

Managing Data using Globus

Rachana
Ananthakrishnan
ranantha@uchicago.edu



globus



“I need a good place to store / backup / archive my (big) research data, at a reasonable price.”



Campus Store



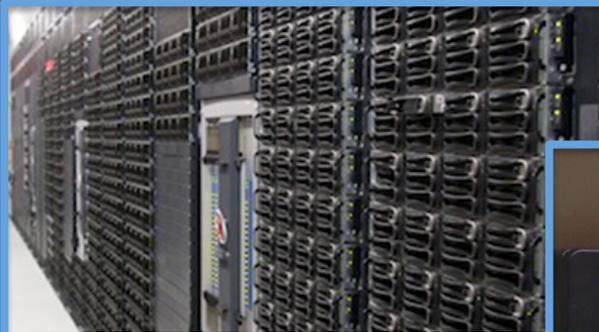
Mass Store



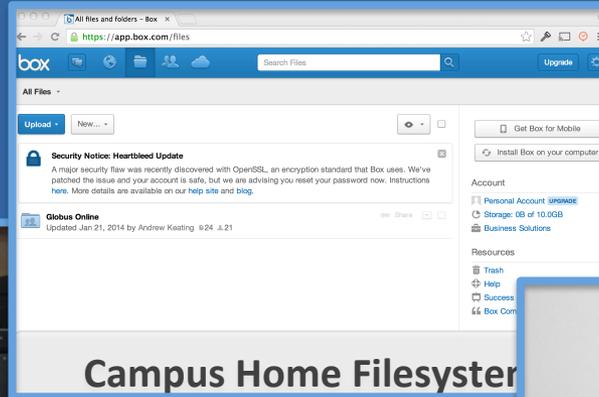
Public Cloud Archive



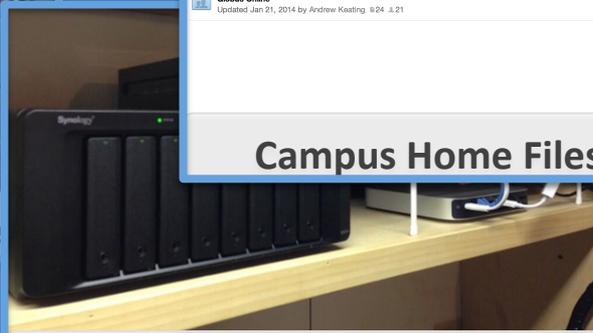
“I need to easily, quickly, & reliably move or mirror portions of my data to other places.”



Research Computing HPC Cluster



Campus Home Filesystem



Lab Server



Personal Laptop



Desktop Workstation



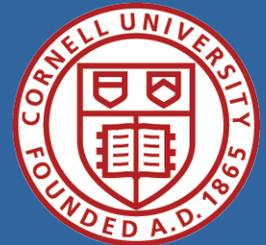
XSEDE Resource



Public Cloud



“I need to easily and securely share my data with my colleagues at other institutions.”



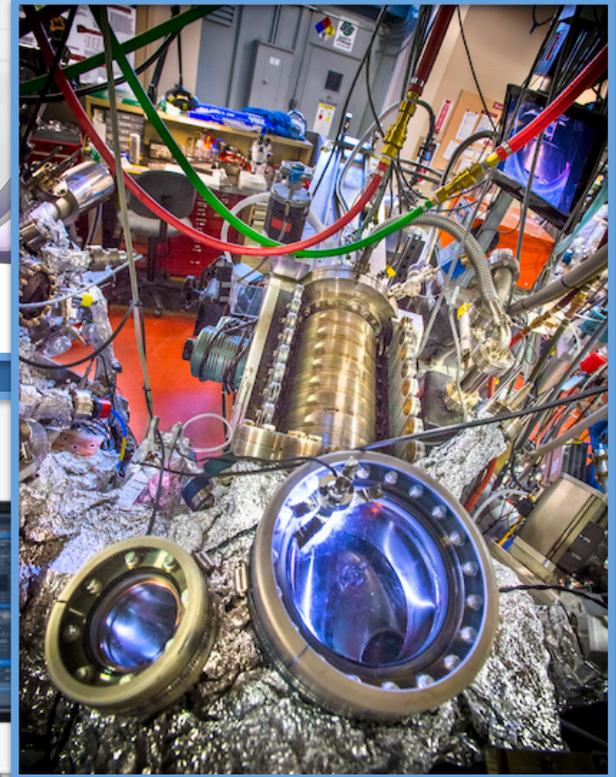


“I need to get data from a scientific instrument to my analysis server.”

