



globus online



Globus Online

Reliable File Transfer. No IT Required.

September 20, 2011

Steve Tuecke, Deputy Director, Computation Institute
University of Chicago and Argonne National Laboratory



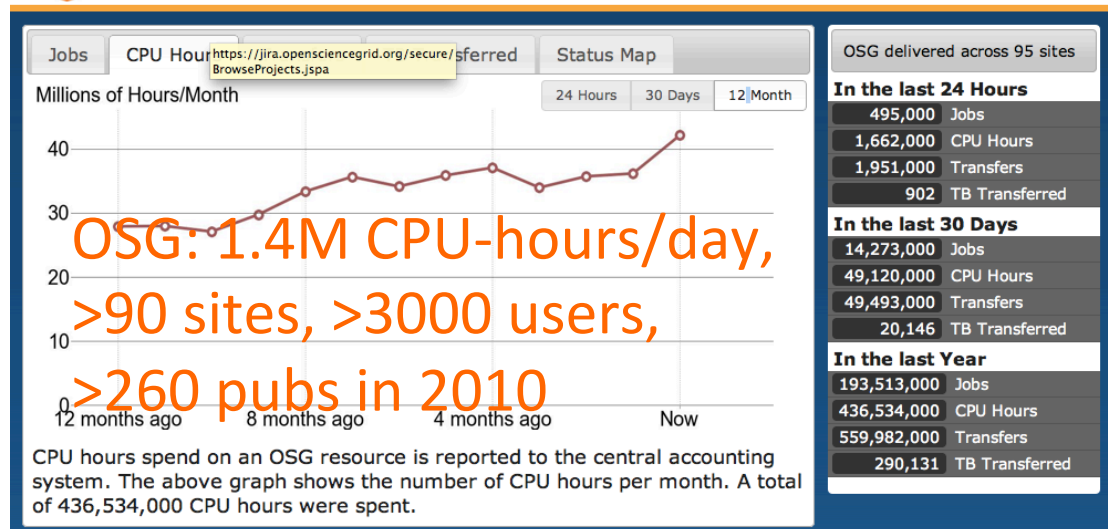
Big science has achieved big successes



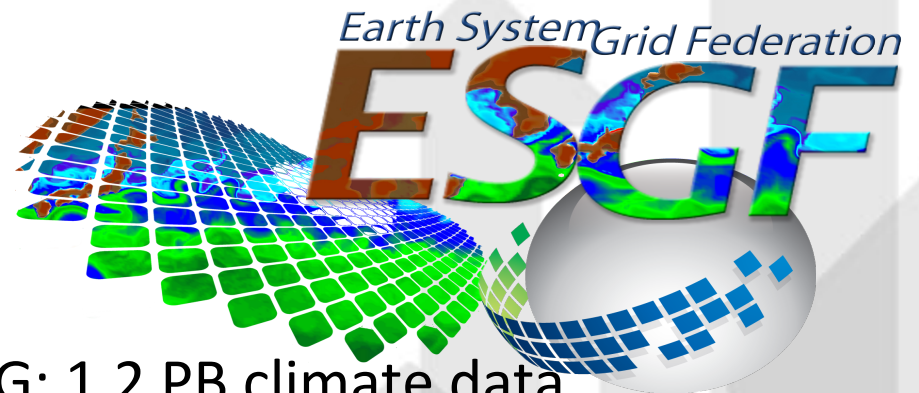
LIGO: 1 PB data in last science run, distributed worldwide



A national, distributed computing partnership for data-intensive research



- Substantial teams
- Sustained effort
- Leverage common technology
- Application-specific solutions
- Production focus



ESG: 1.2 PB climate data delivered to 23,000 users; 600+ pubs



All build on Globus Toolkit software



Globus Toolkit Update

- **GT5 focus on stable, sustainable, production components for building grid solutions**
 - Support and evolve existing communities
 - Deployed on XSEDE / TeraGrid
- **GT v5.2 coming soon with native packaging**
 - To be included in next version of VDT this Fall
 - Fully compatible with GT v5.0.x
- **Initiative for Globus in Europe (IGE) supports European Globus community**



