OGSA-DAI: Two Years On

Mario Antonioletti[1], Malcolm Atkinson[2], Rob Baxter[1], Andrew Borley[3], Neil Chue Hong[1], Brian Collins[3], Jonathan Davies[3], Neil Hardman[3], George Hicken[3], Ally Hume[1], Mike Jackson[1], Amrey Krause[1], Simon Laws[3], James Magowan[3], Jeremy Nowell[4], Norman W Paton[5], Dave Pearson[6], Tom Sugden[1], Paul Watson[7] and Martin Westhead[1]

1. EPCC, University of Edinburgh, JCMB, The King's Buildings, Mayfield Road, Edinburgh EH9 3JZ, UK.
2. National e-Science Centre, Universities of Edinburgh & Glasgow, Edinburgh EH8 9AA, UK.
3. IBM United Kingdom Ltd, Hursley Park, Winchester S021 2JN, UK.
4. UK Grid Support Centre/EPCC, JCMB, The King's Buildings, Mayfield Road, Edinburgh EH9 3JZ, UK.
5. Department of Computer Science, University of Manchester, Oxford Road, Manchester M13 9PL, UK.
6. Oracle UK Ltd, Thames Valley Park, Reading RG6 1RA, UK.
7. School of Computing Science, University of Newcastle upon Tyne, Newcastle upon Tyne NE1 7RU, UK.

The OGSA-DAI project has been producing Grid-enabled middleware for almost two years now, providing data access and integration capabilities to data resources, such as databases, within an OGSA context. In these two years, OGSA-DAI has been tracking rapidly evolving standards, managing changes in software dependencies, contributing to the standardisation process and liasing with a growing user community together with their associated data requirements. This process has imparted important lessons and raised a number of issues that need to be addressed if a middleware product is to be widely adopted. This paper examines the experiences of OGSA-DAI in implementing proposed standards, the likely impact that the still-evolving standards landscape will have on future implementations and how these affect uptake of the software. The paper also examines the gathering of requirements from and engagement with the Grid community, the difficulties of defining a process for the management and publishing of metadata, and whether relevant standards can be implemented in an efficient manner. The OGSA-DAI software distribution and more details about the project are available from the project Web site at http://www.ogsadai.org.uk/.

Background

The *Open Grid Services Architecture – Data Access and Integration* project started in February 2002 as a collaboration between various UK academic institutions and industrial partners. The project is funded and motivated by the UK e-Science community, whose project requirements contribute to the goals that OGSA-DAI aims to satisfy. OGSA-DAI has produced several releases of Grid-enabled middleware implementing DAIS (*Data Access and Integration Services*) specifications. This middleware is based on the GGF-defined OGSI specification [OGSI] and layered on top of the Globus Toolkit 3 OGSI implementation (GT3 Core). Now that OGSI has effectively been deprecated, OGSA-DAI will migrate to take account of changes in the Web and Grid services standards space, to allow access to and integration of data resources in an OGSA Grid environment. By a data resource we mean any object that can source and/or sink data and its associated management framework. The GGF DAIS Working Group, authors of the DAIS specifications, currently restrict their scope to relational and XML database systems. There is exploratory work underway within DAIS to incorporate data access for files. OGSA-DAI has conducted preliminary work to provide file access using BinX[1] and other methods. As part of work to refactor OGSA-DAI, the project will investigate using the current WS-Resource standards [WS-ResourceFramework], [WS-ResourceLifetime], [WS-Notification] and [WS-ResourceProperties] (as well as those others that are scheduled to be released in the near future) which future releases of the Globus Toolkit will support. OGSA-DAI will also consider other emerging standards.

OGSA-DAI also intends to deliver a reference implementation of the DAIS Working Group specifications, which are intended to standardise the interfaces required to access data

[1] See http://www.edikt.org/binx.

resources within a Grid context [GDSS], [GDRR], [GDXR]. Once DAIS becomes a proposed GGF recommendation, OGSA-DAI aims to produce one of the two inter-operable implementations required for the GGF DAIS standards to become a full GGF recommendation[2].

