	
<b>Q8.3 What programming languages have you used in the past year?</b>	C, C++, Java	
<b>Q8.4 What workflow tools do you use in your work?</b>	None except Condor DAGman	
<b>Q8.5 What parallel computing tools do you use in your work?</b>	MPI, a lot of different local schedulers, and PVFS	
<b>Q8.6 If the need for new software-based functionality arises in your work how do you acquire it?</b>	We are trying to reuse as much open source software as we can. That's the first thing we look for, otherwise we'll write it.	
<b>Q8.7 How do you share software with others?</b>	Most of it is published on our FTP site and documented in the OSG wiki	
<b>Learning about the user's problems</b>		
<b>Q9.1 What challenges do you face today in accomplishing your work-related goals?</b>	Communication is a big issue. In a project as distributed as OSG is, you spend a lot of time in meetings and writing emails, just communicating issues back and forth. So I think that's my main problem. If you want to be a part of something you have to invest a lot of time. But the big problem is how distributed the project is. There are so many sites, there are so many projects and experiments and they all have slightly different agendas. So something that might be important to you, nobody else might care about (or the other way around.) Or you're getting pushback from people on something that doesn't make sense to you. I think you get this on all large projects, so there's not really a way around that.	

Interview ID=11 23 July 2007	ANSWERS	ANNOTATIONS
<b>Q9.3 What technology-related obstacles do you currently encounter?</b>	<p>Not that many... there are the standard problems in distributed computing. You know, resources come and go due to planned downtime or something breaking. It might be hardware failure, or a filesystem going away, or hanging, etc. That's the thing I deal with everyday.</p> <p>The difficulty for us is not that things break, the difficulty is in detecting that something's broken. I may not even know who owns the site – it's just a black box to me – and something went wrong. Now I have to figure out what went wrong. So I do a little bit of probing, and then either tell the remote site what went wrong, or fix my stuff.</p> <p>Most of the times it is easy for me to figure out what is going wrong once it is detected. Once you start zeroing in on an issue its easy. But to know that there's an issue – I think that's the problem.</p> <p>The method I use to detect problems is mainly just getting a feel for the run and seeing how it's progressing. If there's a site that doesn't have any long-running jobs, for example, that's probably an indication that something is bad and jobs are dying. I have a few scripts that help me with detecting those problems. They will notify me if it looks like the site is responding too fast or too slow. In general it's about knowing the jobs and seeing how the sites are behaving.</p> <p>So when a new OSG site pops up that we've never run on before, we send only a couple jobs at a time to the site for a week or so. If after a week the site is running the jobs without problems, we'll slowly increase the maximum number of jobs placed there. This approach works pretty well, actually.</p> <p>A lot of sites want us to be on mailing lists to receive announcements about downtime. At this point we don't even care about receiving those messages. It makes no difference to us whether it's planned downtime or if it's a failure.</p>	
<b>Q9.4 By contrast, can you provide examples of technologies you find very useful today?</b>	<p>Computers ☺</p> <p>I have a lot of tools in my toolbox. Scripting (like perl or shell) is probably my number one tool because I try to automate as much as I can. That makes my life easier.</p>	
<b>Q10.1 Can you think of any work-related tasks that decrease your productivity?</b>	Not really. I try to stay out of them. I think communication is slow or boring or time-consuming. But there's no better way of doing it.	
<b>Q10.2 Describe repetitive tasks associated with your work</b>	Not much, that's where the scripting comes in. If there's a repetitive task we'll script it up.	
<b>Q10.3 Describe time-consuming phases of your work</b>	<p>When we first start meeting a new user, that's time consuming because we're on different pages. Trying to get them up to speed on new technology they've never seen before, and at the same time you're trying to learn about their model or simulation. So you're on two very different pages and you're trying to get closer to each other.</p> <p>So that's time consuming. But after about a week or so you get to a point where things start going forward and it's very rewarding directly after that.</p>	
<b>Learning about the Globus user experience</b>		
<b>Q11.1 Which Globus data components do you directly interact with in your work today?</b>	GridFTP and RFT	
<b>Q11.2 Did you install the &lt;component&gt; client yourself?</b>	yes	
<b>Q11.3 Did you install the &lt;component&gt; server yourself?</b>	yes	
<b>Q11.4 How many people currently use your &lt;component&gt; server</b>	<p>GridFTP: Sixteen users at least for my VO. At my institution there are many more, but I don't know the exact number.</p> <p>RFT: Three or four using it day-to-day</p>	
<b>Q12.1 Which Globus security components do you directly interact with in your work today?</b>	<p>GSI wrapped around the VOMS clients</p> <p>A little bit of the delegation service when we run GT4 jobs</p>	

Interview ID=11 23 July 2007	ANSWERS	ANNOTATIONS
<b>Q12.2</b> Did you install the <component> client yourself?	yes	
<b>Q12.3</b> Did you install the <component> server yourself?	Yes, though system administrators at my home institution installed a base package that includes some GSI infrastructure; I did additional configuration.	
<b>Q12.4</b> How many people currently use your <component> server	My sixteen users use the VOMS server to get GSI certificates with VOMS extensions. Delegation: Three or four	
<b>Q13.1</b> Which Globus execution components do you directly interact with in your work today?	GRAM2 and GRAM4	
<b>Q13.2</b> Did you install the <component> client yourself?	yes	
<b>Q13.3</b> Did you install the <component> server yourself?	yes	
<b>Q13.4</b> How many people currently use your <component> server	GRAM2: sixteen GRAM4: three or four	
<b>Q14.1</b> Which Globus information components do you directly interact with in your work today?	A little bit of WS MDS when we do GRAM4 jobs (for notifications and such), but not really much else.	
<b>Q14.2</b> Did you install the <component> client yourself?	yes	
<b>Q14.3</b> Did you install the <component> server yourself?	yes	
<b>Q14.4</b> How many people currently use your <component> server	Three or four	
<b>Q15.1</b> Which Globus common runtime components do you directly interact with in your work today?	XIO and Java WS Core, but indirectly due to use of other services	
<b>Q15.3</b> Did you install the <component> server yourself?	yes	
<b>Q15.4</b> How many people currently use your <component> server	GridFTP: sixteen Java WS Core: Three or four	
<b>Q16</b> Why do you use <component> instead of an alternative technology?	GridFTP: first, I don't think there's anything that's really comparable. But the main reason is probably that it fits with the authentication infrastructure we are using. RFT: because GRAM4 uses it; we have done some independent transfers with it, but more day-to-day it is used because of GRAM4 VOMS-wrapped GSI: that is the standard for OSG Condor-G: mostly because that's what ReSS came out using, and it included information published about the sites. That's a good deal for us, because we don't have to manage the information ourselves. GRAM2: because it's a standard on the Open Science Grid GRAM4: because the staging support is better. It's a pretty big improvement over GRAM2 in that sense, because you can do smarter staging (like the whole filelist). That maps much better to our Condor description files. And in general the architecture is better.	

Interview ID=11 23 July 2007	ANSWERS	ANNOTATIONS
<p><b>Q17 What are the major challenges you face using &lt;component&gt; today?</b></p>	<p>GridFTP: none. It is stable.</p> <p>RFT: I think the user interface could be a little nicer; the commandline clients are a bit messy I think. But other than that – once it’s running, it’s fine.</p> <p>VOMS-wrapped GSI: the main challenge is that it hasn’t been working on 64-bit machines, so we have to move it to a 32-bit machine. I think they have fixed that, but there has been a couple of major bugs.</p> <p>Condor-G: the main issue we have is we would like to have more some more hooks into it. So, for example, when jobs fail we would like to have callouts that enable us to handle that. But that’s something that we are talking to the Condor team about, so maybe that’s something that will be improved soon.</p> <p>GRAM2: not much, just the old standard ones. Sometimes we get weird errors that don’t really reflect what’s going on. Other than that it’s fine. Weird errors like, “Error code 17” that supposedly means one thing, but is most commonly due to something else. For example, it says, “could not create job description file” when the real reason is the user doesn’t exist. But you get used to it.</p> <p>These types of problems are well known in the community.</p> <p>GRAM4: most of the OSG sites are not supporting it in the same way they’re supporting GRAM2. Even the sites that are saying they’re supporting it – it’s a pretty low priority right now. So I think that’s what prevents us from switching right now. Hopefully the next version of the OSG stack will take care of it.</p> <p>By “support” I mean:</p> <ul style="list-style-type: none"> <li>- the time the site administrator spends setting up the service</li> <li>- if you email them about GRAM2 they will be very responsive, but if you email them about GRAM4, their response is only best effort</li> </ul> <p>I think it is this way right now because GRAM2 is the production version for OSG. If GRAM2 is failing for them, the site is considered to be failing. If GRAM4 is failing it is not that big of a deal for most people.</p> <p>Java WS Core: I feel like this area could be a good opportunity for Globus, but at the same time, more dynamic IP address handling is needed. You know, how the container handles the network coming and going needs work. I think the container’s notifications could be a good fit for us, but we need something that works better in that environment.</p> <p>So the use case I’m dealing with (in a different project) is where there’s a sensor somewhere connected by a GPRS cell phone. The sensor gets different IP addresses every time it connects. It’s just up for a few minutes and then goes down again. The current notification framework doesn’t really work well in that dynamic scenario.</p> <p>There’s also the issue of the guaranteed delivery of notifications. Let’s say we have a sensor that needs to aggregate some data. So every now and then it pops up and says, “Ok, here’s my data for the past hour” and sends the data to a service. At the same time it would check for any pending updates from the service, so it would process notifications sent by the service while the sensor was offline.</p> <p>So the dynamic IP address is one issue, and notification guarantees is another. Every so often I think, “Oh we could use notifications for this” but then I remember these issues and realize it is not a good fit. Instead we’re writing our own services using Axis. It’s not that far from Globus, so there’s an opportunity there.</p>	

## D.12 The right approach is to be highly collaborative with domain specialists

Interview ID=12 24 July 2007	ANSWERS	ANNOTATIONS
<b>Establishing context</b>		
<b>Q1.1 Please provide a one-minute overview of your project</b>	The CNARI project is aimed at using database technology to store data that is electrophysiological or neurophysiological in nature. The data consist of hundreds or thousands of timepoints in a time series. Each timepoint consists of a large amount of data itself, such as a brain image or an electrophysiological recording from different sites on the brain. Using database technology to store the data is a way of allowing much more useful access to the data. Also through interactions with Grid computing experts at the University of Chicago, we found that it's a much better way to interface with distributed and high-powered computing devices in order to process the data. This is all done in the context of research in the recovery from stroke, which is a disease of brain vessels that is very devastating.	
<b>Q1.2 What is the project's name?</b>	CNARI	
<b>Q1.3 Which agency funds the project?</b>	The project is funded by the agency that is particularly interested in strokes that affect language function, which is what I mostly do for a living. The agency name is the National Institute on Deafness and Other Communication Disorders, which is part of the National Institutes of Health.	
<b>Q1.4 What field does your project belong to?</b>	Medicine, Neurology, Neurobiology, Psychology, Speech Pathology Also much of the work on this project for the first two years deals with the field of Computer Science, because we are building an infrastructure to do this representation.	
<b>Q1.5 What is your job type?</b>	Scientist	
<b>Q1.6 How long have you been a &lt;job type&gt;?</b>	30 years	
<b>Learning about discipline-specific goals and approach</b>		
<b>Q2.2 How will the success of your project be measured?</b>	The project has very specific milestones for success: The first two years of the project are aimed at developing an infrastructure for representing these large datasets of time series data. The last three years of the project aims at: - recruiting research subjects to participate in experiments - having their data represented and processed with the infrastructure - disseminating this information throughout various sites in the United States and elsewhere to demonstrate that the infrastructure is not only useful for representing and processing data efficiently, but also for sharing information.	
<b>Q2.3 What are the professional measures of success for you?</b>	Publication.	
<b>Q4.2 How do you work?</b>	At the moment we're building infrastructure. Our work is not hypothesis-directed. We keep track of interim progress on a wiki.	
<b>Q5.1 In what ways do you interact with simulations in your work?</b>	In the distributed computing aspect of what we're doing, we sometimes don't have access to as much of the distributed resource as we need. Prior to making use of expensive and/or harder to access distributed resources we will try out our computational ideas on individual computers or on smaller subsets of computing nodes, before we move things to larger sets of nodes. I guess that's one way we use simulations. I'm sure my group uses simulation in other ways, but I can't tell you.	

Interview ID=12 24 July 2007	ANSWERS	ANNOTATIONS
<b>Q6.1 Describe how you interact with data in your work</b>	<p>I am a scientist. Scientists collect data. We make observations and try to make inferences from those observations. Everything that a scientist does has to do with the data.</p> <p>We record from the human brain using magnetic resonance imaging [MRI] or electroencephalography. We store the MRI data, which I'm most familiar with, as four-dimensional arrays with <math>X</math>, <math>Y</math>, <math>Z</math> spatial coordinates and <math>T</math> for time. We store the data on large filesystems. Data are archived on CDs, DVDs and tape, and are retained until the media disintegrate.</p> <p>We use visualization tools, we run analyses on the data, and we write about the results so other scientists can learn from us.</p>	
<b>Q7.1 What resources do you use in your work today?</b>	<p>We use electroencephalographic machines and magnetic resonance imaging machines. Those are two kinds of machines that record brain activity: one records electrical activity and the other uses magnetic methods to record blood flow in the brain.</p> <p>As far as other sensors: we use our eyes and ears and response boxes to ask people questions and have them answer questions. So we have additional forms of data, which are not large-scale arrays of numbers. <i>[prompt asking if the interest in Grid technology comes from a desire to scale the number of processors used for analyses]</i></p> <p>I did my PhD thesis in 1980 in distributed computing and I'm quite interested in the fact that many, many kinds of computing processes can be divided up into little parts and spread across lots of computers. As technology has improved over the last twenty-five or thirty years, it has become clear that people develop software that enables:</p> <ul style="list-style-type: none"> <li>- the automatic division of jobs into sub-parts (so that it does not have to be done manually)</li> <li>- the allocation among lots and lots of processors (so you don't have processors waiting for each other to complete partial tasks before completing what they have to do)</li> </ul> <p>This has made it possible to take voluminous sets of data (brain images from MRI scanners or electrographic time series from electroencephalography) and divide them up into little pieces so that the processing can be done faster. This becomes very important and valuable when you have procedures that are highly iterative, and where the different components can be performed without a significant amount of interactions.</p> <p>And so we're taking advantage of that through Grid computing. I didn't realize that this would be available to us now. But through the collaborations with the Grid experts at the University of Chicago, it's become clear that we can serve as a test case for certain kinds of computing that are both interesting to our collaborators and make some of our tasks much more soluble. Particularly those tasks that are recursive, iterative, and other tasks on datasets with spatial components that are non-interacting.</p>	
<b>Q8.1 What software do you currently use in support of your work?</b>	AFNI, Matlab, R, gcc, Java, Java++, C++, IDL, SPSS, SPM and probably hundreds more.	
<b>Q8.6 If the need for new software-based functionality arises in your work how do you acquire it?</b>	We usually beg, borrow and steal from other laboratories around the world. Or we hire programmers.	
<b>Q8.7 How do you share software with others?</b>	We give away source code and we get source code from others.	
<b>Learning about the user's problems</b>		
<b>Q9.1 What challenges do you face today in accomplishing your work-related goals?</b>	Finding good personnel. If you have smart people you can accomplish anything.	



Interview ID=12 24 July 2007	ANSWERS	ANNOTATIONS
<b>Q9.3 What technology-related obstacles do you currently encounter?</b>	<p>I would say there are not particular technology-related obstacles. Everything could be faster. Everything could be easier. The learning curve for some of the software is quite hard. It takes me five months to train people to use some of our software. That's a long time. But I'm not sure it's related to the software/technology. I think it is that some of the concepts are difficult.</p> <p>Basically I'm not one who thinks machines are really critical. I think human beings are critical – smart human beings with creative ideas. I guess one technology-related obstacle is the lack of technologically trained neuroscientists. I've been trying to hire a post-doctoral fellow who's a neuroscientist to do computer modeling for two years. So one technologic obstacle is training more computationally sophisticated biologists.</p>	
<b>Q10.1 Can you think of any work-related tasks that decrease your productivity?</b>	<p>I administer people who probably could talk about some things that are difficult in that realm. I'm sure that certain brain imaging analysis tasks could be done in a more automated way. That's part of what our collaborators are cooking up, in terms of workflows.</p> <p>We do lots of processing by taking these large datasets and finding out what's in them. There are tremendous advantages to be gained by some of the workflows that are being developed for us by the Grid experts at the University of Chicago. Also the possibilities of provenance tracking they talk about would be incredibly valuable. If we took some of our number crunching tasks and managed to create workflows with provenance tracking over the next five or ten years, it would be highly beneficial.</p> <p>But I couldn't give you examples of specific things, other than the various stages of analyzing the brain images, which are done right now by a bunch of mundane csh scripts. Some of that could be done automatically. Also some of the parameters currently need to be set by hand, and maybe some of that could be inferred.</p> <p>It's possible that some of those things could be automated, but I think it would be beneficial for you to talk to some of the people in the laboratory who do that work. I just do most of the conceptual things, like writing grant proposals, working on papers, etc.</p>	
<b>Wrapping-up</b>		
<b>Is there anything you'd like to say to the people who build software for use by people like you?</b>	<p>I think in order to build good software systems for any particular field, one has to gain a significant amount of domain knowledge in that field. If you don't acquire the domain knowledge, you are not going to be able to do as good a job as if you do. The best software we use is written by people who have tried to achieve the kinds of goals that we investigators are trying to achieve. They improved upon existing software that isn't as good.</p> <p>Computer scientists and software engineers develop better code than the scientists themselves. The scientists themselves develop useful code that is less efficient and doesn't make use of valuable computer science techniques like distributed computing. When computer scientists actually try to use existing software to do the tasks they're writing new software for, they develop software that is highly domain-relevant.</p> <p>I don't know whether acquiring this domain knowledge is something that needs to be done by the team itself, or whether they need to have close collaborations. I think the collaboration that we have with the University of Chicago Grid experts may be very valuable in that respect.</p> <p>As neuroscientists we've had our frustrations, and those frustrations are being solved by some of these new approaches. Actually, that's not fair to say. They're not being solved, but we're working towards solutions. We'll see in five or eight or ten years whether we've really had a good effect.</p> <p>So I agree with the University of Chicago Grid experts' approach, which is to be highly collaborative with the domain specialists. I applaud that and think it's the right way to go.</p>	

## D.13 ● We play a strong bridge role in connecting people with technology

Interview ID=13 2 August 2007	ANSWERS	ANNOTATIONS
<i>Pre-interview question: Do you interact directly with Globus software in your work today?</i>	Yes	
<b>Establishing context</b>		
<b>Q1.1 Please provide a one-minute overview of your project</b>	I work on a statewide Grid called TIGRE (the Texas Internet Grid for Research and Education.) It is a project of the High Performance Computing Across Texas organization [ <i>HiPCAT</i> ], which is a collection of ten research and teaching universities throughout the state. We're charged with bringing up Grid applications in three targeted application areas: biosciences and medicine, energy exploration, and air quality modeling. These areas were chosen as examples of the application of Grid technologies to economically useful and interesting topics.	
<b>Q1.2 What is the project's name?</b>	TIGRE	
<b>Q1.3 Which agency funds the project?</b>	The State of Texas	
<b>Q1.4 What field does your project belong to?</b>	We are charged with demonstrating applications in the three areas I mentioned (biosciences and medicine, energy exploration, and air quality modeling). But we are also given the charter of enabling applications for any Grid-ready field. So we work with a wide variety of fields. It's hard to classify. If you had to classify us it would be an Education and Outreach project.	
<b>Q1.5 What is your job type?</b>	Scientist	
<b>Q1.6 How long have you been a &lt;job type&gt;?</b>	Two years	
<b>Learning about discipline-specific goals and approach</b>		
<b>Q2.1 What are the main goals of your project?</b>	Our main goals are to demonstrate applicability of Grid technologies to a wide variety of economically interesting and intellectually useful activities in the state. Our project was an outgrowth of the larger organization that I mentioned, HiPCAT, which consists of the high performance computing centers in the larger institutions through the state. Lately we've been getting involved with further education and outreach. So our basic goal is to take Grid technology deeper into the academic infrastructure than it has gone so far. Our particular project is targeted at the State of Texas, but we're also working with the Open Science Grid, SURA (the Southeastern Universities Research Association) and several regional organizations.	
<b>Q2.2 How will the success of your project be measured?</b>	We have milestones and deliverables for our project that we report on quarterly to the State through its Department of Informational Resources. At the outset of the project we specified certain milestones in terms of numbers of application areas targeted and delivered, and qualitative characteristics of that delivery. For example: the existence of a project-wide scheduler, and the integration of a security architecture. So we're measured primarily by the achievement of these milestones. The project was composed as a demonstration project in the sense that we have to demonstrate the capabilities in these areas. But we're just now transitioning into creating the conditions for a production-scale Grid. Having proved the technology we then would actually go into the business of working with our various regional partners to provide cycles and storage using this infrastructure. So we presented this to the state as a construction project: we would build it as in building a highway. And then operate cooperatively under the standard Grid model once it was in place and working. So the project really has as its goal producing production-quality infrastructure that will take less effort to maintain after it was created. Nevertheless it takes resources to maintain, update, upgrade and keep secure any architecture. So right now we're looking at the various options for keeping that going.	
<b>Q2.3 What are the professional measures of success for you?</b>	My personal job evaluations have hinged largely on the success of this project. So having it meet the milestones, achieve the deadlines, and deliver on the functionality.	

Interview ID=13 2 August 2007	ANSWERS	ANNOTATIONS
<p><b>Q3 What are you investigating?</b></p>	<p>I have spent the majority of my time making sure that I understand and am not blindsided by developments in authorization and authentication technologies. This has led me to be an active participant in the International Grid Trust Federation. And to interact strongly with people in the Shibboleth and related projects of Internet2. Though we don't implement those we certainly are looking to integrate them in the future.</p> <p>We have created an accredited certificate authority here in the state of Texas to serve our needs. Being accredited by the IGTF means being accepted worldwide, which is a high achievement for our project.</p> <p>I also spend time making sure I understand the development of metascheduling technologies.</p> <p>And then whatever remaining time I have (as well as a colleague) is spent engaging with scientists and users and exploring their needs. The old model of engagement was to go to the scientists to find out what they need and then implementing it. We find that approach to be insufficient by far in the modern interconnected world.</p> <p>We spend a lot of our time investigating, studying the needs derived from interviews and work with those people. But then we also go out into the broader context and find out how the needs are met at other locations and in other cyberinfrastructures. Many times we come back to the researcher and have a suggestion or tool that they were unaware of, or simply lacked the ability to implement on their own. So we find that this bridge role is very important in both directions: both for the researcher and for us to learn which way to go.</p> <p>In a nutshell we find that in order to establish the best practices for a given field of study (say, for atmospheric sciences) it is insufficient to interact only with the researcher. You actually have to interact with other providers of tools for the infrastructure. Find out who the major providers of frameworks and middleware tools are and talk to them.</p> <p>Sometimes simply connecting the researcher with an existing virtual organization can be a big benefit. Most of the time the scientists and researchers we talk with don't know anything about Virtual Organizations. And quite often their work doesn't really map well into that paradigm. So we play a strong bridge role in connecting people with technologies in both directions.</p>	

Interview ID=13 2 August 2007	ANSWERS	ANNOTATIONS
<p><b>Q4.1 What is your method for investigating &lt;phenomena&gt;?</b></p>	<p><i>Regarding work on authentication and authorization:</i></p> <p>Prior to joining this project I had spent a fair amount of my time in this area for the Open Science Grid. First in their policy management group and later was one of the charter members of The Americas Grid Policy Management Authority [TAGPMA]. The TAGPMA organization is the North, Central and South America partner of the International Grid Trust Federation.</p> <p>So when we started this project I was aware that the International Grid Trust Federation existed. I was also aware of the value of having an accredited certificate authority, which could then be distributed by various means. It is distributed by all of the major academic Grid projects. It's a component of the Virtual Data Toolkit [VDT], it's adopted by the LHC computing Grid, and EGEE (which is a big player, of course). So we were very early on aware of the value to our users of having high-quality credentials available.</p> <p>This is an area that I think the Globus Toolkit has been extremely weak (perhaps by design so it would develop in the community.) But the standard Globus instructions basically lead the new user into an exercise of SimpleCA and building their own X.509 capabilities. These instructions are essentially useless in the context of any large-scale deployment where you actually have to trust each other and you need to build a foundation for trust.</p> <p>So we knew early on that there would be a high value, at least for the State of Texas, in having a central certificate authority. When we approached our participants we found that one of them (the Texas Advanced Computing Center) was also a TeraGrid partner. They already had a certificate authority that was already being distributed by one subset (the Virtual Data Toolkit), though it wasn't yet accredited. So it seemed to us the best path was to take that as a starting point. Go forward in the context of TAGPMA. Do all the necessary work to make that certificate authority high quality and accepted. By doing that we would then be able to offer all of our participants in the State of Texas a Grid credential that would be accepted worldwide.</p> <p>So this seemed to be a very efficient path for us. We are actually surprised that other projects have not followed that path. Having an accredited CA really is a great barrier dropper for any subsequent partnerships and collaborations you want to make with other organizations. So it has been a great benefit to us.</p> <p>It's involved some education of our users. Everyone wants to know, "Why can't I just log on with my username and password?" And this is, of course, an area that the whole Grid infrastructure is struggling with. But we took the point of view that the most efficient path forward for us would be to get the highest quality Grid credentials we could into the hands of our users. And we found that to be a very achievable goal.</p> <p><i>[prompt asking for further description of the barriers lowered by having an accredited CA]</i></p> <p>When you have an X.509 credential – or a strong credential of any kind – the real question is, "What is the trust anchor of that credential?" Everyone encounters this these days by going to websites that have put up self-signed certificates, or certificates signed by CAs at institutions that have no commonly distributed trust anchor. So the International Grid Trust Federation exists for the purpose of accrediting CAs to a common standard, so that identity-proofing, operation of the CA, etc. is done in accordance to commonly-agreed profiles. And then distributing that set of accredited CAs to the large-scale Grid projects worldwide. This was an outgrowth of the OSG Computing Grid initially, but now has become a cornerstone of many, many large-scale Grid collaborations.</p> <p>So if you have an IGTF-accredited CA that's enough, because other large-scale projects throughout the world get these sets of trust anchors. So they know whether or not to trust the credentials of your CA, and on what basis. They know that you have, for example, been in-person identity-proofed by someone in the chain. They also know the CA is run in a method that does not allow a graduate student to walk in and issue their own certificates.</p> <p style="text-align: right;"><i>[continued next page]</i></p>	

Interview ID=13 2 August 2007	ANSWERS	ANNOTATIONS
<p><b>Q4.1 What is your method for investigating &lt;phenomena&gt;?</b> [continued]</p>	<p>So since these things are widely distributed and commonly accepted, it's very easy to start a virtual organization. We always make the point that authentication is not authorization, but it's a starting point. They can then know the quality of the Grid credentials coming in and use that as a basis for signing membership to the Virtual Organization. It becomes barrier lowering because I can accept a certificate issued in Czechoslovakia, for instance. I'm not going to accept just any certificate – only those that have come from an IGTF-accredited CA.</p> <p>This is the basis of many large-scale Grid projects such as caGrid, OSG, PRAGMA, the European Grid projects, etc. So this is an area that has emerged out of community practice to solve the problem of distribution of trust anchors. You can tell I do actually spend a lot of my time on this work ☺</p> <p><i>Regarding user engagement work:</i></p> <p>This is one of the most fun parts of my job. It's really a payoff for the drudgery, and is the thing that attracted me to the project in the first place. It involves the challenge of engaging with people in fields very different from my own – I'm a particle physicist by background. I don't know anything about air quality modeling or petroleum engineering. In fact, I knew this would be such a demanding thing that when we hired our second person for this project, we essentially made this his full-time job. He's done a terrific job at it.</p> <p>So you find yourself one day going to a doctor's office and talking to a radiation oncologist about application of Grid computing to radiotherapy dose estimation, and being very seriously engaged in understanding the issues there. And then that afternoon you can be going and talking to an atmospheric scientist who needs access to an earth modeling framework from the National Center for Atmospheric Research. So it has both the advantage and challenge of bringing a lot of variety to your job.</p> <p>So we have to search for efficient methods for doing this. We can't do people's science, and we make that point to them explicitly. But we can look for ways in which connecting them to Grid infrastructure is appropriate. Many times a scientist is perfectly happy with a desktop if they can just get the data they need. But many other times they need access to TeraGrid, or they really need free computing somehow and they don't know how to go about getting that. We might convince them that there are ways, and that maybe they can even join organizations in their discipline that are devoted to solving that problem. This is many times a revelation to them.</p> <p>The promise of Grid computing is access to shared resources, but there is a big gap between here and there, and we work to fill that gap. We didn't set out to do that. We really set out to fulfill the charge to demonstrate applications in these areas that were chosen by our steering committee. But to do that we found we had to go through the process of repeatedly visiting scientists, interviewing them (much as you're interviewing me now), finding out their needs. But then not stopping there: actually going out and talking to other supercomputing centers, other Grid projects and saying, "Hey – how do you serve the needs of this subfield?"</p> <p>And then we go back to the scientists or researcher with our findings. Many times not only are they very interested in participating, they also find themselves involved in something they felt they never could have done on their own. So based on our profession interactions with other large-scale centers, we can quite often implement a framework that would have been beyond the reach of a researcher left to his own devices. That is very satisfying.</p> <p>It isn't really a Grid topic per se, but it intersects a lot with the whole idea of shared resources. And many times we find that there are specific Grid projects in particular areas. Some of them, of course, are extremely well developed. Some of them are to the point that all we really need to do is point the scientist in that direction. But most of the time there is a considerable gap to be filled even to make that possible.</p>	

Interview ID=13 2 August 2007	ANSWERS	ANNOTATIONS
<p><b>Q4.2 How do you work?</b></p>	<p>[Prompt asking for more detail about how the bridge to a user is built]</p> <p>We sometimes will write a little bit of code for them. We always strongly emphasize that they have to do their own research, but we're here to do what we can to help them in a connective role.</p> <p>One of the easiest engagements for us is when the scientists have a running application but they don't know how to run it on the Grid. In fact I have to say for many middleware developers, they often think that's the only thing left to do after you install their middleware. We have to gently point out that application porting is a mere starting point that occupies a small fraction of our time, because it is in fact so easy. It is by no means the end of the story.</p> <p>If we just have someone who has a computationally intensive application that needs a lot of CPU hours, then we hook him up with our Grid, TeraGrid or OSG, and we're done. But very often there's a great deal of social interaction to be done.</p> <p>There's a great deal of organization building.</p> <p>Sometimes there are terribly, terribly intrinsic issues to deal with. For example in the petroleum engineering field we find they have extremely powerful, well-developed expensive codes that the providers are happy to give you almost free academic licenses for. Really shocking how open they are with their code – you can download it almost like you would a piece of shareware. Extensively developed code. But then if you turn around to a particular researcher and say, "Ok, let's put this on the Grid." You find that they stop like a mule at a door because they won't let go of their data.</p> <p>It's the data that's important in that field. They are highly proprietary, having to do with detailed field measurements of oil-bearing strata. They are absolutely unwilling to let that part from what their perception of what a secure space is. So we have to spend a lot of our time working with them to assure them about data security and implementing tools to make sure that they always feel in control (to the degree that they're willing to do it at all.)</p> <p>That's a very difficult problem to solve. We're still struggling with it – we haven't completely solved it. Part of the problem is lack of familiarity with the tools. The same people who are probably logging on with cleartext passwords to POP email accounts react with great skepticism when you approach them with an absolutely locked down X.509-secured, strong cryptography solution for controlling access to their data.</p> <p>So it's partly familiarity with the tools, and – who knows – maybe in a few years it will be easier. I'm not optimistic on that score. I think we're going to have to go and do some implementations that then become standard practice in that field because they're better. That will take detailed work that we're not chartered to do under this grant. We can work with people and explain the issues, but at a certain point you have to stop and move on to the next person (hopefully not without hitting the milestone.)</p> <p>In other areas that you might think would be difficult from that point of view, such as medical applications, we encountered less resistance. We're working in the field of radiation modeling for cancer therapy, and there's a proton accelerator here in the state that has been built by the M. D. Anderson Cancer Center. This large-scale \$120 million facility has huge modeling needs. We thought it would be a very hard problem to move the medical data around. But we found that there are tools in the caGrid software stack that are not only well-suited, they're actually explicitly written for the purpose of moving medical image data around with high security using Grid tools.</p> <p>So that is an area where we thought we'd need to do a lot of development. Instead we found a complete working infrastructure that we just didn't know about until someone asked us. It's fascinating, I have to say. I really am having more fun at this than almost anything I've ever done up to now.</p>	
<p><b>Q4.5 How do you document your results?</b></p>	<p>We issue quarterly reports.</p> <p>As far as documenting our interactions with users, my colleague does keep detailed notes. Interactions are informally documented. We do have a phone site with use cases gathered, though the site is gathering a little bit of dust now. Driven partly by milestone pressures we simply move on to trying to get the implementation out there as soon as we understand it. There's certainly material there that could be gathered by a dedicated person who had an interest in that stuff.</p> <p>We also keep notes of our weekly developers meetings and distribute them to our steering committee. They seem satisfied with those, so that helps keep people informed.</p>	

Interview ID=13 2 August 2007	ANSWERS	ANNOTATIONS
<p><b>Q5.1 In what ways do you interact with simulations in your work?</b></p>	<p>A large percentage of our users are simulation users since we exist to make connectivity happen to large-scale computing. Not all – we have people analyzing data from the field. But for instance, our weather modeling application is almost entirely simulation (driven a little bit by experiment.)</p> <p>We have distinct classes:</p> <ul style="list-style-type: none"> <li>- the massively parallel applications</li> <li>- the ones that can run in a simple method on clusters (people often say embarrassingly parallel, I sometimes say stupidly parallel)</li> <li>- and a comparatively small batch of applications that are suited to cycle-scavenging Grids</li> </ul> <p>Our favorite venue is when we can find someone who has an embarrassingly parallel application that simply needs cycles or data transfer (we're getting pretty good at data transfer too.) We simply implement that. We have a project-wide scheduler. That's where a lot of our work is going now: getting people's jobs to the point where their workflow models are tuned up to handle their workflows. Much of our work right now is in that stage of it. We have a working application, we need it to run simply and reliably on multiple resources, and we're trying to leverage our project-wide metascheduler for that purpose.</p> <p>Science work is really often about workflows. You have a whole sequence of operations you want to do to the data – it isn't just one program. Maybe in fact it involves analysis plus simulation, or maybe it's simulation followed by analysis with a tool that would also be applicable to analysis of real data, or maybe it's pure simulation. But quite often there are a number of steps that need to be carried out. Scientists build elaborate, and I have to say fairly rickety, methods of doing these things, sometimes involving a lot of hand operations on the terminal to do their data processing.</p> <p>So we always have to make a judgment call how much of the workflow to try to put into tools designed for workflows. You know – do this, and that, and the other thing – and wait for these jobs to finish and then process them all. So we have some tools that are doing that, and we're beginning to explore in that direction now.</p> <p>We have ten developers, which is a lot if you think about it. So we spend our programming time making the infrastructure work so that the science workflow proceeds fairly easily:</p> <ul style="list-style-type: none"> <li>- engage the scientist</li> <li>- taking that scientist's work as an example</li> <li>- then adapting our portal or submission mechanisms so the example can be carried out easily</li> </ul> <p>It isn't a perfect approach. Like almost all science-driven work you end up solving the set of problems that are in front of you. Everyone wants the general-purpose tool, but that's not so easy to produce.</p>	

Interview ID=13 2 August 2007	ANSWERS	ANNOTATIONS
<b>Q5.5 By what mechanisms is access to your simulations controlled?</b>	<p>You must have a credential from an IGTF source. We issue credentials through the Texas Advanced Computing Center CA to anyone in our project who needs one. We also have a Virtual Organization Registration Management Service (VOMRS). This is a front-end to a Virtual Organization Management Server, used to generate gridmap files, or to control access with more sophisticated tools (such as the GUMS and PRIMA tools used by Open Science Grid.) Some of our sites are Open Science Grid sites, others are not.</p> <p>We distribute a compact stack of tools based on the Web services-only implementation of the Globus Toolkit. It is based on the Virtual Data Toolkit (VDT) – a very small subset of it compared to the Open Science Grid. It is supplemented with small-scale tools that we provide. So it's a very compact stack. It can go in, certainly in an afternoon if you've never done it before, and in fifteen minutes if you have done it before. It has a client stack. And we're supplementing that now with a non-VDT based simple Java client that lets people interact with the project metascheduler.</p> <p>We've not yet found it necessary to create a separate VO for each application. We do, on the other hand, map accounts uniquely so that we don't have to worry about people all landing in the same account. The philosophy we take, which I think is very common with Grid applications, is that we control the authentication of a person. We make sure they have an established credential. But it's up to the local resource owner to map them. We simply ask them to map these people uniquely. So in principle it would be possible for a resource provider to map everybody to a generic Grid account. We discourage that. We ask them to make a separate entry for every DN if they are going to use simple gridmap files. We do have several sites (including my own) that have far more sophisticated infrastructures where this mapping is done dynamically, persistently and automatically.</p> <p>So if you say to me, "Hey, I want to join TIGRE." When we get that approved, you send me your DN (or I point you to the webpage and you register your certificate.) Then when you come on our site you will get mapped on the fly, but persistently, from then on to the same account drawn from a pool. It's a generically named set of accounts, but you'll get your own.</p> <p>So this is a modest step towards security in the sense that we are operating TIGRE as kind of big statewide VO. But we don't want people to have to worry about sharing each other's stuff. If necessary (and we have done this in some cases) we can control access to certain resources (storage, for example) within the VO. We can put these locally generated mappings into groups such that only people who are members of a certain subgroup can actually interact with the resource. In fact this is one of my pet peeves: we don't really create a user account. We create an account mapping. But it gets implemented as unique accounts quite often.</p>	
<b>Q6.1 Describe how you interact with data in your work</b>	<p>The easy application area to mention, of course, is high-energy physics. The field is very data hungry, whether for simulations or real data, and both on the input side as well as the output side.</p> <p>To pick a slightly more fair example:</p> <p>When engaging the air quality modelers, and we found that they were a highly unique and diverse lot, each with their own science application. But they all needed certain types of model output. And those that cared about real-time modeling needed that to be very current. We also found that in this field they had implemented a method of distribution which could get you the data with a fairly straightforward tool to install.</p> <p>What was unclear was how to map that to making the sets of data available for Grid processing. If you want to be able to opportunistically take advantage of cycles that are available all over the state, you don't know ahead of time where to move the data.</p> <p>So we implemented a method that's running at three places now and we hope to bring it up on a fourth soon. Our approach allows us to have a non-Grid data distribution method to Grid-enabled storage. That way when you run your simulation job or analysis job and you want to consume a certain dataset, you can pretty much count on it being available in a timely fashion at each of our Grid resources. So this is a case where we ended up using a hybrid approach of Grid and non-Grid methods.</p>	



Interview ID=13 2 August 2007	ANSWERS	ANNOTATIONS
<b>Q7.1 What resources do you use in your work today?</b>	<p>TIGRE is primarily a compute Grid right now. A defect in the original project design is that it did not explicitly address storage. We found that storage and data transfer are in strong demand to implement our targeted application goals. So we've done something for now, and we'll probably go back to our parent organization and recommend that we do more.</p> <p>By "something" I mean that we've implemented some data transfer mechanisms for particular subfields (like atmospheric modeling) that move data around, sometimes by non-Grid methods. At my home institution we've also implemented some Grid-aware data transfer tools, and we're looking to encourage other sites within our project to adopt them. So we probably didn't do a good enough job of spec-ing the data transfer parts of the application initially.</p> <p>In terms of seeing an overview of resources you can just go online to <a href="http://tigreportal.hipcat.net/">http://tigreportal.hipcat.net/</a> and get an instantaneous readout. If I were to do that now I'd probably see seven or eight sites, with a variety of compute resources (from very small to very large).</p> <p>In some cases the resources are owned by the application groups. More commonly they are provided out of university-based central resources, since this whole project was started by an organization composed of large-scale computing centers at several different research universities. But there are several interesting examples, including our first TeraGrid application demonstration. It was actually probably too easy:</p> <ul style="list-style-type: none"> <li>- we went to a scientist who was highly computationally bound working on a small set of machines</li> <li>- we created a portal environment to encapsulate his workflow</li> <li>- and in the two week demo got that scientist far more compute cycles than he had been able accumulate to date</li> </ul> <p>He said he got publish-able work out of it. We were very happy that he told our steering committee that. So we still use that example.</p> <p>That scientist has since gone on to get research funding to buy clusters, and then contribute those clusters to our project. So that's an example of people bringing their own resources. That's comparatively rare now.</p> <p>Here at my university we do actually buy machines based on that model. We will talk to people and say, "Look. You don't need to run this in your basement, subjecting your grad students to all that fan noise. You can put your machines in our machine room, or better yet buy into the next cluster we're going to buy. We'll give you your fraction of the resources on a guaranteed basis, and you contribute all your idle cycles to the common pool."</p> <p>And of course, you know the rest. That's the Grid model. So we implement it by getting people to buy in – actually to send us money to help buy the next cluster. To make that attractive to them, we in the IT division will typically chip in also just to get more cycles. So everybody actually does get more than they put in, which always is a good trick.</p> <p>To guarantee their time we have queue structures, which give them priority over their fraction of the resources. So if they asserted on their fraction of the cluster they can bump other jobs. But most of the time they don't. And so there are idle cycles. We have one person who hasn't paid at all, but runs around using up all the idle ones. So it's the usual mix of users you get in Grid projects.</p>	
<b>Q7.4 How do you locate available resources for use in your work?</b>	<p>Our project has a portal, and the portal has some sensors. If you're not careful you will get me on an hour-long diatribe about the monitoring penchants that people have. I usually joke that there are only two types of Grid applications: there are schedulers and there is monitoring software. That's of course not true, we try to make sure there's science software too. But people have a tendency to write and rewrite monitoring applications as if forgetting all the enormous amount of work done on this topic by people before them. I find this infuriating.</p> <p>So we actually implemented a very lightweight monitoring layer for TIGRE that simply encapsulates the basics on availability of resources. It isn't perfect. It's MDS-based, and we're hoping for developments in this field that will make our life easier.</p>	

Interview ID=13 2 August 2007	ANSWERS	ANNOTATIONS
<b>Q7.5 What types of information do you need to know about a resource in order to determine if it is suitable for your work?</b>	<p>Typically people will want to know the architecture, the number of processors per core, for example.</p> <p>Sometimes they'll want to know the interconnect architecture. The massively parallel applications will only want to submit to shared memory machines or, more commonly, InfiniBand-connected multi-core clusters with very high interconnect bandwidth.</p> <p>We use the simply parallel (or embarrassingly parallel) applications as fillers, so they simply know if their application will run.</p> <p>We're trying to get people into a mode where we can compile and go, so you can even run on windows if you happen to be encountering a cycle-scavenging Grid. I should say we have integrated our cycle-scavenging Grids (at least at the portal level) into the pool that we make available.</p> <p>And of course our simple monitoring approach makes it a little hard to get everything we want to the users. So there's a certain amount of manual interaction that takes place.</p>	
<b>Learning about the user's problems</b>		
<b>Q9.1 What challenges do you face today in accomplishing your work-related goals?</b>	<p>Well, the old joke in academics is the fact that we only really care about parking ☺</p> <p>There's a tremendous variety in response on the part of individuals when you go and try to work with them and talk with them. They have a tremendous range of background. They sort of fall into three categories, and I'll talk about them in increasing percentage of the frequency we encounter them.</p> <p>The first category includes people who are perfectly ready for Grid technologies for access to highly distributed computational resources. Those who say, "I've been waiting for you to come along." They have an application that needs cycles or needs to move data or somehow needs a cluster.</p> <p>Dealing with these type of folks is so easy. We enjoy it so much. We give them credentials, we adapt their application. Ours is Web services only, so we wrap it in Web services submission script. Maybe we even build them a small service. We drop the tech in and they're happy, and we don't hear from them again except if they later want a slight tweak. And often we don't hear back for that because they figure it out. That's great. That's a few percent of the people we encounter.</p> <p>The next category is folks of the sort that I mentioned earlier. We have to spend a lot of time talking with them to find out their needs. They aren't really completely sure of the full set of cyberinfrastructure tools they need to accomplish their goals. Sometimes we can connect them with best practices in their field by simply having them talk with colleagues. Other times we do a little building for them. So with some effort we can get them – and maybe the whole group of people – running on the Grid. Maybe we even get them to form something that will develop along the lines of a Virtual Organization. So one third of our time is spent dealing with all of the above categories of people.</p> <p>An embarrassing fraction of the time (I'd say more than half of our time – and I'm really stuck on this) they drop the whole thing at your feet. I kid you not. They say, "Oh good. You're here. You can do this now." And you have to say, "No! That's not what we're here for. We're happy to help, but you have to keep doing your science."</p> <p>The group is populated by a mix:</p> <ul style="list-style-type: none"> <li>- people who are just getting started</li> <li>- people who haven't figured it out yet</li> <li>- people who really don't know how to use the tools</li> <li>- the senior professor who is now going to ask you from this day on how to log on to his email (I kid you not!)</li> </ul> <p>So more than half of the time when we try to work with people who ostensibly are researchers and scientists in their field. They say, "You're so much more qualified than I am to do this that it's hopeless for me to do more. I have not figured out how to deal with this problem.</p> <p>Education is clearly one of the approaches. Having people learn more about their own fields maybe, but boy that's a hard one. And it really happens. They can't distinguish between their domain science and your infrastructure: Middleware, Grid logins, infrastructure, clusters, their particular science application. Maybe they've even been given a science application from someone they're working with. They can't distinguish among them – it's all "The Computer" to them. Almost every day I think about this problem.</p> <p>Obviously what you do with that (and it's not a good solution) is you ignore these people. Well – you don't ignore them – but there's not a lot you can practically do. So it's a big problem. I stew about it a lot.</p>	

Interview ID=13 2 August 2007	ANSWERS	ANNOTATIONS
<b>Q9.3 What technology-related obstacles do you currently encounter?</b>	<p>Educating my home institution about the Grid infrastructure itself. I spend a fair amount of my time making sure that things we need to implement to make Grids work aren't going to get tripped over by the folks who do security. This is, of course, a very common theme. We spend a lot of our time making sure there's adequate communication there so that nobody shuts down my Grid servers because they don't have username/passwords expiring every ninety days (we don't use username/passwords). So we run interference at the institutional level.</p> <p>There's tremendous chaos in the identity management area. Everybody thinks they're in charge of identity management. Everybody! It's like when I first started teaching, I went home and told my wife, "Everyone thinks they're my boss: students, the dean, my funding agency." The problem is that none of them are wrong.</p> <p>Certainly your university thinks they're in control of all of the computer identities associated with you. The Virtual Organizations that you work with all want control. EDUCAUSE and Internet2 think they've got a good scheme. TeraGrid has its own thing and they want to be able to decide who in your university can log on to their resource and they're not interested in your opinion about it.</p> <p>So identity management is a mess. I advise strongly that you stay the heck away from it. ☺ No, of course you can't. I see some signs of progress:</p> <ul style="list-style-type: none"> <li>- authorization infrastructures are developing</li> <li>- the fact that there is an IGTF</li> <li>- the fact that there are coherent ways of getting trust anchors</li> </ul> <p>Another technology-related obstacle I encounter is the issue of coherence of a given set of software. It is not possible to implement just one piece. Even all of the Globus Toolkit clearly is one piece. So the technology obstacles are ones of keeping the different components into a compatible state. It has driven our project to the approach we've taken: one of leveraging an existing stack but making it simpler.</p> <p>So then when we go to add pieces like our metascheduler, if that falls out of synchronization with some features of the Globus Toolkit it can cause us problems. Nobody owns these problems. We have to solve them because they're our set of choices of what to include. So we try to keep that set fairly simple in order to achieve our mission of education and outreach.</p> <p>There are other organizations that would have different challenges because they are trying to do far more sophisticated science. We're not in that category but we have to work with them. So version consistency, standardization – that's clearly the name of the game here. The pace of change of some of the software is dizzying.</p>	
<b>Q9.4 By contrast, can you provide examples of technologies you find very useful today?</b>	<p>We'd be lost without X.509 authentication. That just solves a whole raft of problems.</p> <p>We really like Web services because it lets us build tools that are better suited to scientific workflows. They really are services – the steps that we need to accomplish. That could be better if we had user-pluggable services. We're not there yet.</p> <p>Communication technologies. We use wikis and other shared access documentation technologies.</p>	
<b>Learning about the Globus user experience</b>		
<b>Q11.1 Which Globus data components do you directly interact with in your work today?</b>	<p>My role in the project is a little bit higher level than that. The basic pieces of the Globus Toolkit that we use are the Web services and GridFTP. We are working towards being able to use toolkits for those, so we look at Introduce and other packages. I don't personally interact with RFT or MDS or RLS. But I think there are people on our team who do spend a lot of time on that. In particular MDS is something we are looking at to improve our monitoring.</p> <p><i>[prompt asking if the project has built custom Web services for its use]</i></p> <p>Well we use GRAM4 for our job submission. That's where we started two years ago. We've had a little bit of struggle along the way – of course it's been improved. We really have never regretted that decision. We will support GRAM2 job submissions to our batch-oriented resources on request, but we've never had a request that couldn't be satisfied by teaching the person how to submit via GRAM4. So there's that level.</p> <p>And then we do have some custom Web services, and we're trying to train our developers in tools for writing more. For example the Eclipse-based tools like Introduce and so forth.</p>	

Interview ID=13 2 August 2007	ANSWERS	ANNOTATIONS
<b>Q12.1 Which Globus security components do you directly interact with in your work today?</b>	<p>We use GSI certificates obviously. We do actually support MyProxy and use it in some cases. It is actually a component that is relied on in our metascheduler GRMS. We are looking at GridWay. We know that there is a great deal of effort being put into incorporating GridWay into the Globus Toolkit. And we have a couple of people in the project who are passionately following it because they prefer GridWay.</p> <p>The most sophisticated authorization pieces that we have are based on Open Science Grid components (GUMS and PRIMA). I know that there's an effort to design improvements to that authorization framework. We're tracking that but we're not participating in that right now.</p> <p>So the European Grid and Open Science Grid use different authorization technologies, and they've been working with folks on the ANL/UC Globus team to design a common authorization framework.</p> <p>In TIGRE we don't use the CAS. I'm not saying we couldn't – we just haven't found a need for it yet.</p>	

Interview ID=13 2 August 2007	ANSWERS	ANNOTATIONS
<p><b>Q15.1 Which Globus common runtime components do you directly interact with in your work today?</b></p>	<p><i>[prompt asking for more details about the client software stack TIGRE has produced]</i></p> <p>From the top-level page of <a href="http://tigreportal.hipcat.net">http://tigreportal.hipcat.net</a> you can click on the link marked “Documentation” and be taken to a page containing a link marked “Client Software Stack”. Also from the top-level page, near the “Documentation” link there is a link marked “Administrators.” This will take you to a page with a link to the “Server Software Stack.” It’s a one-page set of instructions; we’ve developed a tutorial based around this. It implements a small subset of the Virtual Data Toolkit. It’s basically Globus Toolkit 4, GSI-OpenSSH, a couple of file transfer tools, Condor-G and MyProxy (for handling Grid credentials and doing job submission.) We are working now with the GRMS client, which is even a lighter-weight layer for the client piece. The goal is to drop a pure Java client onto someone’s (MS Windows, MAC or Linux) desktop and have them able to submit to our common scheduler without installing the Globus Toolkit. Our CA is accredited, so it comes with the VDT when you install their CA certificates package. So we don’t have to do any special steps to get authentication working.</p> <p>We could script these client and server software stack installations more tightly than we have. We could make them really one-button, no-interaction installs. But we found for education purposes it’s actually valuable for people to see the pieces as they go in. So early on we just made the set of instructions, which are those links I mentioned, that tell you how to answer some questions. We found it more valuable to let the new administrators answer the questions than to make that invisible to them. It’s always a judgment call on how much should be magic. We might in the future make that all completely automatic.</p> <p><i>[prompt asking if the users install the software stacks on their own resources]</i></p> <p>To participate in TIGRE we ask that you either install our stack or provide a work-alike set. For example, my home institution is an Open Science Grid site, and we so have several resources that implement Open Science Grid stack. OSG is a superset of the TIGRE stack (with the exception of GSI-OpenSSH, which I drop that in manually.) If you install the minimal stack that is enough to be a member of TIGRE. But unlike most other Grid projects, this stack is built so you don’t have to join anything for it to work.</p> <p>We actually did the exercise of taking the Globus QuickStart Guide and recasting it using the TIGRE stack (in place of the manual compilation and installation of GT.) We simply substituted those steps with installing our stack. It condensed the tutorial from its present ~twelve pages down to three or four pages. And it completely wiped out the section on X.509 that you otherwise have to struggle through, because it’s built-in here. So we actually have a tutorial that is based on your tutorial but simplified because we rely on our stack to provide the components. So our start is quicker ☺.</p> <p>We really took the philosophy that if you really wanted to go and use this without ever talking to us, that should be possible. And we think this is a departure from many other Grids. Many other Grids take you in detail through registering with them and all this other junk. And we thought, “Let’s just leave it so that you could put this in and call up another friend and be off and running.” We’ve had people install this and send us email – one person from Australia.</p>	

Interview ID=13 2 August 2007	ANSWERS	ANNOTATIONS
<p><b>Q15.1 Which Globus common runtime components do you directly interact with in your work today?</b> [continued]</p>	<p>And then we've gone through the exercise of casting a version of our stack in other organization's terms, but otherwise the same stack. And we've had success at several institutions that have previously struggled through trying to do it manually. We've had some pretty complimentary responses.</p> <p>When you install our stack you have a full Web services (also pre-Web services if you choose to turn it on) implementation. You have the ability to set up the ends you want to have access. You can log in by GSI-openSSH to explore the environment. You can submit jobs, you can move data around. You can store proxies with MyProxy. And now with the new central scheduler you can choose to submit to TIGRE. And if you're an authorized user of TIGRE, you can submit to the central resources with the client tools (even the lightweight ones). If you don't want to join TIGRE you can simply figure out who it is you want to work with and you're off and running.</p> <p>We don't make you use TIGRE to use the stack. And it's all based on other people's work anyway: the Virtual Data Toolkit folks, Pacman and Globus. What is it that Picasso said? "Good artists borrow and great artists steal."</p> <p>[prompt asking for the number of users of the interviewee's software stack]</p> <p>There are a handful of non-TIGRE people who have downloaded and installed the stack. We're actually trying to use this as an outreach component. We think this is a great tool to take to a place that's never done any kind of cyberinfrastructure before. Because if you have any experience it takes fifteen minutes to install, as long as it's one of the supported platforms (and that's a fairly broad list.) So with a certain amount of preparation we can go into an institution that's never done this before and bring them up on either a small-scale resource or a client stack that gives them access to other resources.</p> <p>If you would put in the client stack there's nothing except Virtual Organization membership that would prevent you from submitting to TeraGrid or Open Science Grid. In fact, that was how it was selected: the minimal components necessary to interact with those organizations.</p> <p>Where I think we have a gap right now: we talked about these accounts and the authorization being left to the local system? It really is time for TIGRE to face up to the need for a common policy. To simply say, "If you are an organization participating in TIGRE you must subscribe to our central VOMS server, or the equivalent through our portal. We're working on that. It's complicated by the fact that some of our institutions have adopted Shibboleth and others haven't, so it goes back to the whole authorization challenge.</p>	
<p><b>Q16 Why do you use &lt;component&gt; instead of an alternative technology?</b></p>	<p>First I'll answer the question, "Why did we build a Grid based on Globus?" : It's certainly the dominant technology. It is compatible with a lot of the larger projects we want to interact with.</p> <p>GRAM4: we didn't see any particular need to support pre-Web services. And so as an experiment we tried just GRAM4. It seemed to have advantages in terms of being stateful and allowing us to interact more closely with our potential services. Certainly from the point of view of just job submission it was relatively trivial to adopt.</p> <p>The challenges that are associated with using Globus (many people feel it's complex and hard to implement) were greatly lessened by our decision to adopt the VDT-based installs. We've worked very closely with the VDT team, including with security updates (a couple of which actually made it back to the Globus repository.) So we've interacted very closely. The integration of the GSI-OpenSSH component – I think we can take some credit for making that easier in the VDT (as opposed to a standalone) install.</p> <p>So we chose it because it seemed to be the dominant technology for Grid services. And we were interested in going the Web services direction. We haven't found a reason to revisit that decision.</p>	

Interview ID=13 2 August 2007	ANSWERS	ANNOTATIONS
<b>Q17 What are the major challenges you face using &lt;component&gt; today?</b>	<p>The only specific thing that I would add is the lack of a GUI-based client for GridFTP is a barrier to some of our users. We've tried this CGFTP thing that's coming out of China Grid in some highly incomplete state. That satisfied a couple of our users. Some of our users like GUIs, and they don't like using the commandline to move things around.</p> <p>We've certainly hit challenges with the road to Web services: writing Java submission scripts that can be submitted through Web services. The XML – that's the place we sit down with the user. We'll just do it with them because many people look at XML and, well, it doesn't fit their worldview. But it's utterly trivial to sit and work with them, saying, "This is how you specify the batch queue." So I really don't want to make a big strong pitch for GUI-based tools, but certainly in the area of data transfer that would make our life easier. So if I could get a hold of the developers of CGFTP and say, "Make this real or make this go away." I'd do that.</p>	
<b>Wrapping-up</b>		
<b>Is there anything you'd like to say to the people who build software for use by people like you?</b>	Thanks. But don't stop.	