MRI



Advanced Light Source



Next Gen Sequencer



Light Sheet Microscope



Challenge: Manage research
data as easily as...

flickr

...our pictures

...our e-mail



NETFLIX

...home entertainment



What is Globus?

Big data transfer, and sharing...

... delivered via SaaS ...

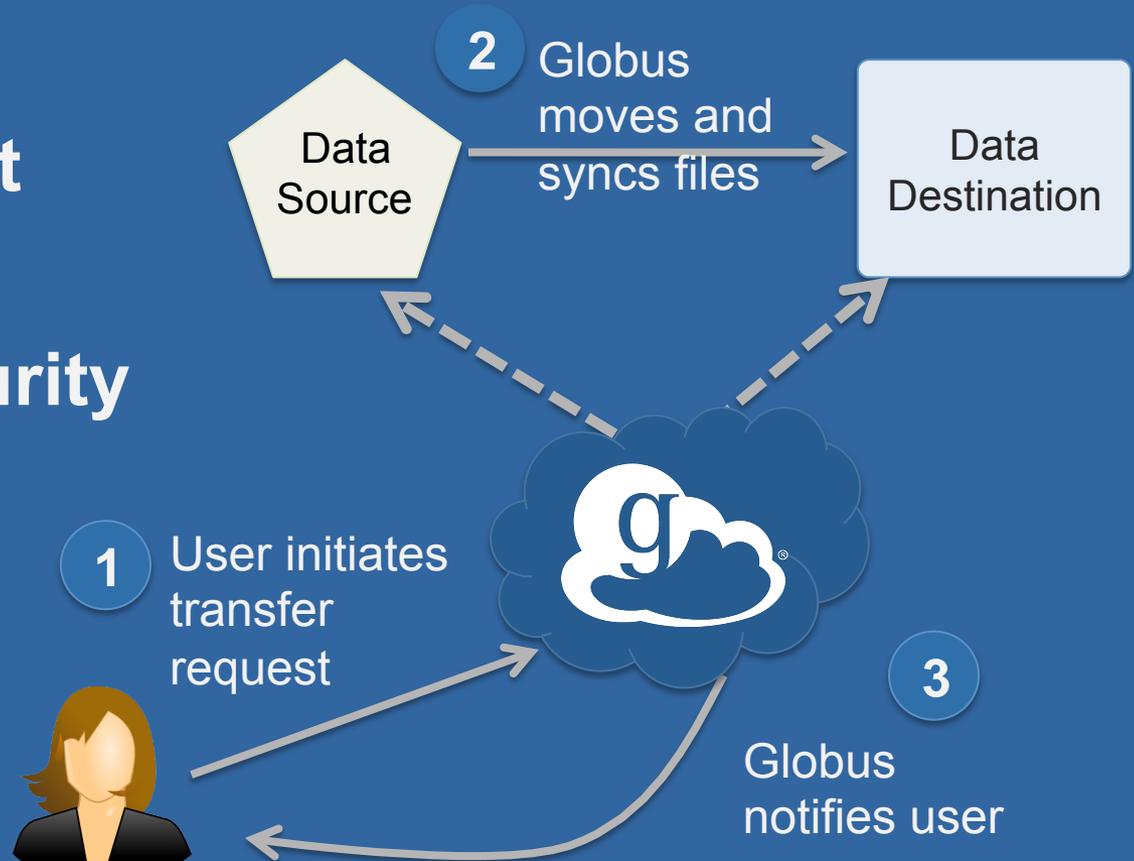
... that is simple, secure, and fast...