In essence, OGSA-DAI is faced with a problem in trying to comply with a moving set of standards while at the same time trying to maintain some backwards compatibility with previous OGSA-DAI software releases which are already being used by established projects. Attempting to encourage projects to adopt OGSA-DAI is in direct conflict with trying to comply with standards that are rapidly evolving and have not stabilised. For now, a middle path has been adopted which, although not entirely satisfactory, allows progress to be made. OGSA-DAI currently implements an former version of a draft GGF DAIS specification and a client toolkit library is under development that will shield technology adopters from the eventual migration of OGSA-DAI to comply with stable DAIS and Grid specifications.

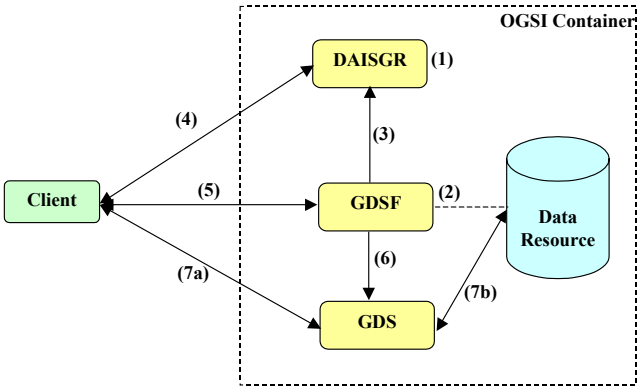
OGSA-DAI Overview

OGSA-DAI currently provides three types of Grid services: these provide for data resource registration and discovery, creating Grid proxies for data resources, and data access and manipulation. The current base OGSA-DAI services are:

- *Data Access and Integration Service Group Registry* (DAISGR) – allows data resources that are represented by services to be registered and discovered. A client can use a DAISGR to identify a resource provider that best meets its requirements.
- *Grid Data Service Factory* (GDSF) – acts as a persistent access point to a data resource and contains additional related metadata that may not be available in the DAISGR.
- *Grid Data Service* (GDS) – acts as a transient access point to a data resource. It is through a GDS that a client interacts with a data resource. GDSs are created by a GDSF.

The GDS and GDSF are based on services that were specified in the GGF7 DAIS draft specification (March 2003). The DAISGR is not specified by DAIS but is based on interfaces that are specified within OGSI. OGSA-DAI will be re-aligned with DAIS once its specifications stabilise – the hope is that this will happen soon after GGF11 (June 2004).

A typical OGSA-DAI usage pattern is represented in the figure below.



[2] OGSA-DAI and DAIS are not currently aligned. Alignment will occur when the DAIS specification is complete.

One or more persistent DAISGRs may be instantiated at container start-up, (1) in the figure. Any service implementing one or more OGSA-DAI portTypes can register with a DAISGR. There is no requirement for the DAISGR to be co-located in the same container as the services that register with it. Services registering with a DAISGR may register their capabilities as well as metadata about the data resource's information content. OGSA-DAI does not prescribe the metadata to be exposed by a DAISGR but instead provides an extensible model that can be developed by a user community to meet their own data resource discovery requirements. Clients may use DAISGRs to discover services that meet their informational or capability requirements – see (4). In addition, clients may subscribe to a DAISGR to be notified of changes to the DAISGR's state.

GDSFs act as a point of presence for one particular data resource on a Grid. These are persistent services configured at container start-up (2). More than one GDSF can be used to represent the same data resource if necessary. Currently GDSFs cannot be dynamically created or configured[3]. On creation, a GDSF may register its service handle with a DAISGR, along with sufficient metadata and capability information to allow service discovery to take place (3). Clients can obtain detailed information directly from a GDSF. A GDSF, in effect, acts as a persistent Grid-enabled wrapper for a data resource but does not provide direct access to that data resource. Access requires the creation of a GDS through the GDSF's Factory portType (6).