## D.14 I start with microbenchmarks and follow-up with real applications

Interview ID=14 22 August 2007	ANSWERS	ANNOTATIONS
<i>Pre-interview question: Do you interact directly with Globus software in your work today?</i>	yes	
<b>Establishing context</b>		
<b>Q1.1 Please provide a one-minute overview of your project</b>	<p>My research area is in providing an implementation of the MPI standard specifically targeted for computational grids, which means running on two or more machines (typically large SMPs, sometimes called supercomputers.) So we provide the MPI library. There are very specific issues in implementations of MPI that only surface when you're trying to couple two or more machines together, and that's where we spend all of our time thinking. We have such an implementation; the current release is called MPICH-G2.</p> <p>With this implementation, we seek scientists or engineers with HPC applications to collaborate with. It's not a strict requirement, but almost all of our collaborators already have an MPI application that they work with. They could write one from scratch – we just haven't encountered one yet.</p> <p>The two requirements for us to engage with these people are:</p> <ol style="list-style-type: none"> <li>1) they must be domain experts in the application area</li> <li>2) the problem addressed by the application is one that cannot be solved on any single computational resource.</li> </ol> <p>So if you can run just fine on a single machine, well, then you should continue to run just fine on your single machine and good luck to you there. But if you can't – if your problem is bigger and literally can't fit on the largest machine you can get your hands on, then that's a good reason to come to us. We'll work with you to try and port your MPI application to our MPI implementation and deal with the issues.</p> <p>One such example application, a blood flow application, is based on a math library called NekTar [<a href="http://www.cfm.brown.edu/crunch/nekstar.html">http://www.cfm.brown.edu/crunch/nekstar.html</a>]. A mathematician at Brown University and his team are the developers. Two years ago he was visiting my home institution and came down to my office to meet me, and we talked about his application and what I do. We felt that there might be a good match and started working together. It turned out that it was a good match. Using his coding skills we were able to solve problems that had not yet been solved before. In fact, he tells me that ours is the first cross-site simulation ever run on the TeraGrid.</p>	
<b>Q1.2 What is the project's name?</b>	MPICH-G2	
<b>Q1.3 Which agency funds the project?</b>	National Science Foundation	
<b>Q1.4 What field does your project belong to?</b>	I would classify the work that we do to refine the tool as Computer Science, but that's not the end of what we do. We also work with application groups (Meteorology, Cosmology, Hydrology, Biology, etc.) that have nothing to do with computer science	
<b>Q1.5 What is your job type?</b>	Professor of Computer Science and Principal Investigator	
<b>Q1.6 How long have you been a &lt;job type&gt;?</b>	8 years	
<b>Learning about discipline-specific goals and approach</b>		
<b>Q2.1 What are the main goals of your project?</b>	We are creating a tool called MPICH-G2, which is a Grid MPI, that can be used to solve computational problems that cannot otherwise be solved. Every time that we have solved a problem that no one was able to do before, we are very happy because MPICH-G2 is enabling technology. That's our goal. And we keep pushing that envelope farther and farther out – as far as we can.	



Interview ID=14 22 August 2007	ANSWERS	ANNOTATIONS
<b>Q2.2 How will the success of your project be measured?</b>	It will be measured by me as how many such problems we solve. Then if we need a finer resolution as a personal thing: it'd be nice to know that it's applying to help humanity. Not necessarily in a direct sense, but for example when I can do something for the medical field, that just personally makes me happy. Not to slam anyone who's trying to build a better bridge or something like that, but I personally get a bigger zing if the stuff that I'm doing helps medicine. That type of stuff.	
<b>Q2.3 What are the professional measures of success for you?</b>	Well, my sponsors and the funding agencies actually hold a pretty similar measurement of success. They would like to see that the product that they funded (the development of MPICH-G, of these Grid MPIs) is finished and actually have an implementation, which we've done. But that's a minimal requirement. Beyond that they really are interested in the same thing I'm interested in: this is a tool – it is not an end. Tell me what you've done and what successes you've had in using that tool. And those are one and the same thing that I've described before. We use the tool to solve problems that could not otherwise be solved. I'm not sure the NSF shares the same personal zeal I have for helping human beings. That's just a side thing. My university measures me by a slightly different metric. They like to see scientific achievement, but they see it in terms of very tangible, countable, measurable things. That tends to be publications and grants and sizes of grants and that type of stuff. That's the coin of the realm, like it is at any other university. And so if I get a lot of those, then they're happy. And they don't really pay too much attention to the science that's being achieved unless it is award-winning science. But they just accept the fact that a peer-review publication counts as legitimate science, and they don't concern themselves too much with the details.	
<b>Q3 What are you investigating?</b>	There are two challenges in the area that we do. Broadly described they are:  1) Harnessing the power. Getting as much power as you can out of each individual computational resource, like these big SMPs. The new challenge today and for the near-term future is these multi-core systems. Before we used to see one, maybe even two CPUs on a single node on these big machines. Now they have eight, sixteen, and even higher in some cases. The challenge is how do you efficiently harness all that computational power? That's an unsolved problem for me and for everyone. Every workshop I go to... petascale, exoscale. That's on the on the topic list; no one has the answer to that. (Note that I use the terms "core" and "CPU" interchangeably. I don't know if well-defined terms exist, but by "node" I mean a single board or blade. Then many CPUs/cores exist on a single node.)  2) The second thing that we spend a lot of time thinking about is how to get the most out of the network – the wide area network that connects these computational resources. When you talk about one gigabit links, that's pretty straightforward. But when you talk about ten gigabit pipes between sites (and in some cases, many of them – three, four, five) – it's not obvious how to fill and use those pipes efficiently. So that's the second place where we spend all our time.	

Interview ID=14 22 August 2007	ANSWERS	ANNOTATIONS
<p><b>Q4.1 What is your method for investigating &lt;phenomena&gt;?</b></p>	<p>I'm a strong believer in microbenchmarks. People come at it different ways. Some people do an implementation and then use the application to see how things perform better. And eventually, we do that too. But our method is to drill all the way down to the basest case, and see what we can do there. A solution at the application level, which is the end game, has no chance of seeing the light of day if you don't have a solid solution at the base. If you try and solve it from the top down – from the application layer down – oftentimes things get in the way or are distracting. So I prefer to use a bottom-up approach.</p> <p>So when we're studying network performance, for example, we will rarely, if ever, work with the application code directly. We'll work with microbenchmarking ping-ponging applications and things like that, which are completely contrived applications. You can consider them like lab equipment in a laboratory, like a test. So that's our method, always working from the bottom up.</p> <p>This method applies to both of our areas of investigation: just as we would use it figure out what we're doing on a network, we would also use it on the compute side. We take a guess as to the types of applications that we'll be running and the types of things that we'll be doing on these nodes. Then we simulate that with a very, very small, contrived program – one that comes close to capturing it (matrix multiplications or something like that.)</p> <p>I might know, for example, that the application will be running one MPI process on one of these nodes that might have 16 CPUs on them. Well, that very often is going to mean a lot of independent local computation, and then at some point the MPI process that's running across all 16 CPUs is eventually going to have to talk to somebody else. Oftentimes that means you will have to combine information and then send it to your neighbor, and vice versa – when you receive information you have to spread it out. So we'll spend that our time handling that problem at a very low level, as close to the board as possible.</p> <p><i>[prompt asking for more information on how guesses are formed about application details]</i></p> <p>When we think about base interactions, we're in the same position as people who are trying to develop like an instruction set for a brand-new computer. What should the instruction set be? What are the users likely to do? Well, they're probably going to want to move information from the memory to a register, so we're going to need some memory-to-register commands. They're probably going to want to add things and subtract things and divide things, so we'll add some of those constructions.</p> <p>We do the same thing. What are the basic level of operations that are very likely to occur on those nodes? So the two sides of the same coin are the ones I just described. The model we think they're going to use is they will run a single MPI process on one node, and then use threads within the single MPI process to get things executing on different CPUS on that board.</p> <p>But eventually, we're going to need to collapse or coalesce the information and ship it off to the neighbor next door or across the way. And vice versa: when we get something from our MPI neighbor running on the computer next door, we have to efficiently distribute that. So that's an example guess for a low level, primitive operations or type of things that these applications will do. And so we try and solve that problem efficiently.</p> <p><i>[prompt asking if the application people see things at this low level]</i></p> <p>They don't yet but they will. What the applications people know is that there is a new type of hardware that's coming out into the world. They know the new hardware is characterized by many, many CPUs per board – much more than there used to be.</p> <p style="text-align: right;"><i>[continued next page]</i></p>	

Interview ID=14 22 August 2007	ANSWERS	ANNOTATIONS
<b>Q4.1 What is your method for investigating &lt;phenomena&gt;?</b> <b>[continued]</b>	<p>And nobody knows how to harness that problem – not the application guys, nor the guys like me who write the middleware code. Addressing the problem will require an ongoing conversation. It's not going to be the application folks coming to middleware folks like me and saying this is what we want. And it's not going to be the middleware folks like me walking up to the application guys and saying, "This is what we're giving you, now change your application." It will be a dialog.</p> <p>In that dialog, both sides will come up with ideas and will find a place to meet in the middle. It's not going to be just between my group and the rest of the world. It will be all the middleware people (and there's a lot of us) working with all the applications. And everyone's going to be talking to everyone up and down and side to side.</p> <p>The odds-on favorite model is the one I just presented. That's the one that's getting traction in the community. Multiple threads operating within a single MPI process. You have one MPI process per node on a machine that has a lot of CPUs, and you use threads to get execution on the CPUs.</p>	

Interview ID=14 22 August 2007	ANSWERS	ANNOTATIONS
<p><b>Q4.3 How do you keep track of interim results, if at all?</b></p>	<p>Well, not very well unless they're regimented. We'll try something, and if it stinks, we try and figure out why it stinks, and work to improve it. If it gets improved, then we don't really write down or remember in any formal sense the thing that we tried that failed and why it failed. We just kind of remember it in our brain, collectively.</p> <p><i>[prompt asking what "stink" means in this context]</i></p> <p>For example, when we were tuning network performance, we were trying different things to find how we could move these bits from point A to point B really efficiently. One of the evaluation measures we would use is effective bandwidth, or latency. Mostly bandwidth is what we're chasing; latency is much harder to solve. And so we would try something and the bandwidth would stink, and we'd try and figure why and how we could we do things better. And then we'd try something different, and the bandwidth utilization would go up.</p> <p><i>[prompt asking how latency is measured]</i></p> <p>We just write our own ping-pong. That kind of stuff, which is part of our microbenchmarks. So we define an area that we want to attack, define a technique for dealing with it, and write and run a microbenchmark to see whether or not we like the numbers. Oftentimes that's only the beginning – it's usually the beginning. I would not be comfortable stopping there. This is just a personal choice; I'm not slamming those who do. I'm not comfortable presenting results unless I actually:</p> <ul style="list-style-type: none"> <li>- start with the microbenchmarks</li> <li>- present those in the paper</li> <li>- and follow it up with a real application that uses the same techniques</li> </ul> <p>I like to demonstrate that the theory we believe in not only bears fruit by empirically observing that it's good at the microbenchmark level, but it also that it works all the way up the top. Because there's some room between the microbenchmarks and the application, and things can go wrong that we don't foresee.</p> <p>So when I read papers that don't have an application, the question that pops into my head is, "This is a fine start, but how do you know that this is going to translate all the way up?" So I have not yet published anything that doesn't include the endgame as well.</p> <p><i>[prompt asking what numbers the interviewee is judging against]</i></p> <p>They're self-referential. And what I mean by that is: prior to us applying the technique, our performance was X. X may have great and X may have been lousy. I don't know. It's just a number. And after applying the technique, our performance is Y. And Y is markedly better than X.</p> <p>Now, Y may still stink, or Y may be great. I don't know where X and Y fit on the absolute scale. But as a relative argument, as a relative observation, Y is definitely better than X, and so we're happy.</p> <p><i>[prompt asking if there are competitors attacking the same interactions as the interviewee]</i></p> <p>Everyone has competitors. I have competitors too. We think of them as colleagues sometimes. ☺</p> <p>There are other people who have numbers we can compare against. In this landscape, we have people who are trying to solve the problem completely independent of Grid MPI. GridFTP is an excellent example of that. Completely independent of running Grid MPI applications, The GridFTP developers said, "Moving data from point A to point B efficiently is a hard and important problem, and one whose solution would be of great interest to the community." So the GridFTP team spent a ton of time hammering out a solution to that problem, and they have a darn good product. A handful of people are working on the same problem (Reliable Blast UDP – now called UDT, FAST TCP, etc.) Some of them have been proposed as new standards to sit side-by-side with TCP at that level in the protocol stack.</p> <p style="text-align: right;"><i>[continued next page]</i></p>	

Interview ID=14 22 August 2007	ANSWERS	ANNOTATIONS
<b>Q4.3 How do you keep track of interim results, if at all?</b> <b>[continued]</b>	<p>In those situations, we're not really direct competitors. They're people who are trying to work on the same problem. And what we tend to do is to leverage all of their work. So we have written code to dovetail GridFTP into MPICH-G2 to see how well that works. We've done the same thing with Reliable Blast UDP, and we've done the same with UDP. We are willing to try anything that comes down the pike. So in that sense they're not competitors. They are people who are working at an even lower level, and we try to leverage their work. We are an application for them. We sit above them. We use them, and developers of the blood flow application use us. So if you're thinking of a vertical stack, at the very bottom is something like Grid FTP. And what sits on top of that is MPICH-G2. We call down to them, and so we are an application of GridFTP. And then our application, for example the blood flow guys, sit on top of MPICH-G2.</p> <p><i>[prompt asking if MPICH-G2 didn't exist, what would be on top of GridFTP]</i></p> <p>GridFTP has other clients that can just move files very efficiently, and there's a need for that in other contexts too. The high energy physics people, for example, produce tons of files in a few locations that everyone on the planet cares about. And they're enormous files. And so you can use a low-level tool like GridFTP and RFT for example to manage the mass migration of volumes of data efficiently.</p>	
<b>Q4.5 How do you document your results?</b>	<p>Sometimes we've thrown some graphs up on the MPICH-G2 website. But we don't keep up with that. I mean that's just something so people can get an initial feel for our work. By far and away the mechanism that we use to disseminate our results is through conferences and journals.</p>	

Interview ID=14 22 August 2007	ANSWERS	ANNOTATIONS
<p><b>Q5.1 In what ways do you interact with simulations in your work?</b></p>	<p>I interact with simulations in my work almost exclusively. The applications we work with are simulation-based:</p> <ul style="list-style-type: none"> <li>- the blood flow people are simulating the flow of blood through the human arterial system</li> <li>- the numerical relativity guys are simulating black holes colliding</li> <li>- the groundwater people are simulating transport – how contaminants move through ground water</li> </ul> <p>Everything we do is simulation.</p> <p>The way our collaborations work – this wasn't by design, it just turned out this way – we'll sit in the room with the mathematicians, or the physicists, or the civil engineers. These domain experts will describe their problem to us in layman's terms, e.g.:</p> <p style="padding-left: 40px;">“We are interested in simulating the flow of blood through the human arterial system. The interesting part to us is where the arteries split. Etc.”</p> <p>Then they will drill down a little bit more and talk about the application they've already written to solve problems. So far, they've always had MPI applications; it will probably be this way for the foreseeable future.</p> <p>As I listen to the MPI application description and I gain an understanding of the problem and their approach, my experience prompts me to ask them specific questions. Like, “Are you using MPI's collective operations here? And are using the asynchronous or the non-blocking point-to-point Send/Receives here?” Because I know where the choke points are when you're running Grid applications. Also, “Why do you need more than one machine to solve this problem? What's your limit?”</p> <p>So at the end of the day, what happens is that I begin to understand their problem and the scale at which they're trying to solve it. I also get an idea of how they're trying to solve the problem through their application. They walk away with knowledge of the potential choke points when you're trying to run on a Grid. What has always happened is that we find places both in the application and in MPICH-G2 that need to be modified in order to make the application run well on a Grid. So the application gets a modification that's good for the application. But what's exciting for me is that MPICH-G2 gets modifications, which not only helps the specific application, but also tends to be of general benefit to all applications. So that's a real big win for us because it's this type of feedback that helps MPICH-G2 to advance.</p> <p>I understand the simulation enough to help them, and can speak at cocktail party levels about what we're doing. However, I can't give a talk on all the issues of “colliding black holes”, for instance.</p> <p>We don't run the simulations. They always do, with us sitting side-by-side (either literally or metaphorically.) So the way it works is we'll meet with them, we'll talk about how things are, we'll get them hooked up with MPICH-G2 so they can run a simple application, then we'll decide what works has to be done both in their application and in MPICH-G2. We both go off and work on things, and then we re-synch, “I'm done. Are you done? Yeah, I'm done. Okay. Go ahead and run the application and let's see how the results look.” And then we move from there.</p>	

Interview ID=14 22 August 2007	ANSWERS	ANNOTATIONS
<b>Q6.1 Describe how you interact with data in your work</b>	<p>I do not deal with input and output specific to the applications at all. But one of the areas that MPICH-G2 can and has been used is to facilitate workflows – sometimes these are called functional pipelines. An example of this is the code we did with the Flash Data Center. There’s a ton of data sitting on disk in San Diego, and you have a cluster sitting in San Diego attached to those disks. The cluster reads in information, does some local computation to refine the data, but then you want to visualize the data. The visualization equipment is over at Argonne. It’s not in San Diego. So now you need to ship a substantial amount of data from San Diego to Argonne.</p> <p>So there is a workflow where you start off data mining, and then the next phase in the workflow is visualization, and so that data gets shipped from San Diego to Argonne. They use MPI do this. And specifically in this case, they used the GridFTP modifications we did.</p> <p>In this case we cared a lot about data. So the extent to which we care about data is all wrapped in the MPI standard. Our concerns and our goal are to make it as easy as possible for the application to ship the data from point A to point B. Not to get it off of a disk – that’s someone else’s job. But once it’s memory-resident in the application and they want to send it from one MPI process to another, we work to make that as easy as possible for them.</p> <p>So we tell them to call an MPI function, MPI Send for example, and we will take it from there and move the data as fast as possible. Furthermore, if you’re moving it from one machine to the other where the architectures are incompatible (like from a big endian machine to a little endian machine) we’ll assume the responsibility of doing the data transformation for them.</p> <p>So we don’t deal directly with any filesystem, such as bringing data from disk or tape to the application. Our area of work is process-to-process communication.</p>	
<b>Q7.1 What resources do you use in your work today?</b>	<p>We use Grids that are already set up – we don’t build them ourselves. They’re too expensive for a low-life like me to build. So we’ll use the TeraGrid. Or we’ll use resources in the UK. We’ll use networks that connect the United States to Amsterdam, for example. We take advantage of infrastructure that requires government-level financing to put in place.</p> <p>By “we” I’m referring to the MPICH-G2 group and the application groups with whom we collaborate. So in the NekTar case, it made a big splash about two years ago because it ran as one application while simultaneously using resources located in the United States and in the UK. From the application’s point of view, it was just one big supercomputer. But this required government-level funding for not only the facilities here in the United States and in the UK, but the networks that connected them. That specific application was a special demonstration project that NSF and the UK decided to fund.</p> <p>But other stuff we’ve done (for example the groundwater transport or colliding the black holes) had no direct funding that provided access to these facilities. Allocation for these cases is a separate issue.</p> <p>In the early days, you would pick up the phone and ask, “Hey, can we run on your machine?” And by and large, they’d be willing to let you do it. But now it’s become more formal where you have to ask for cycles, for example, of the TeraGrid. It’s like currency. They’ll give you so many service units based on the description of your application, measured against other applications that come in. They dole out the resources of the machine accordingly.</p> <p>But that’s a process that’s independent of grant-getting. We have not yet written those into any grants. It’s stuff you do after the fact.</p>	
<b>Q8.1 What software do you currently use in support of your work?</b>	<p>Globus rules. It’s gonna save the world.</p> <p>I’m not kidding. We rely quite heavily on Globus. Globus does all process management, the start-up, the security. In Globus we’re using IO for all inter-machine communication. Globus is the one software that we use across all applications. You can’t build MPICH-G2 without having a Globus installation on the machine first. Other libraries may differ from application to application. But the one that’s ubiquitously required for all applications is Globus.</p> <p>And that’s it. Nothing else. Oh, I’m sorry, I made a mistake. The MPICH software project is another one. MPICH-G2 is a module that plugs into the larger MPICH software framework. And there’s a lot of MPICH code that we also take advantage of.</p> <p>So these two software groups are the two that allows MPICH-G2 to exist. Both the MPICH project, which was developed by Computer Scientists at Argonne and the Globus project, which was developed by a million people.</p>	

Interview ID=14 22 August 2007	ANSWERS	ANNOTATIONS
<b>Q8.2 What scripting languages have you used in the past year?</b>	I personally don't use any. But we have to use stuff like <i>autoconf</i> quite a bit in order to configure the stuff. I don't know if you consider that a scripting language or not. I don't have any experience with it, but the people who work on the MPICH-G2 project with me know about it. We also use shell programming. I can't remember which shell it is; probably the Bourne shell – I'm guessing – to launch the jobs and that kind of thing. But that's at a modest level. It's serious but it's not where the lion's share of our work is.	
<b>Q8.3 What programming languages have you used in the past year?</b>	C, C, C, C, C, C – C for the whole thing.	
<b>Q8.6 If the need for new software-based functionality arises in your work how do you acquire it?</b>	We haven't worked that way. That's not to say that it won't happen tomorrow. But so far the way we work is we assess the landscape, and we see what we can do with what's out there. So as an example of that, eight years ago we saw MPICH and Globus and thought, "We can do something interesting here." Subsequently things like GridFTP, Reliable Blast UDP, UDT... fast data transport technology came up and we said, "Hmm. These are good things. What can we do with these?" Recently, these threads packages are coming up, we're looking at these and we're saying "Hmm. What can we do with these?" So it's not the case that need forces the exploration of a tool. The way it has been so far is as we become aware of tools, we think about how we can use them.	
<b>Q8.7 How do you share software with others?</b>	We do share because it's nice to share. Our software is distributed by the MPICH group. They have an enormous release mechanism and mailing lists and Web pages and we just ride their coattails. We are distributed as one of the modules that they release.	
<b>Learning about the user's problems</b>		
<b>Q9.1 What challenges do you face today in accomplishing your work-related goals?</b>	It falls under two categories: personal and professional. From a personal side, it's time. As you get older, more things of asked of you because you've acquired more experience. You get to know more people. You're asked to serve on more committees and more work groups, etc. It becomes an issue of time and that's a serious problem. Managing it is not as easy as it looks – at least not for me. It's hard to say no when someone asks you to review a paper, serve on a task force, etc. That's one challenge. The other challenge is the cutting edge of the double-edged sword that MPICH-G2 has taken on: On the one hand, leveraging large software projects like MPICH and Globus (and even to a lesser degree GridFTP or UDP or UDT) is great because it's a tremendous leg up. You leverage it. There's a bunch of code that's there, and you apply a little bit of work, and you get a tremendous benefit from it without having to do all of the work. But the cutting edge of that same sword is when you need changes or modifications or improvements it's not under your control. You do not directly control the developer resources to get those changes done. And so you have to go back and ask them, and you don't really have much to offer in return other than the greater good of what you're trying to do. We've been doing well operating under these constraints, but it hasn't been easy all of the time. There have been times when we've been told no, flat-out, "No, we aren't going to do it; I don't care how much you need it." And then other times when we've been told, "Yeah, but it will take a while." And in some cases, it takes a long while. Also there are other times when you get it right away. No one's out to get you, but they all have their own agendas. Your requests need to be fit into larger priorities, and sometimes the requests are not given as high a priority as you would like. It's not that they're trying to hurt you; it's just that they have other bells to answer. That's hard. There's no way around it; I don't know how else to put it. It can be a problem at times.	



Interview ID=14 22 August 2007	ANSWERS	ANNOTATIONS
<b>Q9.3 What technology-related obstacles do you currently encounter?</b>	<p>There's something that is lacking in the field: debugging tools. There are a lot of debugging tools already out there, and we make use of the ones that are out there. It's not a problem with a particular tool; the problem is the absence of a tool.</p> <p>It's a hard problem – I'm not shaking my head saying, "Why doesn't a solution exist today?" But it would be wonderful to have a tool that would allow us to debug things in a multi-threaded environment at large-scale. That's a really hard problem, and I'm not at all surprised that there's not yet a solution. Once there is one we will be very happy.</p> <p>[prompt asking for more detail on the types of debugging information needed]</p> <p>When you write multi-threaded code, bad things can happen. Deadlock. Another condition that also appears in serial code, but is perhaps more pronounced in multi-threaded code: accidental memory overwrites. A tool to handle that at large scale would be tremendously useful. By large-scale I mean hundreds of distributed processes – even thousands.</p> <p>We can't use a tool that produces a three-kilobyte file of error messages when something bad happens. Something happens and then all the processes start sending messages saying, "I can't do this. I can't do this. Etc." We can't manage that. That's too much information, which is just as useless to us as no information. So a tool for use at large scales that can solve or help report and diagnose problems is the type of thing I'm talking about.</p> <p>I don't just need this myself. A lot of people do.</p> <p>As far as wide-area network debugging we're doing ok. We have things like iperf or we can write our own microbenchmarks. Those are pretty straightforward, and I think the reason is that network performance monitoring has been around forever. People have made sure that they can do it. And monitoring is a pretty straightforward technique. It's really not that tough. And so the tools that are out there are pretty good. And what you don't have at your fingertips, you can probably quickly write something that's pretty close to what you need.</p>	
<b>Learning about the Globus user experience</b>		
<b>Q11.1 Which Globus data components do you directly interact with in your work today?</b>	GridFTP	
<b>Q12.1 Which Globus security components do you directly interact with in your work today?</b>	<p>Here you're going see how ignorant I am on how things work in Globus. We do use Globus security, but I don't know specifically which items we actually touch out of that laundry list you just gave me. Our use of security is based on whatever happens when a user types in grid-proxy-init and globusrun. So I know security things happen at both of those layers, but what items on your list get triggered, I couldn't tell you offhand.</p>	
<b>Q13.1 Which Globus execution components do you directly interact with in your work today?</b>	<p>globusrun and globusrun-ws</p> <p>We do use MPICH-G2 ©. And in MPICH-G2, we do make calls to the GRAM2 client library to manage the processes of it.</p> <p>We do use GRAM4 – we've started to use GRAM4. We also use a new thing, which I can't tell you exactly where it sits. It is called the Rendezvous Service. It may be its own service; it may sit inside of GRAM4 – I don't really know. But it was critical for us as we moved from pre-Web services Globus to Web services Globus.</p>	
<b>Q15.1 Which Globus common runtime components do you directly interact with in your work today?</b>	<p>We did use Globus IO, and we are moving to Globus XIO.</p> <p>I think – I don't have the terminology down because I'm not very Web service savvy – but I think the Globus developers had to write some special C hosting environment code. They had to write some special infrastructure in order for MPICH-G2 to access the Rendezvous Service because MPICH-G2 is written all in C.</p> <p>There's a data conversion library in Globus that we use as well. I think it's just called "DC". "something"_DC. It sits in core. It's just a set of functions, but it's the data conversion library. That's all I know.</p>	

Interview ID=14 22 August 2007	ANSWERS	ANNOTATIONS
<p><b>Q16 Why do you use &lt;component&gt; instead of an alternative technology?</b></p>	<p>GridFTP: Oh, we don't use GridFTP "instead of" an alternative. We just try all possible solutions. GridFTP is a good example of this. We use GridFTP because it's a great tool to move data fast from point A to point B without having to think a lot about the issues that the GridFTP developers have already thought of. But an equally good tool is something like UDT, and we've used them both. Now, why would I pick one over the other having the ability to use both?</p> <p>Well, the way we use GridFTP is TCP based, and UDT is UDP based. And UDP can be a little more hoggy and less of a good community network user. So what I'm trying to say is that GridFTP is TCP-based. TCP inherently is a nice network user. When things get congested, it will back off and not hurt everyone else on the network. UDP is fundamentally different. With UDP, you can write an application that hogs more of the resources than you are entitled to as a shared community member.</p> <p>So if we're on the TeraGrid, we'll tend to use UDP-based solutions, but if you're on some open wide-area Grid, we'll just tell people you should really just be using GridFTP because it's TCP-based.</p> <p>GRAM: I'm not aware of any viable or equally good alternative. GRAM is great. In one fell swoop it allows me to specify jobs in a single language, and hides all the details of every local job manager (which nobody wants to learn.) Plus, it has the entire security infrastructure built-in. That's a hard, hard problem to solve. The GRAM developers have solved it. They did such a good job at solving it, and GRAM has been inspected and reviewed and embedded in the community. Now it is widely accepted.</p> <p>You just walk into a sysadmin's office these days and say, "We're using GSI – the Globus security infrastructure." and they get it. They know that it's been looked at by many eyes (maybe even their own). They trust it, so you don't have to convince them it is secure. Alternatives like <i>rsh</i> or <i>ssh</i> would raise eyebrows. Is that secure? Is that really good? And so there's just no alternative out there.</p> <p><i>[prompt asking if there is a difference in the interviewee's mind between GRAM2 and GRAM4]</i></p> <p>Functionally, no. Not from our perspective. But the things I need to do from a syntax perspective are completely different between GRAM2 and GRAM4, and require a rewrite of my stuff. And the RSL versus the XML is completely different. All that stuff is completely different. But functionally, no.</p> <p>Rendezvous Service: The way the Rendezvous Service came into being is a great story.</p> <p>I came forward to the Globus developers with a need. In pre-Web services Globus, there were two things that we used: GRAM2 and DUROC. GRAM2 was great at launching a job at a single site, but DUROC was needed to oversee multi-site launches. DUROC would take the entire job specification, break it up into its pieces of single-site jobs, issue the GRAM commands to launch them at each individual site, and then let you know once the jobs were all up and running.</p> <p>The second thing that DUROC did was to give us a bootstrapping communication library. We didn't need to hand around host/port pairs. DUROC gave us a very primitive Send and Receive library, which we used to bootstrap the communication. We used it to distribute host/port information in an all-to-all exchange after it was up and running.</p> <p>Okay. In the GRAM4 Web services world, we still needed the same functionality, but no Web services counterpart to DUROC existed. We didn't have anything that could act as an overseer, nor did we have anything that was an all-to-all exchange.</p> <p><i>[continued next page]</i></p>	

Interview ID=14 22 August 2007	ANSWERS	ANNOTATIONS
<p><b>Q16 Why do you use &lt;component&gt; instead of an alternative technology?</b> [continued]</p>	<p>And that's where the Rendezvous Service came in. The Rendezvous Service was written to provide all-to-all exchange of information in the Web services context. We were able to whittle it down to a reasonable API. And it was even a great exercise because I showed up with a need that the GRAM developers didn't foresee. We talked with them and nailed down the requirements, and it was done. This was about a year or two ago. It was great.</p> <p>XIO: I love XIO. It is the follow on to Globus IO. IO was very good, but XIO is even better because it allows us to build protocol stacks. The protocol stack design will make it much, much easier for us to introduce new technologies in the future.</p> <p>So when it was all Globus IO-based we had to go through the exercise of pushing GridFTP into MPICH-G2. It didn't kill us, but it wasn't easy. We had to munge the code pretty heavily to shove that stuff in. And we had to go through the same exercise when we had to push UDT into MPICH-G2.</p> <p>With Globus XIO, all of this can be very neatly wrapped into an XIO module that either I or someone else writes. That's the recipe for rapid prototyping. I won't have to mess with the MPICH-G2 code at all anymore. I'll just write an XIO module, and one line of code to activate the module in MPICH-G2, and it's done. And you get it for free. That's a big step forward for us. That's a big help.</p> <p>The data conversion library: The Globus data conversion library is indispensable. We need it. If it were to go away, we'd have to write it from scratch ourselves. MPICH-G2 is responsible for doing the data conversion between big endian and little endian machines, for example. The MPI application is not going to give a damn about that. I need to care about that. It's an ugly job that no application should have to write. One library should write it once and then provide it. We provide it in MPICH-G2 because the Globus developers wrote it.</p> <p>At the very endpoints, there are some cases that currently aren't handled quite right. Okay. They're missing, but so far that's not really been a problem – at least we haven't encountered one. The problematic cases have to do with the IEEE floating point stuff. Something related to conversion of a long-double, which is part of the ANSI standard. You can't convert from a long-double to a double-double to something else.</p> <p>Some of these bizarre endpoint cases are missing. If it ever becomes a problem, we can roll up our sleeves and try to solve it. But the core of that library is a big leg-up for us.</p>	

Interview ID=14 22 August 2007	ANSWERS	ANNOTATIONS
<b>Q17 What are the major challenges you face using &lt;component&gt; today?</b>	<p>GridFTP: Frankly, none. It's really simple. It was simple to integrate. They wrote a wonderful API. I had to sit down and understand it. But that wasn't much of a challenge – that was just an exercise.</p> <p>We use GridFTP not only in the microbenchmarks, but at the application layer as well. We started using it at the microbenchmark level. But the way we've dovetailed GridFTP into MPICH-G2 is:</p> <ul style="list-style-type: none"> <li>- the application calls a simple MPI function to configure the GridFTP pipes</li> <li>- then from that point on whenever they call MPI Send, underneath the sheets they're getting GridFTP for free.</li> </ul> <p>So we used it initially at the microbenchmark level because I wanted to make sure that at the lowest level I would get the performance I needed. So I just wrote some code that ping-pongs some data and monitored the performance. It did perform well and so I said, "Great. Let's write the code to integrate it with MPICH-G2." We did, and then we ping-ponged it again through the MPI layer. It again performed well. And then we used it in a real application. And we compared the performance to not using it; again, it performed well.</p> <p>GRAM: Both GRAM2 and GRAM4 are lacking in the same thing, and that's the ability to do co-scheduling. That's the biggest problem for us.</p> <p>Both GRAM2 and GRAM4 are great for saying I need 10 nodes on that machine, and I want you to run this application when you get them. And I don't want to worry about specific scheduler syntax. I don't care. I'm just going to specify the job in XML and let GRAM talk to the scheduler for me. GRAM is great at that, negotiating to put you in the queue, notifying you when the job is running, etc. That's perfect.</p> <p>But we don't run jobs like that. None of our MPICH-G2 jobs run like that, meaning on a single machine. All of our MPICH-G2 jobs necessarily run on two or more machines. It is imperative that the jobs are co-scheduled: that each job is launched is launched near the same time.</p> <p>It does us no good, in fact in some cases it does us harm, if one job actually gets through the queue executes on machine A, and then two hours later the second job gets through the queue and begins running on machine B. It doesn't do us any good. We need to make sure that they both get through the queue and hit the nodes at the same time.</p> <p>The term we use for that is type of job submission control is <i>co-scheduling</i>, and that's lacking. And that's our biggest challenge. We're working on it, but it isn't done yet.</p>	
<b>Wrapping-up</b>		
<b>Is there anything you'd like to say to the people who build software for use by people like you?</b>	<p>I'm not one to necessarily bite my tongue. If I need something, I'll ask. I don't demand it, but I'll ask it. If something's wrong or can be improved, I'll politely point it out. So I've already said everything that needs to be said. And it'll probably be that way in the future. So there's nothing unspoken; I can tell you that for sure.</p>	

## D.15 ● We provide mechanisms for sharing video, audio and applications

Interview ID=15 23 August 2007	ANSWERS	ANNOTATIONS
<i>Pre-interview question: Do you interact directly with Globus software in your work today?</i>	yes	
<b>Establishing context</b>		
<b>Q1.1 Please provide a one-minute overview of your project</b>	The Access Grid is for real-time collaboration among groups of people at multiple locations. Increasingly, research at the lab and worldwide is becoming a collaborative effort. Our expertise is distributed and collaborations leveraging that expertise must bring together people from diverse locations. The Access Grid attempts to provide an environment where people can interact as naturally as when they're in the same room. This is a more natural and effective method of communication than alternatives like telephone and video conferencing, where you don't see the other people or you only see a subset of them. In addition to that, we try to provide mechanisms for sharing not only visual and audio input, but also their interactions with applications. We've also looked at allowing people to interact with remote instruments and computation.	
<b>Q1.2 What is the project's name?</b>	Access Grid	
<b>Q1.3 Which agency funds the project?</b>	DOE Office of Science, the National Science Foundation, Microsoft Research	
<b>Q1.4 What field does your project belong to?</b>	Collaboration technologies It has relevance to the fields of computer science and engineering. To most people it would look like an engineering effort. But given the extensibility options we have built in, there's room for people to explore the Access Grid from the perspective of more than just an application. There's a wide enough array of technologies involved that we're trying to provide a platform for exploring some of those. And those could be security, streaming media, or things as unrelated as human interaction.	
<b>Q1.5 What is your job type?</b>	Project Lead and Developer	
<b>Q1.6 How long have you been a &lt;job type&gt;?</b>	On this project: Project Lead for 2.5 years, Developer for 5 years	
<b>Learning about discipline-specific goals and approach</b>		
<b>Q2.2 How will the success of your project be measured?</b>	The thing that we're trying to enable is for people to interact better. That leads to measuring productivity, which is difficult. We measure the success of it anecdotally in the comments that we hear from people. They tell us know people they've interacted with on the Access Grid even though they haven't met face-to-face. When they actually do meet face-to-face for the first time, they're already good colleagues. I've had people say they know the people they interact with over the Access Grid better than the people down the hall from them. Beyond personal interaction, it has to do with sharing applications and data. We try to do this in a really natural way. I know we've done some things in a way easier than the alternatives, and in some cases the AG enables stuff that wouldn't have been possible otherwise. Whenever we do that we're succeeding. We also see consistent growth in the community. I see new people coming all of the time. One way to measure it more analytically is by looking at numbers like software downloads. Also because it is server-based we can track server accesses. Those have continually grown.	

Interview ID=15 23 August 2007	ANSWERS	ANNOTATIONS
<b>Q3 What are you investigating?</b>	<p>The project is not only developed at my home institution. There are developers working all over the world.</p> <p>One of the active areas is to bring in higher quality video. That effort has involved a lot of work by other groups: digging down into video codecs and streaming media. It touches on access to open source libraries for these things and dealing with licensing.</p> <p>Audio is also an issue. There are people looking to use the Access Grid for music and music instruction and the audio quality's not good enough for them. So video and audio certainly fit into the things we need to pursue because in trying to be a natural environment for people to interact in, we want to give them the best possible environment. Better environments are possible now because of better video and audio standards, and more bandwidth.</p> <p>The user experience is also important, so constrained networks are a problem for us. If participants are on a low-bandwidth link in the middle of nowhere, it is problematic both for them and for their collaborators. Their networks would be incomparable, some of them on very fast networks and others on very low bandwidth networks.</p> <p>There are always human factors in production issues, although we're not actively researching those. And I don't know that we could solve them anyway really. Humans are really sensitive to disruptions in audio and it's really, really hard to get that right. If everybody involved were an audio engineer, it'd be much simpler. But they're not and they have limited budgets. There's only so much that you can do, and audio quality suffers.</p> <p>I'd really like to see us improve things in terms of application sharing. We often get people who find out about a meeting they're supposed to join with too little time to prepare. If I can get the people who:</p> <ul style="list-style-type: none"> <li>- need to share data and an application with others in a meeting,</li> <li>- can grab the software and immediately set it up without requiring any depth of expertise</li> <li>- can interact in a way that makes sense to the user (in terms of handing data files off to others, guiding people through a tour of visualization data, sharing equations, etc.)</li> </ul> <p>If people could successfully do all of those things and walk away, I'd feel pretty good about it. Some of those things we can do easily. Some of them we'll always struggle with, such as the audio production quality. These are problem areas that others in the community are targeting more than we are.</p> <p>Our focus is to look at the AG [<i>Access Grid</i>] as an infrastructure on which to build applications. An example application is to enable collaborative visualization and venues backended by a cluster. This is one of my priorities.</p> <p>My role in the efforts lead by other groups is as an integrator. For instance, we are not working on new video tools but other people are. So we're working as advisors to fit those things into the AG and make them available to the users.</p> <p>So my focus is to provide the Access Grid as a platform and infrastructure for many people (including my group) to build on.</p>	

Interview ID=15 23 August 2007	ANSWERS	ANNOTATIONS
<b>Q4.1 What is your method for investigating &lt;phenomena&gt;?</b>	<p>Given our experience it's pretty easy to recognize the areas that need to be addressed. I'd like to think I'm a pretty good judge of a priority of those things, but I'm not the only judge of them. I also have to stay in touch with the user community to know what they need. I do that a lot. I know pretty well what they need, so I mesh my view of the priorities with the priorities according to them.</p> <p>Another dimension to this is that some of the things users want may be very different from priorities we have. Because the AG tries to be a research project, there are also research type priorities that we need to pursue. How do we execute that? We try to get money, decomposing the problems into workable subsets.</p> <p><i>[prompt asking how the problems are identified]</i></p> <p>From direct interaction with users in email and at meetings (whether they're over the AG or in-person meetings), through our bug tracking database, through mailing lists. I guess there's a bunch of avenues by which we get information from users. I'm a user myself, so there are things that I perceive as problems. Some of them are widely perceived by the user base as problems that need to be addressed. Others are improvements that users don't even perceive but when the problem is solved they're happy about it.</p> <p>Also in terms of the research aspects:</p> <p>We've spent a lot of time working on engineering issues; our focus lately hasn't been on research. There's a lot of interest in the engineering area, not only in the lab community but in the commercial sector too. We share many of the same goals, but there are other things that I think we should be looking at too. I'm referring to problems that arise in these environments that we need to think hard about how to solve. For example, we currently lose eye contact with people on the remote side. I think we could do interesting work in terms of eye tracking with an array of cameras and sending images of the person based on which camera they're looking into. I'd like to see us introduce gesture-based interactions with the environment. Those are research things that we would have to elaborate and propose and get funded to work on them. There are certainly things that we as researchers perceive but users probably don't think about.</p>	
<b>Q4.3 How do you keep track of interim results, if at all?</b>	<p>There is a software cycle that includes pre-releases and releases. Those happen regularly enough that people know there is progress.</p> <p>Software is delivered to people, they try it and they give feedback. We're in a continuous feedback loop with the users through all the channels I mentioned earlier.</p> <p>There are smaller activities that are independent of releases. Somebody may have a particular problem that is worked on and patched. Creating patches happens more often than releases do.</p> <p>There are external developers who are always working on things, so there are always interactions. Examples include asking questions about code or submitting changes that then must be integrated.</p>	

Interview ID=15 23 August 2007	ANSWERS	ANNOTATIONS
<b>Q4.4 How do you test work-related hypotheses?</b>	<p>As a curious developer I perform exploratory tests of ideas, such as participating in a joint effort where somebody wants to try something and we put together an organized scenario. Lately the closest we've been to this is proof of concept work in preparation for proposals.</p> <p>We also test software releases. There's a lot of in-house testing and we always issue a beta release. Sometimes we have multiple beta releases before a final release. I insist on really high quality, so the finals don't go out until the quality is sufficiently high. We know the functionality and how it should work and how it might break. So we do as much internal testing as we can to assure ourselves of the quality.</p> <p>Users of beta releases explore the code in environments that are different from ours. A lot of times I identify the new functionality in a release and provide some testing guidelines. I encourage beta testers to submit bug reports liberally, whether or not they are sure it's a bug. They don't have to worry about whether or not it's already been reported, or whether it's actually a bug – even if they lack time to track it down and describe it completely. I encourage people to submit those types of bugs liberally.</p> <p>There is fairly verbose logging included with the software so people can submit bug reports directly from within the running software. It grabs portions of those log files and includes them with the bug report.</p> <p>We have a suite of automated tests but it's really horribly out of date and essentially unusable to us right now. I consider that a horrible shame and I wish I had time to go through and make a full-fledged test suite because I'm absolutely certain that in the long run it would save us time. But it's just hard to find that time up-front.</p>	
<b>Q4.5 How do you document your results?</b>	<p>With releases there's always a description of the new functionality. There's always a testing guideline, and a list of the bugs that were fixed in the release. That stuff is all tracked historically too, so I can look back at multiple releases and see which bugs were fixed in which release.</p> <p>This takes me back to regretting that we do not have a test suite. We really should be adding tests to a test suite to do regression testing for bugs, so that we don't reintroduce bugs that have already been fixed. Unfortunately we're not doing that.</p> <p>Lately we've been doing has been primarily engineering work, so recent publications have been technical reports published by the laboratory. When we shift focus to more research-oriented tasks we produce papers for journals.</p>	
<b>Q6.1 Describe how you interact with data in your work</b>	<p>In the Access Grid venues users can share data. They can upload data to a venue and other users can download it.</p> <p>In the recent engineering work we have integrated a couple of data transfer protocols, as well as dealt with issues relating to back-end storage.</p>	
<b>Q6.3 By what mechanisms is access to your work-related data controlled?</b>	<p>The protocol is FTP and goes over SSL.</p> <p>You can only access the data if you are in the venue where the data resides. Access to the venue can be controlled by X.509 certificate-based authorization.</p>	
<b>Q7.1 What resources do you use in your work today?</b>	<p>Most of my time is spent working on my desktop computers, which include the three platforms that we support.</p> <p>The visualization work, because it's back-ended by a cluster on the TeraGrid, is making me touch TeraGrid machines more than I have previously.</p> <p>I spend some time in one or more virtual venues every week; most of the venues are hosted on machines at my home institution. Very infrequently I go elsewhere. I have been in a virtual venue running on a server in Alaska in the past month, as well as one on a server in the UK. A little longer ago but I was on a server in Australia.</p>	
<b>Q7.3 By what mechanisms is access to your work-related resources controlled?</b>	<p>The virtual venue access mechanism is the certificate and PKI-based authentication.</p> <p>The authorization mechanism is our own implementation. The most recent reengineering concentrated on bringing in Internet standards. So that's why we did the FTP work and moved to Jabber for chat. It is also why security is over SSL. I'd really like to see us using some standard certificate-based authorization. I don't remember finding one. If one is available now, it's a serverside component, so we could swap it in.</p>	



Interview ID=15 23 August 2007	ANSWERS	ANNOTATIONS
<b>Q7.4 How do you locate available resources for use in your work?</b>	Typically there's some pointer to a venue included with a meeting request. Other than that, venues are well known. There are a couple of venue scheduling mechanisms: There is the AG scheduler, which is developed at NCSA. You can click a link on their site and it will put you in that venue, so you don't have to know anything about the address of the venue. And then there's some integrated scheduling based on RSS. Anybody can publish a meeting feed. One can subscribe to the meeting feed and join directly from the AG client.	
<b>Q8.1 What software do you currently use in support of your work?</b>	Python, naturally. PHP somewhat lately. I end up touching C and C++ on a regular basis. We maintain a community Web site that is based on drupal [ <i>drupal.org</i> ], so I've had to become familiar with that. There are of course compilers on various platforms. Installer toolkits for building installers. Debugging tools: things like GDB on Linux, Visual Studio on Windows, and something called DebugView on Windows. There may be others.	
<b>Q8.6 If the need for new software-based functionality arises in your work how do you acquire it?</b>	If there's a good open source solution I'll use it. If there's only a commercial solution then I'll ask for my home institution to purchase it. We develop our own tools, in terms of test scripts and stuff. In terms of adding new capabilities to the Access Grid itself: The Access Grid is licensed similarly to a BSD license, which means any dependency must be compatibly licensed. So any additional functionality for the Access Grid itself must satisfy those criteria. We sometimes evaluate existing solutions as a proof of concept stage, just trying to get an idea how things work and the limitations of the tool. If you find significant limitations very early, then you're forced to either extend them or build your own.	
<b>Q8.7 How do you share software with others?</b>	The source code is available in a public CVS repository. With the appropriate permissions people can also write to the repository; there are a number of people around the world who have that permission. Getting permission involves getting an account at a DOE lab. That's difficult because a number of people are foreigners, but it's not impossible. Probably the most important thing is for one of us to become convinced that it's worthwhile for somebody to have that permission. Then we make the case to the lab that they should have an account. So basically getting to know the people through experience in the community.	
<b>Learning about the user's problems</b>		
<b>Q9.1 What challenges do you face today in accomplishing your work-related goals?</b>	I think boils down to a time challenge. There are multiple levels on which I have to function. Broadly I could say there are technical challenges and non-technical challenges; it's a give and take between those two. Sometimes the non-technical challenges are time-consuming enough that it's hard to find time to dig deep down into technical issues to resolve them. Some examples: Non-technical ones involve keeping servers up, interfacing with users (in terms of email or problems), fielding bug reports, trying to reproduce problems, giving demos, etc. There's a lot of that kind of thing that takes time away from pursuing technical challenges. Technical challenges tend to be deeper and require more thinking. More time is needed to dig down into debugging or understanding what's going on at a deep level in the code. There has to be a compromise between being available for the non-technical things and shutting that off so you can pursue technical problems. It's always a give and take. Given infinite time or resources I'd be able to wrap it all up.	
<b>Q9.2 What types of information do you need in order to address the challenges you face today?</b>	Information? No, I don't think so. The things that would help me are probably a full test suite, more time and more people.	

