Software-as-a-Service for Research Data Management
Steve Tuecke
Deputy Director, Computation Institute
University of Chicago & Argonne National Laboratory
Big Science built on Globus Toolkit

Cancer Biology
Informatics Grid

Earth System Grid

LIGO data grid

LHC
Thinking about “small and medium labs”

• Big projects like LHC, LIGO, ESG, etc., can run **resource-level** services reliably—and build and operate effective **collective** services

• Small labs and collaborations have problems with both

• They need **solutions**, not toolkits—ideally **outsourced solutions**

Can we harness the power of the cloud to scale access to the grid?
Time-consuming tasks in science

- Run experiments
- Collect data
- Manage data
- Move data
- Acquire computers
- Analyze data
- Run simulations
- Compare experiment with simulation
- Search the literature
- Communicate with colleagues
- Publish papers
- Find, configure, install relevant software
- Find, access, analyze relevant data
- Order supplies
- Write proposals
- Write reports
- ...

www.globusonline.org
• Most research performed in small laboratories
• Researchers are trained in their field, not in IT
  – They are not experts in collecting, moving, storing, indexing, analyzing, mining, sharing, updating, publishing, and archiving massive amounts of data
• Only limited capital is available for them to spend on data and IT support
• Investment is spent on traditional research tools (e.g., microscopes)—but the world is changing
  – Now need substantial and sophisticated IT to perform research, data manipulation, data mining, collaboration
Globus Toolkit
Build the Grid
Components for building custom grid solutions
globustoolkit.org

Globus Online
Use the Grid
Reliable file transfer Software-as-a-Service
globusonline.org
Time-consuming tasks in science

- Run experiments
- Collect data
- Manage data
- Move data
- Acquire computers
- Analyze data
- Run simulations
- Compare experiment with simulation
- Search the literature
- Communicate with colleagues
- Publish papers
- Find, configure, install relevant software
- Find, access, analyze relevant data
- Order supplies
- Write proposals
- Write reports
- ...
Starting with data movement

Discover endpoints, determine available protocols, negotiate firewalls, configure software, manage space, determine required credentials, configure protocols, detect and respond to failures, determine expected performance, determine actual performance, identify diagnose and correct network misconfigurations, integrate with file systems, ...
Globus Online In Action

Astrophysics simulation data generated in Tennessee, moved to Illinois for visualization (Enzo, UCSD; Futures Lab, Argonne)

28.6 Terabytes
31,000 files
56h 44m
No human involvement
Globus Online highlights

Fire-and-forget data movement
Many files and lots of data
Third-party transfers
Performance optimization
Across multiple security domains
Expert operations and support

Web interface

Command line interface
ls alcf#dtn:~
scp alcf#dtn:~/myfile
nersc#dtn:~/myfile

HTTP REST interface
POST https://transfer.api.globusonline.org/v0.10/transfer <transfer-doc>

GridFTP servers
FTP servers

High-performance data transfer nodes

Globus Connect on local computers
Globus Online architecture

- GridFTP server
- GridFTP server
- Notification target
- Worker
- User gateway
- Profiles & state
- User
- User
- User
- User
Globus Connect Installation

Globus Connect allows you to use Globus Online to transfer files to and from your computer.

Need Help? Click Here

Step One: Choose Your Download

Step Two: Get Your Globus Connect Setup Key

Endpoint Name: lan-laptop
Description: (optional) Enter a description for this new endpoint

Setup Key: 432e8ba5-45cf-442b-a374-5a8d1cfa75cb

Step Three: Finish Globus Connect Setup

Copy the setup key displayed above. Run Globus Connect and paste the key into the Initial Setup window when prompted. This setup key can only be used once.

Ready to use your endpoint? Click here to start a transfer.
What’s next?

- Run experiments
- Collect data
- Manage data
- Move data
- Acquire computers
- Analyze data
- Run simulations
- Compare experiment with simulation
- Search the literature

- Communicate with colleagues
- Publish papers
- Find, configure, install relevant software
- Find, access, analyze relevant data
- Order supplies
- Write proposals
- Write reports
- …
Our goal: To leverage software-as-a-service (SaaS) to accelerate the pace of discovery and innovation worldwide, by providing millions of researchers with unprecedented access to powerful research tools.

“Civilization advances by extending the number of important operations which we can perform without thinking of them”

Alfred North Whitehead, 1911
www.globusonline.org

Reliable, high-performance, secure file transfer

Move files fast. No IT required.

HOW IT WORKS
Learn more about the service

GET STARTED
Setup profile in 3 easy steps

Globus Online makes robust file transfer capabilities, traditionally available only on expensive, special-purpose software systems, accessible to everyone.

Learn more

Why Use Globus Online?
See how easy file transfer can

For HPC Resource Owners
Enable Globus Online

For Developers
Integrate with Globus Online