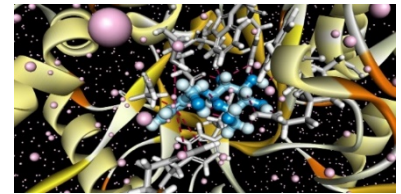




## Cornell Center for Advanced Computing (CAC)

Services for Cornell faculty, staff and student researchers

- Proposal preparation
- Technical consulting
  - Software development
  - DB development and data management
  - Optimization and workflow
  - Cloud and container help
- Computational support
  - Cloud computing
  - Cluster management
- Leased and archival storage services



Visit: [cac.cornell.edu](http://cac.cornell.edu)  
Questions: [help@cac.cornell.edu](mailto:help@cac.cornell.edu)



Cornell University  
Center for Advanced Computing

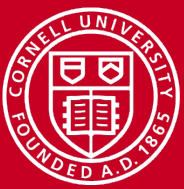
# Using Globus to Transfer and Share Data

Ben Trumbore

Computational Scientist

Cornell University Center for Advanced Computing (CAC)

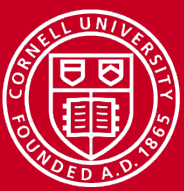
[ben.trumbore@cornell.edu](mailto:ben.trumbore@cornell.edu)



## What is Globus?

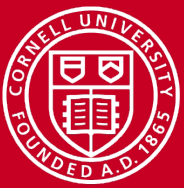
- Non-profit organization providing services for managing research data
  - Information at <http://globus.org>
- Globus can help you:
  - Transfer data
  - Share data
  - Develop data applications
  - Publish data





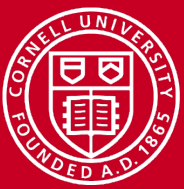
## Globus is Non-Profit

- Some common Globus features are free for individuals
  - Performing simple data transfers
  - Transferring data from your personal computer
- Other features require a subscription for organizations
  - Creating a “managed” server for data transfers
  - Data transfers to/from cloud storage
  - Sharing data
- Cornell has a Globus subscription, shared by
  - Center for Advanced Computing (CAC)
  - Biotechnology - BioHPC
  - Weill Medical



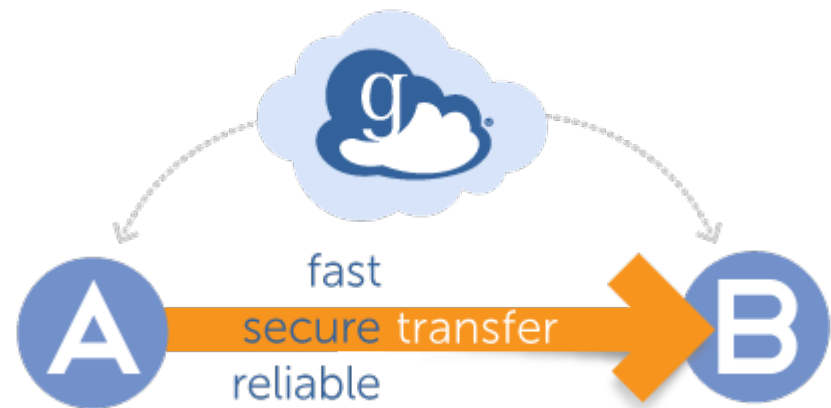
## Who Uses Globus?

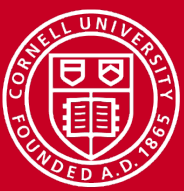
- Over 50,000 registered endpoints at over 500 institutions worldwide
- Researchers at:
  - Universities
  - National laboratories
  - Government facilities
  - Some commercial companies
- HPC facilities
  - XSEDE
  - Blue Waters



## Transferring Data

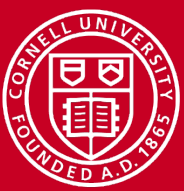
- Globus data transfers are:
  - Fast
    - Concurrency (GridFTP processes) and parallelism (TCP streams)
    - Pipelining (FTP commands)
  - Secure
    - Direct connection between endpoints
    - Encryption available
  - Reliable
    - Failure recovery
    - Validation available





## Comparing Globus to Other Data Transfer Tools

Tool	Parallel	Recovery	Encrypt	Software	Usability	Sync
SCP	no	no	no	CLI	low	no
WGET	no	no	no	CLI	low	no
rsync	no	no	no	CLI	low	manual
SFTP	no	no	no	CLI, GUI	med	no
Box	no	yes	no	explorer, web	high	auto
GridFTP	yes	yes	yes	CLI	low	manual
Globus	yes	yes	yes	web	high	manual



## Globus Endpoints and Collections

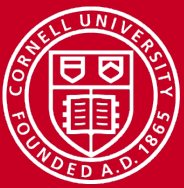
- Endpoints are computers running Globus Connect software
  - Institutional servers run “Globus Connect Server”
  - Personal computers run “Globus Connect Personal”
  - Endpoints can support multiple collections with differing permissions
- Collections are Globus network locations that provide data access
  - Most collections are located at Globus endpoints
  - Some collections are based on “connectors” to cloud storage





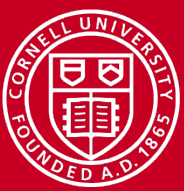
## Globus Connect Personal

- Install free “Globus Connect Personal” software on your system
- Register your system with Globus as an endpoint
- Configure which folders to include and set their permissions
- A user needs “Globus Plus” status in order to:
  - Transfer between two personal endpoints
  - Share a personal endpoint