Interview ID=15 23 August 2007	ANSWERS	ANNOTATIONS
<b>Q9.3 What technology-related obstacles do you currently encounter?</b>	<p>It's not only bugs in the code – it's technical challenges in delivering the next thing.</p> <p>If we were to pursue the collaborative visualization stuff, there's work that you have to think about. To plan, design and execute.</p> <p>It's open source work so sometimes there are problems in other people's code that need to be dealt with in order to achieve your own objectives.</p> <p>There is a fairly mixed bag of technical and non-technical issues and it's always hard to strike the right balance. It'd be a real shame if all my time were consumed with non-technical work and shallow technical challenges. I wouldn't want to live in that world.</p>	
<b>Q9.4 By contrast, can you provide examples of technologies you find very useful today?</b>	<p>I love my Mac.</p> <p>I love the tools that I use all the time, but Firefox and Thunderbird don't always perform as well as I'd like them to.</p> <p>For convenience sake I have begun to use more online Web-based tools for things because you can access them through a bunch of different devices.</p> <p>Keeping information online allows you to read and write it from my Mac, any other machines I have, from my phone or whatever. I would like for the AG to operate that way.</p>	
<b>Q10.1 Can you think of any work-related tasks that decrease your productivity?</b>	<p>I'd to class all meetings in that group. On average I'm sure I have less than ten hours a month of meetings, so I try to limit the time I spend in meetings.</p> <p>I used to run meetings for people on the Access Grid, but we got away from that. Now because the software is easy to use the administrative staff runs them.</p> <p>The ten hours per month spent in meetings is dedicating to testing and a monthly meeting that's an hour for North America and an hour for Asia-Pacific.</p>	
<b>Learning about the Globus user experience</b>		
<b>Q12.1 Which Globus security components do you directly interact with in your work today?</b>	<p>A previous version of the Access Grid, version 2, used Globus tech inside. At that time I worked with the libraries a fair amount. All the connections in the Access Grid were GSI. We used the GSI proxy certificate code and manipulated certificates using OpenSSL directly.</p> <p>We still use SimpleCA for managing our CA.</p>	
<b>Q13.1 Which Globus execution components do you directly interact with in your work today?</b>	<p>I have very lightly touched GRAM for job submission and tried to use the CoG Kit on Windows for job submission from the desktop.</p>	
<b>Q16 Why do you use &lt;component&gt; instead of an alternative technology?</b>	<p>SimpleCA:</p> <p>I wasn't involved in that choice. It was made at the same time that we were choosing to use Globus inside of the Access Grid. Perhaps it was chosen as a companion tech, or because that's what we knew, or frankly the way we use it is pretty simple so maybe it was chosen for its simplicity.</p>	

Interview ID=15 23 August 2007	ANSWERS	ANNOTATIONS
<p><b>Q17 What are the major challenges you face using &lt;component&gt; today?</b></p>	<p>GSI: Probably the lack of platform support, given that Windows was our primary platform. Aside from that, on the platforms where it was supported it was always a challenge from a build and packaging standpoint. For example, we have some scripts for building Access Grid packages that included some fraction of Globus. But because of the way Globus is packaged we ended up shipping a lot more of Globus than we needed to. Personally we were okay with that because we wanted to support Grid computing through the AG. But that added a significantly long step to our build process. When we took that out, one of the comments from the Australians was that building an AG package went down from 50 minutes to 3 minutes. Our bundling of the Globus code was in June 2003 and we removed it in late 2005. We were using the GT2 C code from the GT3 distribution. Another challenge related to GSI: There was a problem in the security code in GSI that would cause connections to get hung up. There was no timeout so the connections just ended up forever hung. They never timed out and our server would hang. It ended up happening under particular network conditions where the MTU size was too big... I don't remember the details. It had to do with firewalls also. At the time support for GT2 had gone downhill. The sun was setting on the GT2 code so there was limited support for it. Either we had to fix the problem ourselves or migrate to GT3. We ended up patching the GT2 code for a while. SimpleCA: We actually have a wrapper script on top of the tools to simplify things, so it might be hard for me to comment on exactly the SimpleCA user experience. We have tools that find unsigned certificates in the request repository and then cycle through them and prompt you to sign them individually. GRAM: The only problem that I've run into in my limited use of GRAM was the varying functionality on different deployments. I don't think that's really a criticism of GRAM, but the specific case was on the TeraGrid. You could specify that you want nodes of a certain type. The types of nodes I wanted were visualization nodes, and there was a way on the TeraGrid version of GRAM to specify those nodes. So I wanted to submit the job to the TeraGrid visualization nodes from the CoG Kit on the Windows desktop. But I was using the regular CoG Kit on Windows; it wasn't the TeraGrid-enhanced version. So it didn't understand those extensions. That was a frustration. I thought I had struck gold when I was able to submit a job from my Windows desktop, but then I hit that limitation and was disappointed. Maybe those extensions have since been rolled into Globus proper. But I was told at the time was that there are as many different flavors of Globus as there are flavors of Linux.</p>	

## D.16 ● We assume a world where lightpaths can be scheduled between computers

Interview ID=16 5 September 2007	ANSWERS	ANNOTATIONS
<i>Pre-interview question: Do you interact directly with Globus software in your work today?</i>	yes	
<b>Establishing context</b>		
<b>Q1.1 Please provide a one-minute overview of your project</b>	The focus of the OptIPuter project is to study what happens to distributed computing in an environment where there is no bandwidth limitation. It assumes a world in which people can schedule dedicated lightpaths between distributed computers, on demand, in the same way that they would schedule supercomputer clusters. We want to understand how those assumptions change application architectures, as well as change application users' perceptions of high-speed networking. In our work we assume the unconstrained bandwidth begins at the TeraGrid to wherever you're residing. With both of the collaborations we're working in, we have 10's of gigabits available to us. So that's the current definition of unconstrained ☺.	
<b>Q1.2 What is the project's name?</b>	OptIPuter [ <i>Optical networking Internet Protocol Computer</i> ]	
<b>Q1.3 Which agency funds the project?</b>	National Science Foundation	
<b>Q1.4 What field does your project belong to?</b>	At its core it is a Computer Science project, but as a reality check we've developed Geoscience and Bioscience applications using the infrastructure.	
<b>Q1.5 What is your job type?</b>	Project Lead	
<b>Q1.6 How long have you been a &lt;job type&gt;?</b>	5 years	
<b>Learning about discipline-specific goals and approach</b>		
<b>Q2.1 What are the main goals of your project?</b>	The overarching goal is to identify the bottlenecks in trying to take advantage of high-speed networking. What are the middleware capabilities that need to be developed that are still missing? What will the endpoints look like that connect to these high performance services? Will they be desktop computers? Will they be browsers?	
<b>Q2.2 How will the success of your project be measured?</b>	I guess it's going to be measured by the number of deployments of the capability at sites beyond the original researchers who worked on the project. Success will also be measured by the papers we generate; those are the normal measurements from NSF's standpoint.	
<b>Q2.3 What are the professional measures of success for you?</b>	If lots of people use it ☺. We track the number of users through a web interface where they download the software. So one measurement of usage is download numbers. We also interact with the people directly because they often want to do remote testing with us once they've got their software set up.	
<b>Q3 What are you investigating?</b>	We are trying to determine the correct computer and software infrastructure for developing a scalable display client that will be receiving high-resolution visualizations from a distributed cluster of resources. We're investigating the right architecture for that. There are several trade-offs associated with the various software architectures. So we're trying to determine the best trade-off, given the limitations of what a computer can do, what the networks can do, etc.	

Interview ID=16 5 September 2007	ANSWERS	ANNOTATIONS
<b>Q4.1 What is your method for investigating &lt;phenomena&gt;?</b>	<p>We look at any prior approaches that might be similar in concept, and think how that approach fits the way end users actually work (as opposed to just solving a technical problem.) And then if we determine there is still a gap, we think about what kind of architecture would also meet the requirements of the end user.</p> <p>A concrete example: consider possible software architectures that manage scalable tiled displays. In the past the governing question has always been, “How do you do distributed graphics to use up the entire tiled display to make one big picture?”</p> <p>But by talking to users from a variety of disciplines, we got the sense that people actually don't want to use these giant tiled displays for displaying a single image. They want to look at a variety of high-resolution information at the same time. So they want to treat it like a giant electronic poster board. In the past we always thought of it as one big, cool TV screen – like a single big visualization.</p> <p>You can think of the old way of doing things as what we used to do in MS-DOS. In MS-DOS you could only run one application at a time. You'd run it, and when you're done you'd quit it, and then you run another one. So, all the supporting architectures for tiled displays before we started investigating were focused on that model.</p> <p>But what we're doing instead is to have distributed clusters do high-resolution rendering completely offsite, such as on TeraGrid. Then we stream the pixels directly onto our tiled display. We then manage all the renderings in separate sub-windows that can be moved around. This approach really amounts to a real-time pixel routing problem. Normally you associate routing with packets in the network. In this case we're routing pixels, so we borrow some of the concepts of packet routing in networks and apply it to pixels.</p> <p>But again, a lot of this work was driven by our study of the way people work in front of walls, and the realization that things had fundamentally changed. This drove our requirements for the software architecture.</p> <p><i>[prompt asking for more detail on how the realization was formed]</i></p> <p>Well we built a prototype of a collaboration environment, and we did a user study. We actually built two prototypes and then we connected them over the network and we had two groups of users, one in each room. We gave the users certain tasks – information foraging-type tasks. We would say, “Find <i>X</i> and correlate the data with <i>Y</i>.” They had a whole bunch of screens and we watched the way they used the space. Based on our observations we realized that, “Oh, people don't want just the one big picture on the screen. They want lots of information all over the place.” We taped several students trying to accomplish several tasks.</p>	
<b>Q4.3 How do you keep track of interim results, if at all?</b>	<p>We started working with experts from the School of Information at the University of Michigan who have done a lot of prior research in group work. These are the people involved heavily with the Access Grid. They worked with us and we taught them how to build tiled displays and how to use our software. Then they worked with their research community within the University of Michigan to deploy some of these displays. The experts also began observing how people worked with the prototype at the University of Michigan, and they started giving us feedback.</p> <p>For example, they had deployed a display with the atmospheric sciences department. The atmospheric sciences people set up a class where the students had to make use of the tiled display for a school project. An interesting thing happened: the students turned that big display into a giant mash-up wall; we now call this “cyber-mash-ups”. So that's now where we see the real value of these high-resolution displays.</p> <p>It's funny because we came to that cyber-mash-up conclusion almost separately from the University of Michigan folks. During a conversation I mentioned, “What I'm starting to see is this evolution of wall displays becoming these kind of mash-up environments”. And he said, “Well that's very interesting, because that's what we are observing at the atmospheric sciences department when the students are using these environments.”</p>	

Interview ID=16 5 September 2007	ANSWERS	ANNOTATIONS
<b>Q4.4 How do you test work-related hypotheses?</b>	<p>The University of Michigan people constantly gave us feedback on user interaction issues. So as a result we changed a variety of interfaces, made the system as a whole more stable, which was important ☺. We were feeding updates back to them, in essentially real-time. They were getting feedback in real-time and they were one of the fastest absorbers of our updates. We started setting up an RSS feed, so every time there was a new software update people who were interested could get the latest version.</p> <p>[prompt asking about turnaround time on updates]</p> <p>Maybe a couple of updates a semester – every four months.</p> <p>I guess the addendum to this also is that the other deployment effort we've been working on is with UCSD. They have embedded this capability into the Rocks distribution. Part of the reason for this is we want to minimize the handholding required for user installs. Rocks has been credited with the ability to manage clusters. So we decided that it would be valuable to put our software into the Rocks distribution. The idea was that people could just buy a cluster, build a cluster, then plug in Rocks and have it rock ☺.</p> <p>The hardware requirements constantly change. Equipment gets cheaper and new stuff arrives, and you think, “Oh, this could be cool.” But the test deployments don't change as rapidly as the software. I think the software is about six months to a year ahead of the deployed versions. It's only in rare cases that people grab the latest version, like the folks in Michigan who want to immediately try stuff out in the community. It's nice for us because they function essentially as a lab for us. So we have both a development and a stable software release stream. The stable releases are more regular, like every six months or every year. For the research version, they tend to be irregular. Like anytime there's some major improvement that we'd like to get out, we make that available. Folks can just go to the subversion server and grab the latest one if they wanted to. We don't expect the average Joe Scientist to ask their system administrator to go and build it. We expect them to go to the Rocks version.</p>	
<b>Q5.1 In what ways do you interact with simulations in your work?</b>	<p>For a different project we interact with simulations. But the project I'm thinking of is unrelated to OptIPuter; it is a museum project.</p> <p>Oh – there is one other simulation-based project I forgot to mention that, while OptIPuter-related, is not visualization oriented. This is in collaboration with NASA Goddard. They work with atmospheric simulations, and we're trying to help them apply OptIPuter capabilities to help speed up their simulations.</p> <p>So in one example, the NASA simulations write all their data into <i>/tmp</i> and then another bunch of codes has to read it and do something with it. The reading and writing to <i>/tmp</i> is so slow that it becomes the bottleneck (because the hard drive is so slow.) So in this other application of the OptIPuter, we gang up clusters of computers, steal their memory, and use the memory like a giant RAM disk.</p> <p>Remember the old RAM disks that we used to have on our desktop computers? Part of the realization that this might be useful was when we thought about the fact that the average Macintosh has a gigabit connector on it. If you run out of RAM while using your laptop, instead of swapping virtual memory to disk, what if you swapped it over the gigabit interface to a neighboring laptop?</p> <p>So this was a compelling way to do things because moving data over a gigabit link is faster than storing it on your hard drive. Your hard drive probably has at the most 200Mb, 300 or 500Mb (if you are real lucky) write speeds. Whereas if you have a gigabit, that means that you can double the write capacity, just by writing it to your neighbor's computer.</p> <p>So we wrote a layer of software that hijacks their IO operation. Whenever they write to disk they are actually writing to our RAM disk. The RAM disk itself is actually a separate cluster of computers connected over a high-speed link. And so they can populate the RAM disk and then another process can quickly start. So that cuts down access times dramatically. Right now we are still doing a lot of testing and we are comparing it against things like PVFS and other standard clustered disk IO systems.</p>	
<b>Q6.1 Describe how you interact with data in your work</b>	<p>For the most part we read a lot of it and then generate pixels to look at it. Then we move it over long distances.</p> <p>In the tiled display case, data originate from a remote parallel disk system and then get transferred to, for instance, another cluster that does the rendering. Then it would take all the pixels and stream it into the tiled displays.</p>	
<b>Q6.3 By what mechanisms is access to your work-related data controlled?</b>	<p>Since we're dealing mostly with experimental data sets in the tiled display case, the access mechanism is through username/password logins on the remote servers.</p>	

Interview ID=16 5 September 2007	ANSWERS	ANNOTATIONS
<b>Q7.4 How do you locate available resources for use in your work?</b>	Everyone who puts together a tiled display and installs the software and rendering services is registered with our central server. So we can actually locate other tiled displays and other rendering sources through this registry. There is currently a single registry for all OptIPuters because the project is currently a research project, not a full deployment project.	
<b>Q7.5 What types of information do you need to know about a resource in order to determine if it is suitable for your work?</b>	I don't know the exact answer to that question because my colleague does all that.	
<b>Q8.1 What software do you currently use in support of your work?</b>	Programming languages like C++ Scripting languages like Python (for some of the lightpath control stuff) We use some portion of Globus – I don't know exactly which portion of Globus. I think we use some of the certificate stuff so could set up a secure way of turning lightpaths on and off without enabling just anybody to flip the mirrors on these networks.	
<b>Q8.4 What workflow tools do you use in your work?</b>	We haven't really focused a lot in that area yet because we are mainly a research project.	
<b>Q8.6 If the need for new software-based functionality arises in your work how do you acquire it?</b>	We get some requirements from people in the community who've installed our system. Both through mailing lists, but also from workshops or events that we've run, such as iGrid. Then there are meetings sponsored by organizations like the GLIF (Global Lambda Integrated Facility). So it's really a tightly knit community where once they start installing the software, they don't go away and bug the hell out of you. As far as acquiring new capabilities: sometimes we write them, and sometimes community members contribute. Like for example we have a group in Korea who wanted to put in support for HD video streaming, and so they wrote something to do that. Then another group in Amsterdam wanted HD streaming with a different piece of hardware, and they contributed by writing a module for that.	
<b>Learning about the user's problems</b>		
<b>Q9.1 What challenges do you face today in accomplishing your work-related goals?</b>	Funding. One challenge is that there is so much development happening now, in so many directions, by so many people, that it's becoming harder and harder to keep track of all the developments. Trying not to reinvent things and trying to leverage what's already there is a constant challenge. You don't know to what degree a particular project has really matured, or what the future of the project is. So let's say you decide that within the software you're developing you want to rely on this particular open source software that seems really cool. You need some assurance that this piece of software will have longevity before jumping into it. Or you may decide that you are better off writing it from scratch, which you really want to avoid if at all possible. That is the biggest challenge – finding compatible collaborators. Another challenge is working with the domain science community and trying to understand their needs. Trying not only to advance their science, but also to advance your own. Because one of the problems we face as computer scientists is that we are seen as the technicians for the domain scientists. The advancement of computer science is seen as secondary, as opposed to something that could be an equal partnership. Establishing this type of relationship takes a bit of education on both sides actually.	
<b>Q9.4 By contrast, can you provide examples of technologies you find very useful today?</b>	Really awesome things today include the mash-up technologies: Google mash-ups, Microsoft mash-up type technologies... I think we are just starting to scratch the surface of that potential. Having new user interaction modalities are really awesome. There's a lot of interest now with tabletop environments, or touch screen environments or multi-touch interaction environments. Coupled with the continuing drop in the cost of high-resolution displays – displays that are capable of doing stereoscopic computer graphics without having to wear glasses – those are pretty exciting. And I think it's also practical especially with disciplines such as the geosciences, where they have a real appreciation for the stereoscopic viewing capabilities. For technologies like the iPhone, it's pretty clear that these touch screen interfaces are here to stay because they are so compelling. We are just sort of scratching the surface with that – at least as far as scientific computing is concerned.	

Interview ID=16 5 September 2007	ANSWERS	ANNOTATIONS
<b>Q10.1 Can you think of any work-related tasks that decrease your productivity?</b>	It depends on how you define productivity. Report writing: that's one problem ☺. Then the usual funding issues where you're writing grants and waiting six months to find out if you did or didn't get it. Especially for solicitations that have less than a ten percent success rate – that is pretty counter-productive.	



## D.17 ● GRAM2 is kept alive by the need to interoperate with European experiments

Interview ID=17 10 September 2007	ANSWERS	ANNOTATIONS
<i>Pre-interview question: Do you interact directly with Globus software in your work today?</i>	yes	
<b>Establishing context</b>		
<b>Q1.1 Please provide a one-minute overview of your project</b>	FermiGrid is a project based at Fermilab. FermiGrid can be thought of as a campus grid – it provides one unified gateway for accessing the Open Science Grid, both inbound from the Open Science Grid and outbound to the Open Science Grid. Another important aspect of FermiGrid is that it provides a unified authentication and authorization structure for Fermilab resources. There are currently seven local clusters in Fermilab included in FermiGrid, with upwards of several thousand batch slots available at any given time.	
<b>Q1.2 What is the project's name?</b>	FermiGrid	
<b>Q1.3 Which agency funds the project?</b>	Department of Energy	
<b>Q1.4 What field does your project belong to?</b>	High Energy Physics and some other disciplines such as Astronomy and Non-Accelerator Physics	
<b>Q1.5 What is your job type?</b>	Computer Professional and Assistant Group Leader	
<b>Q1.6 How long have you been a &lt;job type&gt;?</b>	Ten months, with over seven years spent at my home institution	
<b>Learning about discipline-specific goals and approach</b>		
<b>Q2.1 What are the main goals of your project?</b>	One goal is to provide a unified point of access for inbound Grid jobs, in order to share Fermilab resources with the Open Science Grid. A second goal is to create a structure within Fermilab so that six or seven interest groups with big pots of computing can share each other's resources. This is by far the bigger focus: to share the resources amongst ourselves, as opposed to sharing resources with people outside Fermilab. A third goal is to have a unified systems support structure. A fourth goal is to provide a unified authorization and authentication structure. A fifth goal is to expose mass storage in a shared way. And goal number six, which is our main focus for this year, is making it all redundant. Making the infrastructure all highly redundant so we don't have any one single point of failure.	
<b>Q2.2 How will the success of your project be measured?</b>	We have various metrics, although not enough – we could always have more. We have metrics on: - how many jobs we host from the outside - how we are doing with regard to uptime - our reliability We also have our own internal user survey for satisfaction.	
<b>Q2.3 What are the professional measures of success for you?</b>	I'm evaluated on how much of our tactical research plan we get done, compared with how much we should have done.	

Interview ID=17 10 September 2007	ANSWERS	ANNOTATIONS
<p><b>Q3 What are you investigating?</b></p>	<p>First of all, I have a list of 250 action items that I need to finish at some point. Actually the exact number changes at any given time, but there are major categories.</p> <p>My area is working with batch systems and gatekeepers – providing the OS administration for those. I configure, set up and maintain the gatekeepers. I also work with users of the gatekeepers and the batch system.</p> <p>So within that, a major focus of my work this year is scalability. Our clusters are big enough now that we're hitting scalability issues. For instance, we have a 2000 [job] slot cluster, and we're having a real problem keeping it full. That is one of the biggest things on my agenda right now.</p> <p>A second focus is on authentication and authorization. Being a DOE lab, there are various requirements that Fermilab has to deal with in this area that other places may not have to. For example, ensuring that the user's GSI authentication gets moved within a job from place to place, and that it works everywhere and provides access to needed resources. Fermilab has long had Kerberos authentication, so it is complicated if we want to put up MPI jobs on the Grid. This is another of our long-term goals. Right now we have an MPI LAN that has nothing to do with the Grid, and we have Grid LAN that has nothing to do with MPI. We'd like to get the two talking to each other, and are in the early stages of that work.</p> <p>My third major focus and highest priority in the current fiscal year is making it all redundant and having failover capacity.</p> <p><i>[prompt asking for more information about the difficulties using all the computing power available]</i></p> <p>It's taking so much CPU power, both from the client's submission side and from the gatekeeper side, that we can't keep our 2,000 batch slots. Before we can load up all the jobs, some of them finish. So we can't get 2,000 jobs simultaneously submitted to the cluster – actually I think we've been able to reach that maybe once or twice.</p> <p>This is a joint scalability issue, in particular between GRAM2 gatekeepers and the <i>condor_schedds</i>. Is it the gatekeeper? Is it the <i>condor_schedd</i>? We're not prepared to say yet. But it's a very real problem and we're currently throwing hardware at it; we're trying to do as much as we can.</p> <p>At this point we're not at the end of our rope. But it's at a point where we really need to push the scalability and get it up there. Until just recently, the stakeholders themselves didn't have a submission machine that could actually throw 2,000 jobs at me, but now they do.</p> <p>This is all pre-Web services. GRAM4 at Fermilab is just in the initial testing stages, really. We don't really have any major stakeholders that have gone to it yet.</p> <p><i>[answer continued on next page]</i></p>	

Interview ID=17 10 September 2007	ANSWERS	ANNOTATIONS
<p><b>Q3 What are you investigating?</b> [continued]</p>	<p>Now let me look at my list and see if there are other things I ought to be investigating ☹... We can never have enough monitoring. There are monitoring issues to work on, but I'm mostly leaving that to other people.</p> <p>Really, at my level a lot of my work is actually sending email to other people and saying, "Please do this. Please do that. Please do the other thing." Prior to December 2006, it was me who was actually doing the things.</p> <p>My fourth area of work is not explicitly related to Globus, but is nonetheless relevant. That is the whole issue of provisioning and keeping everything running. The challenge of maintaining a very big and very complicated software stack on more than 3,000 machines is very difficult. The solutions we have in place now for managing this are not adequate. I send the instructions to the sysadmins and they say, "What? This is crazy." and I say, "Yeah, I know, but it's all we've got right now." So getting a very complicated software stack distributed and running on all these machines is difficult. But this is mostly not a Globus problem.</p> <p>In the two-and-a-half years of FermiGrid there has not been a time when we've had the latest software versions installed on all of our machines. We are still not up-to-date. And it is turning into a situation where you can't even use a distributed file system to get the software out there. There are more and more requirements, and more and more stuff has to be pushed out locally to every single compute node.</p> <p>Of course the Grid was sold in a totally different way when it first came out. It was supposed to just live on your batch system host and you wouldn't have to worry about it. The nodes wouldn't have to know about the Grid. In practice on the OSG this is not the case.</p> <p>The OSG stack for every single worker node these days includes all the Globus clients, such as globus-url-copy and the Web service equivalent. It includes Grid security certificates and certificate of authority files, which are used for authenticated file transfers. And then there are many more things the OSG has on top of Globus, the latest of which is gLExec, which is used for pilot [<i>technology from gLite</i>] jobs. Several of the big virtual organizations have this technology. You might have one guy sitting at FermiLab sending out Condor Glideins all the way across the Grid, and pilot uses gLExec to determine the appropriate userid the jobs should run under. In the big picture gLExec pushes responsibility for authentication and authorization to every single compute node. The software stack to support this is very complicated.</p> <p>So as an OSG gateway we get a software stack from the OSG and it stops at my door. I then propagate that into our own local space, sending email to people within Fermilab saying, "Please do this and this and this and this and this." And then send the email again, "No, you did it wrong. Do it again. This, and this, and this, and this."</p>	

Interview ID=17 10 September 2007	ANSWERS	ANNOTATIONS
<b>Q4.1 What is your method for investigating &lt;phenomena&gt;?</b>	<p>Working with batch systems:</p> <p>The design of our site gateway basically emulates the Condor job manager that comes with Globus. So a big part of building the site gateway is writing something that accepts jobs like Condor does and then forwards them by means of Condor to our local clusters. So the gateway doesn't actually execute the job itself; it takes the GRAM2 job and resubmits the job to one of the sub clusters. That's the basic technology. Myself and another person in my group wrote most of the current implementation.</p> <p>So as to the day-to-day work, we fix the squeaky wheel when we see jobs in the gateway are getting held because of something. We monitor and when jobs begin to fail, we look to identify the cause. One common problem is when new users make assumptions about how Condor ought to work that we didn't anticipate. In some cases this uncovers problems in our code, which we diagnose and then fix. We have a series of test jobs, so when problems arise we first run our tests and make sure that our tests pass. If our tests fail that helps us identify what's broken.</p> <p>In addition to monitoring the system, we also work directly with users. They come to us and say we have XXX problem. A significant amount of user problems related to Globus or associated Grid stuff is simply that the client is configured incorrectly.</p> <p>Our approach for managing a deployment as big as ours is to have only a few approved client installations that receive full support. We promise they will work, and we promise to answer questions after the user works with them. People are welcome to install clients on their own desktop if they want to, but they will get better support on the machines where we installed the client and can log in and watch their job from start to finish.</p> <p>As to the scalability work:</p> <p>When we hit a bottleneck we try to figure out where it is: the CPU, the disk, or the network – it's usually one of those three.</p> <p>Then we look and see if anyone else has an installation as big as ours. We look to see what they are doing and then try to do the same thing they did, step by step. We work until either we resolve the particular scalability issue, or we throw more hardware at it.</p> <p>For instance just last week we split off the batch master from the gatekeeper – that was our first step. Second step if that doesn't work will be to upgrade the Condor version. If that doesn't work then we have to get bigger and faster hardware on our gateway.</p> <p>So we take an incremental approach to tackling issues. Has anybody else done it this big? If so, how did they try to address it? If no one else has done it this big, then we also ask ourselves if we should be trying to do it that big.</p>	

Interview ID=17 10 September 2007	ANSWERS	ANNOTATIONS
Q4.2 How do you work?	<p>We don't document our work as well as we should, but we do have an electronic logbook that we use to keep track of what we did.</p> <p>Within Fermilab we meet regularly. The other big clusters are at Fermilab – the guy maintaining a bigger cluster than me is down the hall, and another one is upstairs. I know what they're doing because I can see them in my monitoring system, and I try to work accordingly.</p> <p>With respect to colleagues outside of Fermilab, the Condor team is available. It's very rare that we need to actually go to the Globus list. Although I am signed up to some number of the Globus lists, my typical use cases are Open Science Grid-based. There is enough expertise within the Condor team (or within the OSG list at large) about our use cases, so we've been able to get the answers we need from them.</p> <p>I'd say within the last six to ten months we've been seeing more Globus developers showing up on some of the OSG calls and being there as a resource, if needed. I know where to find them if I need them.</p> <p><i>[prompt asking about the approach for working on authentication and authorization tracking]</i></p> <p>That's not so much tracking as it is new development. So when I'm talking about authentication and authorization tracking, I'm not talking about what's included in the Globus Toolkit itself, because we have a pretty good handle on that. (Although we do have people in my department that are in dialogue with Globus people regarding the next version of XACML. That discussion is above my pay grade ☺.)</p> <p>What I'm talking about is authentication within the batch system, particularly for most of our clusters using Condor. Condor out-of-the box doesn't authenticate the daemons. But you can make it do it and it's recommended. So that's one of the projects we have on the table right now: to make our Condor daemons use GSI authentication to talk amongst themselves. We're trying to figure out the best way to do that. We're investigating if we can do it without getting a GSI host certificate for every single node. We think we have a solution, but we haven't tried it yet.</p> <p>Also we're working on getting MPI authenticated jobs in place. We haven't yet decided our approach for this. We're considering two major alternatives:</p> <ul style="list-style-type: none"> <li>- Is it cheaper to just buy extra hardware and put up a private NAT over which to do <i>rsh</i> among the MPI jobs like everybody else does?</li> <li>- Or should we send a Kerberos along with the job and do a Kerberos authenticated thing?</li> </ul> <p>This is one key thing at Fermilab that most other labs don't have: every single worker node at Fermilab is on the public Internet. There are no NATs and very limited firewalls. The way we've survived up until now has been requiring Kerberos 5 authentication on everything we do. But the Grid obviously doesn't use Kerberos 5 authentication.</p> <p>So you have to figure out some way to let these things talk amongst themselves. Will it be some kind of a VPN that's GSI authenticated? Those are out there – they exist. Should we put up a real private network? Or perhaps we should just send Kerberos authentication around like we used to do two years ago before we started doing Grid stuff. So it has yet to be determined what we will do. It's a big investigation to figure out what's the best way to do that.</p> <p><i>[answer continued on next page]</i></p>	

Interview ID=17 10 September 2007	ANSWERS	ANNOTATIONS
<b>Q4.2 How do you work?</b> [continued]	<p>[prompt asking for more detail on the redundancy and failure-related work]</p> <p>We're going to have to have a meeting with the Globus folks pretty soon on this, actually. Somebody else tried the approach I was planning and found a couple of roadblocks in the way.</p> <p>The first part of the redundancy approach we have the hardware for and are in test mode already. We are targeting our authentication services, such as VOMS, GUMS and a FermiGrid service called SAS, the Site Authorization Service. For redundancy we're setting up two physical machines, each with four virtual servers. We're using Xen to do the virtual server. We talked to the leader of the Globus Virtual Workspaces service and received some advice in the early stages that helped us figure out what to do there. It was very helpful.</p> <p>So the idea is no matter whether they're virtual or real servers, we're going activate all of them. So we'll hide a MySQL server behind there. We tested all of the pieces and we're just now in the integration stage of putting it up and trying to see if it all works together.</p> <p>The next stage is doing failover for the Globus gatekeeper and the Globus GRAM4 gatekeeper. That problem's tougher, as it can only be an active/passive thing. You would have one gatekeeper that's active all the time, and another OS instance that could read that filesystem and be ready to take over should the first one fail. This work is only in the planning stages at this point; we haven't done much active work on it yet. But we have to get it done sometime before September of '08, according to our project plan.</p> <p>In the early stages of writing this plan I did talk to a couple of people. I talked to the Globus GRAM lead and the Globus administration lead a couple of times. We had at least one meeting before that. We'll need to meet some more because right now we only know how to forward GRAM2 jobs with our inbound gateway.</p>	
<b>Q9.1 What challenges do you face today in accomplishing your work-related goals?</b>	<p style="text-align: center;"><b>Learning about the user's problems</b></p> <p>My challenges are about half organizational and about half technical. So half of it is figuring out what is the right technical thing to do and the other half is figuring out who are the right people are to talk to. Trying to figure out a way to get the technology out there and keep it deployed.</p> <p>From a technical standpoint, debugging information could always be better. Documentation could always be easier to find. In the effort of deprecating GRAM2, the Globus documentation has been made very hard to find. At least it was the last time I looked a few months ago – I haven't even checked recently.</p> <p>The other thing is Globus' nasty habit of (at least one time in three, and sometimes more) deleting the file you would like to see before you can get at it. This is with regard to debugging GRAM2.</p>	

Interview ID=17 10 September 2007	ANSWERS	ANNOTATIONS
<b>Q9.2 What types of information do you need in order to address the challenges you face today?</b>	<p>In general, it would help us to know the assumptions that Globus developers are making on the various files. I'm referring to what are they doing with locking and where the state of the gatekeeper is living (for both Web services and pre-Web services.) I know the broad strokes, but we'll need to know a lot more detail when we do the redundancy work. We'll be emulating the service, and we need to know as much as much of nitty-gritty implementation details as we can. Right now we find these things out by trial and error.</p> <p><i>[prompt asking for more information on what it means to "emulate a service"]</i></p> <p>Take our work with our job forwarding as an example:</p> <p>What we did is we took a file that lives down in a Globus library, condor.pm, and rewrote it to do what we wanted it to do. We're basically emulating the pre-Web services "2119/jobmanager-condor" interface <i>[fragment of the conventional network address for remote Condor jobs]</i>. If you send a job to that on FermiGrid, it's not really Condor underneath the covers, it is our own proprietary system.</p> <p>This works most of the time, but we've picked up quite a few surprises and over time we've actually been able to implement a lot more of the functionality than we thought we would. Basically the user says, "I need this." And we say, "Okay, we'll try." And sure enough, it works.</p> <p>One problem we have in our current implementation:</p> <p>It seems like always somehow NFS is involved in some very sticky way when we're dealing with GRAM2 and it's not a pleasant experience. The biggest hurdle that we overcame to get to where we are now is by throwing hardware at the problem and getting a BlueArc NAS server, which is a very, very high capacity NFS appliance. Before that, NFS was crashing more than monthly, triggered by the kind of NFS activity that GRAM2 does. We still crash every once in a while – maybe once every other month or so – but nowhere near as bad.</p> <p><i>[prompt asking if they understand what's triggering the failures]</i></p> <p>We have some ideas. In short, GRAM2 is doing hard links across NFS, and either the NFS client-of-the-day or the NFS server-of-the-day is not always reliable enough to implement that right.</p> <p>With the BlueArc in place the crashes have to be caused by the NFS client because there is no OS on the BlueArc server at all. These things are fixable if you have time to customize your kernel and work on it a month or two, but we don't have time.</p> <p>We're well aware that GRAM4 has gotten around this problem, but given our stakeholders it's unlikely that we'll be rid of GRAM2 any time in the LHC era. I expect we'll have to keep it going for at least five years, maybe more.</p>	
<b>Q9.4 By contrast, can you provide examples of technologies you find very useful today?</b>	<p>Globus itself is useful.</p> <p>Another example is the one I just mentioned is the BlueArc NAS server. That's the difference between being able to do what we do and not.</p> <p>Eventually I want my grid software to be like my Web browser. I don't want to have to go out and find the Grid and download 800 megabytes of code and build it from source. I want it to just be there.</p> <p>In the early days of the web you had to go and download your browser from NCSA, build it from source, and then add the hosts you were going to visit into your <i>etc/hosts</i> file. Nowadays this is seamless and you don't have to compile code. The browser that comes with your machine pretty much works, and it's in the OS and nobody is actually making money on it any more.</p> <p>In ten years if we are with Globus and Grid where Web browsers and search engines are now, I think it will be good.</p>	
<b>Q10.2 Describe repetitive tasks associated with your work</b>	<p>Well, there are a lot of repetitive tasks by the nature of my job. We have some methods of automation, but not as good as they ought to be. Such tasks include:</p> <ul style="list-style-type: none"> <li>- hunting for dead processes</li> <li>- cleaning out dead files</li> <li>- ensuring all of software on the compute nodes is in synch</li> <li>- cleaning dead jobs out of the batch system</li> <li>- distributing authorization userids everywhere</li> </ul> <p>Those are certainly things that take up a lot of my time right now.</p>	
<b>Learning about the Globus user experience</b>		
<b>Q11.1 Which Globus data components do you directly interact with in your work today?</b>	<p>GridFTP</p> <p>RFT because we're just beginning to use that because GRAM4 uses it. So with the testing we're doing, we're using RFT.</p>	
<b>Q11.2 Did you install the &lt;component&gt; client yourself?</b>	<p>It comes as part of the OSG stack (and did in the last release too). It pretty much installs itself. I haven't had to do anything to make it work – it just worked.</p>	

Interview ID=17 10 September 2007	ANSWERS	ANNOTATIONS
<b>Q11.4 How many people currently use your &lt;component&gt; server</b>	RFT: of our user base, very few. To my knowledge, it's the people who are testing GRAM4 in the OSG. I don't know of any production user group that's actually gone to GRAM4 yet.	
<b>Q12.1 Which Globus security components do you directly interact with in your work today?</b>	delegation service, MyProxy, GSI certificates I'm not sure what the Community Authorization Service is. I'm interpreting the term "delegation service" to mean what happens under the covers when the GRAM2 gatekeeper delegates proxy. If you're talking about the GRAM4 equivalent, at this stage I've only used it as part of GRAM4 testing.	
<b>Q12.4 How many people currently use your &lt;component&gt; server</b>	MyProxy: We have quite a few, actually. There are a couple of beginner communities using MyProxy. The CMS experiment and the D0 experiment use MyProxy to send proxies out to their jobs across the Grid, so it's fairly widely used.	
<b>Q13.1 Which Globus execution components do you directly interact with in your work today?</b>	GRAM2, GRAM4 We're thinking about MPICH-G2 as a possibility for authenticating MPI jobs internal to a cluster. We're not yet sure if it's the right solution for us.	
<b>Q13.4 How many people currently use your &lt;component&gt; server</b>	GRAM2: Many. I think there are approximately 400 unique users that have run a job against their server since we started. GRAM4: Still at a level of 10 or 20 users at the site.	
<b>Q14.1 Which Globus information components do you directly interact with in your work today?</b>	We're still sort of using a stub of MDS2. We have MDS4 but haven't really looked at it yet. And likewise with WebMDS. I have never heard of the Trigger service.	
<b>Q15.1 Which Globus common runtime components do you directly interact with in your work today?</b>	My only interaction with Python WS Core was trying to install it and seeing it fail. I know how to fix it, but haven't gotten around to it. We're using the Java WS Core indirectly. Some parts of Condor are using it embedded. We use the C WS Core through our use of globusrun-ws.	
<b>Q16 Why do you use &lt;component&gt; instead of an alternative technology?</b>	GRAM2: One of the big things keeping GRAM2 alive is interoperability with European experiments. They are not going to GRAM4, as far as I know. Another issue has been that it's taken a while to get all the various special OSG authentications working seamlessly with GRAM4, although I believe they all are now. I refer to the various callouts and whatnot.	
<b>Q17 What are the major challenges you face using &lt;component&gt; today?</b>	GRAM4: We just went through an issue that turned out not to be a GRAM4 issue, but a Condor-G issue. It took us two or three weeks to debug that and identify the problem. It turned out that: - Some authentications and authorizations didn't play nice together. - Also Condor-G was making calls when it ought not to (or not making calls when it should). So one GRAM4-related challenge would be working with the external callouts that are common in OSG. The challenge that we have to solve eventually is try to figure out what the GRAM4 analogue of the GRAM2 forwarder will look like. How are we going to implement in GRAM4 what we've done for GRAM2 for FermiGrid: - Will we just put GRAM4 in front and keep GRAM2 in the back? - Will we try to do a GRAM4-to-GRAM4 thing? We haven't decided yet. We also understand there is an issue when the GRAM4 state thing is mounted on a shared file system. This could really put a crimp in what we were trying to do with our high availability. So that is also a big issue that we've got to get a handle on quickly in the next few months. Also GRAM4 is a huge resource hog. It takes 700 megabytes of memory to sit there and do nothing. There is the other thing with GRAM4 (and GRAM in general). It has taken some work to get the OSG accounting interface to play nice with it. I'm not sure if we've got it right yet or not. There are also issues that we have with GRAM, be it 2 or 4, with regard to job auditing. It always takes investigation into at least three log files to get a full picture of what has happened with a job. Not all of the authentication information is in the right place always, etc. There could be more information.	
<b>Wrapping-up</b>		



Interview ID=17 10 September 2007	ANSWERS	ANNOTATIONS
<p><b>Is there anything you'd like to say to the people who build software for use by people like you?</b></p>	<p>Better testing, make it more scalable, and give us the hooks to make it redundant if you don't do it yourself.</p> <p>We've done a lot here at Fermilab with site penetration testing, and sent a few bugs back to Globus along the way. There is more to be done there. I'm very suspicious that there are security holes in the clients, be they GRAM2 or 4 that we'll find eventually.</p> <p>Those are the main things.</p> <p>I guess the other thing I would add is to keep GRAM2 around in addition to GRAM4, especially in the Open Science Grid at large. The Open Science Grid doesn't have an information system to reasonably send GRAM4 stuff around.</p> <p>Our whole information system right now is tied to GRAM2.</p>	

## D.18 ● We provide an appliance for each cluster that acts as a parallel head node

Interview ID=18 10 September 2007	ANSWERS	ANNOTATIONS
<i>Pre-interview question: Do you interact directly with Globus software in your work today?</i>	Yes	
<b>Establishing context</b>		
<b>Q1.1 Please provide a one-minute overview of your project</b>	<p>My group supports the HPC infrastructure for chemists, physics, engineers and other groups on campus. Our users run jobs on multiprocessor machines like Beowulf clusters. Historically we have helped users debug and improve the performance of their code. Now we are using the Globus Toolkit software to enable users to submit jobs into the multiprocessor system in a way they is both familiar to them and easy to use.</p> <p>The Globus portion of the project is known as the UCLA Grid [<a href="https://grid.ucla.edu">https://grid.ucla.edu</a>]. UCLA Grid is a portal where people can log into a Web interface and choose any particular cluster where they have access. Once they choose a cluster then a series of applications will be listed. It could be their own application, or their department's applications or commercial applications. The users choose the application, the number of processors, memory requirements and time. Then they submit the job. The Globus-based Grid submits to that particular cluster and then returns the output to them when it is completed.</p> <p>We also have something called "pool users" where you don't need an account on a specific cluster. This UCLA Grid itself determines where pool user jobs run. The job executes and the user gets the output when it is completed.</p> <p>People can also use the Web interface to monitor the status of their jobs.</p>	
<b>Q1.2 What is the project's name?</b>	UCLA Grid Portal	
<b>Q1.3 Which agency funds the project?</b>	University of California Office of the President	
<b>Q1.4 What field does your project belong to?</b>	Information Technology	
<b>Q1.5 What is your job type?</b>	Programmer Analyst	
<b>Q1.6 How long have you been a &lt;job type&gt;?</b>	Five years	
<b>Learning about discipline-specific goals and approach</b>		
<b>Q2.1 What are the main goals of your project?</b>	<p>We want to share resources among the cluster users because in UCLA there are so many clusters. We have 15 nodes here, 15 nodes there, 20 nodes over there, etc. Individual researchers own them; we found that they are not used all the time. The usage of a cluster goes up when people have an interesting project to do and after they are done they do something else, like analyze the data. During those times, lasting maybe days or weeks, the clusters are not used.</p> <p>So the thought is if we integrate all these clusters through the Globus Toolkit with a Grid-based portal, then users and the cluster owners can effectively share their resources without having to create a user ID for some unknown person. If the Grid portal can take care of everything through a guest user account then all the security details can be taken care of with Globus. That is our primary goal.</p> <p>So the goal is to share resources that would otherwise not be available to ordinary researchers. They don't have to go and buy their own equipment. They can get their research done even if they don't have lots of money. Another thing is that software like Abacus, MATLAB, etc. have expensive licenses. If you only need a resource for two months then you don't want to buy that license for a year. If a cluster has that license then the cluster owners can let them use it without having to pay for an additional license. These things will be much easier in a Grid environment.</p>	

Interview ID=18 10 September 2007	ANSWERS	ANNOTATIONS
<b>Q2.2 How will the success of your project be measured?</b>	<p>We check the number of users logging in and what kind of things they do. Using the Grid portal you have access to multiple clusters and people can transfer files easily among them. So we measure how they are using the Grid and how many jobs they submit. We do monthly accounting. Basically we just need to prove that it's being used; our funders don't dictate usage requirements.</p> <p>We are still developing more features based on user input. For example, some users have input files for the application runs because they may be dealing with hundreds of variables. We have freshman users so it's possible for them to make a mistake in their input file. In this case they submit the job and then after one hour the scheduler runs it and they get an error message saying that the job could not run because of conflicts in the input. So we built an input file creator GUI where they just choose parameters and then the Grid portal will assist them in putting the correct combination. Users told us this feature would be helpful to have so we have been adding that feature into our Grid Portal recently.</p>	
<b>Q2.3 What are the professional measures of success for you?</b>	<p>How much people use what we develop. That is a measure. If the project sees a lot of use I will also have succeeded.</p>	
<b>Q3 What are you investigating?</b>	<p>We investigate general user requirements. What kind of applications does the user need? What kind of architectures do they need? One thing we do is that we benchmark users' executables and applications for various architectures and give them the best architecture for their application. For example, their code may run better in AMD Opteron, as opposed to an Intel architecture. Those kinds of things we do as an investigation.</p> <p>We also write software as part of the portal work. In the UCLA Grid Portal software we use open source software like the Globus Toolkit, Apache Tomcat, Java, Gridsphere, MySQL and Perl. Our Grid Portal software written by us combines all of these together. The main developer is one of my co-workers; he writes the Java part of the software. I work on the linking part of the software – integrating everything between the portal and the cluster nodes.</p>	
<b>Q4.1 What is your method for investigating &lt;phenomena&gt;?</b>	<p>Whenever a department gets a grant they come to us for advice, explaining that they want to buy some equipment for computing. They seek our expertise. We give them our benchmarks and the types of processors and types of interconnects. For example if it is a parallel job we ask them to buy an InfiniBand, because compared to GigE it will perform better.</p> <p>Most of the people already know their applications so the only information they need from us is architectural recommendations. Others want to know which scheduler will be best for the machine, so we also provide advice on that. So we give them suggestions on what to buy and what kind of compiler they need. Once they order the equipment we will put the operating system on it, assemble the cluster and run it for them. So we do everything from hardware to software.</p> <p>Another aspect of our work is adding new capabilities to the Grid portal in response to user requests. Every Tuesday the IT team has a Grid meeting to discuss our opinions and experiences; often people will report on what users have been telling them about the features they like or would like to have. The team then has a general discussion and if a feature is feasible we include it, otherwise we keep it waiting until it is feasible.</p> <p>We have been working on the Grid portal for the last 2.5 years. We continue to add new tools but these days change happens mostly when software updates become available, such as new releases of Gridsphere or Globus.</p> <p>We maintain a production system as well as a parallel test system. We continue to double up on our test system and when it comes at a certain stage we integrate changes into our production system. That's how we operate. We keep our production system untouched for six months at a time and disturb it only if there is a security issue or something like that.</p> <p>So we are currently in maintenance mode with occasional software updates and new user tools being integrated into the system. At this point updates mainly happen if something fails at the operating system level, such as some new clusters coming in with a Fedora Core version we haven't yet seen. So if we see failures during our tests we put the fix immediately into the production system.</p>	
<b>Q4.3 How do you keep track of interim results, if at all?</b>	<p>Not in electronic way because my coworker and I do most of this work so we know our progress. Occasionally some people also jump in, and between our emails we keep track of things.</p>	

Interview ID=18 10 September 2007	ANSWERS	ANNOTATIONS
<b>Q4.5 How do you document your results?</b>	<p>Records of everything we maintain are stored in a database, so we have complete details of what applications are there. When we're in the midst of integrating new applications we correspond directly with users via email. We keep this correspondence between ourselves because it is not of much importance to other people. Such communications involve requests to change executable paths and things like that; nobody else needs to know those types of things.</p> <p>So everything about each application goes into a MySQL database. The application name, its path, where it is installed, whether it is serial or parallel and other common fields. A view of this data is accessible to end users via the Web. If you have a user ID on the portal, once you log in you can click the application button and get information about the application. This allows the user to see what applications are available.</p> <p>End users don't see information like the application path because they don't need to know that. End users don't manually run the application; we take care of that part. All we ask the users to provide is their input file, the number of processors and the desired time. So the database is back-ending the Grid portal. The paths of the executables periodically change, such as when a new MATLAB version is installed. If a user needs a specific version, multiple applications are available (like MATLAB 7, MATLAB 6, etc.) so users can choose.</p> <p><i>[prompt asking how updates are made to the application data in the database]</i></p> <p>There are two types of logins into the portal. One is the Grid admin login and the other one is the end user login. Grid admins have privileges that allow them to update the database. Admins manually go into that table and update it. We don't actually give it MySQL command or anything because those are all coded into our applications, our portal button. The SQL commands are wrapped in Java.</p>	
<b>Q5.1 In what ways do you interact with simulations in your work?</b>	<p>We run simulations to see how much load the system can take. We do things like submit many jobs at a time to see whether the system breaks. We mostly do this on our test system. We normally don't mess with the production system; we let it run as it is.</p> <p><i>[prompt asking if the same resources are behind the test and production systems]</i></p> <p>The clusters are the same for the test and production systems. We have something called an appliance node for each cluster. They're like parallel head nodes. Our policy is that by joining the UCLA Grid, cluster owners should not be asked to change anything on their cluster. It should run how it was running previously. So we ask resource owners to install a parallel appliance node, which is similar to the head node.</p> <p>That appliance node is accessible only to two machines: one is our Grid portal machine and other is our test Grid portal. Everything coming from the Grid Portal and the test Portal is very secure and their cluster is not exposed to anything by joining the Grid Portal. Since all the appliances are open to both portals, we can do the testing from our test portal as well as from the production portal.</p> <p>The appliance node is actually a physical machine that we provide to them so they can join our Grid. We give them the hardware because otherwise they would have to provide it themselves, and they might think of it as losing a compute node. So this way we give them some incentive. The flipside is that if something goes wrong, they can't blame us because we didn't alter their system.</p> <p>One interesting side effect is when their cluster head node goes down sometimes the only way the users can access their data is through our Grid. Because the Grid appliance is still connected to the home directory and the compute node they can continue to work as if nothing has happened. This situation has happened many, many times. Sometimes users do something on the cluster head node and then it dies (because of hardware failure or things like that.)</p> <p>So it takes sometimes a day for the head node to come back. And if the node is down and I'm a user who needs research data right now, like for a presentation, the only way I can get it is through the Grid. You're supposed to back up everything but most people don't do that right. So the appliance idea seems to be very good. Appliance nodes normally don't fail because they don't do anything. The only thing an appliance node does is run the Globus Toolkit and talk to the Grid Portal. Occasionally it talks to the index service. No user directly runs anything on it so it doesn't get overloaded or anything like that, whereas people are compiling and running test jobs on the head node. The head node can be very busy.</p> <p>So none of our appliance nodes have failed in the last three years. Some of the appliance nodes are still running Fedora Core 2 because we never had to change anything.</p>	
<b>Q6.1 Describe how you interact with data in your work</b>	<p>We don't normally have anything to do with big data assets. Our data is normal application data. Nothing in the terabyte scale; it's gigabyte scale.</p>	

Interview ID=18 10 September 2007	ANSWERS	ANNOTATIONS
<b>Q7.1 What resources do you use in your work today?</b>	<p>We own some of the resources in the Grid, but not all. We ask the cluster owners when they join the Grid that what resources they have and what they want to expose. It is their choice if they want their Mathematica license, for example, to be exposed to other people. And even if they expose it, we give them the choice of specifying how many nodes guest users have access to. They seem comfortable with that level of control.</p> <p>The cluster owners have admin rights to the appliance. We tell them it's their choice whether or not they want us to maintain the appliance. Most of the people let us run it, though some run it themselves.</p> <p>The NFS server for the cluster is cross-mounted on the appliance so it has all information about the cluster. And the appliance has an index provider that is transmitting information about all the applications. So the Grid Portal knows everything about the queue system and the people who are running in the queues. The Grid Portal gets information from all the appliances.</p> <p>We also have a VNC interface. One thing that people can get through our Web interface is X Windows. If they want run a GUI or if they to plot a graph, if they want to plot a molecule – in most portals they can't do such things. We added the xVNC service so they can see that through the Grid portal now. So they can get an ssh-like window as if they have logged in to that machine through ssh.</p>	
<b>Q7.5 What types of information do you need to know about a resource in order to determine if it is suitable for your work?</b>	<p>We need to know what kind of scheduler they are running on the cluster, because depending on whether its Sun Grid Engine, PBS or LSF, we need to create appropriate command files. And also when we submit the Globus GRAM job we need to know what kind of scheduler is on the other end.</p> <p>We also need to know the full path of their application.</p>	
<b>Q8.1 What software do you currently use in support of your work?</b>	The operating system of all our systems is Linux except one cluster that runs Mac OS X. The appliances are Linux because it doesn't matter what software they run. Most of the software the users provide themselves, though we provide common software like compilers, MATLAB, Mathematica, etc.	
<b>Q8.2 What scripting languages have you used in the past year?</b>	We use shell and perl. We don't use much Python.	
<b>Q8.3 What programming languages have you used in the past year?</b>	Users ask for C, C++, FORTRAN, and very rarely Java.	
<b>Q8.4 What workflow tools do you use in your work?</b>	Not anything in particular.	
<b>Q8.5 What parallel computing tools do you use in your work?</b>	<p>We use two main APIs. One is MPICH for GigE and the second is MVAPICH from Ohio State for the InfiniBand.</p> <p>We are experimenting with Open MPI. We used to use Lava MPI. So we have two kinds of interfaces because that one is for InfiniBand and other is for GigE. We are advising all of our parallel users to switch to InfiniBand because of the high performance you can achieve with it.</p>	
<b>Q8.6 If the need for new software-based functionality arises in your work how do you acquire it?</b>	<p>If the new application is open source it's no problem. We install it almost the same day of the request.</p> <p>But if it is commercial software somebody has to pay for it. In that case we need to determine how many people will be using it. If there are many people who are going to use it and it is expensive, then it will take some time.</p>	
<b>Learning about the user's problems</b>		
<b>Q9.1 What challenges do you face today in accomplishing your work-related goals?</b>	<p>Performance is a big thing. Mostly computing-related performance – GigE was a bottleneck for a parallel computing. With InfiniBand we seem to have solved that part.</p> <p>And another thing is that the bandwidth speed of NSF reads and writes is still an issue. We don't have much experience with parallel filesystems. That is one area we will be experimenting in future to see if it solves our problem. For our new clusters we are going to experiment with PVFS and Lustre.</p>	
<b>Q9.2 What types of information do you need in order to address the challenges you face today?</b>	Most of the information these days you can get from Google. This allows you to learn from different people's opinions and expertise.	
<b>Q9.4 By contrast, can you provide examples of technologies you find very useful today?</b>	Of course one thing we like is Globus and Grid. It has come a long way to help us. And one thing that which made the parallel computing much better performance is the InfiniBand technology. I find it very useful these days.	