GDSs are transient services created at the request of clients who wish to access a data resource. Data resource access is done through a single document-based operation provided by the GDS[4] (7a) (7b). A client submits a so-called perform document to the GDS, this document consisting of a sequence of activities to be executed. These activities will be executed by the GDS on the data resource and must be from the set of activities supported by the GDS. Activities expose capabilities of underlying data resources and also provide additional functionality at the service level, e.g. transformations and third party delivery. A simple data flow operates between activities and so allows clients to request the execution of a sequence of tasks, thereby avoiding both data communication overheads between client and service and the unnecessary storage of data at the client-side. In addition, the activity composition in a perform document may be optimised at the service end by the enactment engine which the GDS employs to process these documents. On the other hand, clients that wish to control things more precisely at the expense of round trips are still able to do so.

The activity framework is extensible, allowing service providers to bundle additional functionality into their GDSs that easily integrates into the existing OGSA-DAI framework. This can be done by providing an XML Schema definition of the activity's syntax, as is presented in the perform document, and specifying a Java class that implements the activity within the GDSF configuration file. The OGSA-DAI distribution documentation describes in detail how to write activities and incorporate them within the OGSA-DAI framework.

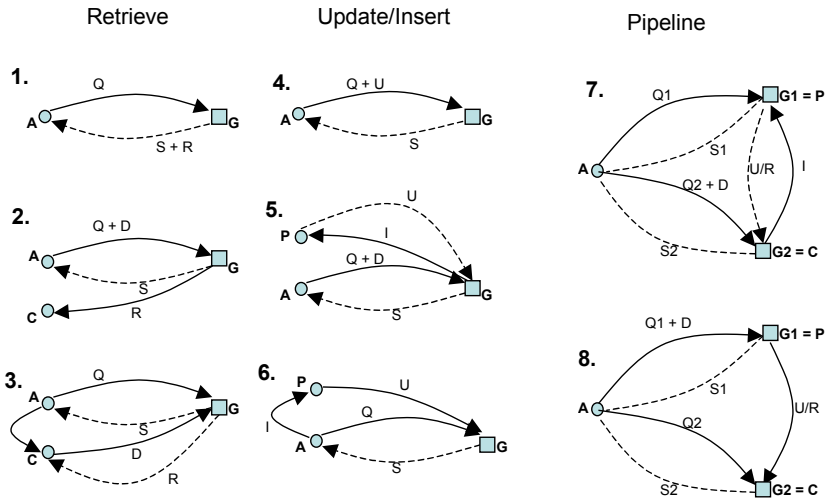
Only one base usage pattern has been described here. More sophisticated usage patterns can be generated once a GDS has been instantiated. For instance, the patterns identified by Greg Riccardi and shown in the figure below are already supported by OGSA-DAI. However, to date, most core development within OGSA-DAI has concentrated on facilitating data access. Work has been done on Distributed Query Processing[5] and it is planned to investigate more data integration scenarios in the near future.

---

[3] A more dynamic framework is desirable and may emerge from work being carried out by the DAIS Working Group.

[4] This framework, originally part of the DAIS specifications, now no longer forms part of the core specifications. The framework may, in a modified format, become a future DAIS informational document.

[5] See http://www.ogsadai.ac.uk/dqp.



Key to Symbols			
Data Flow			
→	Call	- - →	Response
Actors			
○	Non OGSA process	□	OGSA process
A	Analyst	G	GDS
C	Consumer	P	Producer
Data			
Q	Query definition	D	Delivery definition
S	Status	R	Result
U	Update data	I	Data identifier (URI)

Shifting Sands

One of the major difficulties in the development of OGSA-DAI has been the lack of concrete standards upon which to build. Convergence of the DAIS standardisation process has been slower than originally hoped. The OGSI interfaces, upon which OGSA-DAI is currently based, are now to be superseded by the proposed WS-Resource Framework standards. WS-Resource Framework is emerging and there is no indication as yet as to how rapidly it will converge. These factors introduce uncertainties in the software development process, which in turn presents difficulties in trying to persuade the user community to adopt OGSA-DAI. OGSA-DAI thus does not get as much feedback on user requirements and experiences as would be hoped.