## Globus IDs

- To use Globus you must have a Globus ID
- Available for free from Globus
- Link Globus IDs with IDs from other organizations that use Globus

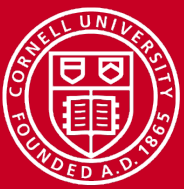


# Logging In to the Globus Web Interface

- Logging in: <http://app.globus.org>

The screenshot shows the top of the Globus login page. The header is dark blue with the Globus logo and 'Globus Account Log In' text. Below the header, the main heading is 'Log in to use Globus Web App'. Underneath, it says 'Use your existing organizational login' with examples like 'e.g., university, national lab, facility, project'. There is a dropdown menu currently showing 'Globus ID'. Below that, a link says 'Didn't find your organization? Then use Globus ID to sign in. (What's this?)'. A blue 'Continue' button is positioned below the link. At the bottom, there are two buttons: 'Sign in with Google' and 'Sign in with ORCID iD', separated by an 'Or'.

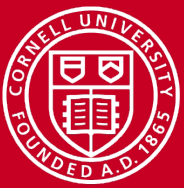
This screenshot shows the same login page as the previous one, but with the 'Continue' button clicked. The page now displays a privacy policy notice in a light gray box. The notice includes a circular arrow icon and text stating: 'Globus uses CILogon to enable you to Log In from this organization. By clicking Continue, you agree to the CILogon privacy policy and you agree to share your username, email address, and affiliation with CILogon and Globus. You also agree for CILogon to issue a certificate that allows Globus to act on your behalf.' Below the notice, the 'Continue' button is no longer visible, and the 'Sign in with Google' and 'Sign in with ORCID iD' buttons are again visible, separated by an 'Or'.



## Data Transfers Using the Globus Web Interface

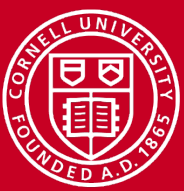
- Between two collections
- Performed asynchronously
- Monitor progress, cancel if desired





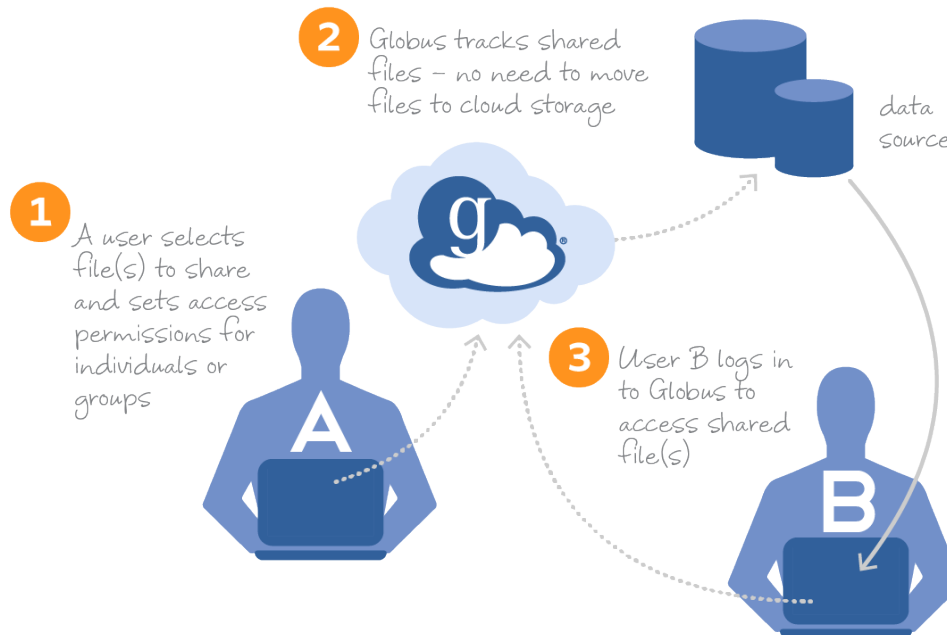
Cornell University  
Center for Advanced Computing

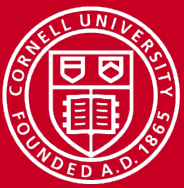
# Data Transfer Demonstration



## Sharing Data

- Select a folder in a collection and share it with others
- Share with Globus members or groups (which you can create)
- Set read and write permissions on the shared data





Cornell University  
Center for Advanced Computing

# Sharing Data Demonstration



## Data Connectors

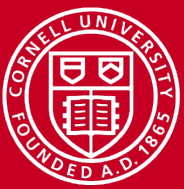
- For accessing cloud-based storage through Globus
  - Amazon S3
  - Google Drive
  - Box
  - Others
- For subscription holders only, requires additional fees
  - Cornell is not currently paying for any of these
- Globus Connect Server runs as a gateway from Globus to the cloud





## Programmatic Access to Globus Functionality

- Command Line Interface (CLI)
  - Python-based
  - Used in shells and scripts
- REST Application Programmer Interfaces (APIs):
  - Auth
  - Search
  - Transfer
- Software Development Kit (SDK)
  - Wraps the REST APIs for Python



## More Information

- Globus “How To” pages
  - <http://docs.globus.org/how-to/>
- Cornell CAC’s Globus tutorials:
  - Basic: <http://cvw.cac.cornell.edu/GlobusXfer>
  - Advanced: <http://cvw.cac.cornell.edu/GlobusXfer2>