Interview ID=18 10 September 2007	ANSWERS	ANNOTATIONS
<b>Q10.1 Can you think of any work-related tasks that decrease your productivity?</b>	Meetings take a lot of our time.	
<b>Q10.2 Describe repetitive tasks associated with your work</b>	Not anything in particular because in this field everything is changing, so we don't get to do many repetitive things. Everything will change in six months. You just need to keep up with the changes.	
<b>Learning about the Globus user experience</b>		
<b>Q11.1 Which Globus data components do you directly interact with in your work today?</b>	GridFTP	
<b>Q13.1 Which Globus execution components do you directly interact with in your work today?</b>	GRAM2, GRAM4	
<b>Q16 Why do you use &lt;component&gt; instead of an alternative technology?</b>	<p>GridFTP: We are using the X.509 certificates for authorization.</p> <p>GRAM4: Because our authorization is through X.509. [prompt asking why the user moved from GRAM2 to GRAM4] Because Globus developers moved away from GRAM2. But we are still supporting the Open Science Grid for the Physics department, and that software still runs the gatekeeper and things like that. It's not part of our direct work; we just helped them install the software. So indirectly we still use it.</p> <p>Index service: Index Service lets us broadcast whatever we want. It's easily configurable and we are comfortable using that. Also we don't need to install any additional software because it is part of the Globus Toolkit.</p> <p>X.509 certificate infrastructure: We find it more secure.</p>	
<b>Q17 What are the major challenges you face using &lt;component&gt; today?</b>	<p>GridFTP: We are OK with it. We don't have a huge data. We don't have much complaint about the GridFTP performance.</p> <p>GRAM4: It could have more features but I can't give you the specific examples off the top of my head.</p> <p>Index service: Nothing I can think of.</p> <p>X.509 certificate infrastructure: We sign and keep all of the certificates ourselves. So we don't have any challenges within our portal. But if you have to go to somebody else's portal than all the trust relationships get complicated because they have to trust you and you have to trust them. It is a challenge if you want to interact with another organization.</p>	
<b>Wrapping-up</b>		
<b>Is there anything you'd like to say to the people who build software for use by people like you?</b>	More documentation is needed. Most of the time the main pages and documentation are good but some applications lack it, so it's a general thing.	

## D.19 I would like sites to serve 100,000 or more end-users per week

Interview ID=19 11 September 2007	ANSWERS	ANNOTATIONS
<i>Freeform pre-interview discussion</i>	<p>Before this job I was a professional web services middleware developer, creating a middleware solution that could be down no more than fifteen minutes in a single year. The application was for manufacturing.</p> <p>I also spent a lot of time developing engineering applications for engineer simulation work. That led me into developing real time systems – real time system simulators, stuff like this. I’ve done web services to robots before. This work included dynamic discovery, where a robot plugs into a line and it will automatically be discovered.</p> <p>My background is user services. I started out at my home institution doing that, working with OSG, getting Globus to run for people with CMS, and so forth. Then I got involved with TeraDRE, which involves distributed rendering. TeraDRE used to be a Maya MEL Script Integration. I’ve taken it to another level in terms of user accessibility with Web Start. The users like to have the local interactivity that a web browser can’t give you (i.e. drag-and-drop capabilities.) We had a portal version of it for a while, and I was just primary working on the Web Start as a technology demonstration. With Web Start we find that yes – there is life outside the browser.</p>	
<i>Pre-interview question: Do you interact directly with Globus software in your work today?</i>	Yes	
<b>Establishing context</b>		
<b>Q1.1 Please provide a one-minute overview of your project</b>	The TeraDRE is a distributed rendering environment. What we mean by rendering is that we take models that are primarily generated from scientific data or from computer graphics and render the frames to make an animation.	
<b>Q1.2 What is the project’s name?</b>	TeraDRE	
<b>Q1.3 Which agency funds the project?</b>	The TeraGrid	
<b>Q1.4 What field does your project belong to?</b>	Visualization	
<b>Q1.5 What is your job type?</b>	System Architect and Programmer	
<b>Q1.6 How long have you been a &lt;job type&gt;?</b>	Six months on this project, 1.5 years at my home institution	
<b>Learning about discipline-specific goals and approach</b>		

Interview ID=19 11 September 2007	ANSWERS	ANNOTATIONS
<b>Q2.1 What are the main goals of your project?</b>	<p>To make it easy for people who are not computer programming experts to render their jobs very quickly. And also to provide a certain level of flexibility to add more rendering technology. Typical users might take a Maya model and render two minutes of animation out of it. Other users might take some scientific data and generate a model out of it. For example, using VMD they might generate a molecule and then watch the molecule's protein fold. They generate a pre-image POV-Ray model out of that and then they would submit this to DRE to actually create the animation. The type of data that's being visualized can be scientific data, but it can also be a student who's doing a cartoon animation. It's the arts.</p> <p>The animation process can take a lot of time to render each frame, depending on the complexity of the model. The DRE is really targeted at complicated models. For example, I rendered a molecule with 90,000 atoms in it using POV-Ray. I changed the atoms to simulate caustics, i.e. they were pieces of glass and I shined light through them. Each frame in that model takes one hour to render. You would not be able to generate an animation on your local desktop computer to do this. Traditionally a lot of visualization and computer artists would be generated using an Apple Cluster or something. But, DRE uses Condor Technology so it's not costing anything to run. Rendering is based on cycle scavenging.</p> <p><i>[prompt asking what the relationship is to TeraGrid]</i></p> <p>Right now it submits against my home institution's Condor Pool. This project is on the TeraGrid visualization page. And it's supposed to be a TeraGrid project, though this project is much more than that. The software and system also can be used outside the context of TeraGrid and has been targeted towards OSG and our local campus grid. The idea is that you can choose where you want to run it.</p>	
<b>Q2.2 How will the success of your project be measured?</b>	<p>Well right now our success is measured by the number of users we have. We're just getting started. We've rendered several animations. So, future success factor would be the number of animations that we could render for people.</p> <p><i>[prompt asking if the number of animations rendered are tracked]</i></p> <p>Right now it's using the TeraGrid model of authentication and project allocation. We do have a mechanism within the system to track all the rendering jobs via Condor, but that reporting system has not been constructed yet. The software also supports local campus submissions as well using a similar authentication and submission mechanism.</p>	
<b>Q2.3 What are the professional measures of success for you?</b>	<p>One success measure is whether or not I am learning. There's been a lot of technology that I've had to learn from looking at Globus source code. There's been a lot of stuff that I've pushed the envelope within Web Start, in terms of the API package.</p> <p>Another success factor is when I get email saying, "Wow, this is awesome." I've had several of those coming in. This has been demonstrated at SIGGRAPH as part of the Gorilla Studio and I really didn't really receive any complaints, very few feature requests, so I'm right on the mark in terms of being where people want to be.</p> <p>DRE can be a complicated tool, but it's simple enough that you can work your way into it. That was my goal: to make something that was not too abstract. It's more designed to get the job done, and hopefully come out with some reusable blocks along the way. Don't spend months and months and months designing something. I would like to see some of the pieces and parts of this project being reused, but that may not happen all the time in the short term, but to reuse the technology and what I have learned about the technology is a must.</p>	



Interview ID=19 11 September 2007	ANSWERS	ANNOTATIONS
<p><b>Q3 What are you investigating?</b></p>	<p>One of the things that I'm going to use the WebStart client for is a GSIFTP client that enables your local PC to participate. A lot of the portals use HTTP to upload and transfer files around, which is can extremely slow and expensive (hence the interest in GSIFTP or other protocols.) The trick is to do this in a way that fits the certificate-based authentication model that we currently use.</p> <p>One of the things I see is that there are too few examples demonstrating the public interface layer of the Globus Java core technology. There is Javadoc that you can walk through, but I don't believe that there's really enough there. I think to myself: if Globus were a company, would I buy the product based on whether or not I could use it? I would tend to say no because I don't have that layer of documentation that I need to get started.</p> <p><i>[Prompt for example of a documentation model that is at the right level and at the right detail]</i></p> <p>Go to MATLAB's website, MathWorks, and look in their toolboxes. That would be an adequate level, where every aspect of their product has an example. MATLAB has all the API documentation just as Javadoc generates, but they also have examples. Globus gives you the API, but without the context of how to use them in more than just the trivial examples – it needs to be a little bit more than that. Of course, you can go to the extreme and do over-documentation, too. Then you're wasting more time doing documentation then you are actually doing something.</p> <p>One of the other things I use a lot is Java Almanac. That moved to Example Depot [<a href="http://www.exampledepot.com">www.exampledepot.com</a>]. Basically it is a repository of how to use various different APIs within Java. You could go in there and examine examples of a package like <i>javax.imageio</i>. Enough to get you over the hump of getting started – that's usually the problem. The examples are compile-able. They're usually small, not significant.</p> <p>Another problem that I find that is also more than just the lack of documentation- it is the ability to debug an application. It would be really handy to have a mechanism that would allow a developer to attach a remote debug utility to a Globus gatekeeper such that a deeper understanding of problems could be obtained. Though I suspect that this is quite possible to do if the gatekeeper is installed locally but its not quite the same a being attached to a production gatekeeper.</p>	
<p><b>Q4.2 How do you work?</b></p>	<p>For small projects is more of a "get it done" attitude. In this case, the technology for the application is evaluated and some sort of beta application is assembled. These types of projects are more related to quickly getting something up and running. But since they in general do not have a long lifespan, it sometimes is beneficial to use them for technology evaluation.</p> <p>One of the things I try to do is stick to a publicly accepted API that I know is not going to change or if that is not present I try to isolate the technology pieces with a very thin wrapper layer. Depending on the technology can become important because I try to manage every changing technology.</p> <p>For projects that are much bigger, spanning several developers and/or have a known lifespan, I try to start with more formal software development practices. I prefer to have a UML model and a definition of the project management rules. By taking more time at the architectural level I can define interface layers where I can isolate developers. I have used the strategy many times where I have many developers working on a project with different skill levels etc. This strategy also facilitates the idea of designing software for test.</p> <p>It's one of my fundamental beliefs that tools are what make software. This is not always a popular belief because developing tools takes additional time and resources but the benefits are that the developer/user base will increase if the tools make the underlying system easier to use. Tools can help by removing the need to manually develop core framework pieces and provides a way to architecturally institutionalize the software development process.</p>	

Interview ID=19 11 September 2007	ANSWERS	ANNOTATIONS
<p><b>Q5.1 In what ways do you interact with simulations in your work?</b></p>	<p>I interact with simulations by getting them to run, basically on the hardware or through whatever technology. My other role at my home institution is providing CMS (Compact Muon Solenoid) project support. I also do a lot of other things on top of these jobs that is basically overhead. I help the physicists get their jobs running. I see both sides of the coin. I've seen the strengths of the VO and the way that they're running their project, and the issues they're faced with doing Grid software in the OSG framework.</p> <p>There are a lot of issues. For example "Globus Error 17", followed by some cryptic and non-meaningful words. When I tried to track the problem down, the trail leads to a log file – syslog, <i>messages.log</i>, and then it goes to another file. I need to track through all these things to find it because the gatekeepers don't remember. I can't query the service as an admin. There are no admin functionalities. There's no way to ask the service, "Hey, this job failed. Tell me where it went. Give me the attributes of that." It's not easy. It's not easy to debug things when things go wrong. And users really don't have a clue. They get back this thing that says "Error".</p> <p>The issue when things go wrong on the Grid is trying to figure out what happened. It can be something on server side – some variable was set wrong. But you have to track it down and be able to replicate it and there's really not a way on either side to replicate that. As to the type of failure I'm describing: Java reported it as a failure, right? It may not be a GRAM failure. For example, we have issues with stage-in and stage-out sometimes (e.g. when a disk dies, auto mounter fails, it's full, or there is an open file handle still) and it's trying to write over somebody's files.</p> <p><i>[prompt asking for what information would be helpful]</i></p> <p>For development, the ability to access real time trace information would be helpful. It's not very helpful sometimes to just submit a ticket and wait for the gatekeeper admin to take a look at it. It is typically the case that the developer knows more about the gatekeeper software than the admin anyway. I guess what I'm looking for in terms of information is the ability to see the log files remotely, or some similar access through a web console or something.</p> <p>From the admin perspective, one of the issues is the ability to capture information when it happens. If a user were having difficulty submitting a job, it would be handy to have a trigger that would capture information when the user tries an action. It's often not the case that it's the middleware that may be having trouble, it could be the backend systems but the ability to capture and repeat the users actions would allow for quicker debugging.</p> <p>It would be helpful if this type of functionality was included as part of a web admin console. Such that both users and admin could see trace logs etc.</p>	

Interview ID=19 11 September 2007	ANSWERS	ANNOTATIONS
<p><b>Q7.1 What resources do you use in your work today?</b></p>	<p>With respect to the DRE, I use the Purdue TeraGrid Condor Pool and scratch file space at my home institution. With respect to CMS, I'm using our CMS cluster hardware and also all of our storage here, which is dCache.</p> <p><i>[prompt asking if CMS-related resources imply OSG resources]</i></p> <p>No. OSG is at the organizational level. CMS is a member of that organization, but not all things that CMS does are on OSG. CMS uses resources outside of OSG as well. For example, we don't allow just any OSG user to push files into our dCache. That's locked down for the most part. You can imagine, right? We don't want to open up storage to everybody. There are different priorities levels within OSG. For us, CMS users are the top priority.</p> <p>I am also the owner of the FPGA (Field-Programmable Gate Array) project here. It's an application accelerator technology. Part of the FPGA is that you create libraries or applications based on accelerated technologies. This is the best way to handle application accelerators for end users, because of the additional complexity. An even higher-level use case would be to use accelerated applications behind web services.</p> <p><i>[prompt asking if the Web services will be based on Globus technology]</i></p> <p>Yes and no. That's an interesting question. Globus is really cool in terms of being on the forefront. But sometimes it is a little harder for people to use. One issue is getting end users used to using certs. Why use certs if we can get a proxy cert from MyProxy with a username/password?</p> <p>The Globus security level that everyone uses is actually more secure than your bank. More secure than your credit card. Why? We're just making it harder. Why are science gateways so successful? Because they hide the complexity of the security. You can create an account and submit a job. Not any job... ah, that's the key. I think issue of security should be posed in terms of levels thus the complexity of the security mechanism can match the needs.</p> <p><i>[prompt asking what the right level is]</i></p> <p>The right level depends on the task. If we want a wider scope of users then we are going to have to make it easier for them. Sometimes it makes sense to increase the security level on access depending on what the user is trying to do. Would you use Google Calendar if it had to have a cert? Honestly? Would you use it? If I had to install my certs on every box I used, that's very insecure.</p> <p>There are some things that I do like about the proxy generation. It's a federated space. I really like the idea of limiting the users. But right now somebody can establish a proxy for some 180 days. There's no clamp. We don't have ways of saying you can only do it for twelve hours maximum at the gatekeeper level. I think Globus would benefit from making security flexible. This would increase the versatility of the Globus software stack.</p>	
<p><b>Q6.1 Describe how you interact with data in your work</b></p>	<p>I have lots of data. In visualization the data consists of images, and in the DRE case there's a lot of it. One of the topics people talk a lot about is how to deal with the data. In the case of the DRE is pretty easy. I'm fetching it; I'm dealing with it in pieces on the local end because I have that capability. But if I go through a portal to a remote gateway, it's not so easy to interact with data. It gets more complicated because the web application programming model is different than the application model.</p> <p>If you look at the user, they want to interact with their data like they do on their own machine (e.g., an explorer window on a MS Windows box.) There's no reason we can't provide them with that, right? I mean that's probably one of the easiest things to do, because in the case of Windows, I already know you can do that with WebDAV. Create a WebDAV drive. Done. I can drag and drop things over to it. There's no reason why we can't create tools that plug into these platforms in this way. I'm referring to WebDAV as just an example of how from a native platform, there users experience is seamless.</p>	

Interview ID=19 11 September 2007	ANSWERS	ANNOTATIONS
<b>Q8.1 What software do you currently use in support of your work?</b>	NetBeans and Ant. I also use the latest and greatest JVM usually.	
<b>Learning about the user's problems</b>		
<b>Q9.1 What challenges do you face today in accomplishing your work-related goals?</b>	<p>The challenge is the different versions deployed at the Grid locations... understanding what you should use.</p> <p>For example, I experienced problems between GT versions 4.0.1 and 4.0.3. It was in the job descriptor – it was a serialization bug. The symptom was, “I cannot deserialize this”, basically. I immediately understood the problem. These things are compiled stubs and the other end wasn't recompiled to match. It was probably a bug that was fixed in one place but not the other. I didn't dig much deeper beyond saying, “Oh A works, B doesn't, go with A.” It's a problem. It's a cross-version compatibility problem, and that is an issue in the Grid world. That's why I just follow whatever is working on the gatekeeper right now instead of using the new features.</p> <p>If we really, really wanted to go after the compatibility problem we need to think about it in terms of building roads and get rid of the buzzwords. Because if we built roads in the same way we use infrastructure and build our current IT infrastructures, our roads would be very scary. We would have bridges where we would drive off. What we're building with Grids is the ability for anyone to be able to get to a resource to do something in a way that's beyond what they can do now.</p> <p>One of the ideas that I thought about to solve this problem is to start changing the grid software stacks into more of a subscription service rather than having the local site admins install it. Basically the idea is that a site puts up a box, does a base container install, and registers with the grid domain and then all the software updates are installed and maintains by some central authority. This same model could also apply to development containers such that developers can keep up to date with the version of software that has been deployed.</p>	
<b>Q9.4 By contrast, can you provide examples of technologies you find very useful today?</b>	<p>WebStart because it allows me to deploy an application that runs on the client and uses the client's resources. Why should I have a \$3,000 PC if the majority of what it's doing is running a web browser? I've written JavaScript code. JavaScript code takes many times as long to develop as Java; Java has proven itself in my mind.</p> <p>WebStart really is a hot thing for me right now because I'm trying to prove to people that there is life outside the browser. We really need to quit this infatuation with being inside a web page, and realize that it's a constriction point and should have a wider scope. We must instead look at things as if they can be plugged in to different UIs – reusable components that can interact with the local graphics card. That's one of the other things that DRE does: it uses your local graphics card. That's why I can get 24 frames a second: not only the data bandwidth, but I'm also rendering images directly to your graphics card.</p> <p>Other thing that I hold dear to my heart is JINI. It hasn't taken off yet. It's still sitting back there. It's not so much the technology of it, it's the philosophy that's behind it.</p> <p>Condor is very successful and very useful technology as well. I feel is a very interesting technology in that its not only can be a scheduler but supplies a means to transfer jobs in a grid-type fashion.</p>	

Interview ID=19 11 September 2007	ANSWERS	ANNOTATIONS
<b>Q10.1 Can you think of any work-related tasks that decrease your productivity?</b>	<p>With regard to Globus, the biggest problem is when somebody has an error and you need to track it down. I mean that's the hardest one. It would be great if a diagnostic tool or monitoring framework existed, but other than that, nothing.</p> <p>The only other thing I do is try to find information. You know: we write information, we write all this stuff in different wikis, and who knows what's right? And getting that information is a nightmare. We're relying on Google to find everything. We're not approaching things in an organized fashion.</p> <p>And there's the issue of scale: we would like to say that we can scale to meet the demand with our current infrastructure. But this is often not the case because of either software or resource limitations.</p> <p>The other side of scale is growing the user base. One thing that I've learned in my years in industry is that if it's not easy then it won't get done. I think that holds true for the infrastructure that we are building.</p>	
<b>Learning about the Globus user experience</b>		
<b>Q11.1 Which Globus data components do you directly interact with in your work today?</b>	<p>GSIFTP.</p> <p>I've tried reliable file transfer and I couldn't get it working on the current package deployed but yeah, I would like to do more. I have lots of ideas.</p>	
<b>Q12.1 Which Globus security components do you directly interact with in your work today?</b>	<p>I currently work in the MyProxy space and Globus certs with respect to the DRE.</p> <p>I haven't worked with CAS yet, it's on my list of todos.</p> <p>GSI-OpenSSH, I use that as my day-to-day login for Teragrid activities</p>	
<b>Q13.1 Which Globus execution components do you directly interact with in your work today?</b>	<p>I use GT2 and GT4 through Condor a lot. The TeraDRE uses WS-GRAM/WSRF packages directly. I wanted something different so I tried the Web services interfaces.</p> <p>Looking into the virtual workspaces is on my todo list because it matches up with my interest in sandboxes for the DRE.</p>	
<b>Q14.1 Which Globus information components do you directly interact with in your work today?</b>	<p>WebMDS.</p> <p>I have looked at this, but the information that it contains really doesn't support any of my needs.</p>	
<b>Q15.1 Which Globus common runtime components do you directly interact with in your work today?</b>	Java	

Interview ID=19 11 September 2007	ANSWERS	ANNOTATIONS
<p><b>Q16 Why do you use &lt;component&gt; instead of an alternative technology?</b></p>	<p>One of the main reason I used Globus components is that I don't have other options on the grid. I can't just install any other technology. But on the server side components that I control I tend to use technologies that are natively installed and are easy for users.</p> <p><i>[prompt asking about the other technologies]</i></p> <p>Well one of the things that I was thinking about is a BitTorrent because I'm interested in getting data back to the client. I don't care about moving it from one high performance system to another. I want to move a terabyte of data back to the client – or have the choice to move it someplace else – but I want that choice.</p> <p>If I'm on my local cable modem connection and I'm downloading stuff it's going to take awhile. Not everybody has gigabit networks – that's the thing we have to realize. There are still universities out here that have 10mbit connections and these other technologies cropping up in the consumer world are handling that. Do you have a gigabit in line at home? Yet you can run Skype at home. I can Skype over wireless, which is kind of interesting. So yeah it is more user focused.</p> <p>I use the globus components because that is currently only supported. So at some level I am forced to use the globus software stack.</p> <p>There are times when I need to connect to the grid elements directly. This is when I used the Globus components to either submit a job or transfer data. I consider Globus a pretty heavy weight tool for the most part. It's not something that I would tend to use without a strong requirement to do so.</p>	

Interview ID=19 11 September 2007	ANSWERS	ANNOTATIONS
<p><b>Q17 What are the major challenges you face using &lt;component&gt; today?</b></p>	<p>GridFTP: One of the things I ran into with the DRE was initially GSIFTP was configured like an xinetd server that's up and down all the time. So for every user call you are creating a process and killing the process. Unlike doing HTTP download where I could basically hit the web server constantly with new connections, GSIFTP really didn't like that and it died. In fact when you hit any box running twenty of these against it, it basically came to its knees very quickly because of the overhead of starting processes.</p> <p>If I am user writing against this service, how do I find out how it's configured? I would like for a service to be able to allow me to connect to it, not really do anything but give me back some information about how it's configured so I can make a choice on how to use it. Am I starting up this process for every connection? Or is there a throttle placed upon me? How many other servers are running right now? Maybe I don't want to run right now – but I don't really have enough information to decide. It's a black box.</p> <p>When I print I am able to connect to the printer spool and it shows me the entire spool. You can inspect the queues. You can inspect the job queue if you had an alternative means but I don't think you can see other people's jobs, or how many jobs are running through GRAM. MDS is supposed to do that sort of thing, right?</p> <p>GRAM: Error messages are the number one thing. With the gatekeeper that's what I usually have had issues with.</p> <p>I think the only other problem I have had in that realm is things getting stuck. Sometimes the state files that are stored get out of sync. And I can't get it back in sync. From a user standpoint there's no reference to what to do. I know the GRAM2 state files are out of sync because it says "stale state" (or something like that) in one of the error messages. I forget the precise details. It was weird – it went away eventually, but I don't know why. I think it had something to do with the way some state files are stored.</p> <p>With respect to that though, this is the only time I'm going to be really negative on Globus: Globus error messages are worse than Microsoft's. Worse in the sense that they really are not helpful at both the administration level and the client level. It doesn't have to be this way, if you look at the stack and the processes that actually do the execution. The functionality that GRAM and the gatekeepers are doing, we've been doing for years. Why is this so hard to do?</p> <p>This is a real challenge ... it's really frustrating for the end user and admin. Because it's not bad enough that you have a problem to solve from a users support standpoint, but it could be a heisenbug. Or it could be repeatable. Identifying it – that's one of the hardest things to determine in any system.</p> <p>Also I really haven't found a clear document somewhere telling me everything that goes on from cradle to grave with job submission. I mean at a level that an admin needs: "When this breaks... when you get this go here." I wouldn't buy a software product without that because it's a requirement for what I consider to be enterprise-class software. Is Globus enterprise class? I expect it from Oracle, Weblogic or an SAP. Either that or a phone number of a helpdesk or my service engineer that I can call. I guess I'm just used to running these things 24/7, 365 days a year and living with a pager.</p> <p style="text-align: right;"><i>[answer continued on next page]</i></p>	

Interview ID=19 11 September 2007	ANSWERS	ANNOTATIONS
<p><b>Q17 What are the major challenges you face using &lt;component&gt; today?</b> [continued]</p>	<p>The older gatekeeper software I think really had a problem with scaling... CMS proved that here. At one point in time our PBS had 4,000 jobs trying to submit against it that blew everything up. That's a scale issue; I think they both suffer scale issues. The gatekeeper becomes completely sluggish and loaded very quickly and I can understand that. It's not so much maybe the software but the software in the box isn't sized correctly or is there a way to provide lateral scaling? Are there actually load statistics available? What's the limit? Another thing: I couldn't find anyplace where I could set a hard limit. One that allows me define the point at which to say, "We're busy go away." I would love to have that feature.</p> <p>I think one of the biggest tools is a lighter weight command and control type of protocol. So why not use what the video game industry is using? They're using command and control protocols that are less heavy. Why can't we use the chat type protocols or those in the JMS? Why don't we make this distinction between command and control versus data.</p> <p>Security: The major challenge I had with MyProxy was debugging. Figuring out problems with the trusted CA within NCSA's MyProxy. And not being able to use the new NCSA MyProxy client portion because of incompatibilities of the Bouncy Castle libraries. That was the hardest thing to debug. I had to generate code around it; partly that's also my own fault because I'm trying to do something other than the standard model. The standard model is to install all trusted certs on the box and go from there. I was trying to prove the point that you could do stuff without installing certs, and fetch them as needed. I really wanted to make this easy. And it's still easy – you don't know they're being installed but they are being installed. I placed the burden on myself to keep them updated. MyProxy has been pretty good; it's pretty easy. I've really not had other problems; really not too many issues with it. The library version issues are highlighted with this example, because Bouncy Castle could not be used with both the NCSA's MyProxy API and the Globus version.</p> <p>WebMDS: Most of the information in there, besides finding the box or queue name, I don't find useful. The reason is I think these information services are storing the wrong information. I believe we need to approach this problem from a different standpoint or path where we can describe the entire system end-to-end. A graphical way of going in and clicking on boxes and so forth and pulling up lists of software.</p> <p>Most people say, "Oh well. You gotta run static linked stuff." Well try statically linking X Windows into your application CMS installs for their distribution when they install an OSG site. It's approximately 4 Gbytes for every version of their software. They take the a parts of a Linux distribution and chop it down because they're trying to maintain consistency. I don't have ways of discovering this from our current information services.</p> <p>eBay was at the last TeraGrid conference. They gave a really good talk that hit a chord with me. It was about asset management. Not asset management in terms of what hardware you have, but what software, what services, what's this, what's that, how are these things connected? This is even more important. I envisioned an information system that's much more than we have now. That allows us to drill into these things and figure out how things are connected: this service talks to that service, to that box and that box has this amount of stuff in it. That sort of thing. To get to that level both from both an admin's perspective and from a user's perspective.</p> <p>[answer continued on next page]</p>	



Interview ID=19 11 September 2007	ANSWERS	ANNOTATIONS
<p><b>Q17 What are the major challenges you face using &lt;component&gt; today?</b> [continued]</p>	<p>Think of how things are dependent. For example, I click on something and it has a reference link saying, "I use Globus." Okay I click that box and it takes me down to Globus and it says all right which part of Globus you are using, which service or whatever. And I click on that and it says, "Oh you're using this set of software under these revisions," or I click on the node and it tells me the node has, "X Windows installed, this version and these libraries installed."</p> <p>It is the asset management. If you talk about asset management in business, there are two camps. One is the actual tagging of a box for tax purposes – that's the bean counters asset management view. If you're talking about the manufacturing engineer's asset management view, he'd want to know where that machine is connected, where the power goes, what the machine requires in order to function. How it plugs into the entire system. How many other processes depend on that piece of equipment being up. Is this mission critical? So from a user standpoint, maybe I could start setting up more complicated requirements. What are my chances of finding something like this out there? Finding it in detail.</p> <p>But especially in a Grid world and academia and so forth, we have such a turnover rate that these systems start becoming beneficial to the actual host environments as well. It gives them a management tool to manage how things are connected. But it is very interesting that eBay is looking at this and they were looking at RDF [<i>Resource Description Framework</i>] as a technology that could do this for them.</p> <p>Again the downside is, we have to not be based on schema metadata. It's got to get down to some high-level metadata categories, but then at some point we get down to natural language processing. But that's kind of cool and I also view that as sort of a learning thing. It's a good way to share information so somebody could go to my home institution and see how I put my Legos together.</p> <p>You can enumerate – and that also works well if you enumerate a resource. That means you can tag properties to that particular resource, down to a box level. Think of it as if you were bar coding everything. I consider a resource at some level a computational resource, not a directory or a NIC card. It's easier to do that because they don't usually change their enumerations. It could be at a level such that resources can only do certain things. That plays into things going forward, such as virtualization, which is a hot topic. Now if a Xen machine lands on my box it can only do so much. It's restricted. That's going to be required.</p> <p>The other reason why this information service would be important is in order to do a successful Grid in a particular Grid domain; some level of central management is needed. The admins are experts in installing and debugging stuff, but also can quickly identify problems. It wouldn't take too many admins to do it. You could centralize them. They would be doing the application management level stuff, not OS level stuff.</p> <p>Java WS Core: Need more examples and figuring out how things work. The existing tests are really good, but that's not enough. That's really not what I'm thinking of in terms of examples. You can see what I mean by looking at like Mathwork's documentation page: how they introduce a concept.</p> <p>[answer continued on next page]</p>	

Interview ID=19 11 September 2007	ANSWERS	ANNOTATIONS
<p><b>Q17 What are the major challenges you face using &lt;component&gt; today?</b> [continued]</p>	<p>In the “Build A GT4 Service” tutorial, there was an example. But hearing the questions asked during that session, a lot of people didn’t get it because it was like drinking from a fire hydrant. You would just uncomment some lines and redo the process again. They didn’t understand what this was. Sometimes. I don’t expect them to understand it. But the thing is it wasn’t clear how these things merge together. It took me the longest time to figure out how the EJB technology worked with the JNDI lookup, with the get a home and get the interface. That took me awhile when I first started. It’s like I had to wrap my head around it because normal C programming doesn’t do these things.</p> <p>I guess the other thing too is that the challenges in instilling a mindset in the community to develop tools – not applications – tools. I don’t think there’s enough work with respect to that, and I don’t know why. I think getting tool-level people engaged is important, because that says Globus is behind a standard. And there are people developing tools to that standard. This eases some of the load because you have a bigger market of tools. Tools are everything.</p> <p>General: Globus advertises itself as being modular, but one of the things I’m finding is there is a lot of overlap in the packages. If I only want to use reliable file transfer, I don’t want to have to use any other stuff. I want a nice stovepipe architecture, with respect to the packages. When the Autojar runs I found a lot of crossover. It’s partly the reason why I run Autojar, to pick out the necessary ones instead of deploying all the JARs in the directory.</p> <p>Eliminating the overlap would really would help the understanding. In my mind I see Legos. I see the client and server portions both being Beans in some sense. That’s what they should be like – components that I can assemble. The Legos may not fit in all situations, so granularity level is a concern. It is hard trying to get that inter-package dependency down a little bit, there’s going to be some. For example the transport- it might be common amongst them all, and I need it. But it’s helpful to identify that component, so I know I need this component for X. It’s not helpful to have just a big directory of JAR files.</p> <p>Another idea that would get around trying to architect a modular jar system is to provide a service that would build a custom jar for you. One could imagine a web site that allows me to select the functions that I needed, which then assembles a jar or a set of jar files that I could download. This would be the al a carte model for deploying software for developers. In this way the interface layers and interdependencies could be better controlled. This would also work with the subscription service model.</p>	
<b>Wrapping-up</b>		
<p><b>Is there anything you’d like to say to the people who build software for use by people like you?</b></p>	<p>If you have an interstate made of gravel roads, people will drive on them if they have no choice. But we know that we can pave the roads and more people can use them. The key is more users. That’s what keeps things alive, both in academia and industry. It’s no different; the economic pressures are the same. If we have something and nobody (or a small number of people) is using it, it’s harder to justify having it around. I would like to have a site that serves 100,000 or more users a week – that would be awesome. How can I do that? I have no idea. But the idea is to figure out how to scale and interact.</p> <p>In closing, I’ll emphasize the importance of examples, documentation and keeping the granularity of the packages such that one doesn’t have to deploy a huge thing. Modularity is key. Those are my recommendations that I would be looking for to make things easier for me as a developer and consumer of packages.</p>	

## D.20 We can provide our users with fresh data more frequently because of the Grid

Interview ID=20 12 September 2007	ANSWERS	ANNOTATIONS
<i>Pre-interview question: Do you interact directly with Globus software in your work today?</i>	Not me, but my group does because we use Globus to access the Grid to run bioinformatics applications.	
<b>Establishing context</b>		
<b>Q1.1 Please provide a one-minute overview of your project</b>	PUMA is the integrated, interactive system for high-throughput analysis of genomes and metabolic reconstructions. In the past decade there were a lot of sequences of different genomes – genetic sequences of different organisms – available and this number is growing. Currently over 1,000 genomes are available, but in the next two or three years there will be thousands of them. To process this data you need to have scalable systems that will enable complex analyses to be performed in some reasonable time. So to satisfy this need we start to use the grid computing to help us with the analysis. We have built a system that is used widely by the community. We have 82,000 individual users. People are using it for analysis of genomes for the needs of a variety of biomedical applications: alternative fuel research, remediation, technology, etc.	
<b>Q1.2 What is the project's name?</b>	PUMA2	
<b>Q1.3 Which agency funds the project?</b>	National Institutes of Health, National Science Foundation Alliance	
<b>Q1.4 What field does your project belong to?</b>	Bioinformatics and Genomics	
<b>Q1.5 What is your job type?</b>	Project Lead	
<b>Q1.6 How long have you been a &lt;job type&gt;?</b>	Seven years	
<b>Learning about discipline-specific goals and approach</b>		

Interview ID=20 12 September 2007	ANSWERS	ANNOTATIONS
<b>Q2.1 What are the main goals of your project?</b>	<p>The main goal of the project is the analysis of large volumes of genomic data. When genomes are sequenced, they are represented as strings of letters, which signify different nucleotides. Once a genome is sequenced you want to know what functions the genes perform and what physiological and metabolic processes the genes are involved in. So starting from just the alphabet soup of the sequence, by the end of the analysis you know:</p> <ul style="list-style-type: none"> <li>- how many genes this organism has</li> <li>- what they do</li> <li>- how this organism lives (because we're reconstructing double helix [?] properties)</li> <li>- does it have any pathogenic or non-pathogenic factors</li> <li>- what does it transport in the cell</li> <li>- and what does it produce</li> </ul> <p>So pretty much by the end of the analysis, not through experiments but by using pure bioinformatic methods, the biologists know quite a bit about the organism already.</p> <p>To do the analysis we take the genomes from a national biology information repository, which is called the National Center for Biotechnological Information [NCBI]. We then analyze a genomic sequence with an array of bioinformatics tools so it will be easier for the researchers to answer the questions that they usually have.</p> <p>Some of our users are interested in the whole organism, genome and the properties and the properties of the organisms. But some of the users are interested in particular properties. There are a huge variety of questions that people can ask PUMA. Mostly it's designed for comparative and evolutionary analysis of the biological data.</p> <p>Our methods require us to maintain a large database backend filled with data we've integrated from various databases, so our current approach is a warehouse model. We parse information from over 20 biological databases and we store it in a central warehouse. This approach allows us to present the user with an integrated view of the accumulated knowledge for a particular genome or piece of data.</p> <p>Then as a second facet of our work, we add value to the genomic sequences. We do this by performing comparative analyses using an array of bioinformatics tools. We employ a number of algorithms used in bioinformatics, and we accumulate the results, adding it to our database. We then make the added information available to the user too. So the information becomes more valuable:</p> <ul style="list-style-type: none"> <li>- not only do we provide access to genetic sequences</li> <li>- not only do we update it with the information from 20 other databases</li> <li>- we also add value through the use of bioinformatics tools.</li> </ul> <p>Analysis of those sequences using bioinformatics tools is something we need to do on the grid. The volume of information is very, very large and some of the bioinformatics algorithms are very, very computationally intensive. If you analyze the sequences that are available now in GenBank on one CPU it will take you approximately 12,000 days. But our Grid-based computational gateway GADU can provide access to 2,000 CPUs, perhaps more. So you can see how much more efficient this analysis becomes.</p> <p>In the future it will be even more critical to have a grid-based computational background because the amount of data is growing exponentially. So if we are not using scalable computational resources we will be totally flooded.</p>	
<b>Q2.2 How will the success of your project be measured?</b>	<p>The success of the project is measured by the fact that our users love the system and they love PUMA too. We know because 82,000 people use it. Also we have very positive feedback from the users for whom we analyze genomes. We know because we interact with these people directly on a collaborative basis.</p>	
<b>Q2.3 What are the professional measures of success for you?</b>	<p>For us success is measured by the number of users of our systems, the positive feedback, and the number of publications written by our users. It's also nice to know that these people are writing something useful. And it's very pleasant.</p>	

Interview ID=20 12 September 2007	ANSWERS	ANNOTATIONS
Q3 What are you investigating?	<p>The major approach in bioinformatics is comparative analysis, and that's what we do. We are trying to understand biological systems through comparative analysis because biology is a reverse engineering science. We have no idea how the systems are built or who built them. Some people think it is God, some people think it is a spaceship with some living creature. Since we don't know how they are built, our method is to compare what is known to what is unknown, and then transfer knowledge from the known to the unknown.</p> <p>Organisms can also be compared, such as genomes that live under different conditions. For example, one might compare microbes that live in boiling vents in temperatures over 100 degrees centigrade with organisms that live at room temperature. The differences can be used to help understand what allows them to live under such harsh conditions. This technique also involves comparative analysis.</p> <p>To do this type of analysis requires comparing large amounts of data against large amounts of data, which is very CPU-intensive. That's why we need a lot of CPU power.</p> <p>So our investigation has three components.</p> <p>First it's integration of all information from the public sources that is useful for notation of biological data.</p> <p>The second investigative component is comparing this information in order to support transfer of information that is known to things that are unknown. For example, if you have two genomic sequences that are very, very similar and you know the function of one of them, you can with a certain degree of certainty say that the other sequence also performs the same function. But one sequence comes from one organism; the other sequence comes from the other organism.</p> <p>The more information we integrate, the more successful and productive our analysis will be. So that's why the first component we're investigating, data integration, is very important for us. Integration is problematic in biology because ontologies are very, very difficult and underdeveloped. The biologists have this culture of naming different things the same and the same things differently. It's a nightmare, you know, and you are limited in what you can accomplish because these naming conventions are a mess. It's driving people totally mad. This area of biology is really difficult.</p> <p>The third component of our work is investigating new ways to look at the system. As you integrate new information, you can look at the data from a new angle and it can lead to discovery. Also, if you apply new algorithms to the data you are able to look at the data in a new way. That can lead to discovery.</p> <p>Different groups are developing good algorithms, but they reside in different places. You can install everything in-house, but it would be much better if you could use Web services and actually access remote services directly. In this case you would have a network of information services and information of services. Users are also distributed between different locations. Pretty much everything in biology is becoming distributed because the projects are getting so large that it's impossible to keep everything in one place.</p> <p><i>[prompt asking if references to the data are stored in the warehouse]</i></p> <p>No. We warehouse all the data. We download all of the information from the public databases, parse it, and store it in an Oracle database. This is perhaps not the best way to handle things. From the public databases we warehouse on the order of 20 gigabytes, but the added data resulting from our analysis runs is larger than that. In the future it will be much, much larger.</p>	

Interview ID=20 12 September 2007	ANSWERS	ANNOTATIONS
<p><b>Q4.1 What is your method for investigating &lt;phenomena&gt;?</b></p>	<p><b>Regarding the data integration method:</b>          We have taken the path of least resistance. We decided it's too difficult for us to develop the federated databases because of the reasons previously mentioned. Federated databases for biology are not very well developed. There were attempts from IBM to provide a middleware, but somehow they didn't work. They are too expensive or they are not efficient. So that's why we're just parsing it. We also warehouse the data to make the queries faster.          We learn of data that should be integrated by being experts in our field. We know what information we need, and what is available. But sometimes we will have no idea about new cool databases that show up if they are not well advertised. So we integrate information mostly from the largest, most popular sources.  <i>[prompt asking if users can add their own data to the warehouse]</i>          No, it's not like that. Pretty much we decide which databases to integrate. With our application it is not possible for a user to add knowledge (such as adding some small database that might help their analysis.) The reason is because it is very difficult to expand the infrastructure, understand the data format and integrate customized data into system. But if somebody would be willing to help do that, it would be cool.          We document the warehoused data with dynamically generated reports sitting on top of the Oracle database.  <b>Regarding data analysis:</b>          We do pairwise comparison of a database, which has six million sequences. So can you imagine how many computations it is? And every analysis run of one sequence produces a substantial output. So it is a data- and CPU-intensive process. Analyzing six million by six million sequences will require on the order of <math>10^{12}</math> jobs. So it's a huge number of jobs.          These comparisons are performed on the sequences stored in the warehouse. Every job applies a popular algorithm in bioinformatics, called BLAST [<i>Basic Local Alignment Search Tool</i>]. It takes one sequence at a time and compares it in a pairwise fashion against six million other sequences in the database to find similarities.          For example, we might start from an unknown sequence that has a high similarity to some sequences already in database. So if they are performing some function, say myzine, then if they are found to be similar then you can say the unknown sequence is also myzine. But in order to come to that that conclusion you must first perform this huge comparison.          The algorithms assign similarity scores, which is included in the output that is produced. We parse that information and load it back to the warehouse. This is the added value.  <b>Regarding new analysis techniques:</b>          We have developed several algorithms and workflow plans to optimize analysis for certain purposes. So sometimes we might work on a new algorithm, but sometimes we create a new workflow that performs an analysis we have invented. For example we created a system called Chisel, which identifies differences between collections of enzymes. It is a sophisticated analysis tool that includes a rule-based algorithm and it executes a multi-step analysis.          And so far we're writing the tools mostly in perl and executing our small workflows for scientific tools also in perl. We started to play with the Virtual Data Language [VDL] for the big jobs, but at the time VDL wasn't supporting recursion. Probably in the future we will pursue this method further – it will be great to express the scientific workflows in a controlled meta language.          We document our algorithms and workflows in papers – mostly papers. And also they are open source, so people can download them.</p>	
<p><b>Q6.1 Describe how you interact with data in your work</b></p>	<p>Well, there are two different ways we interact with the data. One of them is we analyze them. So this is where we execute analysis and this is done on my own machines.          But as a user, I interact with the data over the Web. So everything that we are doing, all of our applications are on the Web. So if I want to do scientific work, per se, (not the preparation for our tools or anything related to the tools and systems development) then it's just over the web.          As a user I am organizing the bug:          - does it have some particular pathogenic factor?          - if it does, what is the best way to design the antimicrobial drug to kill it?          This kind of analysis is mostly done over the Web using PUMA or GNARE through the browser.          Most of the bioinformatics applications are actually accessible over the Web. Some of the applications can be installed locally, but bioinformatics users prefer to use something that is Web accessible for analysis of the data of interest.</p>	

Interview ID=20 12 September 2007	ANSWERS	ANNOTATIONS
<b>Q6.3 By what mechanisms is access to your work-related data controlled?</b>	We don't have proprietary data, but we have restricted data because some of the users don't want to make their data public before publication. And that's we have an authentication system. Sometimes the researcher doesn't want to permit anybody to look at his favorite data. So it's just for him. Sometimes they want to look at the data in groups and in this case we're supporting group access. But sometimes they're just saying okay, now we are ready to make this data public. In this case all of the restrictions are pretty much lifted.	
<b>Q7.1 What resources do you use in your work today?</b>	We're not using data sensors, unless you consider the public genomic databases that we are monitoring to be a type of data sensor. For data storage we are using mostly our own resources at our home institution. We are using Jazz [ <i>a 350-node computing cluster</i> ] and we have started to explore a Blue Gene/L. And we're using the grids: simultaneously Open Science Grid and TeraGrid.	
<b>Q7.2 How do you share work-related resources with others?</b>	They get it through FTP or the Web (through an application that sits on top of GNARE.)	
<b>Q7.4 How do you locate available resources for use in your work?</b>	This is just based on expert opinion. We are looking constantly at what's new and if there is something, then we use it. But I think we talked a little bit about this, that sometimes if something really good, really cool is developed, we just don't know about it.	
<b>Q7.5 What types of information do you need to know about a resource in order to determine if it is suitable for your work?</b>	First from the biological point of view: - what type of information is it useful for with respect to the type of research we want to do? And second type is about the data: - how well structured is its format? Sometimes if the data format is sloppy then it's too time consuming to actually acquire this data. Sometimes we invest time, but it can be really bad.	
<b>Q8.1 What software do you currently use in support of your work?</b>	We use domain-specific bioinformatics software. We also use Globus, Condor and VDL.	
<b>Q8.2 What scripting languages have you used in the past year?</b>	Not me. I am not a programmer, but most of our systems are written in perl.	
<b>Q8.5 What parallel computing tools do you use in your work?</b>	We use Globus for everything.	
<b>Q8.6 If the need for new software-based functionality arises in your work how do you acquire it?</b>	We have been doing ad hoc work. But after we start collaborating with the Globus group, we talk to them and see if something is available.	
<b>Learning about the user's problems</b>		
<b>Q9.1 What challenges do you face today in accomplishing your work-related goals?</b>	Well, funding. In terms of workflows, several things would be interesting for large bioinformatics applications. I will characterize them in general terms: The first thing is flexible language for expression of the workflows. And VDL is good, but the problem with VDL was that it wasn't very stable. But then currently we are running pretty well with it. So I don't know what the future of it will be because I know that now it is called Swift, and probably it will go into the next generation. It's wonderful when new things are developed, but every time there is a new tool available, it means that in awhile we will need to rewrite the whole system to accommodate the new release. And this can be a problem. I understand that new technologies are being developed and that's why they are getting better and better. But it's a little hard on the application developers when new versions are not compatible with the other parts of the system. So that could be difficult. <i>[prompt asking what stability means]</i> It should be a stable service to our user community. A reliable deliverer of services. This probably means we shouldn't use the development versions of the software tools. We probably should use only the production versions. If the production versions are compatible it will be much easier on the application developers.	