As a strategy to isolate users from inevitable interface changes, OGSA-DAI is developing a client toolkit which, through suitable abstractions, will be able to absorb most of the changes in the infrastructure. Developers who use the client toolkit library will hopefully have some stability when OGSA-DAI begins to converge with standards. The toolkit also make the development of OGSA-DAI client applications a lot easier. It is straightforward to extend the toolkit to incorporate additional activities that have been added to the core OGSA-DAI release – in essence a Java object has to serialise itself into a valid XML fragment for the relevant activity. The client toolkit has been made available as a technical preview in release 3.1 of OGSA-DAI and will be fully supported and distributed in future releases.

Attempts are also being made to engage directly with the user community to establish requirements and ascertain their experiences in using the OGSA-DAI technology. A users group meeting is to take place in Edinburgh in April 2004 and a mini-workshop is planned for the UK e-Science All Hands Meeting in August/September 2004 where technology users will be able to describe their use of OGSA-DAI and engage directly with OGSA-DAI developers. User support has been provided through the UK Grid Support Centre to help those trying to use the technology and a user mailing list allows users to share their experiences and difficulties in using OGSA-DAI. On the whole, users have given back positive and constructive feedback to date.

OGSA-DAI believe that the key to reassuring technology adopters about the viability of using OGSA-DAI within their own infrastructure is to provide a strong directional road map for future development of OGSA-DAI, and to exploit the use of structured APIs to absorb changes in standards. In part, until the standards have stabilised and OGSA-DAI is able to run under a greater variety of platforms, OGSA-DAI is tied to changes in the Globus Toolkit.

Metadata

A particular area that has received little attention within OGSA-DAI is that of metadata. At present, OGSA-DAI publishes metadata about an underlying data resource that is exposed, its capabilities, the service capabilities and the connection infrastructure that OGSA-DAI uses to communicate with the data resource. However, the responsibility for determining the metadata describing the content of a data resource is devolved to the data publishers. The static metadata model is extensible, it is provided within a GDSF configuration file. Dynamic metadata can be generated by associating, in the GDSF configuration file, a call back routine with service data. When the service data is accessed the call back routine is used to generate its content. This is fully extensible and can be used by technology adopters. Although the role of OGSA-DAI should not be to define metadata, a metadata container framework is required to facilitate uniform service discovery across disparate application areas.

This raises a number of interesting questions: does a suitable metadata framework exist? How should this be developed to allow improved matching of registry queries with data resources? What metadata is required to allow data integration, and higher-level services to fully and automatically utilise Grid Data Services?

Efficient implementation

It is of vital importance, if OGSA-DAI is to be regarded as a success and widely adopted, that the overhead introduced by adding an extra middleware layer is minimised. Both the ISI [Benchmarking1] and Indiana University [Benchmarking2] have been invaluable in providing benchmark results based on existing applications ported to use OGSA-DAI. This, combined with our own profiling and benchmarking – see [Benchmarking3], is allowing us to investigate the scalability and performance of OGSA-DAI as well as the suitability of some of the technologies (e.g. JDBC), which are being used in this project. Improving performance and scalability is an on-going effort within the OGSA-DAI project.

Benefits of OGSA-DAI

OGSA-DAI must demonstrate the benefits and additional functionality it provides over other data resource access technologies, for example direct JDBC connections. OGSA-DAI provides for third party delivery and server-side transformations and has an extensible framework to support the addition of functionality at the service end. This is in addition to being able to map Grid credentials to database roles. OGSA-DAI is designed to operate in a dynamic framework where resources can be registered and discovered using application-

specific metadata. Clients can transparently obtain data from a data resource unknown to them at the start of their session. A homogenous mode of operation is employed across heterogeneous data resources and technologies. OGSA-DAI provides wrappers that can facilitate the creation of higher-level services or middleware which exploits OGSA-DAI and the data resource capabilities it exposes.

In addition, as OGSA-DAI is currently the main test platform for the DAIS standards, it is important that feedback on the suitability of the specifications for efficient data access and transport is gained. If DAIS is judged a success it is expected that database vendors will directly embed DAIS interfaces within their products. In this scenario, OGSA-DAI would employ these directly and provide higher-level functionality.

