# Globus for Research Data Management

Stuart Martin
Software Development Manager,
Globus
sjmartin@uchicago.edu





Globus delivers...

(big) data transfer, sharing, publication, and discovery...

...directly from your own storage systems



"I need to easily, quickly, & reliably move or replicate my data to other systems."



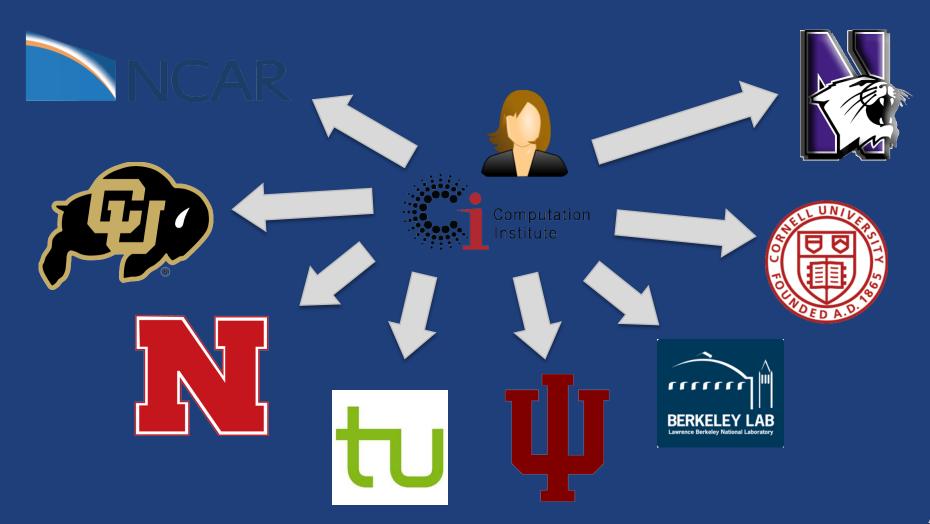


"I need to get data from a scientific instrument to my analysis system."





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





"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 publish my data so that others can find it and use it."





Scholarly Publication

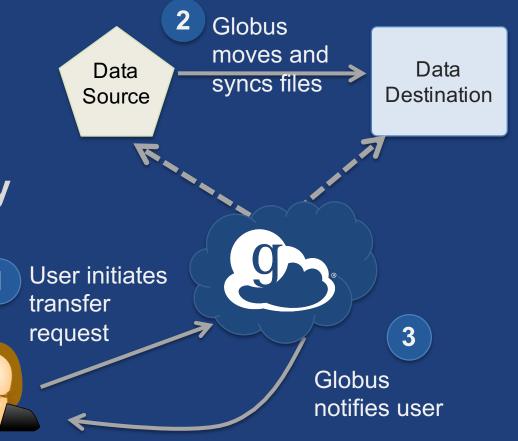


Active Research Collaboration



# Reliable, secure, high-performance file transfer and replication

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





## Storage connectors

- Standard supported storage connectors (Posix)
  - Linux, Windows, MacOS
  - Lustre, GPFS, OrangeFS, etc.
- Premium storage connectors
  - HPSS
  - HDFS
  - S3
  - Ceph RadosGW (S3 API)
  - Spectra Logic BlackPearl
  - Google Drive