Interview ID=20 12 September 2007	ANSWERS	ANNOTATIONS
<b>Q9.2 What types of information do you need in order to address the challenges you face today?</b>	<p>When new software tools become available it would be useful to know:</p> <ul style="list-style-type: none"> <li>- what the new features are</li> <li>- whether the features are compatible with previous versions</li> <li>- where the incompatibilities might impact the other parts of the system</li> </ul> <p>Because in this case we'll know what to troubleshoot.</p> <p>It would be absolutely great if there were some information system – actually I guess it's probably not for Globus, because it's probably domain-specific knowledge. The information system would enable finding the services, finding the information, and somehow linking it in a simple way. In this case it could be distributed services and distributed data, but this is probably too much to ask for. <i>[prompt asking for further description of the domain-specific services]</i></p> <p>There are a lot of different bioinformatics tools and currently we are mostly installing them locally to run them. Sometimes we are submitting it on the network, but probably we just need somehow a more developed system in bioinformatics for Web services. So it should be probably Web-service based, the whole infrastructure.</p> <p>Note that we don't have any distributed algorithms. All of the data and all of the parallelization we are doing is embarrassingly parallel.</p>	
<b>Q9.3 What technology-related obstacles do you currently encounter?</b>	<p>There are several obstacles related to data.</p> <p>The first obstacle is the transfer of large volumes of data. Sometimes the volumes of the data are pretty large. I remember a long time ago we ran an analysis on the grid for the SEED Project. We wanted to transfer this data across the disks within the same building, and it took approximately 14-15 hours to transfer these data. A lot of bioinformatics data is getting huge and it will only get huger. So time is a problem. Also correction of the data because if you are transferring large data sets sometimes they get corrupted.</p> <p>We transferred the data using FTP. I know that some of the large bioinformatics applications, like in CAMERA [<i>Community Cyberinfrastructure for Advanced Marine Microbial Ecology Research and Analysis</i>] projects, they have amounts of data that are 100 times larger than what we deal with, and they have no idea how they will deal with that. We were trying to download at least small chunks of their data and it was taking hours and hours.</p> <p>Sometimes we can use GridFTP, and sometimes you can't. <i>[prompt asking why using GridFTP is sometimes not an option]</i></p> <p>Because in some cases we don't have a GridFTP client available to us.</p> <p>The second data-related obstacle is the data may be of different quality. Some if it can be dirty and some of it isn't.</p> <p>The third data-related obstacle is definitely ontologies. <i>[prompt asking if the interviewee encounters compute-related obstacles]</i></p> <p>Currently probably no because my developer has written this scheduler containing a site selector for the grid and now we can pretty much use quite a bit of resources. But I think it will be better to ask him because he's a developer.</p>	
<b>Q9.4 By contrast, can you provide examples of technologies you find very useful today?</b>	<p>It's the Grid. It's definitely the Grid for us because it just completely – it has made a qualitative difference in what we can do.</p> <p>Like, for example, to analyze the data in preparation for a new release of PUMA: If you want to do it on the cluster sometimes it's very difficult to get nodes – even on Jazz. And on 40 nodes the analysis will run for weeks. But on the grid we can immediately do it and it will take so much less time. We can provide our users with fresh data more frequently because of the Grid.</p>	
<b>Q10.1 Can you think of any work-related tasks that decrease your productivity?</b>	<p>Yeah, that's why we're trying to automate everything. The analytical pipelines [<i>scientific workflows</i>] are very complex and that's why we need to express them in a meta language. We were doing it without a meta language for a while, but under no conditions will anybody return back to those because it's too unreliable, too time consuming and wastes resources.</p>	
<b>Learning about the Globus user experience</b>		
<b>Q17 What are the major challenges you face using &lt;component&gt; today?</b>	<p>VDL: Our Grid gateway uses VDL. We haven't transferred all of our domain-specific applications to VDL. Some time ago there was no recursion, but I think the issue may be addressed in Swift.</p> <p>If there will be some continuity between the releases that would be helpful. Ideally we would not need to rebuild everything in our system to accommodate the new changes. It would be really good to somehow lighten the burden of transitions to new releases.</p>	
<b>Wrapping-up</b>		
<b>Is there anything you'd like to say to the people who build software for use by people like you?</b>	<p>Oh, I just wanted to tell you thank you so much because somehow the use of the grid and the use of the Globus really, really, really made such a huge difference in what we are doing. We can do so much more science after using the grid. So just my deepest, deepest, deepest and sincere thank you.</p>	



## D.21 We work to enable discovery, access and synthesis of distributed datasets

Interview ID=21 13 September 2007	ANSWERS	ANNOTATIONS
<i>Pre-interview question: Do you interact directly with Globus software in your work today?</i>	no	
<b>Establishing context</b>		
<b>Q1.1 Please provide a one-minute overview of your project</b>	The primary project involves an architecture we call PASTA, which stands for Provenance Aware Synthesis Tracking Architecture. As part of the LTER system there are 26 sites, which are spatially distributed across the continental United States, two in Antarctica, one in Tahiti and one in Puerto Rico. Each of the sites is collecting scientific data. The goal of the architecture is to pull data from them in a seamless way, based on both the metadata records and open access to the actual data file. These data are brought into a centralized data warehouse - an expanded data warehouse, if you will. Because of the distributed nature of the sites, we're exploring different types of mechanisms to do this in a seamless fashion.	
<b>Q1.2 What is the project's name?</b>	PASTA	
<b>Q1.3 Which agency funds the project?</b>	The National Science Foundation through the LTER (Long Term Ecological Network Office)	
<b>Q1.4 What field does your project belong to?</b>	Ecoinformatics	
<b>Q1.5 What is your job type?</b>	Lead Scientist, System Architect, System Administrator, Developer	
<b>Q1.6 How long have you been a &lt;job type&gt;?</b>	3.5 years	
<b>Learning about discipline-specific goals and approach</b>		
<b>Q2.1 What are the main goals of your project?</b>	Basically to enable data discovery, data access and synthesis of distributed datasets within the LTER network.	
<b>Q2.2 How will the success of your project be measured?</b>	There are a number of metrics that will demonstrate success. One of them is: - being able to access data at a remote site - pulling it into a data warehouse - making it available as a derived or a synthesis product to other scientists Another metric of success is how seamless the process is. "Seamless" means minimizing the amount of effort required at an individual site to make this happen. In other words we want to leverage resources that are already in place without having to implement new workflows, new processes, new techniques for the site itself. <i>[prompt asking about how the quality of seamlessness will be measured]</i> Mostly through direct feedback of the community. We have an ongoing communication process with the information managers at each of these sites. These information managers are responsible for documenting their data and making the datasets available. In general there's a two-year period within which a site must make their datasets available to the general public. So we try to enable that without putting an extra burden on the information manager. So any undue work – especially the unfunded mandates – we hear back quite loudly on those types of efforts.	
<b>Q2.3 What are the professional measures of success for you?</b>	Success for me is to see this process being adopted within the LTER system. Adoption is measured in part by seeing the data flow from the site to the central warehouse. The other aspect of adoption is having end users actually exploiting that data in a good way.	

Interview ID=21 13 September 2007	ANSWERS	ANNOTATIONS
<b>Q3 What are you investigating?</b>	<p>We're investigating different techniques and informatics in terms of</p> <ul style="list-style-type: none"> <li>- how to access data</li> <li>- how to document data well enough (using metadata standards) to be able to read in a data table, for example.</li> </ul> <p>This is still a difficult problem: understanding the different semantics and syntaxes of these diverse datasets.</p> <p>So basically I am investigating different technologies. For instance, about a year ago we did a pilot study to investigate different types of general Grid technologies and different middleware layers. We were specifically looking at different security models using Globus and the MyProxy – using X.509 certificates to enable pulling in different distributive resources without requiring end-users to authenticate at each site.</p> <p>So we look at various technologies out there to make this whole process work.</p>	
<b>Q4.1 What is your method for investigating &lt;phenomena&gt;?</b>	<p>Our work is partly an integration project. We're also developing home grown applications.</p> <p>Right now the ecological community, and specifically LTER, has pretty much standardized on a metadata standard called the Ecological Metadata Language [EML]. It's an XML-based language that allows somebody to document a dataset. We're building tools around the standard so that we can:</p> <ul style="list-style-type: none"> <li>- read those types of metadata documents</li> <li>- understand the types of datasets that are being described</li> <li>- access the datasets</li> <li>- and load them into, say, another relational database.</li> </ul> <p>So we have a number of homegrown open source tools that we make publicly available. Almost all the tools are being supported through the <a href="http://ecoinformatics.org">http://ecoinformatics.org</a> website; most of this work will be contributed back to ecoinformatics.org.</p> <p>It is the individual sites' responsibility to document their data. Once the data have been documented using the standard, we try to access it. Many times the data is local to the site, so we have to develop specific protocols to access it. That's again the whole issue of authentication and access control to those datasets.</p> <p>On the other side – the flip side of that:</p> <p>Once the data are actually centralized we develop interfaces to enable users to explore the data. I think one term is “exploratory”. So we're developing web-based applications that include discovery interfaces, plotting routines, different types of data download mechanisms, and allowing end users to integrate different datasets so they can generate more or less synthetic products on the fly.</p> <p><i>[prompt asking if the end users would then create new data themselves, which the interviewee would then store]</i></p> <p>In theory we will be. That's part of our long-term plan is to store as a cache so to speak and then perhaps reinsert that back into the main data warehouse so that it now becomes a secondary or third-derived product.</p>	

Interview ID=21 13 September 2007	ANSWERS	ANNOTATIONS
<b>Q4.2 How do you work?</b>	<p>So we have these two distinct lines of work:</p> <ul style="list-style-type: none"> <li>- talking with the remote sites, trying to understand their data, and agreeing upon how to get it from them</li> <li>- maintaining the data warehouse</li> </ul> <p>It's a very modular process. There are a number of different parts of the entire workflow process that can be implemented and enabled independently.</p> <p>So right now we're focusing on the data warehouse and the exploitation routines. Basically on the interface to the derived datasets. About a year ago we worked on the component that would actually read the metadata document (the specification for the data table), and load that into a relational database. We haven't yet integrated those two parts yet.</p> <p>There is still work to be done with pulling data from the remote sites. Many issues we're starting to address with the sites are related to quality assurance: both in terms of the quality of the metadata and the quality of the actual data in the tables. We're also working to ensure that there's good correspondence between how data is described within the metadata and how the data is structured within whatever medium it's being stored. So there's still work to be done there also. <i>[prompt asking if the data are normalized in any way, for instance by addressing cross-data source naming issues]</i></p> <p>It depends. Again this architecture we're working on is really a model. It's not an off-the-shelf type of application, so the answer to that question varies depending on how the model is applied. So in the very basic sense of the architecture, data is loaded from the remote site is replicated in the same structure and format (other than the fact that it is stored in a relational database.)</p> <p>Once that – we'll call that raw data, if you will – once that raw data is actually localized within the data warehouse there can be a number of workflow steps applied to it. So you get these additional products, which we call derived products. The derived products can cascade on one another also. There can be naming issues involved in that. <i>[prompt asking who defines the structures of the site-specific (raw) data]</i></p> <p>That is really up to the site. This question points to one of the community, or social, issues we address. Instead of forcing or mandating that all sites map to a specific standard or schema, we allow them to adopt their own. Our requirement is that their data be well described in the metadata documents.</p>	
<b>Q4.3 How do you keep track of interim results, if at all?</b>	<p>We do not use a formal approach. Most of the documentation is through our project management software. We just annotate our tasks. We annotate when we have success, and describe our failures related to new approaches.</p>	
<b>Q5.1 In what ways do you interact with simulations in your work?</b>	<p>Directly at this point not at all, but perhaps in the future as more of this data become available to the broader community. One way we phrase it is at the network level, as opposed to the site level.</p> <p>Part of our goal is to make this next leap from science that takes place at the site, to science that takes place at the LTER network. This would be a national, if not global scale. So the anticipation is once these datasets become available to end users, that simulation and modeling will begin taking place. That's probably on the horizon within the next 1-5 years.</p>	
<b>Q6.1 Describe how you interact with data in your work</b>	<p>The data is transferred from the remote sites to the warehouse as a batch process. An analogy would be that we run a cron job that goes out and polls the datasets. If the metadata indicates that a new dataset is available, then we'll pull it in. The transfer happens with Java Servlets over the HTTP protocol.</p> <p>This warehousing enables more people to access the data than if they had to go to the remote sites individually, especially since a lot of these derived products will be in a standard format so it makes it much easier. I think it gets back to an issue of integration at that point.</p>	

Interview ID=21 13 September 2007	ANSWERS	ANNOTATIONS
<b>Q6.2 How do you share work-related data with others?</b>	<p>Right now they're just primarily human interfaces [<i>as opposed to programmatic interfaces</i>]. A Web browser interface that allows people to</p> <ul style="list-style-type: none"> <li>- explore the warehouse</li> <li>- find data of interest</li> <li>- plot it in a dynamic fashion</li> </ul> <p>And then if the data look interesting they're able to download it and pursue their research interests.</p> <p>As kind of the side effect of that whole process though is doing things like managing provenance of the datasets: the provenance from how the original data was collected to how is it derived or integrated into the secondary or tertiary products (in addition to the whole issue of citation of authorship and things like that.) So these side effects are not the primary goals of the project but they're certainly valid secondary goals.</p>	

Interview ID=21 13 September 2007	ANSWERS	ANNOTATIONS
<p><b>Q6.3 By what mechanisms is access to your work-related data controlled?</b></p>	<p>Well it's very open. The only thing that's required of the LTER is that you abide by what we call a data access agreement. The purpose of the agreement is for citation purposes; you agree to notify the author (the owner of the data) if you're going to use the data for publication or research.</p> <p>Data access requires a single registration process in a local LDAP. Future plans might include using the GSI and MyProxy to make this even more seamless. It's a one-time registration and accepting of what we call the LTER data access policy. Other than that it's completely public accessible.</p> <p><i>[prompt asking if each user gets an X.509 certificate as part of the registration process]</i></p> <p>No. Right now it's a single point system. In other words there's no distributed processing, there's no distributed resources. Once the data is in the warehouse it's all localized, so there's no need for a certificate. It's basically a one-time authentication. When a person comes to the website we will use tokens like cookies or something like that to actually track session use. It's not a strict authentication.</p> <p><i>[prompt asking if the seamlessness with the GSI is in the data transfer from the remote site to the warehouse]</i></p> <p>That's part of it. Looking into the future one of the things we'd like to be able to do is make different types of resources available to end users of the system. For example one scenario would enable a scientist to find a number of datasets within our data warehouse, but also through the same interface be able to run a simulation using those datasets on a high performance computing system elsewhere.</p> <p>So that's when I start thinking more about the middleware approach, making those types of resources accessible to end users without having to have separate accounts and separate logins or going through these out of band approaches to trying to make that work.</p> <p><i>[prompt asking who owns the data – the original site owner? the warehouse?]</i></p> <p>Good question, and that hasn't been worked out yet so I don't have an answer for you. The issue of ownership becomes a very gray area as you produce these new products. With a product containing a provenance path back to the original owner, one could say that the person still owns that data. But if I perform an invertible process where I can't get back to the original data, then technically I think we become the surrogate owner of that data.</p> <p>Another example is in the commercial satellite business, and how they consider their data proprietary. In their case it doesn't matter. Many companies consider any derived product from their original product still under license of their business.</p> <p>Let's say I have a derived dataset that goes through a number of stepwise sequences and I create a new product. If I can still go backward and produce the raw product then there's still some level of ownership from the original data owner within my derived product. There's still a component of somebody else's work and toil in my derived product.</p> <p>However if I create something based on a derived product, there is no way I can distinguish what percentage of the value can be attributed to the original owner. I think eventually as time goes on if the product morphs enough, eventually the original attribution somehow fades away. That attribution list can become quite large if you're not careful.</p>	

Interview ID=21 13 September 2007	ANSWERS	ANNOTATIONS
<b>Q7.1 What resources do you use in your work today?</b>	<p>Well as far as computing infrastructure right now, most of the work on the development and production systems takes place on quad core processing blade servers. We generally don't delve into cluster computing.</p> <p>The storage capacity within the LTER here at the network office is on the order of tens of terabytes, not hundreds – and not anything near petabyte scale. As more of our data is used for simulation runs and results need to be stored, our storage needs may scale quickly.</p> <p>With regard to network capacity:</p> <p>We have a national LambdaRail [NLR] connection at my home institution, so we're at the gigabyte Ethernet scale. Do we actually take advantage or need that? Not at this point.</p> <p>In terms of real-time collection of sensor data:</p> <p>Historically the LTER network has handled very small data sets – on the order of thousands to maybe a million observation points per year. The types of data collection are quite varied within the LTER. I can't really tell you that much about the science we do here because I'm actually a geologist, not a biologist.</p> <p>There is a huge multi-million dollar funded project from NSF called NEON, which stands for the National Ecological Observatory Network. The basic idea of NEON is to put lots and lots of sensor arrays out into the environment to monitor a bunch of different ecotones or regimes.</p> <p>Putting sensor networks at all the LTER sites has become quite common now. So the expectation is that streaming data collection will increase over the next one to five years. However, I don't believe the data will be on the scale of what the physics community produces. So even though our storage needs will increase I don't think they'll reach petabyte scales any time soon.</p>	
<b>Q7.2 How do you share work-related resources with others?</b>	Access to the data warehouse is via a Web browser or a Web service, unless somebody requests a tar dump to put it on some other type of physical medium.	
<b>Q7.4 How do you locate available resources for use in your work?</b>	<p>I go to my supervisor and tell him that we need more resources ☺</p> <p>In terms of finding other high performance computing resources is proof of concept work.</p> <p>We have working relationships both with NCSA and the San Diego Supercomputing Center, and so we always keep a line of communication open to others in the field that can help us out in that area.</p>	
<b>Q8.2 What scripting languages have you used in the past year?</b>	Well in terms of web server scripting languages it's primarily PHP. We use perl as kind of the workhorse in terms of doing systems type of file manipulation, regular expression searching and things like that.	
<b>Q8.3 What programming languages have you used in the past year?</b>	We use Java	
<b>Q8.4 What workflow tools do you use in your work?</b>	I don't per se in my own work here but a lot of the research is through another project within LTER. The workflow tool of choice then would be Kepler.	
<b>Q8.5 What parallel computing tools do you use in your work?</b>	Not yet actually but we did a pilot project almost two years ago with NCSA working on one of their Tungsten Clusters. The purpose of that was to do some digital signal processing on environmental recordings specifically looking for bird signatures in these unknown recordings.	
<b>Q8.6 If the need for new software-based functionality arises in your work how do you acquire it?</b>	We generally write it. We develop our own software here. We try to reuse tools that are available, that are functional and we can leverage. Unequivocally we try to leverage open source software because funding is hard to come by. We can't spend a lot of money on commercial software apps and they tend to be very expensive.	
<b>Q8.7 How do you share software with others?</b>	We have an open CVS server.	
<b>Learning about the user's problems</b>		

Interview ID=21 13 September 2007	ANSWERS	ANNOTATIONS
<b>Q9.1 What challenges do you face today in accomplishing your work-related goals?</b>	<p>Probably we have more tasks than we have time allotted to work on them. There are only a small number of us working on these projects. And we need to spend time not only on this project, but a number of other side projects that always seem to spring up.</p> <p>For instance one project that's of ongoing is – well, which one to pick from?</p> <p>There's one effort called the Unit Registry Project. Its purpose is to identify the scientific units utilized within the ecological community and develop a registry of them. This provides a mechanism for different people to vet the efficacy of the units: whether a unit is useful, or if it should be deprecated and replaced with another unit. So there are many ongoing social issues in the project, in terms of developing an infrastructure that supports a vetting mechanism for these units.</p> <p>Another sub-project that's just starting up is an effort to develop an automated metadata creation system. Part of our work here at the network office involves collecting satellite imagery. There's a University of New Mexico program called Create that collects a type of satellite imagery that specifically subset the areas over the LTER sites.</p> <p>Our part of this effort is to develop a system to automate the metadata generation of these collections, because they're created on the order of one or two-dozen every day. That means daily you're creating 20 or 30 different metadata documents.</p> <p>These projects are all related, but it certainly stretches your capacity to its limits.</p>	
<b>Q9.2 What types of information do you need in order to address the challenges you face today?</b>	Not really. I think the technology challenge that I face is the learning curve involved in using different software.	
<b>Q9.3 What technology-related obstacles do you currently encounter?</b>	<p>Just the lack of time to learn all the technologies that I think we need to use.</p> <p>Specific obstacles? Not really. Well, maybe the answer is I'm unaware of them at this point, and I'm sure we're going run into them. So far there haven't been any insurmountable technical problems, though it does take effort sometimes to figure out the best approach or the best solution to a given problem.</p>	
<b>Q9.4 By contrast, can you provide examples of technologies you find very useful today?</b>	<p>From a development perspective the Eclipse IDE is great because it's so extensible.</p> <p>The formalization of Web services I think is an incredible technology to allow machines to communicate with machines in a very standard and structured way. Though there are still questions on what to use in the Web services area: Do I use SOAP? Do I use REST?</p> <p>Honestly those types of remote procedure calls have been around for a long time in the Unix environment using sockets, but they were unique to those platforms. I think the convergence now with the standards is really making things fly. It's wonderful.</p> <p>The whole web mash up concept, the whole thing with Google and Ajax and all that stuff I find fascinating. The ability to pull things together so easily now using a web interface.</p> <p>But at the same time this opens up problems, because it really allows anybody without a formal background in software development to develop these applications. The concern is similar to my pet peeve with Visual Basic: the resulting code seems very fragile and you have to be very careful how you use it. Things break, or the maintenance of those types of applications become a nightmare at times.</p>	
<b>Q10.1 Can you think of any work-related tasks that decrease your productivity?</b>	<p>Oh, project management by far. Just trying to document the work phases.</p> <p>We use a modified version of the Rational Unified Process. Doing all the paperwork that associated with the whole iterative to development cycle – all the documentation and stuff like that.</p> <p>Although it's very useful it's also very time consuming and boring. Team members use these documents for directing their work. People outside of the project use it as proof that we're actually doing something. So it's a good documentation trail.</p>	
<b>Learning about the Globus user experience</b>		

Interview ID=21 13 September 2007	ANSWERS	ANNOTATIONS
<b>Q17 What are the major challenges you face using &lt;component&gt; today?</b>	I have no direct experience other than as an end user of some of the products that were developed long ago with Globus. The technology that we used was effective. It worked the way it was supposed to. However, the scuttlebutt on developing with GT3 was that it was something nobody wanted to work with because of all of the problems that people experienced with it. I'm just speaking from hearsay from others who've worked on projects that we're trying to use the Globus Toolkit 3.	
<b>Wrapping-up</b>		
<b>Is there anything you'd like to say to the people who build software for use by people like you?</b>	Keep doing it. The work is incredibly valuable, and I just don't think we're at a point where we can stop. It's like we're building a car and we haven't put the engine in, so we can't start it yet. When the time is right we're certainly going to start utilizing those services and those applications. Exactly how and when and other details have yet to be decided. But yeah, go for it.	



## D.22 Our goal is to bring attribute-based authorization to the Grid

Interview ID=22 20 September 2007	ANSWERS	ANNOTATIONS
<i>Pre-interview question: Do you interact directly with Globus software in your work today?</i>	yes	
<b>Establishing context</b>		
<b>Q1.1 Please provide a one-minute overview of your project</b>	The project is called GridShib. GridShib is a technology that enables attribute-based authorization. GridShib includes a plug-in for Globus Toolkit (GT) and partially separate from GT. In other words there are components of the GridShib framework that fit into the Globus Toolkit and others that work outside of it, providing a comprehensive set of tools for doing attribute-based authorization.	
<b>Q1.2 What is the project's name?</b>	GridShib	
<b>Q1.3 Which agency funds the project?</b>	Initially it was funded by the National Science Foundation, but that ended in December 2006 Now our work is funded by the TeraGrid GIG	
<b>Q1.4 What field does your project belong to?</b>	Grid Security Middleware	
<b>Q1.5 What is your job type?</b>	Middleware Architect	
<b>Q1.6 How long have you been a &lt;job type&gt;?</b>	Four years	
<b>Learning about discipline-specific goals and approach</b>		
<b>Q2.1 What are the main goals of your project?</b>	The idea is to bring attribute-based authorization to Grids. In the past authorization has been somewhat of a weak point in Grid middleware. It was previously centered on the idea of identity-based authorization (i.e., the grid-map file). As we now know, that doesn't scale. So in order for Grids to grow, we need a new approach to authorization and we think that attribute-based authorization is one possibility.	
<b>Q2.2 How will the success of your project be measured?</b>	The success of the project boils down to whether or not people use it. So it really has to do with deployments and the user community and whether or not people see this as being a valid and worthwhile approach to authorization. If they do, they'll use it; if they don't, they won't. We don't currently have very good metrics for measuring usage. I think that's really difficult to gauge in most cases – not just in our project, but many have a similar problem. You can estimate this based on downloads, based on user feedback... but it's really a difficult thing to quantify.	
<b>Q2.3 What are the professional measures of success for you?</b>	Fortunately my goals are aligned with GridShib goals, at least at this point in time. So as long as GridShib is successful then I feel as though I've done a good job.	

Interview ID=22 20 September 2007	ANSWERS	ANNOTATIONS
<b>Q3 What are you investigating?</b>	<p>From the beginning we have thought and hoped that we could leverage existing approaches to attribute-based authorization. Not necessarily Grid-based ones, but some existing approaches.</p> <p>Shibboleth is a good example of what I'm talking about. In the higher education community, Shibboleth has had quite a bit of success as a project and a technology for web-based applications. The idea is to leverage Shibboleth technology in the community. And so that's been a focus: how can we leverage technology like attribute infrastructure that already exists on campuses today? The challenge is that Shibboleth as a technology addresses Web-based resources such as portals, Web applications – things that sit behind web servers. And the whole technology is built around that particular use case. It has been a challenge to leverage that technology because it doesn't translate directly into Grid-based resources. We've tried a number of things to address this issue, some of which have worked others have not. Still we continue to do our best to take advantage of existing technology and existing infrastructure that's built up around Shibboleth.</p> <p>I think the translation problem really boils down to the end user. I mean the science user or person who wants to use a Grid-based resource as opposed to an ordinary Web-based resource. In order to use a Web-based resource, one uses a Web browser. To access a Grid-based resource, though there are exceptions, one doesn't normally use a Web browser. So really the distinction is how the end user accesses the resources. Whether it is browser-based or non-browser-based. Shibboleth addresses the browser user, whereas Grid users are not necessarily of that type, and that's the basic problem.</p> <p>GridShib as a project is almost three years old now, so what that means is that there have been a number of avenues that we've gone down to attempt this problem. Some have born fruit and others have not. So here we are, over two and a half years into it, and we address some use cases, and have solutions to certain problems.</p> <p>Not all problems in this space are solved – not by a long shot. So there still is research to do. But we do have software solutions that address particular use cases today. We have solved some problems and are making it available to users. We are working with user communities to get this incorporated into their infrastructure. And at the same time we are actively developing and enhancing and refining because we're not quite there yet. It's not a finely tuned product because it's still very much research, I think.</p>	
<b>Q4.2 How do you work?</b>	<p>Initially we had so many different use cases, we didn't really have a feel within the project team which ones we should be concentrating on. We talked with our users and went to conferences and gave presentations and tried to get feedback from the community as to what the important use cases were. Quite honestly we didn't really get the feedback that we were looking for. So I guess you could say we initially floundered trying to determine a path forward.</p> <p>In the summer of 2006 it became clear what our primary use case was. Since that time our path has been very clear, and I can still see six to eight months out into the future: that's how clear it is at this point. But for the first year and a half it wasn't at all clear, so it was a matter of finding our way in the dark, so to speak.</p> <p><i>[prompt asking why it was difficult to pin down the use cases]</i></p> <p>I really don't know. Even within our team there were differing opinions as to what the important use cases were, so we couldn't even agree internally where we should be concentrating our limited resources.</p> <p>And like I said, we talked to our user base and entered into dialogue with folks outside the team to see what they thought, and actually, they were basically looking to us for guidance. So we went back into our shells and kept searching for the right use case where we could really make a dent in this problem. It just didn't jump out at us until the summer of 2006.</p> <p><i>[prompt asking if the TeraGrid GIG helps define a use case]</i></p> <p>Yes, it does. There is a type of TeraGrid science gateway that has a Web interface; you could think of it as a portal for Grid users. So there's where the overlap is. So we're back to where Shibboleth likes to live, with browser-based users. That's good because now, perhaps, we can really leverage Shibboleth in the way it was originally intended.</p> <p>And then on the backend of these gateways, you have traditional GT4 Web services. So there's the Grid-based stuff all wrapped up into one package. So it's a perfect use case because it really brings to bear both technologies. Now we have to figure out how to just get it all to work together as smoothly as possible.</p>	

Interview ID=22 20 September 2007	ANSWERS	ANNOTATIONS
<b>Q4.3 How do you keep track of interim results, if at all?</b>	<p>Within our team and with our users, being a Globus Incubator project, there are certain things that all Incubator projects do. Using mailing lists and Bugzilla are two such things that immediately come to mind. And so we document our work in Bugzilla and use the mailing list to communicate both internally and externally.</p> <p>We talk to each other as developers on the mailing list for a couple of reasons. It's actually good, I think, for the community to see what the issues are even before they're nice and solved. I think that's healthy, so we do a lot of our discussion on the mailing list, and then we use traditional email and IM on the back channel for other kinds of discussions.</p> <p>As far as TeraGrid is concerned, we definitely use the TeraGrid wiki a lot. So TeraGrid has a wiki resource, and we use it to document our statements of work and progress. So not only do we have the Globus wiki and the internal wikis at NCSA, but TeraGrid has one. That's pretty much how we keep them apprised of our progress and any issues that might come up.</p>	
<b>Q5.1 In what ways do you interact with simulations in your work?</b>	<p>Very little, in terms of scientific simulations. We don't really work with data or generate or utilize data. Remember we're a middleware project, so to end users the stuff that we do is kind of boring. To an end user the middleware is doing what it's supposed to do if you're not even aware that it's there. So we have to be as silent and unobtrusive as possible.</p> <p>But there is one simulation that we do, and it's a development test that a colleague has put together. It simulates a communication between the Globus Toolkit and a Shibboleth attribute authority. It does this all in code. It's a very nice simulation because it allows us to run our software through the paces without having to stand up Shibboleth services and configuring them to interact with GT services. We do all of this in software, so that's the closest we get to performing a simulation.</p>	
<b>Q5.5 By what mechanisms is access to your simulations controlled?</b>	<p>The test is available for general use; it's in CVS. When you download the software, you download the test. You can use Ant script to install and run it. We are also working with a <i>dev.globus</i> infrastructure person to get the test incorporated into the build-and-test environment. But we still have a ways to go before we work it out – we have to work out some details. I haven't been able to get it to work in B&amp;T quite yet, but I think we will eventually, especially with the <i>dev.globus</i> infrastructure person's help.</p>	
<b>Q6.1 Describe how you interact with data in your work</b>	<p>Very little. Looking at it from a scientist's perspective, we really don't get involved in data at all.</p> <p>Metadata's another thing, however. When we use the term "metadata" we almost always mean SAML metadata. SAML metadata is important because it describes SAML entities and these entities are what make up the GridShib system.</p> <p>When you install GridShib for GT into Globus Toolkit you create a SAML entity. You need metadata to describe that entity so other entities wishing to interact with you know something about you. It also serves as a basis for trust, so SAML metadata is fundamental to the SAML trust model. So we have SAML metadata in all of our components and we tend to rely on it even more as time goes on. We continue to build infrastructure around SAML metadata. It's an integral part of the GridShib process.</p>	
<b>Q7.1 What resources do you use in your work today?</b>	<p>Aside from the development tools that I've already mentioned (mailing lists, Bugzilla, and various wikis) my development environment is mostly self-contained on my laptop.</p> <p>I do have access to a number of Unix systems here at my home institution that I use for testing the various components we ultimately roll out. And I have access to systems at Argonne mainly to maintain our website and, of course, the <i>dev.globus</i> CVS repository and so forth.</p>	
<b>Q8.1 What software do you currently use in support of your work?</b>	<p>I depend on Globus Toolkit, so whenever anything happens in that area, I'm always quick to download it and try it out with our software.</p> <p>As far as development tools are concerned, I use Eclipse and then various Windows-based tools, editors, etc.</p>	
<b>Q8.2 What scripting languages have you used in the past year?</b>	<p>I've written a number of shell scripts.</p> <p>Ant scripts are very important for our software. Everything is Ant-based because it's all Java.</p> <p>Lately I wrote the GPT scripts with the help of a Globus Toolkit packaging expert.</p>	
<b>Q8.3 What programming languages have you used in the past year?</b>	<p>Mostly Java. Some PHP. But far and away – Java.</p>	

Interview ID=22 20 September 2007	ANSWERS	ANNOTATIONS
<b>Q8.4 What workflow tools do you use in your work?</b>	none	
<b>Q8.5 What parallel computing tools do you use in your work?</b>	none	
<b>Q8.6 If the need for new software-based functionality arises in your work how do you acquire it?</b>	<p>If we need some new piece of functionality in GridShib the first question is, "Has somebody already done the work?" Because if someone has, I'll take advantage of that rather than rewrite it myself, assuming it fits into the environment and meshes well. Certainly I'll look at what has already been done. But as I mentioned earlier, it's still very much a research area. There hasn't been a lot of software that I've been able to access directly in terms of incorporating existing libraries. That hasn't happened all that much.</p> <p>We do have dependencies, of course – many of the same dependencies that Java WS Core and Globus Toolkit has. So we have dependencies, but as far as developing the functionality for GridShib – in other words, bringing attribute-based authorization to grids – that's been a matter of writing a lot of things from scratch.</p> <p>We do borrow as much as possible, especially from the Shibboleth and the OpenSAML projects. I guess I'll mention those two as being significant contributors to the GridShib project. We have really leaned on those two projects heavily and used their code bases whenever possible.</p> <p><i>[prompt asking what it means for external software to "mesh well"]</i></p> <p>There are languages requirements, first of all. If a library only exists in C++ and you're developing in Java, that's a mismatch. And there are also compatibility issues in terms of what version of Java is required. That's always a question. And there can be conflicting dependencies. When you look at somebody else's open-source software, they have a set of dependencies and you have a set of dependencies. The first question you have to ask is, "Are there any major conflicts in terms of those two sets of software dependencies?" Because if there are you need to resolve those conflicts before you can even begin to leverage the open-source package.</p> <p>So things don't always work together as you would like, and that's just the nature of the beast. You have to explore each one and see if it's worth the effort of incorporating it into your project or rewriting it from scratch. I'm all for incorporating things where possible, but sometimes it's not worth the effort and you end up reinventing the wheel.</p>	
<b>Q8.7 How do you share software with others?</b>	<p>We have a website – <a href="http://gridshib.globus.org">gridshib.globus.org</a> – that has a download page. Users can visit the page and download various GridShib components from there. Those components are stored in <a href="http://cvs.globus.org">cvs.globus.org</a>.</p> <p>So we announce the release of new versions in the mailing list, and people download them from the website. They can access the source code anonymously from CVS as well.</p>	
<b>Learning about the user's problems</b>		
<b>Q9.1 What challenges do you face today in accomplishing your work-related goals?</b>	<p>The main challenge is it's difficult to concentrate on one thing because I'm spread so thin. When we made the transition from being funded by NSF to being TeraGrid-funded my involvement in GridShib went from full-time to half time. That means there is less time that I can devote to that development.</p> <p>Now that's actually timely in one sense because, as I mentioned earlier, the path forward seems clearer. So it's less research now and more heads-down development. That is conducive to a half-time involvement and so it's not prohibitive. We continue to make progress, but I certainly could make more progress if I could concentrate my efforts on it solely. But that's not an option.</p>	
<b>Q9.3 What technology-related obstacles do you currently encounter?</b>	<p>I can't think of any technology obstacles at this point. Our use case is well defined. The path forward is well defined. It's just a matter of finding enough cycles to pull it off.</p> <p>It really boils down to human resources. The path forward has never been so clear as it is now, so there really aren't any design or research hurdles that need to be overcome to move forward. It's a matter of finding the correct funding and people cycles to do the work. Since that's limited, that means we won't be able to move forward as quickly as we would like, but that's the way it goes.</p>	

Interview ID=22 20 September 2007	ANSWERS	ANNOTATIONS
<b>Q9.4</b> By contrast, can you provide examples of technologies you find very useful today?	<p>One is OpenSAML. We rely on the OpenSAML libraries extensively. Without those we wouldn't be as far along today as we are. That makes our lives a lot simpler. And as OpenSAML grows, we'll be able to grow too.</p> <p>And also, Shibboleth. Even though Shibboleth addresses a different use case, the Shibboleth code base is tremendously useful. And there are large portions of this code base that we have incorporated or otherwise leveraged inside the GridShib project. So we owe quite a bit to those two technologies: OpenSAML and Shibboleth.</p> <p>I haven't been saying that much about Globus Toolkit because I've just been assuming that we don't exist without Globus Toolkit. Our project just simply doesn't exist. I haven't mentioned the Globus Toolkit because it's the core of what we do.</p> <p>I'll say that in the past six months I've had the opportunity to dive into Java WS Core and understand that deeply. I really do appreciate the effort and expertise that went into building that code base, and we've leveraged it significantly in GridShib.</p>	
<b>Q10.1</b> Can you think of any work-related tasks that decrease your productivity?	<p>The first thing that comes to mind is the wiki: it's a double-edged sword. I think it's a great invention, but the problem is it's maybe too good for its own good. Because now there are so many of them I can't actually keep track of where all the information is located.</p> <p>I have access to so many wikis. It seems like when a new subproject comes online, a new wiki comes online to support it. This is a problem. For me it turns out to be a drain on my time because I just really can't keep track of where all the information is and how to keep it current. So it seems to me some consolidation of wikis is in order at some point.</p> <p>One example of a possible solution to the problem is something like Confluence. Confluence is one wiki, but it has a very refined notion of a space. So you can have multiple projects inside of one Confluence instance. Logically they're separate, but physically it's all one Confluence instance. There's a single searchable database. It's separate and combined at the same time, which makes more sense to me. Internet2 uses Confluence quite a bit, which is how I've come to understand this, and I think it's a great improvement over MediaWiki.</p>	
<b>Learning about the Globus user experience</b>		
<b>Q12.1</b> Which Globus security components do you directly interact with in your work today?	<p>Most of them except the delegation service. I haven't used that or studied it closely.</p> <p>And interestingly, I only know of CAS by reading the documentation. I've never installed it or used it, which is somewhat ironic because it is the one piece in Globus Toolkit that uses SAML. A lot of GridShib is based in SAML, so you would think that I would know CAS inside and out. But it's a very different use of SAML. I have not looked at it as closely as I probably should.</p> <p>I'll just mention that of all the Globus security components, the authorization framework is nearest and dearest to my heart.</p>	
<b>Q15.1</b> Which Globus common runtime components do you directly interact with in your work today?	Java WS Core	

Interview ID=22 20 September 2007	ANSWERS	ANNOTATIONS
<p><b>Q16 Why do you use &lt;component&gt; instead of an alternative technology?</b></p>	<p>MyProxy: MyProxy issues proxy certificates, and a GridShib component called the GridShib SAML Tools also issues proxy certificates. So in some sense these two are competitors, but in another sense they're complementary. Part of the work we're doing is to enhance MyProxy so it can embed SAML into proxy certificates in the same way that GridShib SAML tools is able to embed SAML into proxy certificates. So I am interested in MyProxy because there's really an overlap that needs to be explored there.</p> <p>GSI certificates: I don't think there is an alternative. I don't know what else I would use. I will say that another GridShib component is called the GridShib CA. The GridShib CA issues end entity certificates as opposed to proxy certificates. And we decorate both of these kinds of certificates in the same way with SAML assertions. So in that sense we kind of do an end-around on GSI and use end entity certificates that have SAML in them.</p> <p>But when it comes right down to it, you know, the Science Gateway model that I referred to earlier is heavily into community credentials and issuing the proxy certificates. So we need to support that. That's perhaps one of the reasons why we haven't considered anything else.</p> <p>Java authorization framework: There is no alternative. Because GridShib is a plug-in for GT we need to support the GT authorization framework. I'm happy to say that it's a really nice framework. And it continues to be refined and enhanced. It works fine with us and there's no point in considering an alternative, even if there were one.</p> <p>I think it's based on the XACML model. Actually, I don't know what came first, but I can see that there's a strong resemblance to XACML as specified in the OASIS technical committee, which is good. So the framework is based on concepts that are already well known outside of the Globus Toolkit. That makes it easy for developers to come to grips with the concepts.</p> <p>So it's built on an existing set of concepts. I don't know what their source is but I'm quite sure that it is outside of Globus, and I think that's a plus. That has made it fairly easy for me to understand and code to the framework because it is based on a standard set of concepts.</p> <p>Java WS Core: I don't think there is an alternative, as far as the scope of our project is concerned. If we're going to be a GT plug-in, then that is the base.</p> <p>CAS: CAS is important to our project and me because it is one component in Globus Toolkit that leverages SAML. GridShib also relies on SAML so there is something to be learned by looking at the CAS implementation. It's a different use of SAML, but it's a use of SAML nonetheless. Though I haven't used it but I've looked at it enough for me to understand what it does and how it works.</p> <p>So I've looked at the codebase and understand it from a conceptual point of view. Part of the motivation was to see if there's something to be learned from a design or implementation perspective. We've employed some of the techniques seen in CAS along the way.</p> <p>So this is an interesting use case for the Globus code in the sense that, as a developer, not only do I build components that depend on Globus code, but I also use it as an example. In fact I've used Java WS Core and even the CoG jGlobus in that way. I've studied those codebases extensively, and they've been a tremendous help in developing some of the GridShib components.</p>	

Interview ID=22 20 September 2007	ANSWERS	ANNOTATIONS
<b>Q17 What are the major challenges you face using &lt;component&gt; today?</b>	<p>Whenever I mention proxy certificates outside the Globus community I get strange looks from people. In fact I've gotten negative remarks. There are people in the Internet2 community who just do not subscribe to proxy certificates, even though they're well defined in an RFC. They just don't buy it. That's a hurdle, and I don't expect to overcome that anytime soon.</p> <p>As far as using them on a day-to-day basis, yes, there are some issues there. Our software depends on Java WS Core. It depends on CoG jGlobus. And that's good, as far as it goes, until I find a problem.</p> <p>I've discovered a number of low-level bugs in jGlobus/Java WS Core. And these bugs don't tend to get fixed very fast. I don't know why. Even though I go through formal channels to report them (they're in Bugzilla) they don't get addressed. So that poses a problem.</p> <p>And so I end up duplicating code, which I hate to do. But to keep my project moving forward, that's been my approach. If I find some code which I think is bugged, I copy it into our code base, fix the bug, and move forward based on that duplicate code.</p> <p>Another problem has to do with software dependencies. When you leverage a technology you need to look at its dependencies and compare it with your own to see if they clash. Java WS Core has a very large set of dependencies. This is not an issue for GridShib for GT, which is a plug-in for GT and sits on top of Java WS Core, because it was built from the ground up to work with Globus Toolkit. But one of our standalone components, called GridShib SAML Tools, has its own set of dependencies because it has its own standalone code base. At one point I was asked to investigate incorporating it into GT. The idea was to have it deploy into the GT codebase in the same way that GridShib for GT deploys into GT. This work is still not finished because I've not yet figured out how to reconcile the dependencies.</p> <p>Java authorization framework:</p> <p>I think the biggest challenge is that it is a moving target. We've had to recalibrate or recode at least two times (maybe more) because the GT authorization framework continues to evolve. It's evolving at a relatively rapid rate. Since we depend on it we have to adapt to changes, and that's created some work for us. That's a challenge, though not insurmountable. We've been able to deal with it, especially thanks to one of the Globus developers who really is on top of things. I guess it hadn't been as bad as it could've been, but it's a moving target.</p> <p>CAS:</p> <p>This is more of a future challenge. CAS has existed for quite some time. It's a framework that relies on SAML. GridShib is relatively new and hasn't been around for so long. It also relies on SAML. As GridShib has been developed and grows and takes form, you can begin to see this tremendous overlap between GridShib and CAS. There's overlap at all levels. But at a functional level, in terms of some of the techniques that are employed – they could share a common code base. CAS and GridShib have so much in common now, they could share a common code base. They could be combined into something. I don't know what we would call it, and I wouldn't venture a name at this point, but there's a lot of overlap and that needs to be exploited at some point.</p>	
<b>Wrapping-up</b>		
<b>Is there anything you'd like to say to the people who build software for use by people like you?</b>	<p>I've had very good luck working with Globus developers and it's been a rewarding experience for me. I can't really think how that could be improved. It's working rather well, I think.</p>	

## D.23 It would take forever for a biologist to get this machinery working

Interview ID=23 21 September 2007	ANSWERS	ANNOTATIONS
<i>Pre-interview question: Do you interact directly with Globus software in your work today?</i>	Yes	
<b>Establishing context</b>		
<b>Q1.1 Please provide a one-minute overview of your project</b>	<p>GNARE is a project for genome analysis. It enables protein sequence analysis of various organisms that the bio-community is coming up with. Before GNARE there was no system that allowed end users to submit their sequences and analyze them. Most systems that take publicly available genomes, do analysis and then publish the results. But there was no system before GNARE that allowed <i>end users</i> to submit their genomes, perform high-throughput analyses and display results.</p> <p>GNARE takes as input from users protein sequences in text files. Then it runs a variety of bio-tools on the input. The tools are computationally intensive; they can take a long time to actually run on a single machine. In order to increase the speed of analyses, we use Grid resources.</p> <p>We use both OSG and TeraGrid for this. We wrote a portal that is connected to TeraGrid and OSG in the background for job submissions. So we run the various tools on the Grid, store results in an Oracle relational database, and then display them back to the users in a very fast manner.</p> <p>Before GNARE no system existed that would return results to users in a couple of hours. Some systems would take input sequences and perform analyses for about a week before returning results. And the transactions were one-to-one, with communication through email. Users would send files to the analysts, who would return them via email. So GNARE was the first portal-based system to enable large jobs to run immediately on the Grid; it organizes the sequence of jobs and delivers the results back to the portal.</p>	
<b>Q1.2 What is the project's name?</b>	GNARE	
<b>Q1.3 Which agency funds the project?</b>	The project is no longer funded. The portal is still running with the last version of the code we released. We have six months left on an LRAC allocation on TeraGrid and it's using OSG resources opportunistically. People are still submitting jobs to GNARE and are using it, but there is no active development at this time.	
<b>Q1.4 What field does your project belong to?</b>	Genomics and Bioinformatics	
<b>Q1.5 What is your job type?</b>	Project Lead, System Architect and Developer	
<b>Q1.6 How long have you been a &lt;job type&gt;?</b>	One year for GNARE and two years for prior work, so three years total.	
<b>Learning about discipline-specific goals and approach</b>		
<b>Q2.1 What are the main goals of your project?</b>	To perform high-throughput analysis of sequence data, to deliver results as fast as possible, and to make Grid resources available to the community. So at a high-level the goal is to create a gateway to the Grid for use by biologists.	
<b>Q2.2 How will the success of your project be measured?</b>	<p>We developed GNARE by allowing individual groups of users to repeatedly submit genomes. Each group submits five or six genomes (or even more). Then the groups annotate the genomes. We measure how many groups are using GNARE. Currently we have about seventy different groups that have genomes in there. And they're doing community curation on the genomes. So it is a user-based system. It's accessible only to the group of users; they can log in and see only their genomes. Each group has more than one member in the group. So there are actually hundreds of people who are registered.</p> <p>Not all of the groups are actively submitting new jobs right now. Most of them are using it for visualization and for looking at the data. They're not submitting any new genomes because once they're done with the analysis they use GNARE for reviewing the results. It has a graphical user interface that includes visual analysis tools. So all of them are using the system, but not all are submitting new jobs.</p>	



Interview ID=23 21 September 2007	ANSWERS	ANNOTATIONS
<b>Q2.3 What are the professional measures of success for you?</b>	Publication.	
<b>Q3 What are you investigating?</b>	<p>The most important aspect has been to create a dynamic interface to the Grid resources, which is a tedious job for a biologist who wants to use them. It would take forever for a biologist to get all this machinery working together.</p> <p>In my experience the most difficult part has been to connect the user interface to the component that generates jobs and submits them to the Grid resources. That was the most technically challenging part of the project.</p> <p>The “dynamic” aspect of the system is in the resource selection logic, which dynamically selects resources from OSG or TeraGrid to run the analyses. This feature has been a big challenge to implement. We had problems in part because there was no existing resource reservation system we could use.</p> <p>The key requirement for GNARE is that all job submissions be automatically handled: to accept genomic sequences from the users and find available resources to process them. Implementing this has been the tricky part. This work has involved handling job failures, site failures, etc. These are the issues in the backend that we focused on.</p>	
<b>Q4.1 What is your method for investigating &lt;phenomena&gt;?</b>	<p>In the first stage of the project we developed a system called GADU that actually handles the job submission. So the basic idea with GNARE is to generate the workflows for these jobs and submit them using GADU. We use VDL (like Swift is used now) for defining the workflows.</p> <p>The user comes to the portal and selects some parameters for the type of analysis desired. Based on the parameters GNARE dynamically generates the workflows. Then the workflows are converted into Condor jobs. Then the tricky part is to do the site selection. We implemented a site selector for OSG and TeraGrid that does dynamic site selection across the Grid and submits the workflows to those sites.</p> <p>Research aspects of the project include:</p> <ul style="list-style-type: none"> <li>- finding out how to run the jobs</li> <li>- identifying the problems associated with selecting sites</li> <li>- developing solutions to those problems</li> <li>- interacting with the different resource architectures</li> <li>- developing techniques for interoperating with them.</li> </ul> <p>So these are the various things we focused on.</p>	

Interview ID=23 21 September 2007	ANSWERS	ANNOTATIONS
<b>Q4.2 How do you work?</b>	<p>Using the example of implementing the site selector: When we first designed GNARE we could see the biggest problem facing us was getting jobs to run on the Grid:</p> <ul style="list-style-type: none"> <li>- How to identify the sites that are available?</li> <li>- How to even know in the first place which sites are accessible?</li> <li>- If a job fails, how should we handle it?</li> </ul> <p>So we came up with a plan by listing all the things we required to run a job on the Grid. From that we understood that we needed to select sites that could authenticate us and would allow us to run jobs. Once a job executes, we understood the need for a tracking mechanism that would track if the job succeeded or failed. So we track jobs using some of the VDL data, the Condor logs and the Condor queue information.</p> <p>The tracking approach is based on observing recent system behavior, not some big algorithm. We look at the Condor queues and at the performance of previously-submitted jobs on the various sites. If one of the sites is putting all of our jobs in its queue, then we know that for whatever reason that site is not accepting our jobs, so it doesn't make sense to submit new jobs to the site. In this case we take the site off our list and send the job to another site.</p> <p>We record the performance of all the sites used in the previous run. So the site selector looks at the previous performance, the current Condor queue and available information services (like GridCat provided by OSG). Based on those inputs the site selector dynamically chooses where to run a job.</p> <p>So we listed all the parameters relevant to selecting a site and implemented a simple site selector. Once we had that we began tracking the job successes and failures, feeding into a resubmission process. We automatically resubmit jobs that fail.</p> <p>The site selector is a daemon that provides a site upon request. Failures are reported to the site selector, which in turn removes it from the list of available sites. I have a Web page that shows the current list of available sites in one column and the unused list in another. As soon as a job fails for some reason (like it's not authenticating) the site is automatically moved from the available list to unused list. This is all done automatically.</p> <p>So we created all these modules and then we put a user interface on top for accepting jobs.</p> <p><i>[prompt asking about the difference between TeraGrid and OSG in terms of the site selector]</i></p> <p>That is an interesting part. Actually we got a publication on this subject when we wrote a paper on interoperability of GADU using OSG and TeraGrid. The point we made in the paper is that Globus actually makes different sites (like the many TeraGrid and OSG sites) appear like they are the same site. I mean Globus just completely shields us from the fact that they are different. In the end it looks like one type of site to us. As long as I have a GRAM pointer that I can submit jobs to, I don't have to worry about what scheduler is there.</p> <p>The only thing I need to worry about is the hardware – if my executables are able to run there or not. Like for example if it is a 64-bit machine I need to send the correct type of binary to the machine. Other than that the whole mechanism stays the same for OSG and TeraGrid. So the site selector has a list of all the OSG sites mingled with all the TeraGrid sites. It doesn't treat them as separate Grids, but as one large pool of clusters.</p>	
<b>Q5.1 In what ways do you interact with simulations in your work?</b>	<p>We don't have any simulation work in it. Its focus is on running the different tools.</p>	