... directly from your own storage
systems



Reliable, secure, high-performance *file transfer*

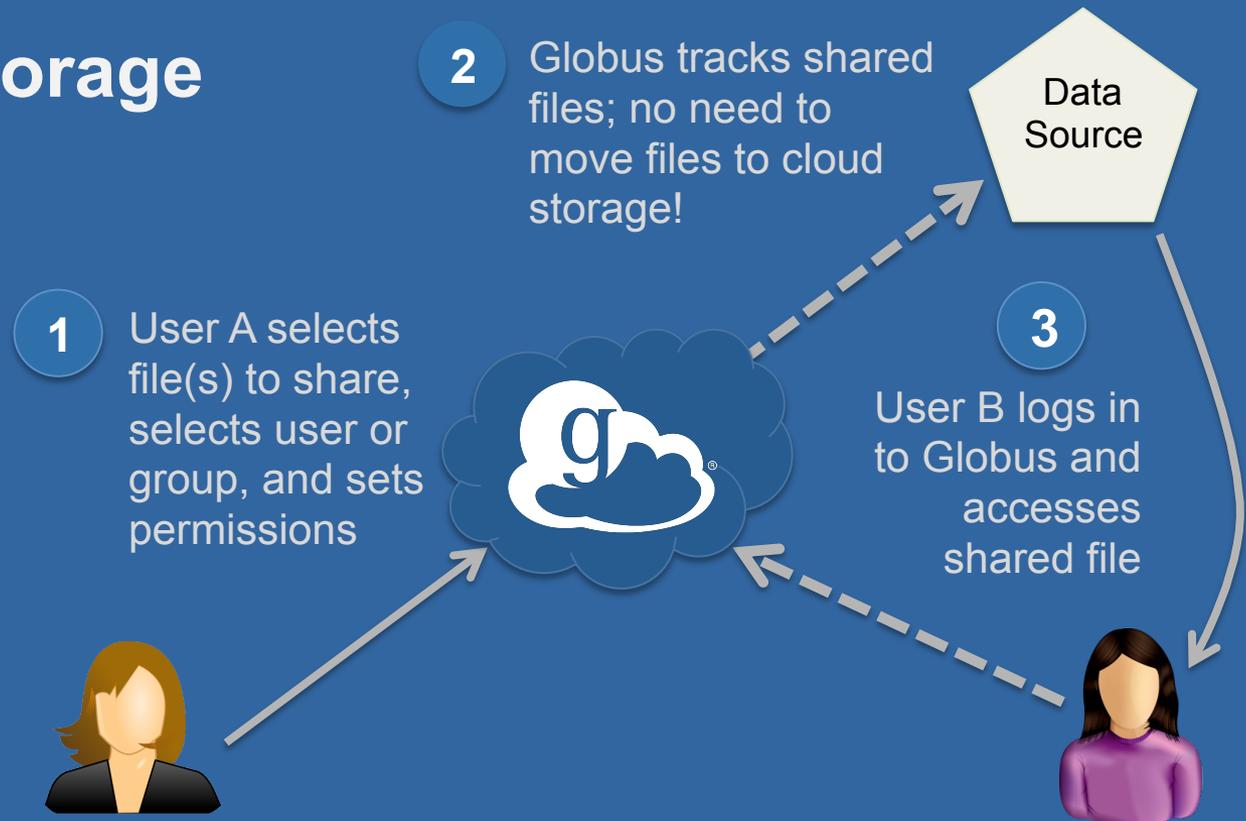
- “Fire-and-forget” transfers
- Automatic fault recovery
- Seamless security integration
- Powerful GUI and APIs





Simple, secure *sharing* off existing storage systems

- Easily share large data with any user or group
- No cloud storage required





Globus is SaaS

- **Web, command line, and REST interfaces**
- **Reduced IT operational costs**
- **New features automatically available**
- **Consolidated support & troubleshooting**
- **Easy to add your laptop, server, cluster, supercomputer, etc. with Globus Connect**