Current Status

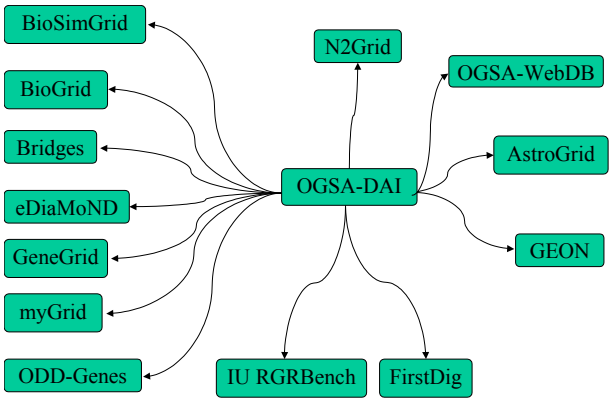
The project history, current release, planned releases, user documentation, tutorials and user experiences can be found at http://www.ogsadai.org.uk.

There have been three major and four minor releases of the OGSA-DAI distribution to date.

Release	Release Date
Release 4	**April 2004**
Release 3.1	09/02/04
Release 3.0.2	26/09/03
Release 3	**31/07/03**
Release 2 interim	11/06/03
Release 2	**15/04/03**
Release 1 interim	28/02/03
Release 1	**15/01/03**

OGSA-DAI currently works under the Globus Toolkit 3 OGSI implementation. There have been over 1700 downloads of the various OGSA-DAI distributions from the OGSA-DAI Web site.

Currently 13 projects, illustrated below, provided details of how they are using, evaluating or have implemented OGSA-DAI within their projects.



More details about these projects can be found in the Appendix.

The development of the OGSA-DAI software is currently funded until October 2005, and major releases are expected on a six-monthly cycle, with release 4 due in April 2004. OGSA-DAI is being distributed as part of the Globus Toolkit, starting as a technical preview component in Globus Toolkit 3.2. OGSA-DAI will continue to be released as part of the UK's e-Science Programme with specific versions being contributed to future Globus Toolkit releases.

Conclusions

In this paper, the current status of the OGSA-DAI project has been described. Salient points, in no particular order, that have emerged though the development of OGSA-DAI have been:

- The importance of engaging the user community to ensure that base functionality meets their requirements and that their opinions are fed back to the standardisation process.
- It is important that OGSA-DAI does not introduce an unacceptable overhead over the use of direct data resource connection methods and that OGSA-DAI provides added-value over these other methods.
- Usability of middleware is of critical importance. OGSA-DAI has endeavoured to produce good documentation to enable users to intelligently evaluate OGSA-DAI. Attempts have been and are being made to simplify the OGSA-DAI installation and configuration process via additional tooling.
- It is important to track and contribute to the standards process. Since the project inception, OGSA-DAI has been in a field that is in an extreme state of flux: the OGSI specification went through a large number of iteration cycles as it gradually converged, in turn reflected by the Globus Toolkit, and thus impacted upon OGSA-DAI. DAIS is still evolving too fast for OGSA-DAI to effectively track it in its releases and now a migration from OGSI to WS-Resource Framework may be required.
- It is important not to alienate the emerging user community. Changes to standards will be hidden behind a client toolkit interface, which will minimise disruption at the client-side. However, change will still be felt at the service-side. If this technology is to be widely and confidently used then a period of stability is required.
- A concerted community-wide effort is required to define metadata containers to aid resource discovery. This task is bigger than OGSA-DAI but one to which it could contribute.
- OGSA-DAI has been successful in data access roles. Some work has been done in data integration scenarios, e.g. Distributed Query Processing[6], but more work needs to be done in this area.

This work has been funded by the UK e-Science Grid Core Programme whose support we are pleased to acknowledge.

References

[Benchmarking1]
E. Deelman, G. Singh, M.P.Atkinson, A. Chervenek, N.P. Chue Hong, C. Kesselman, S. Patil, L. Pearlman, M-H. Sui. *Grid-Based Metadata Services*. Submitted to 16th

[6] See http://www.ogsadai.org.uk/dqp for details.

International Conference on Scientific and Statistical Database Management, June 2004.

[Benchmarking2]
D. Kodeboyina, B. Plale. *Experiences with OGSA-DAI: Portlet Access and Benchmark*. Presented at the Designing and Building Grid Services Workshop at GGF9.