Interview ID=23 21 September 2007	ANSWERS	ANNOTATIONS
<b>Q6.1 Describe how you interact with data in your work</b>	<p>There is a huge amount of data movement in this project.</p> <p>The input for every bioinformatics tool is a protein sequence file. Bioinformatics is based on comparative analysis. The biologists take a set of sequences and compare it against another set of sequences. And most of the time the database used in the comparisons is huge – containing millions of sequences. For example the database we use is a non-redundant database containing about five million protein sequences. This has to be installed on all the Grid sites.</p> <p>There are two ways of handling this. One approach would be to send the database with every job that we submit. Another approach is to preinstall the database and point to it within the job, sending only the input sequences with the job.</p> <p>We use the second approach. Whenever we get a new version of the database, we preinstall it on all the Grid sites across TeraGrid and OSG. We have an environment variable that is set in all these sites. So our jobs just use that variable to point to the database.</p> <p>We move the data local to that site because the database does not have remote access. The jobs that perform the comparative analysis actually read the input sequences and compares them with the sequences stored in the database. So all the tools that we run actually need the datasets; they are inputs to these tools. It's not possible to run these tools with a remote access database. These tools need the data on the local file system in their own flat files basically. When I say "database" in this context it's not like a DBMS or anything. It's just a local flat file.</p> <p>So users provide the input data. They upload it via the portal. And then that data is transferred to the Grid site with the jobs using GridFTP.</p> <p>Once jobs are done, the data is then pulled back onto the host machine using GridFTP. Then we store those results in an Oracle database. And then the portal actually queries the database to display the results to the users.</p>	
<b>Q7.1 What resources do you use in your work today?</b>	<p>It's mostly compute cycles; we use OSG and TeraGrid for our computational requirements.</p> <p>And then we have a database that's maintained at my home institution. It's an Oracle database. All of the GNARE data is stored in this central place.</p> <p>We also have all these outputs and inputs in flat file format, but references to them are stored in the Oracle database. So all portal accesses go through the Oracle database. We have the flat files saved to ship around with the jobs and also so the users can download them.</p> <p>The portal code resides on machines at my home institution alongside the Oracle database.</p>	
<b>Q7.4 How do you locate available resources for use in your work?</b>	<p>That's what the site selector actually does for us. Particularly whenever a new job comes in it gets translated into a huge, parallelized workflow. Then we submit it in chunks to different sites, with the site selector basically picking one site at a time. We're not using ClassAds for this. The site selector has a daemon process that runs small remote sleep jobs every five minutes on all the sites to make sure they are reachable.</p> <p>Let's say that GNARE is completely idle and no jobs are being run. When a new job comes in, the selector chooses one of the sites that successfully ran a sleep job in order to run the new user job. The site selector doesn't really examine the remote site. If a site does not accept our sleep job, then we consider it as not available. Maybe the site is completely free, but if it won't accept the jobs for some reason (or my job is sitting in the queue) then I don't submit new jobs to the site.</p> <p>The way of looking at the queue is through the Condor queue. It's probably not the best way, but it works. So we submit one job to a site. We look at Condor's information and see what happened to the previous job. If both jobs have run, then we can consider that site to be okay. So we'll submit new jobs to that site.</p> <p>The way we know the site exists in the first place is VDS. We use VDS for workflow management. In VDS we maintain a catalog of all the sites to which we have access. Whenever we generate a new workflow, the workflow contains site-specific information. The VDS maintains a catalog of all the sites that can be used for executing this workflow. Basically the site selector uses this catalog to get a list of sites that we have access to.</p>	

Interview ID=23 21 September 2007	ANSWERS	ANNOTATIONS
<b>Q7.5 What types of information do you need to know about a resource in order to determine if it is suitable for your work?</b>	<p>We have a basic sanity check where we see if the site can authenticate us. We had a GADU VO for OSG. On TeraGrid we just used one certificate for all the jobs, so we would just check for authentication.</p> <p>So one of the first things we check is if the site authenticates me. And the second thing we test is to run a small GridFTP job to make sure the GridFTP interaction is working. And then the third thing is to run a small sleep job to see if there is any failure, for example sometimes we encounter IO errors. These are some basic sanity checks.</p> <p>But there have been many instances when all these checks would succeed but the actual job would fail. Basically then we ignore those sites and resubmit the job to another site.</p> <p>All jobs are submitted using GRAM2.</p>	
<b>Q8.2 What scripting languages have you used in the past year?</b>	The whole thing is implemented using perl.	
<b>Q8.3 What programming languages have you used in the past year?</b>	Perl.	
<b>Q8.4 What workflow tools do you use in your work?</b>	VDS, which is Swift now.	
<b>Q8.5 What parallel computing tools do you use in your work?</b>	<p>We use specifically Condor.</p> <p>We use Globus.</p> <p>And we were actually using the VDT toolkit that's provided by OSG. It has a lot of tools in it, but the major ones we use are Condor, Globus, and GridFTP.</p>	
<b>Q8.7 How do you share software with others?</b>	<p>We haven't because the site selector is really specific to our architecture. It is difficult to make it more generic because it depends on things like Condor queue and VDS for site lists, which is not in standard use by everyone. People don't use Condor all the time and very few people use VDS.</p> <p>There was some discussion at some point about writing a site selector based on the MDS information services. Like we could use MDS to maintain a list of sites instead of using a VDL site catalogue. But we never found time to work on this idea. Once we got the portal up-and-running our immediate focus was on building the user community for GNARE.</p>	
<b>Learning about the user's problems</b>		

Interview ID=23 21 September 2007	ANSWERS	ANNOTATIONS
<b>Q9.1 What challenges do you face today in accomplishing your work-related goals?</b>	<p>The most difficult thing was implementing the site selector, as I mentioned previously.</p> <p>No matter what we did we always ended up having jobs that would fail or just sit there not doing anything. And we always encountered a new set of problems that were not taken care of in the previous implementation. We had to keep changing, keep looking.</p> <p>I can say that we came across all possible problems and fixed them one after another. One difficult part was that we have a huge number of sites, both from OSG and the six or seven TeraGrid sites. We had access to so many sites, but then the challenge was to do the selection across such a large pool. If I have seventy different sites to choose from, then I need to understand:</p> <ul style="list-style-type: none"> <li>- where to submit the job</li> <li>- how to know which sites are available</li> <li>- how to know that a site has failed</li> <li>- what should happen if a site fails</li> <li>- how to know that a site will be going down</li> </ul> <p>When we started working on this we didn't have any information services built-in to these Grids. So there was no GridCat. There was a version of GridCat but it wasn't up-to-date.</p> <p>When we were first implementing the site selector we were relying on remote site information. We ran a daemon on every remote site that would in turn report back to the site selector. But that didn't work out so well because if the remote site went down then the whole system went down.</p> <p>Another problem scenario:</p> <p>Let's say we had a daemon running on each of the remote sites, inspecting the queues. Further, let's say that a given daemon saw that a hundred nodes were free, and reported that back to GNARE. But this still didn't guarantee that the jobs will be run. The hundred nodes might be reserved for somebody else. Also some sites had restrictions based on the VO; some sites would only allow N jobs at a time to be run by our VO. And some sites were dedicated to supporting a specific VO so they would restrict our jobs. So even when the daemon saw free nodes our jobs might just sit in the queue forever.</p> <p>There were other issues problems we encountered with the daemon approach that we weren't able to resolve.</p> <p>So we thought it would be best to ignore the remote sites and keep all of the site selector implementation on the portal site. I submit one job at the remote site, then I try to submit the next job. If the site cannot run either job then I don't submit any more jobs there. We eventually settled on this approach after our experiences with the many problems we faced.</p> <p><i>[prompt asking if there were any challenges that could not be overcome]</i></p> <p>There were other challenges that required us to come up with work-around solutions. For example, there were some sites that would show us as running forever. The jobs would have the status as running, running, running... and nothing was really happening.</p> <p>We had another problem that we could never solve. There was this one site where all the resources were dual-CPU nodes. When we submitted BLAST <i>[bioinformatics tool]</i> jobs one of the jobs would be assigned to one CPU. If the next job were assigned to the second CPU on the same node it would crash due to memory problems. The problem was that neither Condor nor Globus would report this type of crash.</p> <p>So in this case half of the job would run, and then if another BLAST job came in, that other half would crash out. So you get an inconsistency, in the sense that incomplete output would come back and not be reported anywhere across all the software layers. Condor didn't catch it. Globus didn't catch it. Nobody caught the error. So we assumed that it was all completely done and we loaded the results into our Oracle database. And then once the user looks at it he saw bad results because of it.</p> <p>And we could never figure out how to fix this problem. We had a long discussion with some of the OSG sites. Then they implemented a policy where they would not submit more than one of our jobs onto each node.</p>	
<b>Q11.1 Which Globus data components do you directly interact with in your work today?</b>	<p style="text-align: center;"><b>Learning about the Globus user experience</b></p> <p>We use RLS and GridFTP.</p>	

Interview ID=23 21 September 2007	ANSWERS	ANNOTATIONS
<b>Q11.4 How many people currently use your &lt;component&gt; server</b>	You can think of it as all of the users who submit a job from the portal – basically all of their transactions use GridFTP.	
<b>Q12.1 Which Globus security components do you directly interact with in your work today?</b>	We use GSI certificates. And then we have this VO thing. I don't know if it is a Globus technology. We have a VO for OSG that puts all the certificates for us on the Grid sites. We have a portal account; we have a proxy there for the portal and then all the jobs are submitted from the proxy.	
<b>Q13.1 Which Globus execution components do you directly interact with in your work today?</b>	GRAM2	
<b>Q16 Why do you use &lt;component&gt; instead of an alternative technology?</b>	GridFTP: Because we're using VDS for the workflow. The workflow that it generates basically uses GridFTP. We didn't have any specific reason for that. It was just because we chose VDS and Globus. RLS: Same reason as GridFTP – VDS uses RLS. GSI certificates: Because of the way this whole thing was set up. That was the most convenient way of accessing these Grid resources, especially OSG. Because we didn't have logins on any of the OSG machines, it was all done using GSI certificates. GRAM2: It is the best thing available right now. It just makes everything easy. It completely hides all the complexity. And one of the reasons is the interoperability it offered between OSG and TeraGrid. All we needed was a GRAM endpoint and that's it. VDS: We everything we needed was integrated: Condor and Globus.	
<b>Q17 What are the major challenges you face using &lt;component&gt; today?</b>	GridFTP: I don't think we have any problems with GridFTP, in fact. RLS: One of the problems we had – there are two components to the RLS program: one is RLI and there is another component. One of these RLS components didn't get immediately updated and we couldn't figure out how to fix this. Whenever we listed a component in RLS, if we immediately queried it we were not getting those components back immediately. So we had this problem and nobody could figure out why it was happening. GSI certificates: It is perfect, in fact. We never have any problems with it. GRAM2: Just from a GRAM point of view, I don't see any problems. VDS: We were a first test user of VDS, so we basically went through all of the bug fixing. Apart from that, once we had the whole system working, it was perfect. Another problem we had was VDS kept changing all the time: from VDS, to Pegasus, and now to Swift. It's been a changing like every year.	

## D.24 ♦ The vast majority of people who could use supercomputing are excluded

Interview ID=24 24 September 2007	ANSWERS	ANNOTATIONS
<i>Pre-interview question: Do you interact directly with Globus software in your work today?</i>	No	
<b>Establishing context</b>		
<b>Q1.1 Please provide a one-minute overview of your project</b>	<p>I'm the technical director of the Network for Computational Nanotechnology. The core effort of the network is to provide online simulation services to a group of nanotechnologists around the globe. We want to move actually nanoscientists to nanotechnology, so we want to put simulation tools into the hands of people that normally wouldn't touch simulation with a ten-foot pole. The target audience is experimentalists that have work to do in the lab and they want to maybe design before they build. They're educators who want to train their students. They're students who want to study nanoscience and simulate structures. And they are potentially industry people as well as government persons. We run the nanoHUB, the online simulation facility.</p> <p>The facility over the last twelve months has hosted 5700 simulation users that ran over 225,000 simulations. They did so without any particular UNIX knowledge. They were able to set up the experiment proactively and run the experiment, and overall we've used around 3 million CPU hours to provide that service. Most of the simulations that the 5700 used are actually very small simulations. They're not generally speaking HPC, CPU-intensive applications. We have built our own middleware to enable this type of interactive simulation and we've built our own toolkit to enable building of graphical user interfaces. This can happen very rapidly, where a user interface can be created in a day or two. We basically host UNIX-like applications for Web browsers and deliver them to the user.</p> <p>We also host tutorial seminars and podcasts on the nanoHUB. Overall 50,000 people use the nanoHUB, so the vast majority of people don't run simulations. They are gathering information about nanotechnology – that's an interesting side effect.</p> <p>There's a similar smaller scale effort in Europe and there are a couple of other nano modeling simulation sites that offer services that are similar to the typical models that people build, which are non-interactive and static. So our focus on interactive simulation and data visualization is unique.</p>	
<b>Q1.2 What is the project's name?</b>	Network for Computation Nanotechnology	
<b>Q1.3 Which agency funds the project?</b>	National Science Foundation is the primary funding source and my home institution, Purdue University, is cost sharing a significant portion. We have associated funding from other government agencies, industries, etc. to do the research nanotechnology part. We don't have core funding from the outside for middleware and Grid infrastructure otherwise.	
<b>Q1.4 What field does your project belong to?</b>	Nanotechnology	
<b>Q1.5 What is your job type?</b>	Associate Director for Technology	
<b>Q1.6 How long have you been a &lt;job type&gt;?</b>	Three months shy of four years.	
<b>Learning about discipline-specific goals and approach</b>		
<b>Q2.1 What are the main goals of your project?</b>	The main goal would be to enable researchers access to computing simulation codes to further nanoscience and nanotechnology. In particular not to address computational sciences and their problems, but people with real problems to solve in laboratories and experiments.	
<b>Q2.2 How will the success of your project be measured?</b>	The ultimate success would be to change the expectations of experimentalists and educators regarding theory and modeling and simulation, and ultimately to change the way they do work. Really seeing the concept of "simulate first, build later" be pursued in several areas of nanotechnology.	

Interview ID=24 24 September 2007	ANSWERS	ANNOTATIONS
<b>Q2.3 What are the professional measures of success for you?</b>	<p>There would be citations in the scientific literature of usage of nanoHUBs by people I don't know.</p> <p>We request when we offer services that people cite the nanoHUB and we try to keep track of it through Google Scholar or various other search engines. Right now I think in the nanoHUB we have 200 citations or so and 60% of them are by people who are not affiliated with NCN (Network for Computation Nanotechnology).</p>	
<b>Q3 What are you investigating?</b>	<p>We work to enable people to use simulation for their real work. If you look at commercial codes that are doing real work in design and engineering, they will have an interactive graphical user interface that allows people to set up their experiments, run their experiments, and to analyze the data that comes out of the experiments, and I mean numerical experiments. So there's a user interface that provides transparency in terms of the processes that are going on.</p> <p>All other Web service-oriented or a portal-oriented science gateways that I know of provide interfaces that look very much like a bank would provide you. The user fills out a couple of numbers, a couple of drop-down menus, and then you run the job. After some time you get the data and static graphs back that you can't do anything with. You have to download it onto your own system and then run another case, download that data, and then you can compare things. In other words you have to have a secondary installation to do real work. You can't do it all in that Web environment. What the real user needs is the ability to run his experiments in a Web browser and support an end-to-end workflow for their work. Unless you can provide that you're just providing a secondary service.</p> <p>We hosted a Web form-oriented system since 1994 and were seeing flat usage numbers of about a thousand simulation users annually. Despite all kinds of arm-twisting the numbers wouldn't grow. When we introduced interactive simulation services the numbers jumped up to 5700, so usage went up five- or six-fold within 2 years. That's a dramatic change in how online simulation is done and how it's offered. So thus far we've seen a dramatic increase in usage numbers and a dramatic drop in the number of source downloads.</p> <p>After the conversion of a tool from a web-based delivery to a fully interactive delivery hardly anybody downloads the source code to install and run it himself. People use it on the nanoHUB. So end users are (generally speaking) not interested in installing s/w on a UNIX system and work with Grids, etc. They want to solve problems. They're not tool builders, they're tool users.</p> <p><i>[prompt asking for more detail about the user interface design]</i></p> <p>We have fundamentally two types of tools on the nanoHUB. I will talk about the geek version first because it is easy to explain. Most simulation engines have a reasonably arbitrary input language that is customized to that particular tool. The scripting language allows the user to describe the geometry and maybe the materials that are in the structure, the algorithms that are being executed and the sequence in which they execute. But you have to be a tool expert to generate your first input file file.</p> <p>So this class of tools allows the user to basically enter free text in a form, allowing them to enter the input script. Then they would hit a simulate button and the middleware takes care of the simulation. Then the toolkit would render the result in a form that allows users to interact with it: they can change fields, compare data, do further runs, etc. So at least you wouldn't have to be a UNIX expert to do these things. There are a couple of tools like this, but unless you're a tool expert you're not going to learn the (reasonably arbitrary) scripting language. On the other hand these tools are very powerful and very general, and they can deal with different topologies or geometries of different devices.</p> <p>The second class of devices is geared towards users who are not tool experts, are not interested in becoming tool experts, but have end-to-end problems to solve. For these users, who actually outnumber the tool experts by factors of 10 or 100x, we generate simpler interfaces. We basically say okay, let's have a class of devices and call it PN Junction. It might only have like 30 or 40 parameters that refer to geometry or the materials, etc. There might be three different tabs in the window allowing a user to change those parameters in an appealing fashion. There are simple integrated pop-up user hints for all the options that users can chose.</p> <p><i>[answer continued on next page]</i></p>	



Interview ID=24 24 September 2007	ANSWERS	ANNOTATIONS
<b>Q3 What are you investigating</b> <i>[continued]</i>	<p>With this approach the user (such as a professor for a class) will be able to teach his students about P-N junctions without requiring them to learn this cryptic language that they would've had to use for this industrial strength tool. So yes the tool is now limited to P-N junctions but it has enough options in there, allowing for some 10-40 different options. If you assume that each option might have 10 reasonable variable choices, you still have a parameter space of <math>10^{40}</math> combinations. So it's a real tool that can do real work and is also useful for research. We have some example applications like these, PN-junction lab, MOSCAP lab, MOSFET lab, and DriftDiffusion lab, which all use the same sophisticated computational engine Padre. The nice graphical user interfaces can be generated by students within a few days of work and do not require fundamental s/w efforts.</p> <p>Although I said it second in the list it is really the primary class of tools. We take simulation engines that are powerful and put a user interface on them that is relevant to an end user without requiring them to understand the guts of the code. We have about sixty-five tools in the nanoHUB and I would guess right now that sixty of them that are in the simple user interface mode and the other 5 might be in a scripted form. You might say, "Well you can't really do research using the simplified tools." I very much beg to differ on that. When experimentalists use a tool they conduct research as well. So it's not geared toward the computational scientist that develops algorithms, but for a person that has a problem.</p> <p>The other requirement is that generally I have plots that show on a log-log scale a straight falling line that has number of users vs. CPU time. And it shows that most users really want fast execution time (meaning minutes or seconds) and only very few users are willing to wait an hour or a day to get their results. In fact most people would have forgotten what the parameter setting was if the results came back a day later.</p> <p><i>[prompt asking how the usage patterns are captured]</i></p> <p>We monitor every single simulation in terms of CPU time being used, etc., and we have a significant effort on mining the data and learning about it. We do not monitor the content of the simulation.</p>	
<b>Q4.1 What is your method for investigating &lt;phenomena&gt;?</b>	<p><i>[prompt asking how the interface to the simulation engine is made user-friendly]</i></p> <p>We develop an infrastructure we call Rappture, which stands for Rapid APplication infrastrucTURE. It is a language that supports data abstraction. The user describes inputs and outputs for the simulation engine in XML. And the tool automatically renders the XML and the description of the input into the graphical user interface. It also manages the data, gathering the output of the simulation. Rappture is easy to use and students can create beautiful interfaces literally with a few days.</p> <p>So all but five of the nanoHUB tools right now are utilizing this Rappture toolkit. In the toolkit the data analysis is built-in, where the user can vary input parameters and at the end automatically compare and tweak the data, asking additional questions even if different variables are involved. Also it imbeds 3D data rendering on a remote hardware render farm.</p> <p>The tool is under active development. It is open source, and is available at <a href="http://rappture.org">http://rappture.org</a></p> <p>Architectural requirements are determined by two designers who have been building user interfaces and tools for the last fifteen years.</p> <p>As far as user requirements are concerned, we did not do any surveys to gather those types of requirements. The nanoHUB team has hired electronic design experts for building commercial user interfaces over the past several years. And I myself have built interactive simulations for the last twelve years as well. So we have worked together to decide what the capabilities of that infrastructure shall be. It is hard to ask users like experimentalists or educators who have been underserved by modeling and simulation who have no expectation for their user interface expectations.</p> <p>Our users drive the decisions to add new capabilities. nanoHUB has a suggestion box where we gather further requirements. And as we add more and more tools and talk to the actual tool developers we find more and more requirements as well that are being added to the toolset.</p>	

Interview ID=24 24 September 2007	ANSWERS	ANNOTATIONS
<b>Q4.5 How do you document your results?</b>	<p>All the projects at nanoHUB.org have workspaces, an associated Wiki and an SVN source repository. The nanoHUB team reviews the status of these projects on a bi-weekly basis. There are 130 ongoing software developments efforts on the developer.nanoHUB.org web page. The developers of these tools are included in the count of 5700 users. (The only people excluded from the usage statistics are the people on the core nanoHUB team.) We have about 65 active tools that are in the end users' hands, and we have some 50 more under development and maybe 10-20 dead ones that are still in the management system. Then we have the Rappture software project as well.</p> <p><i>[prompt asking for clarification on the relationship between Rappture and the tools hosted on nanoHUB]</i></p> <p>For these nanoHUB projects Rappture is a library that enables tool developers to utilize the IO. Some tool developers (5 out of the 65 tools) decided to build their own user interface and that's fine. We can host any X11 UNIX application that runs on a Linux box as is, so we don't require people to use Rappture. But most people don't have an adequate user interface and we help those developers build user interfaces rapidly. Rapidly means that an undergraduate who knows nothing about the science can have at least a rudimentary graphical user interface up-and-running in two or three days.</p> <p>Active tools on the nanoHUB and the Rappture infrastructure undergo continuous improvements. We keep track of this using a tracking system, which is where users can also file bugs.</p>	
<b>Q5.2 How do you share simulations with others?</b>	<p>Some use scenarios:</p> <p>There's a faculty member in the department of electrical engineering typically or physics who is teaching a class on semiconductor devices. He's interested in teaching P-N junctions and wants to teach students about the basic principles of device operation. He can direct his students to the nanoHUB. He might have found homework assignments geared toward nanoHUB's P-N junction tool. He can download these homework assignments and modify them at his leisure and then assign them to his class. So the class will sign up on the nanoHUB like regular users. They will run the simulation tools and they will do their homework. So that's I would say a rather typical workflow that might happen. Each tool can be shared with other users, so the students can type in the nanoHUB login of their faculty member (if they know it) and the faculty member can actually be on his computer in the shared area at the same time with his students. He can be on the phone or through e-mail and say, "Well here's your problem."</p> <p>Another scenario among my favorites is the student at Stanford who was taking a class in 2005 on nanotechnology. In the course he learned about basic nanotechnology device concepts. Later he used one of the tools from his class h on the nanoHUB to get device characteristics that were then put into his circuits. His use of the nanoHUB was not the primary goal of his work, but it was a means to an end (the end being the circuit design.) The student published an IEEE Transactions paper on alternative circuits designs. The case is interesting because he was not a computational scientist interested in improving the codes. He was using nanoHUB to learn, gain insight into a new field, and applied that insight to a related area.</p> <p>Another story I like is we have a tool called CNTbands that can simulate electronic band structure. In the original design we felt it was a pure educational tool because it visualizes carbon nanotubes, giving some pretty rudimentary information about properties of nanotubes. But we have a strong support letter from an experimentalist who used the tool to sort out his experimental data. This is neat because a tool that may not be the sophisticated enough for a computational scientist might be quite useful for an experimentalist.</p> <p>An example I like to cite as part of our tool development is a tool called NanoWire built by our own group. A year and a half ago we put it on the nanoHUB. After two weeks of being out in public, it was used by 50 people without us having to do anything. People just found it and ran it. It's a parallel tool that is now using the Open Science Grid to provide computation. After a year and a month or so there have been 690 users and – I'd have to look at the details – but I think 9,000 simulations have been run with it. One person who ran about 2,800 simulations with it published a paper in IEEE that cites his nanoHUB usage. He's a theoretical person, but again not a tool builder but a tool user who uses the tools to explore potential design spaces. He looks at what happens if you can't make these nano wires exactly the same. What happens if they have variations in them? What are the consequences for design? So again it's a way to get a tool that was being built for computational science and research put it in people's hands to make further use of it.</p> <p>So there are four different scenarios, three rather concrete with people and one more general in terms of the classroom.</p>	

Interview ID=24 24 September 2007	ANSWERS	ANNOTATIONS
<b>Q5.5 By what mechanisms is access to your simulations controlled?</b>	<p>End users can fill out a form that's online. It takes 2-5 minutes. They push the "submit" button and virtually instantaneously receive an e-mail. They click on a link in the e-mail (to verify that they're actual people and not a robot) and they get an account within ~5 minutes. No allocation process, no proposals to write – no convincing required.</p> <p>If you are a developer as part of a development team then you can in principle get access to what we call workspaces. These are UNIX workspaces that run inside a browser and provide connectivity to OSG and TeraGrid. We have a separate approval process for workspace accounts, so access is not automatic for computational scientists wanting to install their own code.</p> <p>But end users are not in the business of installing code. End users access the resources through their use of a tool. With deployed code such as NanoWire, which uses the Open Science Grid, the end user isn't really aware of the resources involved.</p>	
<b>Q6.1 Describe how you interact with data in your work</b>	<p>We're not data-centric right now. There's no data sharing, as such. Our current focus is application sharing. Basically our applications generate data that can be examined. We might give power users more disk space, but for general users we do not retain the smaller simulations. We just throw them away.</p> <p>The user sees the results of the simulation in their Web browser and they can choose to download the results. They can also mount their nanoHUB disk on their desktop and drag files in and out, so users can manage their own files if they pull them out of nanoHUB.</p> <p>You can conceive of an application that acts as a big dataset and crawls through the data, creating abstractions based on that. We're starting to talk about that; it's just not something we currently do from nanoHUB. But we're transferring the HUB idea into other disciplines; ideas like the importance of interactivity and service availability. And also the focus on serving end users, as opposed to computational scientists.</p> <p><i>[prompt asking if data management is an issue for people in the nanotechnology field]</i></p> <p>Generally not in electronics. We cover three sub-areas: electronics, mechanics, and biomedical devices. It's not like earthquake science or atmospheric modeling or things of that nature, where there are huge datasets that are unique or difficult to acquire so people spend time exploring them. That's not the case in our three sub-areas. I suspect there might be people looking at biomedical aspects of nano that involve huge amounts of data, but we're not serving that community.</p>	
<b>Q7.1 What resources do you use in your work today?</b>	<p>The vast majority of our 5700 simulation users are being served by a ~40-node cluster that serves simulations interactively, and then a smaller subset of users is using the GRAM service out of the Open Science Grid's cycle service.</p> <p>Then at the NCN itself we have research students and computational scientists who are developing the next generation tools utilizing the software development environment and local hardware. We have roughly 1,000 cores available for nano-type simulation at my home institution.</p> <p>Then we have a resource allocation on the TeraGrid of 250,000 SUs that are not yet used in production mode.</p>	
<b>Q7.2 How do you share work-related resources with others?</b>	<p>We're part of Open Science Grid, so we share our compute resources with the community.</p> <p>We are also prototyping capabilities where users install virtual machines on their local machines and those local machines act as agents for the nanoHUB. That is not in production. We're playing with that.</p>	
<b>Q7.3 By what mechanisms is access to your work-related resources controlled?</b>	<p>The access mechanism to these resources is through the nanoHUB. Their requests get translated into a nanoHUB community account in order to submit their simulation to the TeraGrid or Open Science Grid. We keep track of who that user is but on the service side from the service providers perspective they see nano1 or nano2 running a simulation. It's up to us to make sure that we document which user nano1 is running as at that particular point. So our users don't require their own allocation.</p>	
<b>Q7.4 How do you locate available resources for use in your work?</b>	<p>Well this is a difficult one.</p> <p>What we try to do is calibrate our estimate of the requirements for a particular application. The bad part is they can vary dramatically depending on the choice of inputs by the user.</p> <p>So on the Open Science Grid we try to estimate the memory requirements and runtime requirements a priori based on empirical data, and put that into the resource request.</p> <p>Requirements vary in terms of both memory and time.</p>	

Interview ID=24 24 September 2007	ANSWERS	ANNOTATIONS
<b>Q7.5 What types of information do you need to know about a resource in order to determine if it is suitable for your work?</b>	<p>Some tools benefit from a true parallel machine, so submitting to TeraGrid and requesting 50 nodes or 100 cores to enable a true parallel run. Having a way to reliably start MPI running will be a key issue. Condor doesn't do all that well for MPI runs.</p> <p>We typically use Condor-G systems for sequences of data. For example a simulation might vary the applied voltage on the device. An individual part would be meaningless – the engineer needs the sequence. But the calculations can be executed independently of each other so we can use the Condor-G engine or a single CPU-type engine.</p> <p>Large computational problems like my research code called NEMO use hundreds of CPU's and scale very well. In fact we showed that it scales to 8,000 CPU's. The submit logic is tuned for specific applications. The way Rappture works is you need to write a wrapper script that captures the needed applications as a simple workflow-type process. The scripting language can be Python or Tcl or Perl. It could be MATLAB. That wrapper would know about some of the resources that it can use: staging information, a particular segment requires a run on parallel resources, etc. So the script is always tool-specific so you can easily build in the tool-specific information. The tool developer writes that script.</p>	
<b>Q8.1 What software do you currently use in support of your work?</b>	<p>The project management and Wiki system are powered by Trac on our site <a href="http://developer.nanohub.org">developer.nanohub.org</a> or <a href="http://nanoforge.org">nanoforge.org</a>.</p> <p>We wrote our own middleware to handle interactive simulations. And we wrote our own system that renders interactive 3D graphics on a rendering farm without scheduling.</p> <p>We wrote our own tool development environment called Rappture.</p> <p>Those elements will be or are open source (Rappture is already open source. The rest will be open source.)</p> <p>The tools come in whatever flavor the original authors want. The one important thing is we don't request anybody to rewrite the software. If they choose to put in the Rappture interface they can. They can do that in their favorite language – Perl, Python, Fortran, C, C++ or MATLAB. Rappture can be integrated in all of these and we don't request or require anybody to rewrite that software. So the applications can be written in almost anything that I can think of these days. We don't have anything in Java actually but we have MATLAB, we have Python, we have Fortran... those kinds of things.</p> <p>Individual authors may publish their own source. They may grant users the write to download the source of it. We encourage people to publish their source but we leave it up to the developer to decide whether a project is open source or not.</p> <p>Each tool has its own tool page. It's sort of like a homepage for each tool on the nanoHUB and there would be links as to the ability to download a certain source.</p>	
<b>Learning about the user's problems</b>		

Interview ID=24 24 September 2007	ANSWERS	ANNOTATIONS
<b>Q9.1 What challenges do you face today in accomplishing your work-related goals?</b>	<p>What I would say is that the Grid as it exists today as a computational resource provider is at the maturity level of the telephone system 80 years ago. What I mean by that is if you wanted to place a phone call to somebody you would call the operator and say, "Tomorrow at noon I would like to place a long-distance call to so-and-so." And you pray that all the connections will work and you are able to make that phone call. The Grid is not yet a service that you can dial up, instantly connect to, and repeat again and again without hiccups. The Grid needs to work more like the telephone network. I just drove 99 miles and I'm almost certain I went through several service providers while I was talking to you on my cell phone, but I didn't have to think about it at all during our conversation. There are reliability issues with the Grid software that's out there. File systems fill up, certificates expire, and jobs fail. Maybe computational scientists are knowledgeable enough to put up with that, but not end users—not experimentalists. With nanoHUB, we cater to people that wouldn't touch computers with a ten-foot pole. They are certainly not experts at GSI-OpenSSH, key generation and other matters related to Grid certificates. They don't know the first thing about web services. We must have an architecture that handles all of that for them transparently. That is to say when a simulation fails it should be retried automatically, and if it fails again or can't be retried for some reason, it should be reported clearly to the user with something more than an obscure error code. We're quite far away from doing that with any sort of reliability. For computing to act as infrastructure, we have to take serious steps to not conduct it as a research experiment, but to do software engineering that lets us succeed in a production sense. A lot of middleware is developed as a research effort and papers are being written about it. I don't see many papers written about latex, for example, because it's actually infrastructure that works. So to me, building an infrastructure means creating and operating something that's useful for people even though there may be nothing novel to publish about the underpinnings. The funding agencies tend to tie the creation of infrastructure into research activities, but they need to fund it and evaluate it differently.</p>	
<b>Q9.4 By contrast, can you provide examples of technologies you find very useful today?</b>	<p>I like the model that OSG has where people bring resources to the table and like to share those resources. I like the BOINC [<a href="http://boinc.berkeley.edu/">http://boinc.berkeley.edu/</a>] model where people donate computer cycles. I think those are good models in terms of service provision. I think the compute centers that we have, the HPC compute centers either on the TeraGrid or even the ones that exist by themselves are serving an elite few. And though those people like to solve real science questions, I think the vast majority of people who could use the supercomputing are excluded. [prompt asking if there are other compelling technologies that are not necessarily distributed computing-related]</p> <p>We've bought into open source-type systems. So whenever we can find an open source solution that is close to what we want to do, if we have access to the source and can modify it to fit our needs, we will adopt it. So the open source model is great. We are using python, Tcl, perl languages that are shared, so that's a great model. We try to use standard languages, etc. So it needs to be open source and it needs to be openly available and it needs to be service oriented. The nanoHUB web site is based on LAMP (Linux+Apache+MySQL+PHP) combined with Joomla, which allows us to expand into new capabilities as needed.</p> <p>When some people talk about service oriented science, I think they're talking about Grid services. I don't think they're talking about things like:</p> <ul style="list-style-type: none"> <li>- who is running these Grid services</li> <li>- who gets paid for being of service to others</li> <li>- who actually puts themselves second to others in order to enable work.</li> </ul> <p>So I am hesitant to use the term "service oriented science". Certainly I think of it as turning my own research, conducted in the last 15 years, into something useful for others. And I'm not necessarily getting paid directly to do that but eventually that service is building up my reputation, which helps me to gain further funding down the line. So it's not that I'm doing this service to conduct research by myself on my own computational work. It's not about me wanting to advance my own research and life and improve research through that service, but literally being of use to others.</p>	
<b>Learning about the Globus user experience</b>		

Interview ID=24 24 September 2007	ANSWERS	ANNOTATIONS
<p><b>Q17 What are the major challenges you face using &lt;component&gt; today?</b></p>	<p>General:</p> <p>I took a Globus Toolkit 3 workshop while I was working at NASA JPL and very much interested in learning about this technology. That workshop opened my eyes that I will not clutter up my application with all the requirements that GLOBUS imposes on my application. I talked with a person that was teaching the course, asking, "Have you ever considered talking to end users of your framework?" The answer was, "No, we haven't done that." I think that was 2003. It was very eye opening.</p> <p>One person in that class actually said I should be rewriting my application in Java. I have a 200,000-line application written over many person-years. I'm not going to rewrite that in Java. Some might say, "That was a misunderstanding." I'm not sure. I mean I certainly asked twice and I think reflected a mindset of, "You users adjust to what we have, rather than we actually do something that you can use." That was a very eye opening experience for me.</p> <p><i>[prompt asking for suggestions on how to improve the situation]</i></p> <p>Don't deploy a toy application of some fishes in a bowl, but demonstrate that you can really host a real application that is actually driven by users requirements. Such requirements might be true interaction with simulation tools, not batch processing. Real interactive science. Then you'll experience what it takes to build a real application that serves not just one specialized user but a whole slew of different users. You'll find that these users don't have the ability or willingness to put in certificates left and right. They don't have the ability to rewrite front-end codes. Users are not as sophisticated as you think they might be.</p> <p>The real impact of computing is not going to be done by computer scientists or computational scientists, but by people who use these tools for solving real problems. I don't think – even though you might try very hard – that you can design middleware in a vacuum. You need a set of candidates who at any given stage of the design can actually operate on that system. The strength of nanoHUB is that we are always in production. We are always serving lots of users.. They tell us what works and what doesn't, and we do our best to listen.</p> <p><i>[prompt asking if NanoHUB is changing that]</i></p> <p>Well it's certainly changing the user experience and I think what we're starting to see now is that we're growing quite a bit and maybe we might be starting to hit the problems that computer scientists might have anticipated to happen in terms of allocation of resources and opportunistic runs, etc. I think the requirements we can provide to computer scientists solving problems would be very insightful.</p> <p>The ideal interaction would be collaborating with a computer scientist who wants to apply the knowledge gained over the last sixteen years or twenty years (or however long they've been in this field) to make themselves useful to others. In my profession I don't know middleware or HUBs. I'm a nanoscientist and nanoelectronics engineer. I want to make my stuff useful to others. I'm looking for computer scientists who don't ask me, "How can I publish the next paper?" Instead i prefer, "I understand the resource allocation problem, I have developed prototypes, I would like to try them to solve your problems. I have solutions for you that will work today and don't require three years of research." That's the computer scientist I'm looking for, and if they develop research questions out of this whole thing that's great. I don't want to say we don't need research in general. What I'm saying is we need little "r", capital "D." Little research and major Development to make this actually useful.</p>	
<b>Wrapping-up</b>		
<p><b>Is there anything you'd like to say to the people who build software for use by people like you?</b></p>	<p>There is a distinction between tool developers and researchers: Given an arbitrary deadline: are you going to finish the paper for the conference that's due on Friday or are you going to make your system work reliably by Friday? I wish more people would answer, "I'm going to make my system work reliably" without publishing another incremental article. That would be my request to people who want to build our infrastructure.</p>	

## D.25 The diversity of the systems we run on is a problem

Interview ID=25 25 September 2007	ANSWERS	ANNOTATIONS
<i>Pre-interview question: Do you interact directly with Globus software in your work today?</i>	Not directly. I use it through Condor-G.	
<b>Establishing context</b>		
<b>Q1.1 Please provide a one-minute overview of your project</b>	NanoHUB is a center for simulation, among other things, and several of these simulations require resources outside the bounds of what we can provide locally. And so we use TeraGrid and OSG Grid resources to accomplish simulation runs or solutions. The bridge between these Grid resources and us is Condor-G, and hence from there, Globus.	
<b>Q1.2 What is the project's name?</b>	NanoHUB	
<b>Q1.3 Which agency funds the project?</b>	The National Science Foundation	
<b>Q1.4 What field does your project belong to?</b>	The electronic aspect of nanotechnology, with the intention of going into biology and other fields.	
<b>Q1.5 What is your job type?</b>	Application Engineer for Scientific Computing In a practical sense, I serve as a bridge between the local resource and the Grid resources; I'm responsible for implementation of applications that require the Grid resources.	
<b>Learning about discipline-specific goals and approach</b>		
<b>Q2.1 What are the main goals of your project?</b>	With respect to simulation, the idea is to put the simulations in the hands of the people who need them and who wouldn't otherwise have access to them. And we feel that simulation itself generally speaking is a very powerful tool to be used by people in the research or industry (or even undergraduate studies or whatever.) It's fundamentally useful across the board. But not everybody has access to everything, so we're trying to fill that niche as best we can. Our users include over 5,000 people using one simulation or another, and they cover the spectrum from undergraduate, graduate, post-graduate, industrial users. Exactly what they're trying to accomplish with these tools is hard to tell.	
<b>Q2.2 How will the success of your project be measured?</b>	That's a tough question. Or it's an easy question, but a tough answer. We measure the nuts and bolts of everything, but that's probably not the true measure. We can report how many times every simulation was run and how long it took, but the real success would be that people would use it and further their own understanding and have their own goals in terms of their own environment. If they're trying to design a new product for a company, or trying to write their thesis, it could be a lot of things. But for us to get direct measures is difficult. We're also starting to measure is publications, and that's another way to quantify success. We count people who write peer reviewed papers and cite NanoHUB as a resource.	
<b>Q2.3 What are the professional measures of success for you?</b>	Well, my task is to not get in the way of people making their tools available for use. So if I'm successful delivering the tools, then that's the goal.	

Interview ID=25 25 September 2007	ANSWERS	ANNOTATIONS
<b>Q3 What are you investigating?</b>	<p>I'm not the application developer, but what the application developer has trouble with is making his tool viable in a Grid environment. So one could say that's my function, is to make the tool viable in a grid environment. So what we're doing is providing an automated, I guess you'd call it a tool, to allow the user to submit their application to the Grid environment. So they don't have to know about Condor commands or Globus commands or any other middleware that you might use. So I develop an application framework and I integrate these new tools as they come in. New tools are always coming in, and the underlying framework that we use is subject to change at any time. So there's sort of a combination of keeping the tools that are already running, running, and also bringing on board new ones (and dealing with whatever new challenges that might bring.)</p>	
<b>Q4.1 What is your method for investigating &lt;phenomena&gt;?</b>	<p>Well I guess the high level overview is that we also work in conjunction with the Condor team in Wisconsin, and so it was decided early on that that would be the major delivery vehicle between NanoHUB and Grid resources. Starting with that, we're putting a wrapper layer around that to hide it from the user in some sense, so he can deliver his application to the Grid as though it was running locally on his computer (using a similar kind of command syntax, context, and so forth.)</p> <p>All the user tools basically have an executable, some input files and some output files, and maybe they have an environment variable or two they need to set. That's about all there is to it. So what you need is to deliver the executable and the input files to where they need to run, and a way to retrieve the results and output. So we can put a pretty generic wrapper around Condor type delivery systems to accomplish that.</p>	



Interview ID=25 25 September 2007	ANSWERS	ANNOTATIONS
<b>Q4.2 How do you work?</b>	<p>Developers will seek us out and we also seek out developers in some cases (if we're aware of the work that they're doing.) In order to integrate the tool the developer would go to the NanoHUB website to initiate the application contribution process. He would fill out a very simple form, giving us a name for the tool and a paragraph description of what it does. If we decide it's a valid tool for NanoHUB, we take that information and create a subversion repository for their source code, tell them where that is so they can access it and assign privileges to that particular user so they can access it. The code is not regarded as open source unless the user wants it that way.</p> <p>The user then puts his code in the repository, and through a Unix environment is able to develop the tool. When he thinks it's ready to deploy to the rest of NanoHUB he'll get in contact with us.</p> <p>We then do a minimal evaluation of the code, making sure that it runs and that it is in some sense a scientific type of code (it's not generating random text or pornography or whatever.) But we don't put it through a grueling test.</p> <p>We then post it on the NanoHUB for others to use. Once it's posted for use, the tool is subject to user ratings. Everybody can comment on any tool. And it is hoped that the worthy tools will become known through this mechanism. People will use them and they'll stick around, and the ones that are not so good will fall by the wayside in time. We actually haven't had to do this yet, but at some point we may actually have to remove a tool, and we can do that too.</p> <p>They are simulation tools. Some of them may run for ten seconds and get a very fast response for a relatively simple problem. On the other end of the scale you may have codes that are parallel, requiring fifty to one hundred nodes that run for six to twelve hours. That's the kind of case where we'd deploy it to a Grid type environment.</p> <p>The other thing that NanoHUB provides is a mechanism to put a user interface around the application for easy input of primers and graphical display of results.</p> <p>By and large, the tools are whole and complete themselves, but not necessarily. People can also contribute libraries, as opposed to complete applications. We probably have just one of those at the moment, but it is possible that people could share sort of numerical solution techniques or other blocks of the simulation rather than having to write the whole thing themselves.</p> <p>The tool developers are quite sophisticated in terms of both programming and science. But not the end-users. The end-users are less strong in the area of coding or computing, but are very strong in their science fields. They're the ones who use the GUIs the tools are wrapped in.</p> <p>So each project may have one developer, two developers, three, four, five developers, whatever, and they can work as a group or a community. But when the tool is made available on NanoHUB, you may find that you have a hundred people who want to use that simulation. And that's the sort of leverage that we're trying to gain.</p> <p>So NanoHUB serves as not only a repository, but also sort of a discovery mechanism for the field.</p>	
<b>Q5.1 In what ways do you interact with simulations in your work?</b>	<p>As part of posting them for use I need to test them all first, so in that sense I see them all. My background is not in nanotechnology or electronics so I would not claim to understand them all. But I do see them all and run them. Part of the validation process is verifying that they do execute. It's often obvious when they fail to execute. One common thing I need to watch out for is the developer will write a tool that only runs in a particular directory. NanoHUB is a multi-user environment; everyone can't run in one directory. One of my jobs is to make it run anywhere, as opposed to one particular location.</p> <p>So I sometimes need to modify either the interface or the wrapper around the simulation to generalize it in terms of where it can execute. Oftentimes, that's what happens. You'll try to run it and it'll say, "Oh, I can't find this file" because it's in the wrong directory, or it doesn't know the path to something. So there's a little bit of work in cleaning that up before it can "go public" for general consumption.</p>	

Interview ID=25 25 September 2007	ANSWERS	ANNOTATIONS
<b>Q5.3 How do you interact with inputs to your simulations?</b>	<p>On the input side, generally once the tool developer has put one of these graphical interfaces around the tool, every input will have a default value. And so in that sense, I don't usually have to interact with the inputs. A reasonable set of default values must be provided; that's part of the tool verification testing. The default case should always run.</p> <p>The same tool may also run in different modes as a result of different input parameters. In this case I should try testing the different modes with different inputs by making sure that they do run.</p> <p>So the goal is that the framework helps simplify dealing with inputs to the simulation.</p>	
<b>Q5.4 How do you interact with the output of your simulations?</b>	<p>The same can pretty much happen on the output side. As long as you're producing the typical XY graph or 2D plots or even 3D volume data, the standard interface package has ways to render that. And it's up to the application developer to build that interface as well. So by the time it gets to me, it's generally already been done.</p> <p>Now if there's something that I feel they could do differently or better, we make that suggestion. Sometimes we change our code, and sometimes the developer makes the revision in their code. And we go forward from there.</p>	
<b>Q5.5 By what mechanisms is access to your simulations controlled?</b>	<p>Not every tool is available to everybody. The first thing they have to do is log in. That's a very simple process – it's just a log in and password. And creating an account is free and very simple as well. They just need to provide us with some demographic information, which we lock in a vault. We count the number of login accounts for NSF purposes.</p> <p>Beyond that some tools can't be run outside of the United States. This could be controlled through IP addresses and the like. Some people would restrict use of a tool to a group of individuals, and we can control this through a licensing mechanism as well.</p> <p>So we can control access at various granularities. So far it's worked out well.</p>	
<b>Q7.1 What resources do you use in your work today?</b>	<p>There are several dozen resources locally that NanoHUB operates on. There are file servers, database servers, and backend compute cycle machines. But it's basically self-contained. In my role I access only three or four of those.</p> <p>Then of course, beyond that, we have access to eight or nine TeraGrid sites. And we can access about a dozen OSG sites – maybe twenty. That provides the potential for a large number of machines and cycles.</p> <p>There are no external data sources feeding into NanoHUB. There may be people generating data in lab experiments, and we would like to enable them to use the data in conjunction with the simulation tools. But they would have their data locally so it wouldn't necessarily be integrated with NanoHUB. Laboratory scientists have access to the simulations through NanoHUB, but unless the data is posted for everybody to see, NanoHUB wouldn't have it.</p>	
<b>Q7.3 By what mechanisms is access to your work-related resources controlled?</b>	<p>The user doesn't directly log in to any compute cycle machine. They log in to NanoHUB and if they want to run the simulation, simulation interfaces deliver it through their web browser, and they interact just with the graphical interface there.</p> <p>So within the graphical interface there's a button that generally says, "Simulate," which they will push. Then through NanoHUB middleware it'll take the input parameters specified and execute the tool he's running on a backend compute machine. This is all transparent to the user. When it's finished the results are delivered back again through the same interface.</p> <p>So far all users are equally privileged, in terms of running on OSG or TeraGrid. That's one of those things that could change at some point. At this time, a user may select a simulation tool that runs on OSG and they don't even know it. The presentation that they see through the interface is the same, whether it's running on OSG or a local backend machine.</p> <p>So far we are not starved for compute resources, but our growth rate is pretty high. I guess if we really do a good job, we might become starved. Right now we're at like 5, 6, or 7,000 users. I guess people talk about 100,000 users, but that's a way off.</p>	

Interview ID=25 25 September 2007	ANSWERS	ANNOTATIONS
<b>Q7.4 How do you locate available resources for use in your work?</b>	<p>We don't have a structured approach for that. Basically, each tool or application will know that it should run locally, or within a Grid environment. That's assigned at integration time, mainly on the scale of its requirement.</p> <p>So a small tool that doesn't run very long (using a single CPU) will just run locally, with the particular backend machine chosen at random from whatever's available.</p> <p>For a tool that needs to run on the Grid, we currently also pick a site randomly from those that we believe are currently operating. We send them either to OSG or TeraGrid, depending on whether it's a parallel code or a sequential code. So we match the remote resources with the application itself.</p> <p>Until recently, it was not really possible to run a parallel job on OSG. On TeraGrid you can run either one. A lot of the TeraGrid resources are commodity clusters or specialized cluster machines, providing access to a few thousand nodes for running MPI jobs. We've pretty much split it so the parallel ones'll go to some TeraGrid site, and the sequential to an OSG site.</p>	
<b>Q7.5 What types of information do you need to know about a resource in order to determine if it is suitable for your work?</b>	<p>There are three classes of information now. We have to match the operating system and hardware architecture with the application. So to run something on say, the Pittsburgh Cray machine you have to compile it on their machine. You can't take something developed on my Dell computer and send it over there. So that's one thing.</p> <p>The other two major things are space requirement and memory requirement of the application. This is especially true of the TeraGrid sites. There's a range of memory available. Most of the OSG sites are pretty uniform, so it will either fit on any of them or none of them, for the most part. They're also of the similar architecture, so we can build something here and distribute it across OSG, running it pretty much on any of their sites without having to do a specialized build of the application.</p>	
<b>Q8.2 What scripting languages have you used in the past year?</b>	We use Tcl, Python, and bash shell.	
<b>Q8.3 What programming languages have you used in the past year?</b>	We have tools in C, and Fortran, Fortran90, and some of them are actually written in scripting languages. I think there's one that's written in Perl, so it's a pretty wide variety.	
<b>Q8.4 What workflow tools do you use in your work?</b>	I don't really use any. Maybe I should be. Maybe somebody needs to create one that would fill the bill.	
<b>Q8.5 What parallel computing tools do you use in your work?</b>	As far as I recall, every parallel tool we have actually uses MPI. We would also support open MP if a developer were to use it. And then of course, included in that or in the associated queuing mechanisms of PBS and so forth.	
<b>Q8.6 If the need for new software-based functionality arises in your work how do you acquire it?</b>	That's a good question. I guess the operating process here is if something is available, it's cheap (as in free) and it does a good job, we would use that. Otherwise, we end up writing it ourselves.	
<b>Q8.7 How do you share software with others?</b>	In the future we're planning to share the NanoHUB framework code with others. There are different parts of it that may be available already. The application interface development system, called Rapture, is already available. People can download it from the website and use it themselves for NanoHUB related work or other work.	
<b>Learning about the user's problems</b>		

Interview ID=25 25 September 2007	ANSWERS	ANNOTATIONS
<b>Q9.1 What challenges do you face today in accomplishing your work-related goals?</b>	<p>I guess the major challenge is still in the Grid aspect of this. It's still difficult to have a job run the first time you submit it anywhere and that shouldn't be the case. So our typical issue is in site selection. As I mentioned earlier, we pick a site randomly from a list that of sites we believe are operating. We believe the site is operational because it was operating an hour or two ago. But it may have stopped working in the interim – stopped accepting NanoHUB jobs. That throws up a barrier that we have to get around.</p> <p><i>[prompt asking how it was determined to be operating an hour ago]</i></p> <p>Currently we probe the sites with a very simple job to make sure that we're still authenticating properly and that the file will transfer back and forth. So it's a quick test, and we run it every four hours or so. We could run it more often, but then we'd be doing all probe jobs and no real work jobs.</p> <p>To do site selection we use Condor matchmaking. And the results of our probe test is fed into the ClassAd for the site, so the matchmaker will only pick a site that was running the last time it was probed.</p> <p>So the problem is that our rate of submission failure is higher than we'd like to see under high loads. In the past we had a situation where our load was considerably higher than it is today, and the failure rate was higher than what was acceptable.</p> <p>Example failures can be seen here:  <a href="https://twiki.grid.iu.edu/twiki/bin/view/VO/NANOHUB_UtilStatus">https://twiki.grid.iu.edu/twiki/bin/view/VO/NANOHUB_UtilStatus</a></p> <p>We don't have any codified standards for tolerance of failure. Of course everybody wants zero failure, but that's not realistic. So what we've done is, if a site is selected and the job fails for whatever reason, the job will then be resubmitted at a different site. And if that job fails, it will be submitted at yet a third site if available. If there is a third failure we come back to the user and tell them the job failed. In a sense, we've said, "If it fails three times in a row, that's not acceptable."</p> <p>Then what's the user supposed to do? He's just going to hit simulate again and the same thing will likely happen.</p> <p><i>[prompt asking if any diagnostic information is provided]</i></p> <p>The user sees a little bit of diagnostic information. Not a lot. Condor and Globus log everything, so there's always a log file that has some kind of error report in it. But based on our experience it doesn't tell the user how to fix the problem. Even if it did, he still wouldn't be able to fix things, because he's the user of the tool, not the developer.</p> <p>So it isn't particularly useful from our standpoint to put "Globus error 43" in front of the user. He will look at that and say, "I have no idea what that is." So in that sense not a whole lot of information is given back to the user beyond an indication that something failed.</p> <p>We do try to deduce the problem from the error report to a level like "the transfer of input files failed." And then maybe suggest that perhaps their file doesn't exist. If we can tell them that much, we will. But by and large the user doesn't see much error feedback.</p> <p><i>[answer continued on next page]</i></p>	
<b>Q9.1 What challenges do you face today in accomplishing your work-related goals?</b> <i>[continued]</i>	<p><i>[prompt asking if there's a helpdesk or some other support mechanism for the user]</i></p> <p>There's a help or support button on virtually every page, which the user can select and give us a short description of their problem. In the case of jobs that completely fail, we trap and record the event in a ticket so we know which jobs are failing and how often.</p> <p>We've not been completely successful yet in capturing the information that would tell us why it failed. But we're working on that as well. So this would happen and the user wouldn't be aware of it. But a ticket would be created in the system and it would have the user's name on it so we could get back to them if we find, "Oh gee, this person is trying really hard to get this tool to run, and it's not." We may discover a problem with the tool that we can fix and get back to the user saying, "Please try again. We noticed you're having a lot of trouble and we think we've resolved the issue. Come back and please try it again."</p>	
<b>Q9.2 What types of information do you need in order to address the challenges you face today?</b>	<p>Usually we need to get back to the intermediate files that were left by the application run. Each stage of the grid process has a log file: for the file transfer there's a log file, for the execution section there's a log file. These are created when the job is run under the user's home directory system. We of course have access to that, so we can go back to where that particular job was run and dig into it a little bit.</p>	

Interview ID=25 25 September 2007	ANSWERS	ANNOTATIONS
<b>Q9.3 What technology-related obstacles do you currently encounter?</b>	There is a problem related to the diversity of systems that we run on, in particular on TeraGrid. We can build a tool on one site, and it'll only run on that site. It doesn't run anywhere else. And this is especially true of the parallel tools. Diversity is kind of a double-edged sword. You may find that an application runs really well on a particular type of architecture and not so well on others. In one sense diversity is a good thing, but the flip side is you have to be able to develop for all. So for us that means logging into all of them, re-porting and building code, and maintaining the application across all those platforms. That's one of those barriers that will stop the casual user.	
<b>Q9.4 By contrast, can you provide examples of technologies you find very useful today?</b>	Well probably one of the tools I use around here more than anything else is Google. And I'm sure that's true everywhere, because if you don't know the answer, you often find somebody else does and they've put it out there. So that's extremely useful.	
<b>Q10.1 Can you think of any work-related tasks that decrease your productivity?</b>	Those repetitive and tedious ones are those meetings that keep getting in the way. There are only so many hours in the day.	
<b>Learning about the Globus user experience</b>		
<b>Q11.1 Which Globus data components do you directly interact with in your work today?</b>	None, directly. We use the Condor Stork product, which uses some or all of those for file transfer. I think that our main mechanism is GridFTP probably or GSIFTP, which I guess is a GridFTP implementation	
<b>Q12.1 Which Globus security components do you directly interact with in your work today?</b>	We're using gridProxy and vomsProxy. Also the suite of tools comes with that: GSI-OpenSSH, scp, etc. NanoHUB uses a community account for user access.	
<b>Q12.2 Did you install the &lt;component&gt; client yourself?</b>	no	
<b>Q12.3 Did you install the &lt;component&gt; server yourself?</b>	no	
<b>Q12.4 How many people currently use your &lt;component&gt; server</b>	That's a good question. Probably less than twenty, I'll say. And that would be either people here or the actual application users who use it without their knowledge.	
<b>Q13.1 Which Globus execution components do you directly interact with in your work today?</b>	Condor-G The probe probably uses a globusrun tool, for the very low-level kind of thing	
<b>Q14.1 Which Globus information components do you directly interact with in your work today?</b>	No, we don't. And we probably should if we could figure out how to do it. I've not personally tried, but it seems like that should be feeding into the site selection process.	
<b>Q16 Why do you use &lt;component&gt; instead of an alternative technology?</b>	GSI: Because it works; it seems to work. What would be the suggested alternative? I don't know of any. We started out using it because it was necessary. It was the way to do it two or three years ago. And we haven't changed. Condor-G: That stems from the collaborative partnership with the MNI initiative between NanoHUB folks here and the UW Condor people in Wisconsin.  Globusrun: The probing script was taken from OSG, it was something that they wrote up, and we just kind of borrowed it and started using it for our own purposes. My feeling is that it had the least overhead to accomplish the goal, which is just to get this quick job there and back. We use the same probe for both TeraGrid and OSG sites.	