But small and medium science is suffering



- Ad-hoc solutions
- Inadequate software, hardware & IT staff
- Data plan mandates



Medium science: Dark Energy Survey

- **Every night, they receive 100,000 files in Illinois**
- **They transmit files to Texas for analysis ... then move results back to Illinois**
- **Process must be reliable, routine, and efficient**
- **The cyberinfrastructure team is not large!**

Blanco 4m on Cerro Tololo



Image credit: Roger Smith/NOAO/AURA/NSF



The challenge of staying competitive

"Well, in our country," said Alice ... "you'd generally get to somewhere else — if you run very fast for a long time, as we've been doing."

"A slow sort of country!" said the Queen. "Now, here, you see, it takes all the running you can do, to keep in the same place. If you want to get somewhere else, you must run at least twice as fast as that!"





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
- ...



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

www.globusonline.org



Globus Online Vision

Goal: Accelerate discovery and innovation worldwide by providing research IT as a service

Leverage software-as-a-service to:

- provide millions of researchers with unprecedented access to powerful tools
- reduce research IT costs dramatically via economies of scale

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

--Alfred North Whitehead, 1911
9 www.globusonline.org



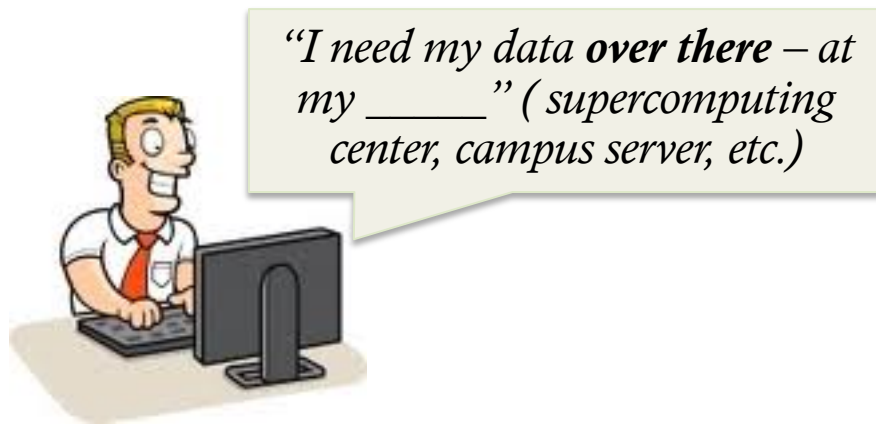
Time-consuming tasks in science

- 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
- ...



The Challenge: Moving Big Data *Easily*

- What should be trivial...



- ... can be painfully tedious and time-consuming





What is Globus Online?

- **Reliable file transfer.**
 - Easy “fire and forget” file transfers
 - Automatic fault recovery
 - High performance
 - Across multiple security domains
- **No IT required.**
 - Software as a Service (SaaS)
 - No client software installation
 - New features automatically available
 - Consolidated support and troubleshooting
 - Works with existing GridFTP servers
 - Globus Connect solves “last mile problem”
- **Initial implementation of XUAS (XSEDE User Access Services)**



“I moved 400 GB of files and didn’t even have to think about it.”

– Lawrence Berkeley National Lab



“It’s just not a big deal to move big data anymore.”

– Initiative for Biomedical Informatics

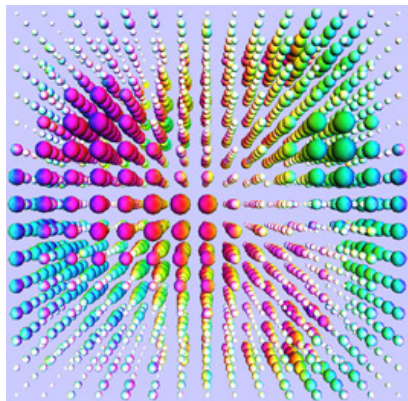


“Fantastic! I have started using globus connect to transfer data, and it only took me 5 minutes to set up. Thank you!”