8,000

active endpoints

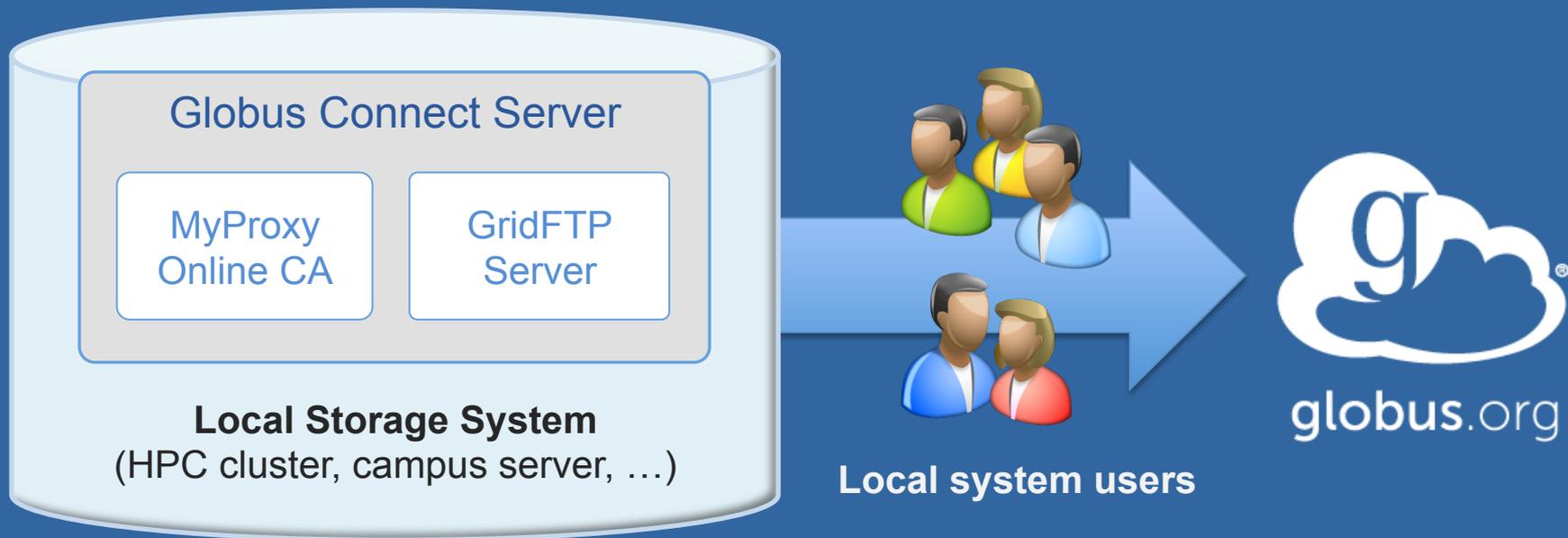
(in the past year)



Demonstration



Globus Connect Server



- **Create endpoint in minutes; no complex GridFTP install**
- **Enable all users with local accounts to transfer files**
- **Native packages: RPMs and DEBs**
- **Also available as part of the Globus Toolkit**



Globus Platform-as-a-Service



Globus APIs



Sharing Service

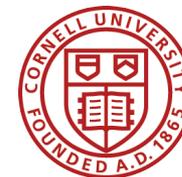
Transfer Service

Identity, Group, Profile Management Services



Globus Toolkit

Globus Connect





globus
genomics

**Flexible, scalable,
affordable
genomics analysis
for all biologists**



Next-gen sequence
analysis SaaS

+

Data management
PaaS

+

Scalable IaaS





Globus is moving beyond
transfer and sharing to
**data publication and
discovery**



Globus Data Publication

(coming soon)

- **SaaS for publishing large research data**
- **Bring your own storage**
- **Extensible metadata**
- **Publication and curation workflows**
- **Public and restricted collections**
- **Rich discovery model**



Enables data to be easily...