Interview ID=25 25 September 2007	ANSWERS	ANNOTATIONS
<p><b>Q17 What are the major challenges you face using &lt;component&gt; today?</b></p>	<p>GSI: It basically functions when I try to use it – most of the time, not all the time. I had a case (actually just a couple days ago) trying to connect to Pittsburgh Center with it, and it would refuse five times in a row and then the sixth time it would be ok, and don't really know why. This was running GSI-OpenSSH. I never got an explanation as to why. And it's not the first time this has happened. The answer normally, "Is wait a few minutes and try again." I don't really have any choice; it's not working. I either have to go somewhere else or try again. And if that's the site you need to get to, well, you wait and try again.</p> <p>Condor-G: I guess it's still a little bit of a black box to us, and that means we have to ask more questions. The other problem is when it doesn't work, we don't know whether it's the Condor layer, the Globus layer or some other layer that's failing. It adds layers of complexity (maybe a little too strong a word) on top of the process you're trying to accomplish. Sometimes it works transparently. Other times it fails and you don't know why, and you're kind of left holding the bag, so to speak.</p> <p>Globusrun: The use of the test probe is now is becoming quite automated, so that's good. The challenge might be in presenting the result and using the result, actually. The test itself is pretty straightforward. We use the result in our site selection process to some degree. And we also can put up a web page that has the status as of the last probe with red and green indicators. But the challenge is not actually doing the test; the challenge is making the test successful, which involves interactions with the various resource providers to resolve why the test might've failed.</p> <p>Also: we can probe all these sites every three, four hours and stick the results in a file or a database somewhere, but unless you use it for something, what's the point? So another challenge is interpreting the results. Because there should be a difference between an intermediate failure that happens in one out of ten probes, versus something that failed four times in a row. There are probably different issues at work there. And that needs to be brought to somebody's attention in a clear way. So right now the interpretation logic is just pass/fail based on the most recent test. We have a history of these tests going back for perhaps months, but we only pay attention to the most recent one. This approach will not catch a catastrophic error that takes out OSG, for instance. Something like this nearly happened last week because our certificate was about to expire. Fortunately this was noticed about 20 minutes before it expired. We were able to sneak one in there and didn't really lose much. But if somebody hadn't remembered and it hadn't been taken care of, we would have lost all sites, basically, because without the certificate you can't get to anybody.</p> <p>And then the question is, well how long would it take us to notice that? That's a good question because if it does happen, there won't be any sirens that go off to tell us that.</p> <p>If you were looking at the queue you would say, "Oh, all these jobs are on hold because authentication failed." You might notice, but you have to go look at the queue. If you look at the webpage you might say, "Oh, everything is red." But again, you have to go look at the webpage. We don't have a mechanism that takes a proactive approach and tells us, "Hey everything has failed here." We obviously need to do something to fix this.</p>	

## D.26 Our goal is to make it easier to troubleshoot Grid applications

Interview ID=26 25 September 2007	ANSWERS	ANNOTATIONS
<i>Pre-interview question: Do you interact directly with Globus software in your work today?</i>	yes	
<b>Establishing context</b>		
<b>Q1.1 Please provide a one-minute overview of your project</b>	I'm working on a troubleshooting effort within the CEDPS SciDAC project, the Center for Enabling Distributed Petascale Science. The goal is to make it easier to troubleshoot Grid middleware and Grid applications, where troubleshooting doesn't just include failures but also includes performance-related issues. In particular, we're focusing our effort on helping the Open Science Grid at the moment. So they're our first customers.	
<b>Q1.2 What is the project's name?</b>	Center for Enabling Distributed Petascale Science	
<b>Q1.3 Which agency funds the project?</b>	The Department of Energy Office of Science	
<b>Q1.4 What field does your project belong to?</b>	Computer science	
<b>Q1.5 What is your job type?</b>	Project Lead of the troubleshooting effort	
<b>Q1.6 How long have you been a &lt;job type&gt;?</b>	One year on this project, ten years in this general area	
<b>Learning about discipline-specific goals and approach</b>		
<b>Q2.1 What are the main goals of your project?</b>	So I've heard from both OSG and from the LHC Grid and from a number of other Grid projects. They all give roughly the same answer: somewhere around 25 percent of their remote job submissions fail. This is a shockingly high number. In general they don't know why the jobs fail – they can only guess why. The top reasons cited for failure are basic authentication problems. You know – the user might not be in the right gridmap file. There are also disk-related issues such as running out of disk space during the act of staging in some input file, or they don't have the right permissions, etc. But then there are a whole lot of other failures that fall into the unknown category. The nature of Grids makes it quite difficult to figure out the source of failures, and much of the underlying middleware lacks the right hooks to make it easy. So the goal of this research project is to figure out what is missing and try to get it added. The Globus team is partnering with us on this project. So Globus software is our first target for some of these new logging techniques and standards that should (hopefully) make it easier to do troubleshooting.	
<b>Q2.2 How will the success of your project be measured?</b>	We're wrestling with that question ourselves; we're writing our first-year report at the moment. The hope is for OSG to report a noticeable drop in number of failed jobs. Also to report a decrease in the amount of time it takes to track down problems. This is a difficult thing to measure. It's fairly abstract. My current approach to measuring this is by talking to people. I know that OSG has some of their own metrics they're worried about. OSG is, of course, being asked a similar question: how successful is OSG as an infrastructure? I know they're working on some mechanisms for tracking metrics, but I don't know the details.	

Interview ID=26 25 September 2007	ANSWERS	ANNOTATIONS
<b>Q2.3 What are the professional measures of success for you?</b>	<p>Well as a researcher, getting papers published and things like that are always important.</p> <p>But I also have a personal interest in trying to make this stuff work as a way of improving usability. I hear a lot of hallway conversation about grids still being hard to use. One of the reasons they're hard to use is that they're hard to debug. So if we can make an impact there, that's what I'm most interested in.</p> <p>[prompt asking if debugging on a Grid is different than debugging a client-server application across two machines]</p> <p>Yeah, I think debugging is often considerably more difficult on the Grid. Even the term debugging can be vague.</p> <p>For many years I have also been interested in performance issues, not just debugging. People think of debugging as a response to program crashes: "Why did it crash?" Well to me debugging includes cases where the program is running slower today than it ran yesterday. There are a lot of issues like that. As another example, the program may run fine on my particular client and server but doesn't run well on somebody else's client and server.</p> <p>Also when debugging in the Grid context, you may not have accounts on all the machines. The developer likely doesn't have accounts on all the machines that the software is deployed on. So mechanisms are needed to collect the log files and get them back to the developers. It's definitely a more challenging problem on the Grid.</p>	
<b>Q3 What are you investigating?</b>	<p>We are trying to get many pieces of Grid middleware to use a common logging format and to log the right stuff. In support of that we've put together a document called the <i>Grid Logging Best Practices Guide</i>, which defines a log format and includes advice on what should be logged.</p> <p>Then we've been working closely with a number of the Globus developers to add this new style logging to Globus; it should be included in the next release. So step one is trying to get the right stuff logged.</p> <p>Step two is creating a logging collection mechanism so they're in one central location (or multiple central locations). In support of this we are using an open source tool called <i>syslog-ng</i>, and have been working with the Open Science Grid to figure out how best to configure and deploy a central log collection facility. We are currently testing to identify scalability issues, reliability issues... all of those things.</p> <p>Once logs are written in a standard format and are beginning to collect in a central location, we will focus on a third step. We're really just starting this work now. The third step is actually trying to analyze these logs: figuring out what can be correlated with what, how one might automatically find anomalies and failures, and other things like that.</p> <p>But we can't do anything until we can get our hands on some logs. So the focus for year one of the project is getting logs into a central location.</p>	



Interview ID=26 25 September 2007	ANSWERS	ANNOTATIONS
<p><b>Q4.1 What is your method for investigating &lt;phenomena&gt;?</b></p>	<p>Our method for defining best logging practices has been to put together a document based on years of our own experience plus talking to a lot of people out there. We went to OSG meetings and EGEE meetings and talked to many people to get feedback. We presented the document at OGF, and people seemed to think it's a good idea.</p> <p>The log format is basically a simple ASCII name-value pair that is compatible with <i>syslog</i>. And everything that might fail is wrapped with a start and end event, which is very, very helpful for troubleshooting. Typically when programmers write their programs, in the debug code they'll log either the start or the end event. It's actually rare for programmers to log both the start and the end. So we're trying to get people doing it that way and show the utility of that.</p> <p>So within the last month almost all the Globus components are adopting this new style of logging. Hopefully once it is pushed out in the next Globus release and people start getting some experience with this type of logging they'll say, "Wow, this is really helpful. This is really useful." And the idea will catch on.</p> <p>We figure we won't get it right the first time – there will be stuff that's missing. It will be a somewhat iterative process to get the exact right logging information in there. To help with that we're putting up an OSG node ourselves here at LBL as part of what OSG calls their integration test bed. We hope to deploy a version of Globus with the new logging style on our OSG site and get some experience with the log files before the official Globus release (whenever that's gonna be – I guess the last date I heard was maybe March.) Hopefully we'll have some of that iteration cycle before the release actually happens.</p> <p><i>[prompt asking for more information about how name-value pairs are defined]</i></p> <p>The best practices guide includes naming recommendations. For instance every log line should have a unique event name. The advice is to use the Java naming conventions, so for a Globus MDS log, an event name would be <code>org.globus.mds.&lt;something&gt;</code>. We're trying to make it as easy as possible for people to think about converting existing logs. There are only two required fields in the new logs: a timestamp and an event name. And we strongly recommend people try to come up with unique event names; and we give some suggestions on how to do that.</p> <p>We found from years of trying to get people to improve their logging that programmers don't like to be told what to do too much. So we're trying not to be too prescriptive. And there's nothing as formal as an official ontology – at this point, it's fairly loose. It's partly a psychological argument, I guess. We just found from experience that if you throw something really complicated at people, they resist.</p> <p>So we wanted to make the initial step as painless as possible. If it turns out later that it's too freeform and we really need a stricter ontology, hopefully by then people will have bought into the concept of unified logging in general. Then they perhaps won't complain quite so much about adopting stricter standards. But I'm not yet convinced it's even necessary.</p> <p>So if it's a Java program and they're using <i>log4j</i> already, it should be very familiar to them. The format is very similar to a <i>log4j</i> log.</p> <p><i>[answer continued on next page]</i></p>	

Interview ID=26 25 September 2007	ANSWERS	ANNOTATIONS
<p><b>Q4.1 What is your method for investigating &lt;phenomena&gt;?</b> [continued]</p>	<p>[prompt asking if there is a tie between the log file and a the central logging facility]</p> <p>Not from the programmer's point of view. The connection can be seen from the system administrator's point of view.</p> <p>Then we have a recommended configuration of the <i>syslog-ng</i> tool that can basically suck up log files from arbitrary places on disk and forward them upstream. Then the forwarders can forward to additional forwarders. The idea is you could have a central site repository, which in turn can forward some or all of the data to the Grid operation center repository. But the programmer doesn't have to think about any of that. He/she just needs to write a log file.</p> <p>Regarding the method for analysis phase of log files:</p> <p>If people wrap everything with a start and an end event one can look for missing end events. For a given operation, you might know that it normally takes one second and worst case it takes 20 seconds. If 60 seconds have passed and you still haven't gotten the end event, you can generate an error event. We have a simple tool that does that.</p> <p>We're also starting to play around with some more sophisticated performance anomaly detection techniques, such as tracking the running averages of GridFTP performance. We just presented a paper talking about some of those ideas at the Grid2007 conference last week.</p> <p>The idea is to store all this stuff in a database so you can keep track of baseline performance. If your current performance deviates too much from the baseline, you can use things like the MDS trigger service to notice that and generate some sort of alarm. It's still at the conceptual phase right now. The devil is in the details in terms of baselines. You can say this program should always take <i>X</i> amount of time, or this program on this architecture should always take <i>Y</i> amount of time, or this network connection... just how specific you need to be for a baseline to be valid is definitely an open question.</p> <p>One thing I forgot to mention is that for a given execution thread of a given program, the logging information must include a unique identifier that allows you to tie together the series of events. So for a given GridFTP data transfer, for example, the log lines representing the user authentication and the file transfer itself would both be tagged with the same unique identifier. It could be a process ID, it could be the file name – just so that it's something unique that ties the series of events together so you know that it's part of one operation.</p> <p>[prompt asking if requirements were gathered from middleware developers, application developers, sysadmins, etc.]</p> <p>Requirements were gathered mostly from sysadmins, I think, because they're the ones who are often are in charge of troubleshooting. We certainly involved some applications developers, some middleware developers... a little bit of everything. But probably the majority were sysadmin types.</p> <p>We're still figuring out exactly what our database schemas are and exactly how to put all this data in a relational database in a useful way. In parallel with this activity we started doing some of the analysis stuff, in terms of the performance of GridFTP. But we haven't thought real hard about how to generalize that yet. That's what we're going to be doing over the next years.</p> <p>[answer continued on next page]</p>	

Interview ID=26 25 September 2007	ANSWERS	ANNOTATIONS
<b>Q4.1 What is your method for investigating &lt;phenomena&gt;?</b> [continued]	<p>I don't think we necessarily yet understand all the failures. We're trying to make a system that will both be useful to Grid administrators and Grid users and Grid middleware developers. It's challenging, because they all come from a different perspective.</p> <p>We're certainly targeting just the ability for a user to know the reason behind the error. An error message that says, "Grid authentication failed" doesn't always make it back to the user. So part of our work is identifying easier ways to provide feedback to the user.</p> <p>[prompt asking if "the user" referred to above would be the initiator of the client, or the sysadmin]</p> <p>I guess that depends on whether they're</p> <ul style="list-style-type: none"> <li>- at a command line prompt typing <i>globus-job-run</i></li> <li>- running something through a portal</li> </ul> <p>There are so many different ways to interact with the Grid right now that abstracts stuff further and further away from the user. The answer is all of the above.</p> <p>[prompt asking how log line items that would be produced will differ as compared to today's typical GRAM logs]</p> <p>Actually, there has been some discussion on that topic recently with the Globus folks. I'm not a 100 percent sure where that discussion ended up – whether it will be one log file or two. I think it will be two because there's a low-level debug mode for WS-GRAM that produces way more logging than one would want to shove through this central log collection facility. So at least the detailed, low-level debug logs should stay separate.</p> <p>We recommend that programmers target anything that might fail. Typically, it's any sort of network callout of some sort. If you're calling out to a VOMS system, or querying a database system, or reading/writing from the disk or network: these are all things that potentially fail. You definitely want to generate a log message indicating the failure. Much of that stuff is already logged, but in a very ad hoc way. We're just trying to formalize how that happens.</p> <p>Some of the Globus MDS Trigger service folks are part of this project, and the vision is to utilize the Trigger service as much as possible for alerting users when failures happen. But it's not clear how easy that will be. You have to write the information provider to get that information into MDS so the Trigger service to do something with it.</p>	
<b>Q4.3 How do you keep track of interim results, if at all?</b>	<p>We have a Wiki that's very... very full of stuff ☺. Perhaps too full.</p> <p>The DOE SciDAC projects have strong guidelines on reporting requirements. We have quarterly reports, and are just in the process of finishing our annual report.</p> <p>We also use the Globus Bugzilla system to help keep track of milestones.</p> <p>And we generate many, many Wiki pages, conference calls notes, etc.</p>	
<b>Q5.1 In what ways do you interact with simulations in your work?</b>	<p>We're putting up an OSG integration testbed site so we can submit fake jobs to it (and maybe even real jobs to it) to get some experience with some real logs. It's very, very, very close to being up. We haven't quite got to this stage yet, but I assume the Globus developers will point us at the right CVS branch for the logging code, and we'll do a checkout and a build.</p>	
<b>Q6.1 Describe how you interact with data in your work</b>	<p>Other than the log data, I don't think there's too much to talk about there.</p> <p>We have had an interest in GridFTP performance for a long, long time. Long before Globus existed, one of my research areas has been TCP tuning and parallel data transfer. I've always had an interest in GridFTP, so one of our first targeted applications is GridFTP performance.</p> <p>But for this project we don't have data per se.</p>	
<b>Q6.3 By what mechanisms is access to your work-related data controlled?</b>	<p>Controlling access to log data is something that we've been talking with the OSG folks about: exactly how to do that, how much of it can be public, and various access control issues. Certainly the long-term goal is to have a fairly sophisticated mechanism based on a person's X.509 certificates, so access to specific types of log data can be controlled. But at the moment, it's just user name and password for the right OSG people.</p> <p>Clearly, a big issue is that a user should be able to have access to his/her own logs.</p> <p>We haven't sorted all this stuff out yet, but they are important issues that we need to start working on.</p>	
<b>Q7.1 What resources do you use in your work today?</b>	<p>None, other than the OSG node we're setting up, plus our laptops.</p>	

Interview ID=26 25 September 2007	ANSWERS	ANNOTATIONS
<b>Q8.1 What software do you currently use in support of your work?</b>	I suppose the main answer is Globus because we are working to get some Grid middleware out there doing the right logging. There's also our own software, some of which we put under the <i>netlogger</i> label, and some of it we haven't figured out what to label yet. But there's a bunch of various tools we've written. MySQL database servers. I mentioned we're using this open source tool – <i>syslog-ng</i> – for the collection of log files; that actually seems to work quite well.	
<b>Q8.2 What scripting languages have you used in the past year?</b>	Many of our own tools are actually written in Python.	
<b>Q8.3 What programming languages have you used in the past year?</b>	Python and C for the most part	
<b>Q8.4 What workflow tools do you use in your work?</b>	We haven't used any of the workflow tools. We're certainly interested in trying to get our logging mechanisms into various workflow packages, but as users we don't really do that.	
<b>Q8.6 If the need for new software-based functionality arises in your work how do you acquire it?</b>	These days anytime before you go off and develop something new you always Google to see if somebody else has done it first. As a matter of fact when we wrote the proposal, we weren't aware of <i>syslog-ng</i> . We thought we would have to write that tool. Then after the project was funded we started digging around and, "Oh, great. That milestone's checked off."	
<b>Learning about the user's problems</b>		
<b>Q9.1 What challenges do you face today in accomplishing your work-related goals?</b>	For us to do the research we want to do we need to have Globus and (ideally) Condor and some other software using the new format logs. We also need to get OSG to deploy the central log file collection stuff. Some of these things are not super-high priority items for everyone involved. So it can take a lot of phone calls and prodding. Everybody agrees it's a good idea. It's just not always on the top of everybody's priority stack. So I would say that it's gone a little slower than I had hoped. But really, given the number of people involved and the number of priorities involved, I think we're on pretty good track. Like I said, I think the Globus stuff is really close to being done. The original goal was to have it done by June 1st. It's not that much later than June 1st, really. We don't directly help people instrument their code. We point them at our document and they change their own code. It's usually a somewhat iterative process so far. People will send us their new sample logs and we'll critique them, telling them, "No, you need to tweak this, this, and this." And sometimes, "You're forgetting about this, this, and this." If you put a 20-page document in front of programmers they just skim it, and then they go off and do it. So that seems to be the best way to do it, as opposed to us trying to figure out their code, or expecting them to read our document closely and get it right the first time. We say, "Here, go try to do this. We'll tell you what you got right and what you got wrong." It's just usually two or three iterations, and then it looks okay. At least until we actually deploy it and then discover, "Oh, we forgot about this thing." There certainly will be stuff that we won't know is missing until we get it deployed in real systems. Perhaps we should also package a tool that users can verify their logs with.	
<b>Q9.2 What types of information do you need in order to address the challenges you face today?</b>	I guess it would be good to really understand what kind of failures Grids like OSG experience today. Most of what I know is somewhat anecdotal. Getting a picture of this is a hard problem. I don't know that they know. TeraGrid is the same way. It would be nice if there were somebody tracking and documenting failures in an organized way. It would be good to know what the current issues are a little bit less anecdotally and more concretely.	
<b>Q9.3 What technology-related obstacles do you currently encounter?</b>	We're worried about things like firewalls and security policy getting in our way. I don't know if security policy is a technology obstacle or not, but it could be considered one. Part of the problem with logs is that there is potentially sensitive information in there, and if you strip out the potentially sensitive information you often lose the ability to do troubleshooting. So there are some tricky issues there that we still need to figure out.	

Interview ID=26 25 September 2007	ANSWERS	ANNOTATIONS
<b>Q9.4 By contrast, can you provide examples of technologies you find very useful today?</b>	<p>As I mentioned I was awfully happy to learn about <i>syslog-ng</i>, and I actually think it's a pretty cool piece of software. Glad we didn't have to go off and rewrite all that.</p> <p>It's very, very flexible tool that supports arbitrary inputs and arbitrary outputs. It can read in TCP, UDP, or files and spit out TCP, UDP, or files based on various filters. It's very, very flexible in what you can do with it, and hence potentially very scalable. You can aggregate and filter in arbitrary ways. For central log collection, it seems like they've thought through the issues, and handles everything we needed it to.</p> <p>We haven't yet deployed it in enough sites to know what its scalability issues are, but so far everything seems to be working fine.</p>	
<b>Q10.1 Can you think of any work-related tasks that decrease your productivity?</b>	<p>I think for the CEDPS project, we've managed to keep the number of meetings pretty reasonable.</p> <p>But when you start coordinating with groups like OSG it's pretty complicated, because they're such a big organization with so many different conference calls. It's hard to figure out. Also just trying to keep up with them in e-mail lists – some of these lists get hundreds of messages a week.</p>	
<b>Learning about the Globus user experience</b>		
<b>Q17 What are the major challenges you face using &lt;component&gt; today?</b>	<p>I don't think I have any strong opinions on the components per se. Thus far I have been mostly using the Gatekeeper and GridFTP. I am going to start using WS-GRAM more.</p> <p>The one comment I would like to make is that sometimes when trying to work with various Globus developers, I get the feeling that there's nobody really in charge. Everybody seems to say, "Well, I don't know. Is that more important than this? Is that more important than that?" And the developers are very hesitant to commit to anything without talking to somebody else first. One week it'll sound like it will be a priority, and the next week the work will get bumped. From the outside perspective there doesn't seem to be a lot of cohesive direction and vision. It seems like a lot of firefighting and jumping around ☺.</p> <p>All the Globus developers we've worked with have been great to work with. But every single time you ask, "Hey, can you add this?" They'll say, "Well, sure, but I've gotta find out if this is more important than that." I always get that response. I'm talking about tasks that take somewhere between a half day and two days.</p> <p>[prompt asking if there's anything else to say about the Globus user experience]</p> <p>We've recently become an incubator project for our <i>netlogger</i> work. There have been no hassles other than trying to convince LBL lawyers that Apache and free BSD licenses are effectively the same thing – which is LBL's problem, not Globus's problem. I've been pretty impressed with the whole incubator process.</p> <p>I like the fact that you get a Wiki, a bug tracker, and a CVS repository. And if you need something configured it seems to happen pretty quickly. The lead <i>dev.globus</i> infrastructure person seems really good. I was impressed with the whole incubator startup process and how smoothly it all went.</p>	

## D.27 The end goal is to automatically detect network anomalies

Interview ID=27 26 September 2007	ANSWERS	ANNOTATIONS
<i>Pre-interview question: Do you interact directly with Globus software in your work today?</i>	yes	
<b>Establishing context</b>		
<b>Q1.1 Please provide a one-minute overview of your project</b>	Our area of research is networking and data mining. The particular project I'm involved with mines network data in order to identify anomalous behavior of the network and traffic moving through the network. We're trying to use distributed technologies to offload some of the processing that the user interactively might request, as well as to offload processing of large amounts of captured network behaviors, mining them for specific events. We're trying to use distributed technologies like Globus, for example, to leverage a large number of CPUs to quickly – almost interactively – process data. Processing will be triggered by user requests, as part of analyzing the data in a specific way using specific parameters. The users might just be interacting with the data trying different parameters and observing the result. They further might make inferences based on the results and refine their searches with new parameters to get a better understanding of what's happening in the captured data.	
<b>Q1.2 What is the project's name?</b>	Angle	
<b>Q1.3 Which agency funds the project?</b>	The National Science Foundation	<i>Note to interviewee: you weren't sure about the funding agency and asked that this answer be flagged to remind you to check</i>
<b>Q1.4 What field does your project belong to?</b>	Computer science	
<b>Q1.5 What is your job type?</b>	System administrator, research programmer	
<b>Q1.6 How long have you been a &lt;job type&gt;?</b>	Seven years	
<b>Learning about discipline-specific goals and approach</b>		
<b>Q2.1 What are the main goals of your project?</b>	The end goal is to automatically detect network anomalies. As a first step toward that goal, we want to be able to interactively analyze the data to gain better understanding of it. This will allow us to devise better algorithms that will automatically do that for us. Such anomalies could include: - a user transferring large amounts of data - the presence of a probe - or some kind of an attack Any kind of anomaly, but we're looking at the problem from a behavioral point of view, as opposed to mining actual content.	
<b>Q2.2 How will the success of your project be measured?</b>	One area of the project I'm involved in is system administration. So I'm involved in providing the data and mechanisms to exchange the data, so I'm facilitating the engineering tasks of the project. I'm not as much involved in the actual data mining research, so I can't answer the question about project-level measures of success. The P.I. of the project and students are responsible for devising the algorithms, measuring the quality of the results, and refining their analytical tools. I create data access tools so that the data-mining portion of the project can be fairly approachable to the students. I set up Globus nodes and work with people who have other Globus nodes that we might be able use. The hope is that the nodes will be used to provide data, archive it and index it.	

Interview ID=27 26 September 2007	ANSWERS	ANNOTATIONS
<b>Q2.3 What are the professional measures of success for you?</b>	<p>As long as my responsibilities are not slowing down their progress, I consider my work to be successful. I don't want technical details of how things are distributed or how hard or easy the data is to access to affect the people doing the research. So as long as I can make these resources available and easy to use for them, I consider that to be success on my part.</p> <p><i>[prompt asking for more detail regarding the technical expertise of the users]</i></p> <p>Our users are Masters and Ph.D. students in, typically, math or computer science. Being students they're not very well versed in different toolkits, technologies and languages. They might not have good judgment in how things should be implemented or used or utilized.</p> <p>Given that most of their goals relate very closely to data mining and algorithm development, not having to deal with engineering details can significantly help the project. I feel that it's important for me to provide them with infrastructure that allows them to easily experiment with algorithms.</p>	
<b>Q3 What are you investigating?</b>	<p>For various reasons, which are not always technical in nature, we have started using Globus, because it is definitely widely deployed in different institutions. Also many people are aware of it – there's some expertise out there in using it. We started setting up some of our own Globus nodes. I have mixed feelings about Globus as it is. It forces the user to implement their code in some very specific ways. So there's a certain mindset that you have to work with. You cannot just take your code and just pop it in there if you really want to take advantage of Globus.</p> <p>Otherwise, you're just putting your own code on Globus using your own socket code, disregarding Globus security, and you're just using Globus to schedule things and gain access to machines. So unless you do it the Globus way, you're not really utilizing it.</p> <p>And the Globus way tends to be restrictive. We use a lot of libraries and toolkits like <i>R</i>, which typically are not included in a standard Globus install. So we can't rely on them being on other Globus clusters. So we're shipping our own precompiled code that maybe has <i>R</i> installed in the home directory. We do things like that to get our code running, but in the process we're missing the point of Globus. So we've investigated that, and we're using that.</p> <p>The other way we're utilizing computational resources is that we have a local cluster of our own machines that run our own software. Very simple, lightweight. It kind of works the way we intend things to work.</p> <p><i>[prompt asking for more detail regarding the difference between "the Globus way" and the user's preferred way]</i></p> <p>To me, Globus is a set of daemons and infrastructure that</p> <ul style="list-style-type: none"> <li>- provides a unified security mechanism with cryptography, key exchange, and authentication on each service using a common set of keys</li> <li>- provides a uniform remote procedure call interface</li> <li>- provides some file transfer protocols using multiple underlying network protocols</li> <li>- has some scheduling capabilities (I guess limited to per node scheduling)</li> <li>- contains a set of standard libraries of tools that one can rely on being available on Globus nodes</li> </ul> <p>In order to start doing something outside of the Globus-provided services but staying within the Globus security network, one has to learn additional APIs and how to code things up. So it seems like there's a high startup cost to use Globus.</p> <p>To me it's cheaper to put 20 CPUs behind a firewall and a private network with no way in except through some gateway node that's well secured. I can then just run whatever I want behind that firewall using the most approachable, easiest to use toolkits with the least overhead and with least restrictions on how we code things up.</p> <p>I know there are many people who truly want to distribute their processing across multiple data centers. To them security will be more important. But as long as our project will fit within our local cluster that we can handle ourselves in the back of our lab, it's too much additional work to do it the Globus way. Our cluster is secured behind a firewall. The only access to it is through a gateway machine. The only way to log into the gateway machine is with a pre-shared ssh key. That is equivalent to Globus security levels. With Globus you have a pre-shared Globus key that is also password secured. If the machine you log into Globus from is compromised, the compromiser has both the key and the password potentially. This is the same amount of information you need to gain access to a cluster running behind the head node, which is accessed with private keys and a password.</p>	

Interview ID=27 26 September 2007	ANSWERS	ANNOTATIONS
<p><b>Q4.1 What is your method for investigating &lt;phenomena&gt;?</b></p>	<p>In an earlier phase of the project my involvement included writing the software that captures network events and sends them to a central location to be archived. They're archived on shared storage so multiple nodes can access it. The information about the archived events is stored in a database. Most students have some background in databases, so they're able to query the data they're looking for. Once they retrieve a subset of the available data from the database, they can pull the files representing those events and start data mining. The students have a number of nodes available to them to do the data mining. With the shared storage they can just directly access the file and start their computation. So, basically their job boils down to:</p> <ul style="list-style-type: none"> <li>- starting some number of processes on a number of nodes,</li> <li>- querying the database,</li> <li>- accessing the right files from the right nodes on shared storage,</li> <li>- and running their algorithm on the data.</li> </ul> <p>We designed our database schema by discussing what types of events or types of queries the students will utilize to select the datasets. Based on their particular needs, the database schema was defined and indices created to support the types of queries that will be taking place. Then we start populating it with the data. So the students help determine what queries the database needs to support.</p> <p><i>[prompt asking for more detail on the data capturing]</i></p> <p>All data are captured on the networks we monitor, and then data mining algorithms highlight events of interest. We have relationships with institutions that are interested in the same line of research, and they provide us with capture feeds of their network activities. So we receive our partner's data plus our own network's data. So the data from those networks is stored and mined.</p> <p>The database has references to all the raw data files. We can quickly query based on location, time period and the network load. So we can find out these very general network properties from the database. Then in the background as these files come in, we're doing cluster analysis on features derived from that data. The clustering results are also stored in the database.</p> <p>So we cluster on a set of features, and these cluster coordinates and some statistics about these clusters are also stored in the database. You can then observe the changes in the clustering results between different sites and between unions of different sites. Based on these how these clusters evolve (the changing number of clusters, the sizes of the clusters and other parameters) the clusters themselves can be mined for patterns.</p> <p>Particular files can be picked out of storage and additionally mined because of interest in either the network conditions at the time of capture, or the clustering results. At the top level if we've identified an event that is linked to several clusters we have a way to track back to the original data (the original data being those data files that came in from the originating sites.) So once we know when the event happened, we can look at the original files and maybe derive further understanding of what happened by running a different algorithm on the same file.</p> <p>The amount of data we currently archive in this phase of the project is on the order of one half of a terabyte.</p>	



Interview ID=27 26 September 2007	ANSWERS	ANNOTATIONS
<p><b>Q4.2 How do you work?</b></p>	<p>So if a new partnership were to form, I would provide them with a tool to run on a computer that's capable of observing traffic entering and leaving the partner's network. I provide them with the software and give them instructions on how to run it. Then they run it, and everything gets sent to us, and we do our thing.</p> <p>At this stage in the project I don't need to make any changes to the database in response to new data coming in. The queries are set for the time being. It would have to be a certain very specific type of query for it not to fit in the current schema. And it might be a big task to reform the database to support that. Thus far we haven't had such a request. We're still at the stage where we're looking into better ways of observing and detecting and highlighting events in the data that we do have.</p> <p>As far as introducing the database to new students, I meet with them and explain the schema to them. I explain what kind of things we can officially support, what kind of queries can be run very rapidly without overloading the server, what to avoid, etc. We go over the documentation of how the database was created and structured.</p> <p>As far as file storage, there's really nothing to document. The clusters and events stored in the database point directly to a file system, a path, and a file representing that set of data. So once students pull it out of the database, the students can access it directly.</p> <p>The captured data files have a specific format, so I explain to students how to work the program that extracts the raw data. The raw data files are converted with this program into an easy-to-use, comma separated value format. This is very approachable to students for parsing, working with statistical tools like <i>R</i>, reading into a spreadsheet or into their own program for processing.</p> <p>So I provide a tool that extracts this data from a PCAP file (which is the capture format the data comes in to us.) As new fields need to be extracted or computed based on the PCAP data, I update the tool and provide the new output format to the students.</p> <p>The entire process of how:</p> <ul style="list-style-type: none"> <li>- the file is captured</li> <li>- the file is stored in the capture format</li> <li>- the file is sent to us</li> <li>- the file is archived on our system</li> <li>- the file is preprocessed</li> <li>- the references to the new pieces of data are stored in the database</li> </ul> <p>The entire process is basically abstracted with a few utility programs the students use to extract the contents.</p> <p><i>[prompt asking about the types of tech support requests students have]</i></p> <p>Sometimes they have trouble running a certain query they're interested in. In these cases I will help them rewrite the query or maybe replace it with a more efficient query, returning more or less what they need from the database.</p> <p>I sometimes might have to modify the extraction program to add additional fields they're interested in that, while present in the PCAP in some form, need to be computed (such as a running average.) So I'll add support for these features in the utility programs that the students use to extract the actual fields that they're interested in.</p> <p>Handling this class of changes made up the bulk of my "ease of use" work when the project started. But we really don't have too many requests now that things have settled down and we have a set of fields that are proven (or are believed to be) useful and representative of the behavior of the network.</p> <p><i>[answer continued on next page]</i></p>	

Interview ID=27 26 September 2007	ANSWERS	ANNOTATIONS
<b>Q4.2 How do you work?</b> [continued]	<p>So right now students just get those fields and do their data mining thing, which involves deciding:</p> <ul style="list-style-type: none"> <li>- which data to mine,</li> <li>- how to score it,</li> <li>- how to cluster it,</li> <li>- what parameters to use for it,</li> <li>- how to look for similar behavior in the database once they observe something</li> </ul> <p>And based on the file they are processing, they can try querying the database to see if similar behavior took place at an earlier time. If so, they can access the files representing the earlier behavior and compare it to the current one.</p> <p>Once the system is running, I just make sure it stays running. My workload currently is focused more on the interactive portion of the project. We have a Web page where the user logs in and can select a type of data mining operation and subset of the data they want to mine. I'm currently working on facilitating that.</p> <p>I'm also facilitating the launching of these data mining tools on multiple nodes, so that Web user queries will get results back in, hopefully, seconds. There are a lot of details of how to launch these jobs and how to move the files to the compute nodes; that's what I'm currently involved with. It's at a higher level than the low-level services that are already established.</p>	
<b>Q6.3 By what mechanisms is access to your work-related data controlled?</b>	<p>Access is only allowed from specific nodes using specific accounts and passwords that are limited to specific tables in the database. So both host control and log in password control.</p>	

Interview ID=27 26 September 2007	ANSWERS	ANNOTATIONS
<b>Q7.1 What resources do you use in your work today?</b>	<p>For this project we'll have available about 36 CPUs with about 36 gigs of RAM total on those nodes. We'll have approximately 30 terabytes of storage across the nodes. We'll have a ten-gigabit link to a StarLight facility and the rest of the world, should we choose to move the data somewhere else for processing or launching jobs.</p> <p>A researcher in our lab developed a tool that will be used on those particular nodes. The system is not ready yet and I have not actually used it, so I can't really comment on it. It's said to be a lightweight framework for file transfers and launching processes on those nodes. And it's also optimized for long distance high latency networks where the data can be sent between U.S. and Japan at multi-gigabit speeds by utilizing protocols other than TCP-based. So this worldwide, distributed framework is future work, some of which will be done (hopefully) by Supercomputing 07. The work is really starting right now. We're actually working with the University of Chicago, ANL and some of the Globus guys. So it's being worked out right now using Globus. Then separately there's the framework that we have.</p> <p>So strictly speaking the resources I use right now are all on my local cluster. Also there are the data sources from partner institutions coming in over the net, but the data is stored on our site. Once it is sent to us, we never access the remote nodes that captured it.</p> <p>Our lab outside of this project does network research. So we have nodes around the world that belong to us or are on networks belonging to our partners, and they're used for some of the networking research. The particular toolkit that I mentioned, which is called Sphere, is based on UDT. UDT is the network protocol for long-distance high latency networks designed to achieve multi-gigabit speeds between, for instance, StarLight and JGNII [<i>Japan Gigabit Network II</i>].</p> <p>We might at some point do some experiments to show that we can launch these jobs and efficiently transfer files back and forth to the compute nodes located in Japan, Korea, Cal Tech, CERN and StarLight Chicago using our toolkit. The files would be transferred and shared using a client-to-client, peer-to-peer file sharing type of software based on UDT. They would be distributed over the nodes, and the nearest site containing a copy of the needed file would transfer it to the central location. We would use the experiments to demonstrate how performance could be improved by distributing a process across dozens and dozens of nodes located around the world. The experiments would leverage our UDT servers, which store copies of some subset of the files, and optimize the server that actually provides the file to the compute node. Sphere involves some of these things, and this particular project that I'm involved with might or might not utilize Sphere to offload processing. Or it might just utilize it in the basic job of distributing the files, with launching the processing happening only on our local cluster. But it would be capable of doing the long distance transfers, as well. I'm not familiar with it enough to answer further questions about it.</p>	
<b>Q7.5 What types of information do you need to know about a resource in order to determine if it is suitable for your work?</b>	<p>Most of the data mining code being written by students is done in Python. Oftentimes <i>R</i> is utilized, which is a standard statistical package. There are also a number of additional <i>R</i> modules that are not installed by default because they're rarely used or are not part of the official <i>R</i> distribution. So Python, we usually assume, is installed. It's popular enough nowadays that we can assume it is installed. <i>R</i> oftentimes needs to be installed; once it is installed we need to make sure the particular modules are installed.</p> <p>The PCAP files the network data is stored in require a library for extracting the contents. That library, <i>libpcap</i>, typically is not installed. So if we want to use our code to extract these features on the remote nodes doing the computation, we need the PCAP library to be installed.</p> <p>Typically it's difficult to go to a regular Globus site and just run our code because we rely on so many external libraries and tools that are not part of the Globus standard install. Sometimes it's possible to request these things to be installed at the site, sometimes not. If it's not, a lot of these tools can be compiled and installed in a home directory and run from there, as opposed to assuming the system has them.</p>	
<b>Q8.2 What scripting languages have you used in the past year?</b>	Most of our things are written in Python or perl.	