[Benchmarking3]
M. Jackson, M. Antonioletti, N. Chue Hong, A. Hume, A. Krause, T. Sugden and M. Westhead. *Benchmarking of Grid Data Service implementations within the OGSA-DAI project*. Paper submitted to HPD-13.

[GDSS]
M. Antonioletti, M. Atkinson, A. Krause, S. Malaika, S. Laws, N. W. Paton D. Pearson and G. Riccardi. *Grid Data Service Specification*. DAIS-WG Informational Draft.

[GDRR]
M. Antonioletti, A. Krause, S. Hastings, S. Langella, S. Malaika, J. Magowan, S. Laws, N. W. Paton. *Grid Data Service Specification: The Relational Realisation.* DAIS-WG Informational Draft.

[GDXR]
A. Krause, M. Antonioletti, S. Hastings, S. Langella, S. Malaika, S. Laws, N. W. Paton. *Grid Data Service Specification: The XML Realisation* DAIS-WG Informational Draft.

[OGSI]
S. Tuecke, K. Czajkowski, I. Foster, J. Frey, S. Graham, C. Kesselman, D. Snelling, P. Vanderpilt, Open Grid Services Infrastructure, Version 1.0, http://www.gridforum.org/ogsi-wg, March 13, 2003.

[WS-ResourceFramework]
I. Foster (Editor), J. Frey (Editor), S. Graham (Editor), S. Tuecke (Editor), K. Czajkowski, D. Ferguson, F. Leymann, M. Nally, T. Storey, W. Vambenepe, S. Weerawaranna. *Modeling Stateful Resources with Web Services*. Version 1.0. 20th January 2004. Available from: http://www.globusalliance.org/wsrf.

[ResourceFrameworkRefactoring]
K. Czajkowski, D. Ferguson, I. Foster, J. Frey, S. Graham, T. Maguire, D. Snelling, S. Tuecke. *Resource Framework: Refactoring & Evolution.* Version 1.0. 12th February 2004. Available from: http://www.globusalliance.org/wsrf.

[WS-ResourceLifetime]
J. Frey (Editor), S. Graham (Editor), K. Czajkowski, D. F. Ferguson, I. Foster, F. Leymann, T. Maguire, N. Nagaratnam, M. Nally, T. Storey, S. Tuecke, W. Vambenepe, S. Weerawarana. *Web Services Resource Lifetime (WS-ResourceLifetime)*. Version 1.0. 20th January 2004. Available from: http://www.globusalliance.org/wsrf.

[WS-Notification]
S. Graham (Editor), P. Niblett (Editor), D. Chappell, A Lewis, N. Nagaratnam, K. Parijkh, S. Patil, S Samdarshi, S. Tuecke, W. Vambenepe, B. Weihi. *Web Services Notification (WS-Notification)*. Version 1.0. 20th January 2004. Available from: http://www.globusalliance.org/wsrf.

[WS-ResourceProperties]
S. Graham (Editor), K. Czajkowski, D. F. Ferguson, I. Foster, J. Frey, F. Leymann, T. Maguire, N. Nagaratnam, M. Nally, T. Storey, W. Vambenepe, S. Tuecke, S. Weerawarana. *Web Services Resource Properties (WS-ResourceProperties)*. Version 1.0. 20th January 2004. Available from: http://www.globusalliance.org/wsrf.

Appendix

A list of projects which have volunteered information about how they are using, have used or are evaluating OGSA-DAI as part of their infrastructure is included below. Up to date information about OGSA-DAI usage is included at: http://www.ogsadai.org.uk/projects.