– NERSC user



Case Study: Lattice QCD



- **Fast:** Reduced transfer times
- **Easy:** Fire-and-forget transfers
 - Automated retry
 - No file pre-staging
 - No complex infrastructure
 - Convenient CLI or GUI interfaces

Indiana University researcher moved ~6 TB from Oak Ridge to TACC in 2 days



“Globus Online frees up my time to do more creative work than typing scp commands or devising scripts to initiate and monitor progress to move many files.”

“I moved 100 7.3 GB files tonight in about 1.5 hours. I am very impressed. I also like the new commands and help system.”



Case Study: Enabling Users @ NERSC

- **Challenge**
 - *“We need to provide web-based ways to accomplish computing tasks – it’s what our scientists expect. and it will make them more productive.”*
- **Solution**
 - Globus Online endpoints maintained by NERSC
 - GO = recommended transfer method
- **Benefits for NERSC users**
 - Drag and drop archiving
 - Easy to use
 - Users can focus on their research (not on IT)
- **Benefits for NERSC**
 - Operations and support outsourced to GO
 - Fast and easy to make endpoints available
 - Automated authentication
 - Reliable performance and support



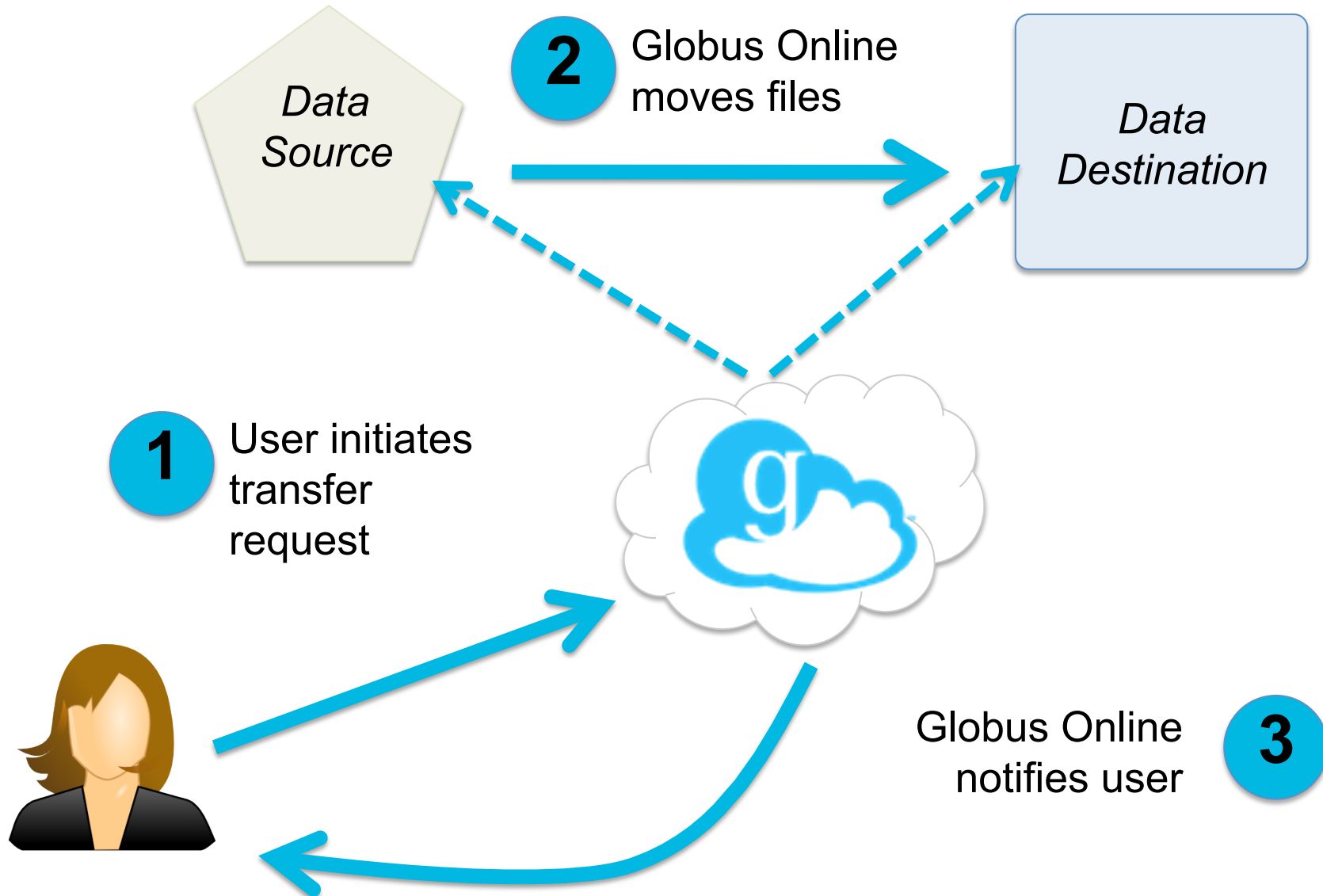
“Fantastic! I have already started using Globus Connect to transfer data, and it only took me 5 minutes to set up. Thank you!” – NERSC user



