



Case Study: Sequencing Center at University of Chicago

Simplifying Data Delivery to Customers



By using Globus Online to deliver results to users' local compute clusters, UC's Sequencing Center is reducing the cost and complexity of DNA sequencing services (and making their customers happier).

Why Globus Online for Data Delivery?

- **Simplicity**
 - Simple logon and authentication
 - Fire and forget usage
 - Web interface for execution and monitoring
- **Globus Connect**
 - Easy addition of sequencing facility or local server as transfer endpoint
- **Reliability**
 - Failed transfer retry
 - Logs to identify reasons behind failed transfers
 - Ability to check for data corruption post-transfer
 - Expert support team
- **Performance**
 - Massive data handling
 - Built for researchers

"Users were tired of all the issues with rsync and other transfer mechanisms – they just wanted to be able to fire off their transfer job and walk away."

–Neil Bahroos, Director, Initiative in Biomedical Informatics

Challenge: Getting Data Where It Needs to Go

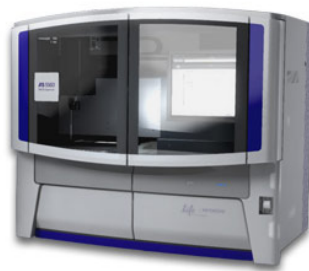
From the sequencing of human genomes which is resulting in massive amounts of new raw data, to the increase in quality (in terms of pixelation and density) of x-ray and MRI imaging files, biomedical researchers are dealing with an onslaught of data that threatens to mire progress even as it promises to enlighten.

For the University of Chicago's Functional Genomics Facility, a next-generation Sequencing Center, this creates a number of challenges. A service provider of DNA sequencing, the Center is used by scientists to sequence the DNA of blood, tissue and other physical samples: A customer ships or hand-delivers a sample to the facility, and the results – raw data in the form of potentially multi-terabyte datasets -- must be delivered back to that scientist (or his/her facility) to be processed and analyzed.

"It's just not a big deal to move big data anymore."

–Neil Bahroos, Director, Initiative in Biomedical Informatics

And therein lies the problem: How to deliver these massive files to the customer?



The UC Sequencing Center uses ABI's SOLiD™ 5500 for next gen sequencing.

While a seemingly trivial problem, the level of secure and reliable data transfer required to support the daily needs of a facility like the UC Sequencing Center is in fact tremendous. Every 1-2 weeks, the Center needs to deliver around 1.2 TB of data, in the form of anywhere from 2 to 1600 files, to up to 16 different customers.

Options are few – and limited in desirability:

- Shipping DVDs or hard drives of data adds risks like file corruption or loss, and introduces security and privacy issues
- Using tools like scp means long waits, a high level of manual intervention, and the constant threat of transfer failure

Solution: Use Globus Online for Reliable File Transfer

Working with the Initiative in Biomedical Informatics (iBi), the UC Sequencing Center is creating a file transfer solution based on Globus Online, a hosted service for secure, reliable transfer of massive data files.

Technicians at the Sequencing Center will use Globus Online to move data from their site to iBi's data center, where it can easily be accessed and analyzed by the Center's customers. The iBi provides informatics support for the Center and offers robust HPC resources including over 350 analysis tools, applications, and software suites to University scientists. Now, thanks to Globus Online, customers can simply access their data results from the iBi as easily as they access other online resources.

"Other sequencing centers should be looking at solutions like this – it's a better way to get their product to customers."

–Brigitte Raumann, Program Manager, Initiative in Biomedical Informatics

Case Study: UC Sequencing Center

Simplifying Data Delivery to Customers

"In a sense, the iBi is a cloud provider for biological scientists, and Globus Online is the tool we use to bring data – like sequencing data from the UC Center -- into our facility," says Neil Bahroos, Director at the Initiative in Biomedical Informatics.

"With Globus Online, file size is no longer a barrier to productivity."

–Neil Bahroos, Director, Initiative in Biomedical Informatics

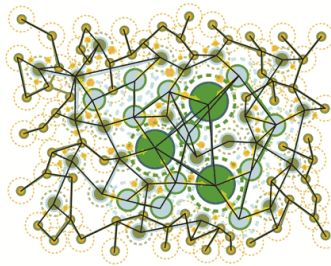
Now, thanks to Globus Online, Sequencing Center personnel can quickly and easily move users' results files from the Sequencing Center to iBi's data center. As Bahroos explains, "Sequencing Center staff are experts in their respective fields – not in IT. With Globus Online, they don't have to be file transfer experts to get their work done."

How It Works

With Globus Online, Sequencing Center technicians simply log in with their credentials, select the data they want to move and its destination, then click Transfer. Globus Online manages the process and automatically retries failed transfers, so users don't need to monitor their requests and can rest assured their jobs will complete barring loss of connectivity.

In addition, Globus Online offers a feature (called "Globus Connect") that allows users to add their own personal computers as endpoints for transfer – so scientists can move results to their local computers or laptops if desired. In the past, this process would have meant a full GridFTP install on both endpoints, with security verification and authentications required.

As Bahroos says, "Users shouldn't have to spend time managing file transfer – we wanted to make the system much easier for them. With Globus Online, the process is trivial and users can move data to the right location with just a few clicks. File size is no longer a barrier to productivity."



The Center provides resources for investigating gene functions with a focus on applying DNA/protein microarray technology.

Results: Benefits for Sequencing Center Customers and Staff

- **Improved performance:** Using Globus Online, the Sequencing Center can get results to its customers faster and more reliably.
- **Improved process efficiency and time savings:** With its simple setup and fire-and-forget ease of use, Sequencing Center technicians don't have to be IT experts to transfer big files.
- **Reduced storage requirements:** With Globus Online, the Center can move raw results files out of its facility faster – so there's less need to store massive files for months at a time. Now, Center technicians can simply transfer the file and then delete it from their system.

"Globus Online can become part of the sequencing infrastructure to support genomics."

–Brigitte Raumann, Program Manager, Initiative in Biomedical Informatics

- **Revenue opportunities:** Sequencing data is costly (currently ~\$10,000 for one human genome), and facilities like UC Sequencing Center don't want to miss the opportunity to perform this service for as many scientists as possible. Now, with a state of the art file movement service, the Center will be better equipped to meet the needs of its customers.

- **User satisfaction:** Scientists who get the data they need faster are happier and more likely to keep using the Center's

services and advocate those services to colleagues.

- **Scope increase:** As the cost to sequence DNA goes down, the demand will surely increase – and sequencing centers will have to be prepared. Says Bahroos, "one day sequencing the human genome will be as routine as performing a blood test – but not until the infrastructure is ready to process all that data in a more efficient fashion. With Globus Online we're headed in that direction."

Key Benefits for Sequencing Center at UC:

- Get results to customers faster
- Simplify file transfer for Center technicians
- Help scientists be more productive
- Maintain cutting-edge IT standards
- Reduce maintenance and training need

"Our users are experts in their respective fields – not in IT. With Globus Online, they don't have to be file transfer experts to get their work done."

–Neil Bahroos, Director, Initiative in Biomedical Informatics

For more information:

- www.globusonline.org
- info@globusonline.org