- **AstroGrid** (http://www.astrogrid.org/)
  AstroGrid is using OGSA-DAI to build the prototype of a Grid data warehouse in which large data extracts from many data centres can be combined for easier analysis.
- **BioGrid** (http://www.biogrid.jp/)
  The BioGrid project has developed a system for federating bio-related databases with Globus Toolkit 3.0.2 and OGSA-DAI, aiming at an application to Drug Discovery. The system has currently bridged 11 databases in heterogeneous communities; such as biology, medical science and pharmaceutics.
- **BioSimGrid** (http://www.biosimgrid.org/)
  A pilot project that aims to enable large-scale analysis of the results of biomolecular simulations. In particular the project will establish generic procedures for comparative analysis of simulations of biomolecules of pharmaceutical interest and integrate simulation data with those emerging from post-genomic approaches to structural biology. To address this challenge of distributed computing on large amounts of simulation data (60 billion of records in over 2 TeraBytes storage for 1000 trajectories), the BioSimGrid project is building an open software framework system based on OGSA and OGSA-DAI.
- **BRIDGES** (http://www.brc.dcs.gla.ac.uk/projects/bridges/)
  BRIDGES aims to incrementally develop and explore database integration over six geographically distributed research sites within the framework of the large Wellcome Trust biomedical research project Cardiovascular Functional Genomics. Three classes of integration will be developed to support a sophisticated bioinformatics infrastructure supporting: data sources (both public and project generated), bioinformatics analysis and visualisation tools, and research activities combining shared and private data. The inclusion of patient records and animal experiment data means that privacy and access control are particular concerns. An exploration of index factories accelerating sequence processing will test the hypothesis that the Grid makes a new class of e-Science indexes feasible. Both OGSA-DAI and IBM DiscoveryLink technology will be employed and a report will identify how each performed in this context.
- **eDiaMoND** (http://www.ediamond.ox.ac.uk/)
  eDiaMoND is using the Grid for X-Ray Mammography. The mammograms stored by eDiaMoND are large binary objects, typically 32Mb each, and are held in a Content Management system. Patient data and image data are stored in a corresponding relational database. OGSA-DAI services are used to abstract both of these data resources. In order to achieve this the project is extending OGSA-DAI to include a wrapper for IBM Content Manager. The architecture for eDiaMoND is built on top of the abstraction layer OGSA-DAI offers. Underneath this layer a number of options for virtualizing the mammography data will be explored: OGSA-DAI has support for data replication and distributed query processing. Alternatively OGSA-DAI can wrap data resources that are federated or replicated through the underlying data management systems. The performance and scalability characteristics of these options can be assessed and matched to the data flow requirements of different application scenarios. Peer to peer, hierarchical and hub and spoke models of breast screening clinics can be explored. The common feature in all cases is the abstraction of mammography data as an OGSA-DAI data resource in a Grid.
- **FirstDIG** (http://www.epcc.ed.ac.uk/~firstdig/)
  The FirstDIG project was a collaboration between First plc, the passenger transport company, and the National e-Science Centre represented by EPCC. The project involved the deployment of OGSA-DAI services within the First South Yorkshire bus operational environment. These services were used to access heterogeneous databases within the organisation. This deployment allowed specific business questions posed by First plc to be answered through a short data mining analysis of the OGSA-DAI service-enabled data

sources. The project successfully demonstrated the deployment and utilisation of OGSA-DAI services in a commercial environment.