Hopper, Franklin and HPSS are among the NERSC resources leveraged by Globus Online.

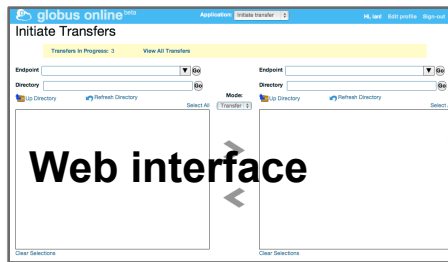


How It Works





Globus Online Interface



Web interface

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

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



GridFTP servers
FTP servers

High-performance
data transfer nodes

Globus Connect
on local computers



globus online



Demo



Logging into the CLI

- Interactive login to command line interface:

```
$ ssh tuecke@cli.globusonline.org
```

- Running commands remotely:

```
$ ssh tuecke@cli.globusonline.org <command>
```

```
$ ssh tuecke@cli.globusonline.org scp -r -s 3 -D  
olcf#/~myfile* mylaptop:~/projects/p1  
Task ID: 4a3c471e-edef-11df-aa30-1231350018b1  
$ _
```

- Using CLI with gsissh:

```
$ gsissh tuecke@cli.globusonline.org <command>
```



Coming Soon

- **GO-User:** Identity and profile management
 - Federated IdP: Shibboleth, OpenID, OAuth, LDAP
 - Federated SP: Shibboleth, OpenID, OAuth
- **GO-Team**
 - User-centric group management
 - Group provider interfaces: REST, LDAP
- **Community Branded Web Sites**
- **Domesticated collaboration tools**
 - Jira, Confluence, Wordpress, email lists, etc.



Coming this Fall

- **Functionality enhancements**
 - File/directory deletion, mkdir
 - Job labels
 - Better handling of community/VO credentials
- **Web GUI improvements**
 - Set transfer sync level, deadline, label
 - Endpoint CRUD
- **Endpoint Partner Program**
 - Globus Online Multi-User (currently in beta)
 - Simplified user sign-up for partners
 - Flight Control for endpoint owners



Globus Online and science gateways

- **GO platform for 3rd party SaaS applications**
- **Some use cases:**
 - Use GO-Transfer to file transfer between community endpoints (using community/VO account)
 - Use GO-Transfer for reliable user upload/download to user endpoints (with Globus Connect)
 - Use GO-User as identity provider
 - Use GO-Team as group provider



For More Information

- **Visit <https://www.globusonline.org/signup> to:**
 - Get a free account and start moving files
- **Visit www.globusonline.org for:**
 - Tutorials
 - FAQs
 - Pro Tips
 - Troubleshooting
- **Contact support@globusonline.org for:**
 - Help getting started
 - Help using the service