Identified

Described

Curated

Verifiable

Accessible

Preserved



...and facilitates rich discovery

Search

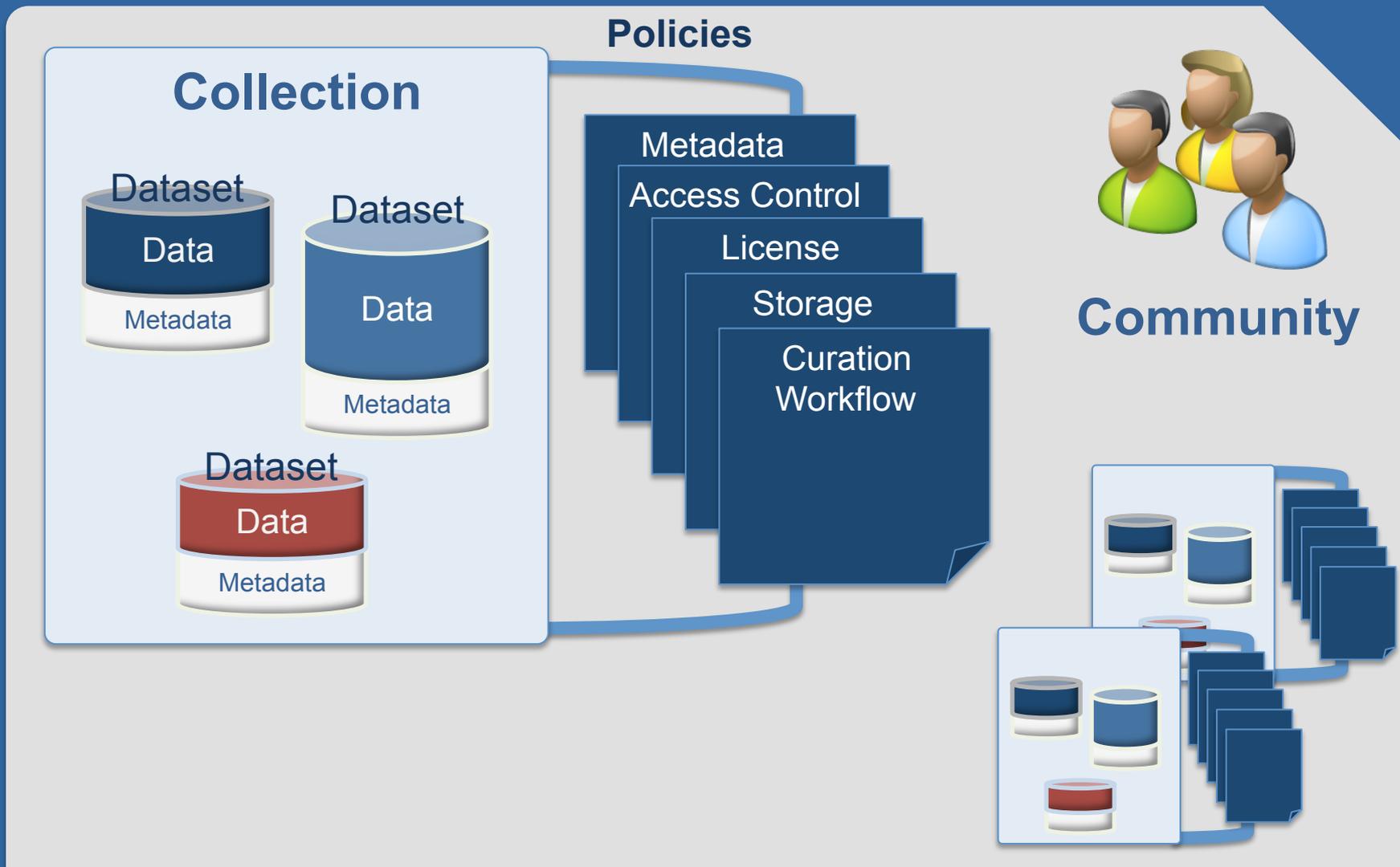
Browse

Access

*...across collections,
endpoints*



Globus' view of data publishing





Exemplar Use Case

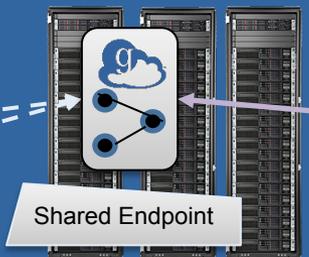
1. Publish Data

2. Describe Submission

6. Download



3. Assemble Dataset (Transfer Data)



Argonne Storage System

4. Curate Dataset
✓ ✗



Argonne Curator



To Learn More...

- **Transfer and share: www.globus.org**
- **Email: support@globus.org**
- **Globus Genomics:
www.globus.org/genomics/**
- **Globus Publication:
www.globus.org/data-publication**



Exercise 1: Account Signup

1. Go to: globus.org/signup
2. Create your Globus account
3. Validate e-mail address
4. Optional: Login with your campus/InCommon identity



Exercise 2: Transfer to/from ALCF

1. Choose ALCF endpoint: `alcf#dtn` and authenticate with your credentials.
2. Move file(s) your ALCF account from) from `esnet#anl-diskpt1`
3. Move file(s) from ALCF account to `go#ep1`



Exercise 3: Transfer, Sharing, Group Management

1. Install Globus Connect Personal
2. Move file(s) from esnet#anl-diskpt1 to your laptop
3. Sign up for a free Globus Plus trial
4. Create a shared endpoint on your laptop
5. Grant your neighbor permissions on your shared endpoint
6. Access your neighbor's shared endpoint
7. Optional: Create group, and grant share access



Exercise 4: Using the CLI

1. Configure SSH public key in Globus profile
2. Log into the Globus CLI: **ssh cli.globusonline.org**
3. Transfer files from alcf#dtn to your laptop using **scp** command
4. Check status of your transfer using **status** command



Thank you to our sponsors!



U.S. DEPARTMENT OF
ENERGY



THE UNIVERSITY OF
CHICAGO

Argonne
NATIONAL LABORATORY



powered by
amazon
web services