- **GeneGrid** (http://www.qub.ac.uk/escience/projects.php#genegrid)
  The GeneGrid project proposes to exploit Grid, existing micro array & sequencing technologies together with the specialist tissue specific datasets for particular disease types being generated from the large data volumes produced through screening services so that all the genes that relate to a disease can be studied insilco. This will be enabled by the creation of a Grid-based framework that will integrate the generation and analysis of cancer and infectious disease specific genetic information from various distributed international sources, public domain data sets and other unique data generated by e-Science projects. One of the important areas this project will have to examine is data integration, and that is what we are hoping to use OGSA-DAI for. OGSA-DAI allows the easy integration of the likes of workflow definitions and configuration data. However, GeneGrid will require integration of a lot of Flat File Format databases that OGSA-DAI does not support yet. Currently examining the potential of expanding the OGSA-DAI functionality to incorporate flat file format database support specific to the needs of GeneGrid project.
- **GEON** (http://www.geongrid.org/)
  GEON IT research focuses on modelling, indexing, semantic mediation, and visualization of multi-scale 4D data, and creation of a prototype GEONgrid, to provide the geoscience community an IT head start in facing the research challenges posed by understanding the complex dynamics of Earth systems. An important contribution will be embarking on the definition of a Unified Geosciences Language System (UGLS), to enable semantic interoperability. The GEONgrid leverages experience gained in the National Partnership for Advanced Computational Infrastructure (NPACI) program, and the TeraGrid Distributed Terascale Facility. We will create a portal to provide access to the GEON environment, which will include advanced query interfaces to distributed, semantically-integrated databases, Web-enabled access to shared tools, and seamless access to distributed computational, storage, and visualization resources and data archives. OGSA-DAI is part of the GEON systems software stack and will be deployed on all the GEON nodes. OGSA-DAI is currently being used to access data from databases running on remote GEON nodes. We are also in the process of evaluating the performance of OGSA-DAI and determine how it fits in the overall GEON architecture.
- **IU RGRbench** (http://www.cs.indiana.edu/~plale/projects/RGR/OGSA-DAI.html)
  IU RGRbench (Indiana University Relational Grid Resources bench) is a Grid information services benchmark/workload used to better understand resource information management in Grid information servers. In the spirit of conjunction with the UK e-Science OGSA-DAI group, we have undertaken a couple collaborative efforts: the evaluation of OGSA-DAI v2.0 using IU RGRbench and the development of a portal interface to OGSA-DAI.
- **myGrid** (http://www.mygrid.org.uk/)
  MyGrid is a research project that will extend the Grid framework of distributed computing, producing a virtual laboratory workbench that will serve the life sciences community. The integration environment will support various patterns of scientific investigation. Scientists will have the ability to customize the work environment to reflect their preferences for resource selection, data management and process enactment. MyGrid's applicability to the bioinformatics community will be tested through use cases our academic and industry partners develop. Minimally, the environment will be able to support activities relating to the analysis of functional genomic data and the annotation of pattern databases.

  OGSA-DQP is delivered as one of the many middleware components within myGrid. Currently myGrid components do not support direct interaction with Grid services, the use of OGSA-DQP up to now has been via adhoc Web service wrappers around the

GridDistributedQueryService(GDQS) for specific scenarios. However, we have a short term plan for:

- Implementing a general purpose (generic) Web service wrapper around GDQS.
- Implementing a tool that generates code for a Web service that implements a specific distributed query session by defining canned queries in its interface.

These, we hope, will provide a well-defined route for increasing the usability of OGSA-DQP within myGrid. Notably they will enable the users to call the OGSA-DQP as part of a scientific experiment either by embedding it into a workflow (myGrid WF enactment engine cannot directly call a Grid service) or by using it as a declarative alternative to what a WF Enactment Engine does.

- **N2Grid** (http://www.cs.univie.ac.at/institute/index.html?project-80=80)
  This project is developing a neural network environment based on the Grid. The Grid is employed as a transparent environment that allow users to exchange information (neural network objects, neural network paradigms) and exploits available computing resources for neural network specific tasks leading to a Grid based, world-wide distributed, neural network simulation system. The approach implements a highly sophisticated connectionist problem solution environment within a Knowledge Grid. The framework uses only standard protocols and services in a service-oriented architecture, aiming for a wide distribution of this Grid application. The N2Grid system is based on the common middleware Globus, with Java CoG and some extensions of Globus by the DataGrid project and the OGSA-DAI project. N2Grid is an evolution of the existing NeuroWeb system. The idea of these systems is to see all components of an artificial neural network as data objects in a database. These components are now being evolved to be parts of the emerging world wide Grid infrastructure.
- **ODD-Genes** (http://www.epcc.ed.ac.uk/oddgenes/)
  ODD-Genes demonstrates the power of OGSA-DAI in enabling researchers in genetics to perform new kinds of data analysis, greatly enhancing their ability to understand the wealth of data in post-genomic bioinformatics.
- **OGSA-WebDB** (http://www.gtrc.aist.go.jp/dbgrid/)
  OGSA-WebDB has been developed to bring existing Web database resources (webdbs) into the OGSA environment. It is an extension to the OGSA-DAI relational implementation. As users can query webdbs with SQL, it is possible to integrate internet webdbs with local Grid databases. Prototype software was demonstrated at SuperComputing2003.