Interview ID=27 26 September 2007	ANSWERS	ANNOTATIONS
<b>Q8.3 What programming languages have you used in the past year?</b>	Some C and C++.	
<b>Q8.6 If the need for new software-based functionality arises in your work how do you acquire it?</b>	If the nodes are under our control and the software's already under license, I just go ahead and install it. It's done within hours and we start using it. If it's somebody else's Globus node, then it becomes much more complicated. Complications could include administrative issues, policy issues, scheduling workload issues. I mean they might be busy and not willing to install it. They might have a policy of not disturbing the standard installation. Basically it involves a lot of phone calls and e-mails. The primary reason behind these complications is that we don't own the resource. With my local Globus nodes (of which I have a number) things just get installed when they're needed.	
<b>Q8.7 How do you share software with others?</b>	We don't restrict sharing of anything we write, so if somebody were interested in something we've written, I don't believe there would be any problem in sharing it. Our subversion server is not publicly accessible, but we could easily export something for somebody on request.	
<b>Learning about the user's problems</b>		
<b>Q9.1 What challenges do you face today in accomplishing your work-related goals?</b>	<p>I'm a sysadmin. I help steer this project and help supervise the students. I solve the engineering problems, and I do everything else that's needed to support this lab. I'm very short on time. If I have a choice between using local CPUs to do work, or Globus CPUs controlled by somebody else to do the same work, I'll use my local CPUs.</p> <p>I know it'll take me an order of magnitude less time to set up something on my nodes to give the numbers that need to be computed, than it would take for me to schedule the use of the Globus nodes on the TeraGrid (for instance). The setup work I'm referring to includes account setup for the students that need to access the resources, scheduling time to run our code on them, and installing prerequisite software. To use the remote resources I need to either:</p> <ul style="list-style-type: none"> <li>- work with the admin of the remote sites to install things, or</li> <li>- devise instructions for the students on how to compile the software in their home directories so they can be ran that way.</li> </ul> <p>Or I can just use my nodes and just get it over with much, much quicker.</p> <p>In the past we've worked with some of the Grid experts at the University of Chicago to run some of our code on TeraGrid. We provided them with scripts and tools that would process our data files – software we would normally provide for each job run locally. However when we tried to run on TeraGrid we found that some <i>R</i> modules were missing, as well as some PCAP stuff. So some of the required dependencies for our software were missing.</p> <p>I had to devise a way to compile it to home directory so it could be loaded from there. Then a University of Chicago Grid expert had to compile it on the TeraGrid. But he had to compile it with different paths based on the machine architecture. He also had to make sure to launch the code appropriately so the proper library paths were defined for perl, Python, and other libraries. He had to compile it for the home directory.</p> <p>So it was a somewhat involved process just to get things running. And he, being very knowledgeable about Globus, was able to do a lot of the scheduling and whatnot to get things running. But it still seems like a very involved process compared to us just launching it on our nodes in the corner of the lab.</p> <p>So the time burdens associated with setting up the application-specific environment on the remote machine is a big challenge. The University of Chicago Grid experts handled the Globus-specific setup, so I can't comment on challenges associated with that.</p>	
<b>Q9.4 By contrast, can you provide examples of technologies you find very useful today?</b>	I don't think there's a silver bullet.	
<b>Q10.1 Can you think of any work-related tasks that decrease your productivity?</b>	<p>I'm a sysadmin so this may not be true of everyone, but I find meetings, collaborating with outsiders, getting everybody up to speed, exchanging docs to be very, very time consuming.</p> <p>So ideally I can have everything in my lab and have my students stop by my office to answer their questions, write something on the whiteboard, and have them start running it immediately. That saves me a lot of time.</p> <p>I'm comparing this to bringing different groups together, having weekly or biweekly meetings, exchanging our little limited views of each other's work, and trying to make sense of how we are going to put things together. That seems like a big, big time drain.</p>	

Interview ID=27 26 September 2007	ANSWERS	ANNOTATIONS
<b>Learning about the Globus user experience</b>		
<b>Q11.1 Which Globus data components do you directly interact with in your work today?</b>	<p>Basically my direct involvement with Globus has been installing a cluster at my home institution.</p> <p>I basically following the Quickstart Guide, and I got it to where I can run GridFTP, and launch jobs using globusrun-ws, and query the MDS server for the services available registered on what I deem to be the head node.</p> <p>So I've used the command line tools for GSI certificates, GridFTP, GRAM4, the Index Service, and Java WS Core. I've never programmatically accessed them.</p>	
<b>Q16 Why do you use &lt;component&gt; instead of an alternative technology?</b>	<p>Clients for GSI certificates, GridFTP, GRAM4, the Index Service, and Java WS Core:</p> <p>My boss requested the project so that we can become more familiar with Globus and to get some experience using it. We basically followed the recommended path suggested by the Quickstart Guide in the Globus documentation.</p>	
<b>Q17 What are the major challenges you face using &lt;component&gt; today?</b>	<p>Clients for GSI certificates, GridFTP, GRAM4, the Index Service, and Java WS Core:</p> <p>For our limited needs, it seems to work.</p> <p>Security infrastructure:</p> <p>It seems that if I were to have 70 Globus nodes under my control (which I have not reached yet) but if I had 70, I can foresee difficulties associated with centralized account management. Key management for Globus seems very complicated. Until I go beyond 20 or 30 nodes it's been suggested to me just to keep the keys locally on all the machines and not try to centralize everything - it's much easier that way. Some heavy Globus users have suggested this to me. A lot of the more advanced configurations and uses of Globus seem to be not as well documented. So for me, that means I'm generating each key by hand for each user, and distributing the signatures to each node to allow the user to log in. It seems very painful and very complicated. And for what I was tasked to do (enable users to copy files and launch jobs remotely) it seems like a lot of work.</p> <p>Now if we were to start using all your RPCs, your Secure MPI, and the secure-wrapped standard libraries that have been modified for Globus use. If we were to use all that plus GridFTP, sharing a common authentication infrastructure, then the burden of all these additional layers and centralized key distribution would seem worthwhile. But if we have eight nodes, and we just want to launch our scripts remotely, that seems like a lot of work.</p> <p><i>[prompt asking for ideas for how to better streamline things]</i></p> <p>I don't see a solution that's totally secure or very easy to set up. I mean if I gain access to somebody's private Globus User Key and I get ahold of their password used to initialize the Grid Proxy, then I basically have local access to all of your Globus nodes. Then I can exploit any local vulnerability of those nodes. To me that is as secure as gaining access to somebody's private <i>ssh</i> key and the password used to unlock it, which also gains you access to all the machines that that user accesses.</p> <p>The fact that authentication between the different Globus services is encrypted has no value to me. Ideally, I would run all these nodes on my own network that's already firewalled and nobody can actually observe the traffic on it. So I don't need all these RPC calls being encrypted or these handshakes to be encrypted between the different services because there's nobody that can hear them. I still have to protect the head node, but that is just as vulnerable as the key that somebody uses to log into Globus. I don't know of a solution to that.</p>	
<b>Wrapping-up</b>		
<b>Is there anything you'd like to say to the people who build software for use by people like you?</b>	<p>Just keep it approachable because we're all very busy. So it's easy to install it, to use it, to run it, to really utilize it.</p>	

## D.28 The production worthiness of infrastructure is of the utmost importance

Interview ID=28 28 September 2007	ANSWERS	ANNOTATIONS
<i>Pre-interview question: Do you interact directly with Globus software in your work today?</i>	yes	
<b>Establishing context</b>		
<b>Q1.1 Please provide a one-minute overview of your project</b>	<p>The VO services project provides user registrations and fine-grained access privileges to resources. We have an infrastructure that serves several virtual organizations and stakeholders.</p> <p>US CMS [<a href="http://www.uscms.org/">http://www.uscms.org/</a>] is one of the founders of the project. The US Atlas project [<a href="http://www.usatlas.bnl.gov/USATLAS_TEST/Physics.shtml">http://www.usatlas.bnl.gov/USATLAS_TEST/Physics.shtml</a>] is also one of the founders. We work with the Open Science Grid and several VOs at Fermilab.</p> <p>We develop infrastructure that implements user registrations and access to resources and plugs into different resource gateways. In particular we provide access to three types of resources. One type is computing elements via different versions of the Globus gatekeeper. Another resource type is storage via the SRM interface and dCache. And the third type is what we call the gLExec, which is a su-like facility that changes users when jobs execute on the worker nodes.</p>	
<b>Q1.2 What is the project's name?</b>	VO Services Project	
<b>Q1.3 Which agency funds the project?</b>	US Atlas, US CMS, Open Science Grid, Fermilab Computing Division	
<b>Q1.4 What field does your project belong to?</b>	High energy physics, non-high energy physics (through OSG), computer science	
<b>Q1.5 What is your job type?</b>	<p>Project Lead</p> <p>I manage various aspects of the project: software releases, requirements gathering, coordinating bug fixes, coordinating deployment, communicating status and ideas with stakeholders.</p>	
<b>Q1.6 How long have you been a &lt;job type&gt;?</b>	1.5 years	
<b>Learning about discipline-specific goals and approach</b>		
<b>Q2.1 What are the main goals of your project?</b>	<p>There are two major goals associated with the project. VO user membership user registration is one of the goals, and the second goal is providing fine-grained authorization access to resources.</p> <p>As part of membership user registration, we provide an infrastructure that enables virtual organization administrators to create a structure inside the virtual organization using concepts such as groups and roles.</p> <p>We then apply this concept in the second goal. We provide access authorization based on this organizational structure, such that users can present themselves with groups and roles. Different authorization privileges and execution environments are enabled, depending on the roles and groups that the users present.</p>	

Interview ID=28 28 September 2007	ANSWERS	ANNOTATIONS
<b>Q2.2 How will the success of your project be measured?</b>	<p>The project is organized into several different sub-tasks and the different sub-tasks take care of different aspects of our stakeholders.</p> <p>For example, in our project we have products such as gLExec that provide context changes of privileges when a job is executed at worker nodes. This is very closely related to workload management systems. So we measure success of that infrastructure with respect to the success of the workload management system on the Open Science Grid.</p> <p>We focus on other aspects too, such as providing access to computing resources. In this case success is measured on how production-ready the infrastructure is. This is a question of whether we can meet the baseline of our users job flows, of access to files, etc. These are the more technical metrics, if you will.</p> <p>Then there are the more political metrics related to how happy people are with the way we conduct our business. Do we have an open process to consider input from different stakeholders? Do we have large groups that are not considered in our requests for input? Etc.</p> <p>Regarding the production aspects of our work:</p> <p>Our infrastructure is deployed at several institutions. Examples include Fermilab, BNL, US Atlas facilities, the Tier-2 facilities of US CMS, and several sites on the Open Science Grid that may not be US CMS or US Atlas facilities.</p> <p>Some sites are particularly careful at measuring the performance of our services. We have a very good example of this at Fermigrid. They maintain metrics on the number of accesses to the services of our infrastructure. They can keep a baseline of these metrics compared to the requirements that were given to us by the various virtual organizations we work with.</p>	
<b>Q2.3 What are the professional measures of success for you?</b>	<p>Wearing my hat of VO services project leader, I am happy when we can collaborate with different groups that work on authorization. Also when our ideas and implementations can be shared and re-used, and when I see that we have traction with other projects doing similar things.</p>	

Interview ID=28 28 September 2007	ANSWERS	ANNOTATIONS
<b>Q3 What are you investigating?</b>	<p>We have a couple of lines of investigation that we are following up with for what we call the phase 3 of the project starting in October 2007.</p> <p>One of them is very important to me personally: authorization interoperability. This is a project we started in collaboration with EGEE and Globus and that later involved also Condor. We want to standardize the protocol used by resource gateways (policy enforcement points in jargon) to communicate with policy decision points (PDPs). PDPs are servers that keep the policies for privileges to those resources.</p> <p>It is very important for us to make sure that these protocols are common so that developments of middleware in the US can be immediately plugged into authorization infrastructures developed in Europe (and vice versa). This will allow middleware developed in Europe to be plugged into the authorization infrastructure that we have here in the US. The implementations are different, but if we achieve a common protocol, then we can achieve interoperability.</p> <p>Globus is doing the development on this; we are working with them to define requirements. Eventually the Globus 4.2 series will have this authorization interoperability protocol plugged into by default. So, this is one of the venues we are following up with.</p> <p>Another thing that is very important to us is providing support to the storage groups in defining what is called the next-generation of storage authorization models. Access to storage is one of the big use cases that we are trying to make right in collaboration with our stakeholders. This includes Open Science Grid and various different storage groups, in particular SRM and dCache. So, we are working in close consultation with them, understanding their use cases for access to files, directories, storage, reservation of space, etc. (There are groups in storage that work with space reservation and the groups work with SRM, too. There are issues related to authorization on those fronts as well.)</p> <p>So we're working to make sure that our infrastructure can support the use cases of interest to those various capabilities. We work both in consultation with them and in collaboration with them to understand what these use cases are, what our current infrastructure can do, and what type of access privileges we can define with our current infrastructure. We are open to extending our infrastructure to include new features if there are use cases that are not covered by our current infrastructure.</p> <p>With regard to workload management systems:</p> <p>We work in close collaboration with workload management systems to enable authorization access to computing for systems that are pull-based. There is a workload management system that submits what we call Pilot jobs to resources. These Pilot jobs get actual payload from some repository. There are all sorts of different security and authorization issues related to this model. We work closely with these groups to provide the appropriate solution.</p> <p>So the three areas discussed thus far are: authorization interoperability, supporting the storage use cases, and the pull-based workload management use case. We also pursue a fourth area, which is the definition and enforcement of policy.</p> <p style="text-align: right;"><i>[answer continued on next page]</i></p>	
<b>Q3 What are you investigating?</b> <i>[continued]</i>	<p>The idea is VOs will be able to define, on top of the organizational structure that they have defined, privileges directly associated with these various groups and roles. Privileges include things like priorities in a filesystem, or priorities in a batch system, etc. So these policies will be able to be propagated to sites, which in the end will be able to enforce them. So, we are working on this with a group called Tech-X, with which we won an SBIR phase I grant. So we have some soft money to look into these aspects of the problem.</p> <p>The privileges provide access to services running on resources. For example, priorities in a batch system, or quota for a certain group and role. They are defined in the structure. The users present themselves with a certain group and role, and they are enforced by the resource providers.</p> <p>So those are the four major things that we are looking into for phase 3 of the project.</p>	



Interview ID=28 28 September 2007	ANSWERS	ANNOTATIONS
<b>Q4.1 What is your method for investigating &lt;phenomena&gt;?</b>	<p>We believe very much in the power of collaboration. There are several groups interested in authorization and we are in contact with a few of them. In particular we are in contact with Globus, people in the Open Science Grid, and with interested groups within EGEE. From EGEE there are various people who are also looking into this area – people from Switzerland, Northern Europe and Holland.</p> <p>We tend to discuss these broader issues when we meet at various fora. For example at the middleware security group is a forum where we participate very actively, and then we have mailing lists that are shared among all of these various participants. The collaborations involve middleware developers and some end users when appropriate. The people involved tend to be field experts.</p> <p>So, we have phone conferences, face-to-face fora and the mailing lists. Once we have agreed on ideas and we have discussed the feasibility of things, we go into more of the technical side of project management. We create working group charters, define a plan, and try to follow up with the plan. You know – the usual project management process control.</p> <p>Our environment is very distributed, especially for the resources that the project controls. Modern development models like Agile software development techniques tend not to work so easily. So we have to employ a less Agile approach often. For example, having stand up meetings as prescribed by the Agile software development method is not easy for us. We tend to have phases in our projects where we define the milestones for the next 12, 18 months with some reasonable checkpoints (so if things come up we can change our plan) and then we execute. Milestones are defined at a pretty high level so we tend not to lose ourselves in the details of things.</p> <p>In addition, one of our major activities is supporting the current infrastructure. If it happens that there are deployments and particular use cases that are not addressed by current features, we need to change our plan accordingly. So we tend to have a pretty high level view of the tasks. With such large distributed collaborations it is not possible to control every little detail of the project. So both delegation and trust within the collaboration are critical.</p> <p>Then there is a whole process for testing. We tend to work very closely with the VDT team. One of our stakeholders is the Open Science Grid. So when we deploy and release new versions for the stakeholders, we go through the whole process defined by them – we work with the integration testbed of the Open Science Grid, we have nightly builds via the VDT infrastructure – so that is very structured.</p> <p>VDT is the distribution mechanism for most of our stakeholders. Some of the products are also distributed by via RPM. At BNL they use the RPM distribution mechanism to keep up with the deployments. But we use VDT for most of the other deployments. VDT's a large suite of software. So they have our infrastructure as well as other middleware. In EGEE currently they are not currently using our software, and this is also why we're doing this authorization interoperability project, so that different implementations can be interchangeable.</p>	
<b>Q4.5 How do you document your results?</b>	<p>We tend to work in the model of the VDT. So we document our features in our own website, and then we have also documentation into VDT and the Open Science Grid pages. So, depending on the stakeholders that we are working on, we follow different documentation paths, depending upon what is most easy for the recipients. We have extra documentation, for example, the Open Science Grid users of the VDT distribution of our software.</p> <p><i>[prompt asking if users report problems with the software to VDT or OSG]</i></p> <p>It depends on the environment. So, if problems occur in the Open Science Grid, then there is a whole triaging and reporting structure with the GOC (grid operation center of the Open Science Grid.) Typically users open a ticket with the GOC and they figure out how to deal with it appropriately. It goes to administrators first, and then eventually to the developers if it is a bug.</p>	
<b>Q5.1 In what ways do you interact with simulations in your work?</b>	<p>We provide the ability for a VO to give different privileges to scientists who do simulation vs. other types of computing processing. But, this is pretty much it.</p> <p><i>[prompt asking about the approach used for testing software]</i></p> <p>We have different processes. We do internal unit tests, such as class-level tests. Then we have nightly builds and a standard test suite that is run by VDT. And then when a new major release comes up we follow the process of the integration testbed of the Open Science Grid.</p> <p>So typically this software is deployed on the order of six sites. Different administrators install and configure it for their infrastructures. We also participate in the weekly conferences and mailing lists and we interact with the administrators. If they find problems we go back in the cycle until everything is fine.</p>	

Interview ID=28 28 September 2007	ANSWERS	ANNOTATIONS
<b>Q7.1 What resources do you use in your work today?</b>	For the development itself we don't need many computer resources. We are not a big team. We have some desktops. We use some tools like Maven, for example, to do the building and documentation. Then for the actual testing, we use the facilities of VDT. They have a cluster that allows running of tests. Then as I mentioned, for the final part of the testing we interact with several system administrators. They have their own test systems and they install our infrastructure on those systems and try to break it.	
<b>Q7.5 What types of information do you need to know about a resource in order to determine if it is suitable for your work?</b>	Actually, generally it is the other way around. We try to be pretty open in the type of platforms and resource configurations that we support. So sometimes we find a particular configuration that we do not support. We generally figure this out via the testing process. It's difficult for us to understand if something is not supported unless it's something obvious. For example we do not support Windows platforms and people in our testing circles do not have this use case. We have a set of requirements and specifications and we try to find those configurations that break them.	
<b>Q8.1 What software do you currently use in support of your work?</b>	We have different pieces of infrastructure: We use Java and Servlet-based servers, run under Tomcat and Apache. We have C for some of the other components (for example the callouts from GT2 gatekeepers, or the gLExec infrastructure.) We use Maven and then some development environments. We are investigating now the use of XACML as a protocol to communicate policies.	
<b>Learning about the user's problems</b>		
<b>Q9.1 What challenges do you face today in accomplishing your work-related goals?</b>	The fact that the collaboration is very distributed. We have fractions of people working on the project. They are scattered around the US (for the base part of the program) and around the world for the other collaborations (like authorization interoperability.) This is clearly a challenge for managing and controlling the development processes. So you have to take a higher-level approach to project management. Delegate and trust the results and have a looser control over what's done. One of the advantages is that you have contacts with different scientific groups in the different universities and laboratories that collaborate with us. So in principle you have access to more ideas and fora, but the distributiveness of our infrastructure is very challenging.	
<b>Q9.2 What types of information do you need in order to address the challenges you face today?</b>	We tend to create project-related structures, such as communication channels. We have status report-type meetings. We have operational mailing lists to address the fact that our userbase is very distributed. These are the venues that we tend to put in place in our project management structure to address the flow of information that we need.	
<b>Q9.3 What technology-related obstacles do you currently encounter?</b>	For projects that are at the forefront of the technologies (like authorization interoperability) we have the challenge that some of the standards we're planning to adopt (such as XACML) do not have a stable and accepted implementation. So, currently for example, there are two implementations of the libraries, one is by OpenSAML community, and the other one is by the Globus team. And both are non-complete, both try to address the same issues. There are different tweaks that the different groups do to the specifications in order to be able to implement things. And so there is always this question of what implementation should we use. In the end, we tend to favor the ones for which we have more personal contacts with. So our first choice, for example, goes to Globus, because we have a long-standing collaboration with them. But this is in principle an issue. Eventually, ideally we would converge into a single implementation, and this would be it.	

Interview ID=28 28 September 2007	ANSWERS	ANNOTATIONS
<b>Q9.4 By contrast, can you provide examples of technologies you find very useful today?</b>	<p>Servlet engines are useful to our environment. For example Tomcat is pretty much the basis of our development. Also Java. There are issues associated with that as well, right? Different versions of Java, different versions of Tomcat for example. But these are things that we find particularly useful for our work deployments.</p> <p>Servlet engines are useful because they let you do development of the business logic without having to spend a lot of time coding, for example, access and data propagation. These are pretty standard technologies, right? I mean they are used in the industries all over; this is also a positive aspect.</p> <p>If we want to look at cool things, then I would have to go back to the implementation of these new specifications. These new ideas to express policies in XACML. Those would be the technologies that I would find cool. XACML is really cool because they represent new ideas that a group of people came up with, and they have formalized this in XML and XML schemas languages. They let you do things that we could not do five years ago. They have an expressive power that we didn't have five years ago. So they make our life easier.</p>	
<b>Q10.1 Can you think of any work-related tasks that decrease your productivity?</b>	<p>I think that we could improve our efficiency with more automated tools for scheduling conferences and meetings. Some of those things are a bit repetitive, for example, setting up a context for an event when we do a meeting. There is some repetitiveness to that.</p> <p>We could improve, for example, our way of querying information on our mailing lists. The tools we have today only let you do a very basic search of information. Sometimes I spend quite some time looking for all the information and have to dig through my personal email. We have some room for improvement there.</p>	
<b>Learning about the Globus user experience</b>		
<b>Q11.1 Which Globus data components do you directly interact with in your work today?</b>	GridFTP	
<b>Q13.1 Which Globus execution components do you directly interact with in your work today?</b>	GRAM2	

Interview ID=28 28 September 2007	ANSWERS	ANNOTATIONS
<p><b>Q16 Why do you use &lt;component&gt; instead of an alternative technology?</b></p>	<p>GridFTP: GridFTP is both a protocol and an implementation, right? So let me qualify the context first. I'm talking about this technology with a different hat now because we do not use any of this technology for the VO services project. We work in conjunction with these technologies, because our user communities are interested in controlling access to resources by via these implementations and these protocols. So I have experience in using them in other contexts. We like GridFTP because it's both a protocol and implementation. The protocol has a specification that led to different implementations of the protocol. For example dCache, which is a storage element developed at Fermilab, has its own implementation of the specification. So the protocol has been demonstrated to be specified well enough to be implemented by different providers. Again, from the VO services project point of view, we interface to it because our users use it, so we have to provide authorization to storage resources via that protocol as well.</p> <p>GRAM2: As for the gatekeeper, it's the same story. Most of our users control access to computing resources via the gatekeeper, so we had to provide authorization plug-ins for the gatekeeper too. We also have authorization plug-ins for the "gatekeeper 4", but our users are still investigating whether this is really the technology they want to buy into. GT2 had a big acceptance in the past years. The entire Open Science Grid, for example, the computing of US CMS here in the US is all based on GT2. So it is very important for us.</p> <p>Java WS Core: We have some information about the technology that comes from people who are trying out the Web services version of the Globus Toolkit. We have implementations of plug-ins for our authorization infrastructure for the core services, but they are not mainstream because they are not in production yet. In OSG they're not in production yet, and for our other stakeholders like US CMS, they are not in production yet.</p> <p>MDS2: I use MDS2 for a different project. We don't provide authorization to information tools from the VO services project. We used it in the past, like five years ago because it was a convenient way of describing information using a tree structure information data model. And then, more recently we have abandoned it for a competing technology called CEMon, developed by gLite. But this is for different project.</p> <p>Security: Oh I love it. It addresses all of the use cases that interest us. It supports digital signature with capability of doing encryption, integrity checks, there is the ability of doing delegation, all the expected steps in the authentication processes, the ability of having control lists, signature policies... it's a very complete suite that does what we need.</p> <p>There are people who are starting to use a different infrastructure, like OpenSSL for example. For the time being we are happy with the GSI infrastructure which is, by the way, deployed everywhere by our stakeholders.</p>	

Interview ID=28 28 September 2007	ANSWERS	ANNOTATIONS
<b>Q17 What are the major challenges you face using &lt;component&gt; today?</b>	<p>GridFTP: We don't have major issues with GridFTP.</p> <p>General: With the GT2 framework in general, the fact that it's pretty much a frozen development can present a challenge. We have contacts with the Globus Toolkit developers, so for exceptional things we can have some features added to the infrastructure.</p> <p>But the challenge that we face is that it is in production everywhere for our stakeholders and it's not actively developed anymore, because the Globus Toolkit has moved to the Web service version. This is challenging to us. And this is also why there are groups that are investigating the new technologies. But before you convince yourself that they are really production quality... well, it takes a long time.</p> <p>This situation affects us most with the GT2 gatekeeper. For example, in the VO services project we would like to pass more context back and forth between our authorization plug-in and the gatekeeper. This would require a change to the GT2 gatekeeper code and our code (to adapt to a different API). And it takes a lot of effort to try to bring this thing up again. We have now the agreement with the developers that they will work with us. But then on our side the people who were following up with that don't have so much effort anymore. So things got stopped. It's a frozen piece of development, so it's difficult to make it alive again if you need to change anything.</p> <p>Security: The fact that the C implementation and the Java implementation don't do exactly the same thing is a problem. I mentioned policy signature files earlier. While the C implementation does consider them to define the namespaces that CAs are allowed to sign, but the Java version does not. So there are inconsistencies between the different releases. So you might have a version of the Globus Toolkit, and you expect the different versions to do the same things and sometimes they don't.</p>	
<b>Wrapping-up</b>		
<b>Is there anything you'd like to say to the people who build software for use by people like you?</b>	<p>One thing that I should like to emphasize is that the production worthiness of the infrastructure is of utmost importance to us. Make sure the software not only has quality attributes like performance or maintainability, but also has quality attributes such as usability and the ability to operate the infrastructure. This implies a need to provide all sorts of bells and whistles all around the software, such as the ability of doing monitoring and operational tools to manage administrative sides of the services. Not having means we must provide such services around the software to make it usable by <i>our</i> users.</p>	

## D.29 Our framework must adapt to changing conditions from the problem & the Grid

Interview ID=29 5 October 2007	ANSWERS	ANNOTATIONS
<i>Pre-interview question: Do you interact directly with Globus software in your work today?</i>	No	
<b>Establishing context</b>		
<b>Q1.1 Please provide a one-minute overview of your project</b>	The project deals with water distribution security. As such we are developing a simulation optimization framework to solve source identification problems in water distribution systems. The purpose is to identify a contaminant source in the system from the measurements that come out of sensors or water quality meters.	
<b>Q1.2 What is the project's name?</b>	Adaptive Cyberinfrastructure For Threat Management In Urban Water Distribution Systems	
<b>Q1.3 Which agency funds the project?</b>	National Science Foundation	
<b>Q1.4 What field does your project belong to?</b>	Civil engineering	
<b>Q1.5 What is your job type?</b>	My job title is Associate Professor and I am the project lead for the project	
<b>Q1.6 How long have you been a &lt;job type&gt;?</b>	For over seven years	
<b>Learning about discipline-specific goals and approach</b>		
<b>Q2.1 What are the main goals of your project?</b>	<p>The main goal is to develop optimization algorithms that are less sensitive to the distributed, heterogeneous nature of the Grid resources and work reasonably well under different conditions. We apply these algorithms to this problem of water security, in which we want to identify a contaminant source and its release history.</p> <p>We will also want demonstrate it to an urban water distribution system. In particular we have chosen the Greater Cincinnati Waterworks system as our demonstration system. We are working in collaboration with the University of Cincinnati and the Environmental Protection Agency (EPA) on this project.</p> <p>So to summarize, the goals are to develop optimization algorithms, to develop the simulation part to work under these environments, and to apply them to this source characterization problem. So these are the three goals.</p> <p>We have optimization algorithm development that will work in the Grid environments. We'll also enhance the simulation tool to work under these conditions, and then demonstrate the work.</p> <p>Comprehensively, we will be developing a prototype system that city authorities can take and apply to their own problems.</p>	
<b>Q2.2 How will the success of your project be measured?</b>	<p>It'll be measured by the performance of the entire framework in a Grid environment. So we will be doing a number of tests to evaluate, under different heterogeneous conditions, how this entire framework works.</p> <p>And then the ability to locate these contaminant sources accurately under different scenarios. So we will be looking at different hypothetical scenarios, and seeing whether our entire framework (including the optimizing and algorithms) can detect these things with the given computational sources.</p> <p>And then third is to demonstrate the approach on a realistic problem in Cincinnati.</p> <p>The word "adaptive" in our project name means that our entire framework needs to adapt to both changing conditions from the problem side and changing conditions from the Grid side, responding to TeraGrid resources that come and go.</p>	
<b>Q2.3 What are the professional measures of success for you?</b>	For me success would mean taking this project to the next level once we develop and demonstrate it. We would like to get the Greater Cincinnati Waterworks or the EPA interested in taking it to the next level. We hope to evolve into a center-type project where we used distributed computing resources for this type of problem.	

Interview ID=29 5 October 2007	ANSWERS	ANNOTATIONS
<b>Q3 What are you investigating?</b>	<p>There are two problems that we are investigating: one is source characterization and the second one is contaminant control.</p> <p>We're investigating these intentional contaminants, primarily motivated by homeland security concerns. If a contaminant is introduced in any location in the water distribution network, then we want to determine where and when it occurred and also the time history of the occurrence. The approach is to take readings that come from water meters at different households and water quality sensors that are placed in the network in different locations. So the problem that we are investigating is what's called "source characterization problem" in water distribution networks.</p> <p>The second part of the problem is to develop control strategies to minimize the impact to consumers. This involves looking at hydraulic control strategies such as turning off pumps or turning on other pumps to minimize the spread of contamination, thereby minimizing the impact.</p>	
<b>Q4.1 What is your method for investigating &lt;phenomena&gt;?</b>	<p>The method that we are looking at is heuristic optimization algorithms, in particular algorithms that are in the class of evolutionary computation algorithms. Our optimization component is based on that.</p> <p>The power behind this method is that it's very flexible in terms of the types of problems that we can solve. They also fit well in the heterogeneous distribution nature of the TeraGrid resources.</p> <p>The simulation code that we are using is based on EPANET, which is developed by the EPA. We enhanced this code to work in the Grid environment.</p>	
<b>Q4.2 How do you work?</b>	<p>There are three groups: the algorithm people, the implementation people and the problem people.</p> <p>The methodology team focuses on algorithms, and is headed by one of my colleagues at my home institution. He directs methodology development. This means he ends up directing some of my students' work, in terms of developing these methods. There is another graduate student who is directed by my co-investigator, and he's doing most of the method development.</p> <p>And then there's an implementation team, which is actually doing the implementation and testing on the TeraGrid. So the computer science student who is directly working under me is actually doing most of the implementation, but he's also familiar with the methods.</p> <p>There's another student who is looking at the problem aspects of things. She is more involved with the EPANET code and setting up the problem. Problem people focus on the application and logistics, basically designing the problem scenarios. They are the ones who communicate directly with the Cincinnati team, which is focused on this area. Another investigator for the project is located at University of Cincinnati.</p>	
<b>Q4.3 How do you keep track of interim results, if at all?</b>	<p>We have a wiki website where we post findings and publications and things like that.</p> <p>Most of the results are not communicated through the wiki. We have another website that the students use to post and share their results.</p>	

Interview ID=29 5 October 2007	ANSWERS	ANNOTATIONS
<b>Q5.1 In what ways do you interact with simulations in your work?</b>	<p>There are three parts to the infrastructure. There's an optimization component, a simulation component, and the resource component. The problem drives these three components.</p> <p>The simulation component is actually based on a tool from the EPA called EPANET. The code is widely used in the industry and is quite well known. We took that code and then ported it to the TeraGrid environment. We added coarse-grained parallelism around it and then coupled it with our optimization toolkit. So we run the simulation on TeraGrid resources.</p> <p>It simulates water movement and contaminant movement in urban water distribution systems. It simulates what we call "the forward problem", because given a source it'll generate observations from the sensors (wherever they're located.) The tool allows us to vary the characteristics of the source. In layman's terms, this allows us to adjust the source characteristics to match the sensor observations. Our optimization algorithm systematically adjusts the source characteristics to quickly match simulated results with the sensor observations. So it is in the simulation component that enables the adaptive behavior on the problem side. Whenever you adjust these source characteristics (inputs) the simulation produces different sets of outputs. Once you find an input that accurately produces matching results, you may have a solution. I use the word "may" because multiple inputs can produce similar results. So we also develop methods to track that.</p> <p>Our optimization component currently uses Python scripts to give us information on resource availability. But a tool is being developed that will give us better information about resource availability. It will provide us with the ability to dynamically change the resource configuration. The type of information we hope to get from the tool includes the number of computing nodes available at our sites. We have three sites right now, and at each site at any given time we will need to know the number of compute nodes that are available. Our goal is to minimize the queue wait time.</p> <p>So the tool will query these sites periodically, getting us the set of available resources from each site. And when we get the set of resources, we will try to use it to the maximum. While the simulation is running, additional resources might come on, depending on queue availability. If that happens, the tool will add the new resources to our pool of available resources. The optimization component will adaptively adjust to both the new resources and the old ones that have gone away (like when the queue time has expired.)</p> <p>It will be a persistent process. Once you start, technically you can keep using those resources for days until a time where there is not a single resource available at all sites. Then it will stop.</p>	
<b>Q6.1 Describe how you interact with data in your work</b>	<p>Input file sizes vary, depending on the network. For example a network of 11,000 nodes (considered large for a water distribution system) will have an input file of 10 megabytes, which is not very big. But then consider that you might perform multiple runs, each with slightly different input data; that would mean multiple input files. Still the storage requirements are pretty small on the input side.</p> <p>Regarding sensor data, all of the data right now is synthetic. We generate synthetic sensor observations for the optimization toolkit to feed into the simulation component. So current data requirements, in terms of the sensor readings, are not that big. They are in the kilobytes range.</p> <p>Ultimately we want to get sensor readings directly from the EPA and dynamically feed them into the whole framework. So whenever measurements are made (every 15 minutes or so) they would come in through some kind of network to our framework.</p> <p>Right now there is a sensor manufacture company that's involved peripherally on this project called Neptune. They have a Visual Basic-Excel interface that'll extract data from the sensors and then put it into an Excel file. We want to make that interface more efficient at a later point; but right now we are not working on that.</p>	
<b>Q8.1 What software do you currently use in support of your work?</b>	<p>We use the basic Globus software, which I think it's GT4.</p> <p>We use MPI at the coarse-grained parallel level inside the simulation.</p> <p>The simulation code itself is written in C. So we don't have any portability issues.</p> <p>The optimization part is written in Java, so there are some issues. We are trying to make that also C so it becomes more performance-oriented.</p> <p>All the scripts that do all the communication between resources and launching the jobs are written in Python.</p> <p>We also have a graphical user interface that's written in Python.</p> <p>So we have Python, Java, C, MPI and GT4.</p>	



Interview ID=29 5 October 2007	ANSWERS	ANNOTATIONS
<b>Learning about the user's problems</b>		
<b>Q9.1 What challenges do you face today in accomplishing your work-related goals?</b>	<p>Well, there are many challenges. You know, one is maintaining an effective communication link between the project partners. That's a challenge because we have to coordinate the work and have regular meetings. The other challenge is getting the students to communicate effectively with one another. So it's primarily communication between the teams that's a major challenge.</p> <p>As far as other challenges, from the TeraGrid side we find there is sometimes a need to write our own custom scripts. To my knowledge there's no TeraGrid resource query tool that provides us with sufficient information to build adaptive behavior into our framework. So we have to write our own scripts to do that.</p> <p>Also we sometimes encounter older software versions on TeraGrid. Unlike some TeraGrid users, we generally want the latest versions. But the latest updates are not always available, so sometimes we need to install the software ourselves in user directories. We've encountered this in the past with MPI2 and Java.</p>	
<b>Learning about the Globus user experience</b>		
<b>Q17 What are the major challenges you face using &lt;component&gt; today?</b>	<p>I think Globus is a very good product, even though I don't care for some of the security things because they hinder some of our work. But it's still a good product.</p> <p>There are some issues for applications that require co-scheduling. That is a big problem in Globus, because we need to reserve resources before running things. But I think the main problem that I have with Globus is the lack of ability to change resources while things are running. Everything depends on RSL scripts. We have multiple servers running at the same time, and there is no way using the RSL scripts for us to change resources while things are running.</p> <p>So the way we are doing it, we are using Python scripts and files to communicate, rather than depending on those more efficient things because that is the most portable way we've found. So basically, we move files, start a new job, and then everything is independent. It's just we have a script that's monitoring the progress of different jobs that are running.</p> <p>Within a user's space, they don't allow you to – at least as far as I know – change resources while things are running. So pretty much once something starts running, that's it. And then if you want to start a new one, you have to submit a new RSL script with these multiple resources. So you pretty much have to submit new things every time, rather than some way of manipulating within the job that's running.</p> <p>Let's say on the application side we would like to handle dynamically changing resources. It would be nice if I could define a job script saying I want to run a job using between 4 and 200 processors. So the job starts running with four processors. Now whenever new resources come along, I would ideally have a mechanism within the application that tells me when sixty additional processors become available. So it's not like you you're specifying a fixed number of resources in your RSL script. That means you are stuck to those resources. Currently if you want to change resources you have to submit a new RSL script.</p>	

### D.30 The scriptable interfaces at various sites are not consistent

Interview ID=30 16 October 2007	ANSWERS	ANNOTATIONS
<i>Pre-interview question: Do you interact directly with Globus software in your work today?</i>	Yes	
<b>Establishing context</b>		
<b>Q1.1 Please provide a one-minute overview of your project</b>	<p>Our project focuses on the identification of contamination sources in water distribution environments. The problem scenario is one where you have a large water distribution network and in one area an intentional or accidental contamination occurs. A water distribution system is a network of pipes, junctions, tanks, reservoirs, etc. If a contamination is introduced you want to identify the location of the source as quickly as possible.</p> <p>In this project we are trying to identify contaminant sources through simulation-optimization. Based on limited information we try to identify the source location to better apply remediation measures.</p> <p>As far as the computational part, we have a simulation component that does the hydraulics and water quality simulations. The simulation component itself is not very computationally intensive, but we need a large number of simulations so it adds up. We also have an optimization component that tries to guess the source location, gets some readings and goes back and forth until it finds the location. This component follows an evolutionary algorithm based approach.</p>	
<b>Q1.2 What is the project's name?</b>	Adaptive Cyberinfrastructure For Threat Management In Urban Water Distribution Systems	
<b>Q1.3 Which agency funds the project?</b>	NSF	
<b>Q1.4 What field does your project belong to?</b>	Environmental sciences	
<b>Q1.5 What is your job type?</b>	System administrator, developer, researcher	
<b>Q1.6 How long have you been a &lt;job type&gt;?</b>	1.5 years	
<b>Learning about discipline-specific goals and approach</b>		
<b>Q2.1 What are the main goals of your project?</b>	<p>The main goals of the project are to find the contamination source as quickly as possible using available computational resources from as many sites as possible. So we try to accrue as much resources as we can to solve a time-sensitive problem.</p> <p>One important component of our project is find out which sites have the most resources available, and try to offload our computations to that site. So applying computation to the science problem is one of our major goals.</p>	
<b>Q2.2 How will the success of your project be measured?</b>	By the end of the project we should have an application framework deployable on the Grid that can adaptively adjust the resource requirements to the problem requirements. The dynamic data driven part of the project comes from additional hydraulics information becoming available. In response we need to run more simulations using new parameters, putting an additional demand on the resource usage. So we have to find new sites or new resources to run those simulations. By the end of the project, we should have a framework that adjusts to the needs of the application, as well as suits the computational requirements for that.	
<b>Q2.3 What are the professional measures of success for you?</b>	We started out with a serial version of the simulation component. Midway through the project we have now parallelized the simulation. We have Grid-enabled the code so it runs on the TeraGrid at multiple sites. And we have a rudimentary mechanism for adaptively picking resources from various sites and driving the simulation through a controller interface. Master's thesis has been on this, so one good sign is that I have successfully defended my Master's thesis. That's one measure of professional achievement I guess. My dissertation is also closely related to this, so even from outside the project itself I can measure my progress through how it is fairing with my dissertation.	

Interview ID=30 16 October 2007	ANSWERS	ANNOTATIONS
Q3 What are you investigating?	<p>I'm working on how to find available resources quickly and use available resources to the maximum capacity possible.</p> <p>Part of the problem is we wanted to use as much of the system as we can right at that instant. We want to start the computation right away. The schedulers on the machines have a backfill window. If you request a number of processes that equals the available backfill window for calculation, your job will get scheduled right away. So that means that your computation will start very fast and you don't have to wait in the queue for a long time. So we want to minimize the queue waiting times for our jobs and try to use as many resources at as many different sites as possible to fill our computational needs. But the scriptable interfaces at various sites are not consistent, so our scripts to interact with the resources need to be site specific. If you want to get the backfill resources available at site A, you have to have a special script to talk with it. So one challenge is that the scriptable interfaces to the queuing systems at various sites are incompatible. And the Globus Toolkit functionality is supposed to hide that, but we've still encountered incompatibilities. This was six-eight months ago that we had to write our own custom scripts to do this task.</p> <p>At that time we tried using Globus job management services at various sites. But the word from one of our collaborators within Globus at that time was that the deployment scenario for certain Globus components on TeraGrid was not on track. So, at that point, we had to resort to writing custom scripts. The scripts went to particular schedulers and not the Globus gateways. And they would just state the input files and then launch the jobs using parameters that would minimize queue wait time.</p> <p>An offshoot of the data-driven part of the project is that we will never know what resource requirements might suddenly present themselves. We have to do some sort of resource requirement forming, but that may change a lot. So we might have a maximum limit and a minimum limit, but we'll be swinging back and forth between those two. So when the hydraulics information kicks in we will need to do adaptive resource management. So that is also one of our challenges. Right now, for example, we may be running on 64 processors, but we need additional computation – we need 128 processors. So when we start a run, we may discover during execution that more resources are needed.</p> <p>We would like to request those and immediately scale the computation. Right now we don't have that functionality. We can't dynamically increase the allocation of a running job. We currently use files to communicate the new requirement and restart the computation. We would like to get away from this cumbersome file-based communication.</p> <p>But right now get the information that we need from the optimization framework and start a bigger job with updated parameters. It works kind of like a checkpoint and restart. But in this case we are writing in order to generate the next set of parameters. So the smaller job finishes, and then the optimization framework generates the new parameters for the bigger job. So right now we can end up submitting jobs multiple times. And that can introduce more queue wait time too.</p> <p>So we have three areas of investigation on the computation side: finding available resources quickly, using available resources to their capacity, and adaptive resources management where we allocate more processors dynamically for the simulation runs. We do the simulations runs only on the TeraGrid, the optimization framework is typically on our local cluster. We have also tried runs using an optimization framework on one site and the simulation spread out across other sites. But the optimization framework is not parallelized. It's a serial java code that just runs on one processor.</p>	

Interview ID=30 16 October 2007	ANSWERS	ANNOTATIONS
<b>Q4.1 What is your method for investigating &lt;phenomena&gt;?</b>	<p>As far as finding available resources for immediate use, the initial strategy is to identify potential sites with minimal queue wait times and a goodly amount of resources. So at the NCSA site, for example, some days it's good and some days it's so busy you don't want to use it. Our approach is to choose a few candidate sites that can be potentially used for a long time. So after selecting those sites I log into each site and see if there are any inconsistencies in the queuing interface or whether our customized scripts will work on that platform.</p> <p>One thing we observe is that instead of using Globus url copy, sometimes using just <i>scp</i> is fast enough for us because Globus increases the latency, sometimes by a factor of two or three. So we are better off copying the files using <i>scp</i> instead of going through Globus for small files.</p> <p>So identifying potential sites, preparing the sites for running the jobs and then refining the scripts. We have a framework that we have given to collaborators who are working with the Java CoG Kit. They will take our scripts, which are currently hardcoded to work at two sites, and will build to make the functionality work in a universal way on TeraGrid.</p> <p>As far as addressing the problem of using available resources to their capacity, at every site we query the resource scheduler about the backfill window using the <i>showbf</i> command (on PBS system.) This shows us the maximum amount of resources available right at that time and the amount of time they are available. This tells you, for instance, that 32 processes are available for the next two hours; the job can start immediately. There are no guarantees, but this is the best source of information we have.</p> <p>Also we use a command supported by the PBS job scheduling system called <i>showstart</i>. If you specify a processor count it will give you, based on the current scheduled requests, an estimated start time of the job. We query based on the requirements of our application. Let's say we say we need 64 processors. We will try to see if 64 processors available at a given site within a reasonable interval. "Using the resources to the fullest" really means, "using all available resources within our time constraints." If not, then we go down in our processor count (to, say, 32) and see if that is available. If it is, we will use that and also look for another site can take the remaining 32 jobs.</p> <p>So job partitioning is also taken care of by us. Every input file to the simulation contains parameters per line and you would have the parameters for 10,000 simulations. When the job is split between 32 processors each on site A and site B, then we split that input such that they contain 5,000 parameters each. There are some common parameters in the input file, so we preserve the common parameters and we split up the problem parameters for site A and site B separated using that script. So we manage the job partitioning by manipulating the original input file into two input files.</p> <p>We have some rudimentary batch scripts as well as some small Python programs. The one that does partitioning of the input file is actually a Python program. It's not just dividing it in half. You can do a more intelligent thing. Say we are looking for 128 processors; we can have 32 one at site A, and the rest on site B. We divide things in a way that equals the number of resources available. So you may have three-fourths of the simulations offloaded to site B and one-fourth to site A. You can specify how much each fragment of the input file will contain.</p> <p>As far as time, an individual simulation will be like 20 or 30 seconds. But we can run thousands of simulations as part of a job. So a job may run for a few hours.</p>	
<b>Q4.5 How do you document your results?</b>	<p>I created a project website for this and an associated wiki. I try to document the structure of the input files, the partitioning scripts, and how they are used to look for the patterns. Using the wiki simplifies things.</p> <p>I intend to post most of the results also on the internal project website for our group meetings every week.</p>	
<b>Q5.2 How do you share simulations with others?</b>	<p>The simulations run under my user name. The custom scripts use my user public ssh keys. But we have plans to move to a more refined solution. That's where our work with the Java CoG Kit comes in. So right now it runs under my userid. The goal is to get at least all the project team members to be able to run it, I guess.</p> <p>Honestly I don't have high hopes for certificates. They may provide better security, but they bring the performance down a lot. So I go with the idea that maybe for eight or ten users, we can integrate the ssh keys into the tool itself. This is a focus of the Java CoG work: to provide the ability to import the user's public key and use this to launch the jobs.</p>	

Interview ID=30 16 October 2007	ANSWERS	ANNOTATIONS
<b>Q6.1 Describe how you interact with data in your work</b>	<p>The optimization framework writes out the problem parameters, and that file has to be copied over to the simulation component. Then the simulation component reads the data and writes an output file, which is then copied back to the optimization toolkit. So the primary interaction with data is this transfer of files across. This is where we must support interactions between multiple sites.</p> <p>We have multiple options for addressing this issue. We can install our public/private key pair and use <i>scp</i> to move around the files. Or we can use the Globus GSIFTP to move it back and forth between sites.</p> <p>Some of the time our optimization module is deployed outside of TeraGrid. I can be deployed on our local cluster, in which case we use <i>scp</i> from our site to the TeraGrid site to copy those files. So we are flexible, in that we have multiple options in copying those files. I think that's the primary data interaction that we have. And the data files we are talking about are small files – only a few megabytes. So even the simulation doesn't generate large data files.</p> <p>The result of the simulation that is most important in our case. The contamination location is something that people need to see. So those results are documented. And some of the output files are also tracked. I have a visualization tool that tracks the passage of the simulation and show the current best estimate of the contamination location. We eventually will build a model, and the visualization tool will show that. So to generate that sort of visualization we may need intermediate files.</p>	
<b>Q7.1 What resources do you use in your work today?</b>	<p>There is the optimization cluster at my home institution and the TeraGrid for running the simulations.</p> <p>The second phase of the project we are supposed to get some real sensor information from a database of the EPA, but we are not at that stage yet.</p>	
<b>Q7.3 By what mechanisms is access to your work-related resources controlled?</b>	We leave it to job manager, using the custom scripts I wrote that interact with the job manager. But we are looking for that functionality to be present in a usable way using the Globus interface if possible.	
<b>Q7.5 What types of information do you need to know about a resource in order to determine if it is suitable for your work?</b>	The simulation component is ANSI C, so we we can run it anywhere. We are able to run it on every architecture we have seen so far. We would prefer resources that are fast.	
<b>Q8.1 What software do you currently use in support of your work?</b>	<p>The main simulation code is in ANSI C.</p> <p>We are porting the optimization component from Java 1.5 to Java 1.4. The reason for this is because when we tried to run the optimization module on some TeraGrid sites, Java 1.5 was not available. So we had to backport it. A post-doc wrote that component, so after week I came into the picture and had to back port it to Java 1.4.</p> <p>So now the code is written both in Java and C. And now we are even working on porting the Java components completely to C so as to improve the performance of the module. So we will have the simulation and optimization modules in C, primarily.</p> <p>The scripts and the glue code are written in Python and bash shell scripts.</p> <p>So C, Java, Python and bash.</p>	
<b>Q8.4 What workflow tools do you use in your work?</b>	The Java CoG Kit has a workflow tool called Karajan that is supposed to provide the workflow management for us, so all these functionalities should be integrated into the CoG Kit. We also investigated Kepler, but never deployed anything. We will be using the CoG Kit in our final deployment.	
<b>Q8.5 What parallel computing tools do you use in your work?</b>	The simulation component uses MPI.	
<b>Q8.6 If the need for new software-based functionality arises in your work how do you acquire it?</b>	<p>On our local site we purchased a large-scale compiler suite for our development work. We also experimented with MATLAB distributed computing engine. So we have a version of this, not on the Web – just on our local site – just to try running our computations. So we use MATLAB for parallel computing. But that is mainly used on our development cluster rather than TeraGrid.</p> <p>Most of the functionality we need, we have to write it ourselves. If we need system-level functionality or scheduler functionality, we try to see if that exists already. And if not, we have to resort to writing custom code for doing it.</p>	
<b>Q8.7 How do you share software with others?</b>	So the custom scripts that we have are application-specific mostly. So the strategy that we use in there was for presenting papers. The code is too specific to our application to be of use to other projects.	
<b>Learning about the user's problems</b>		

Interview ID=30 16 October 2007	ANSWERS	ANNOTATIONS
<b>Q9.1 What challenges do you face today in accomplishing your work-related goals?</b>	<p>Incompatibility between the sites is the biggest challenge. Whatever we do, we have to make sure that it works on every site. So we have to do the manual listing ourselves. So if there were a compatibility layer that ensures the resource allocation mechanisms work on all sites as expected, that would eliminate a lot of testing on our part.</p> <p>The problems I experienced with the Globus job submission mechanism [GRAM4] happened around a year ago, so some of the information may be dated. But after that experience we are waiting for the word from the CoG Kit group to give us the go ahead to try Globus again. But at that time they told us they were not able to submit the jobs properly and were having certificate issues and so forth. So when we talked to the CoG Kit folks, they say that the TeraGrid deployment schedule is delayed so we need to wait another six months before that stuff becomes available on all sites. So they were acknowledging the incompatibilities at that point.</p> <p>Another problem is that sometimes we even have to investigate what is the best strategy for communication. We are using file-based communication and there are more than a few methods to transfer files between sites. We need to investigate which one works for us because we tend to have smaller files, so latency rather than bandwidth is an issue for us. Even on TeraGrid we have <i>tgcp</i>, <i>gsi-ftp</i>, <i>scp</i>, and a couple of other file transfer mechanisms. So we have to pick and choose which mechanism works best between the sites that are available to us. And I don't if there is an easy way out from that. In some places <i>scp</i> will be the best way and in other places <i>globus-url-copy</i> will be the best.</p> <p>Another issue:</p> <p>We had to resort to this file-based communication because we can't directly communicate through a running job. We would like to stream in new input into a running process. My application really needs that. The simulation component produces results that feed into the optimization framework. We had to resort to using files to communicate between the two because we can't directly communicate with the running process. We are exploring ways to eliminate file based communication wherever possible.</p> <p>The MPI2 functionality, like connection sockets and dynamic process management and spawning new processes: this functionality would be useful to us. But right now it is not available on many TeraGrid sites. For security reasons, no compute node is allowed to communicate with outside machines as far as I know.</p>	
<b>Q9.2 What types of information do you need in order to address the challenges you face today?</b>	If the MPI2 functionality was available and deployed on all the sites, and if Globus worked as advertised that would ease our lives a bit.	
<b>Q9.4 By contrast, can you provide examples of technologies you find very useful today?</b>	Wow, that's a tough one. HPC has a dearth of software. I think if Globus does what it does without degrading the performance. I think that would be a good model. But right now, it doesn't do as much as it should do and there is a big performance penalty. So my ideal product would be something that provides the Globus functionality without the big performance penalty. [see answer to Q17 for more information on the performance issue]	
<b>Q10.1 Can you think of any work-related tasks that decrease your productivity?</b>	For repetitive tasks we write shell scripts, if it can be scripted. So I try to avoid repetitive tasks by scripting them away. That's at a computational level. At the macro level, we would like to have the workflow system could capture some experiment, such that if you wanted to repeat it you could just push a button or just run one command and rerun the experiment. So that is functionality we are looking for from the workflow. That is the idea to be implemented with help from collaborators within our project.	
<b>Learning about the Globus user experience</b>		
<b>Q16 Why do you use &lt;component&gt; instead of an alternative technology?</b>	<p>GridFTP:</p> <p>We are just trying to find the best way to move our files across multiple sites, so whatever works is fine with me. GSI-FTP seems to work well between ANL's TeraGrid site and NCSA's TeraGrid site.</p> <p>Performance varies though, even at different times in the day. For some trials we are able to get better performance just using <i>scp</i> instead of GSI-FTP and other commands. It just depends on the load on the Globus service, I guess. So we had to model even that, and that's an extraordinary level of information that we don't need to deal with.</p>	

Interview ID=30 16 October 2007	ANSWERS	ANNOTATIONS
<b>Q17 What are the major challenges you face using &lt;component&gt; today?</b>	<p>Security:</p> <p>The issue with security is that we are not able to directly communicate to a running a process due to security concerns. In order for us to send in the next set of parameters to the simulation component, we have to copy a file from the optimization framework through there. We can't connect directly through sockets. No compute node is allowed to communicate with outside nodes.</p> <p>So the bad performance I alluded to earlier is due to having to transmit data via files. We should be able to stream the information directly to the compute node rather than resorting to writing our stuff to a file, copying the file, and then redoing this over and over. And though they are small files, the latency itself is a killer for us. So it's more that the performance is bad because of the file-based communication, not having to do with the technology itself.</p> <p>GRAM:</p> <p>I think the CoG Kit folks maybe have a better idea what to say here, because one year ago they couldn't get GRAM4 to work on all the TeraGrid sites we had, and then I didn't try it yet for this project.</p>	
<b>Wrapping-up</b>		
<b>Is there anything you'd like to say to the people who build software for use by people like you?</b>	<p>I think that I'm not a typical representative of the distributed computing community because I come from the HPC community. So I put a premium on performance, whereas the stuff that has been coming out of the distributed computing community has not focusing on performance. So that's one area where I would like to see some improvement. I would like to see those things perform well and with less bloat. Globus GSI-FTP and others need to do the authentication through using the certificates. They tend to take a longer time. I think people might say six seconds is not a big latency, but when you have many interactions, six seconds adds a lot to that. We don't say six seconds is not bad at the MPI level.</p> <p>I understand there are technology challenges but I think there should be less cumbersome methods for authenticating the requests.</p>	

**About Argonne National Laboratory**

Argonne is a U.S. Department of Energy laboratory managed by UChicago Argonne, LLC, under contract DE-AC02-06CH11357. The Laboratory's main facility is outside Chicago, at 9700 South Cass Avenue, Argonne, Illinois 60439. For information about Argonne, see [www.anl.gov](http://www.anl.gov).

**Availability of This Report**

This report is available, at no cost, at <http://www.osti.gov/bridge>. It is also available on paper to the U.S. Department of Energy and its contractors, for a processing fee, from:

U.S. Department of Energy  
Office of Scientific and Technical Information  
P.O. Box 62  
Oak Ridge, TN 37831-0062  
Phone (865) 576-8401  
Fax (865) 576-5728  
[reports@adonis.osti.gov](mailto:reports@adonis.osti.gov)

**Disclaimer**

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor UChicago Argonne, LLC, nor any of their employees or officers, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of document authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof, Argonne National Laboratory, or UChicago Argonne, LLC.