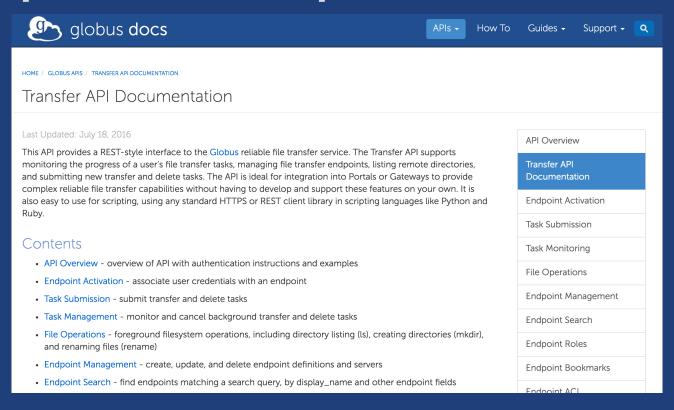


# Demonstration: Sign Up File Transfer



## Globus Transfer API

## Nearly all Globus Web App functionality implemented via public Transfer API



https://docs.globus.org/api/transfer/

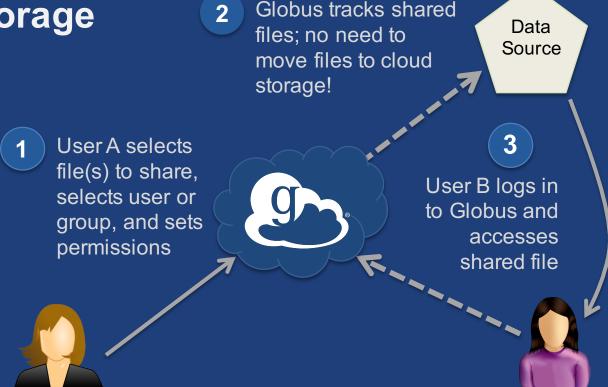
## New Globus CLI (coming soon)

```
$ globus transfer
Usage: globus transfer [OPTIONS] COMMAND [ARGS]...
  Interact with Globus Transfer API. Transfer, Delete, List, and Rename
  files on Endpoints, manage your Endpoints and Shares, and monitor your
  ongoing Transfer Tasks
Options:
  -F, --format [json|text] Output format for stdout. Defaults to text
  -h, --help
                           Show this message and exit.
  --version
                           Show the version and exit.
Commands:
  acl
                 Manage Endpoint Access Control Lists
  async-delete
                 Submit a Delete Task
  async-transfer Submit a Transfer Task
  bookmark
                 Manage Endpoint Bookmarks
  endpoint
                 Manage Globus Endpoint definitions
  ls
                 List Endpoint directory contents
 mkdir
                 Make a directory on an Endpoint
                 Rename a file or directory on an Endpoint
  rename
                 Manage asynchronous Tasks
  task
```



# Simple, secure *sharing* off existing storage systems

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





# Demonstration: Globus sharing Group Management



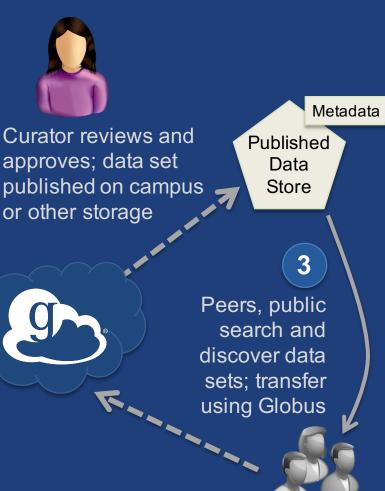
# Curated *publication* of data, with relevant metadata for *discovery*

- Identify
- **Describe**
- Curate
- Verify
- Access
- Preserve

Researcher assembles data set; describes it using metadata (Dublin core and domain-specific)

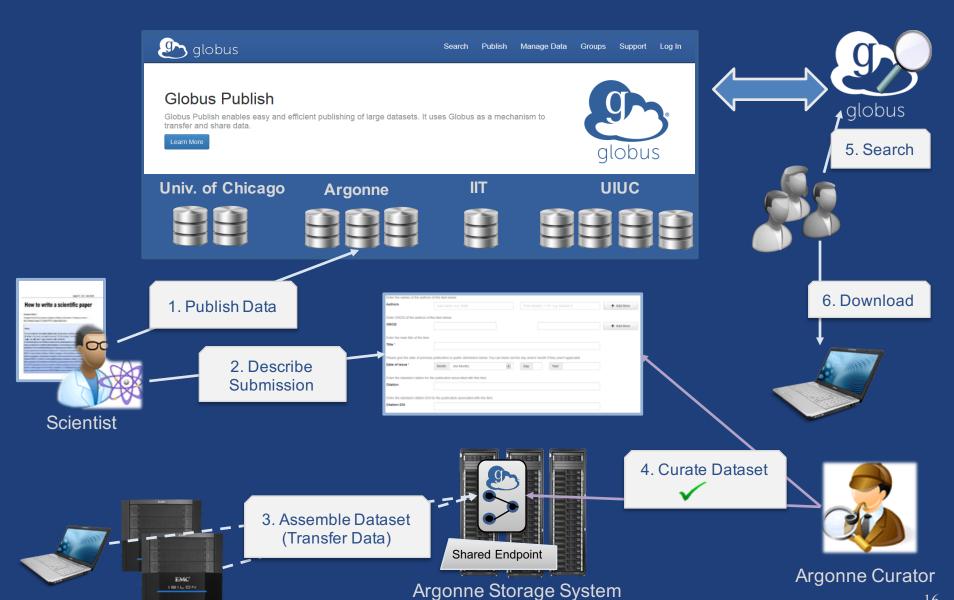


15





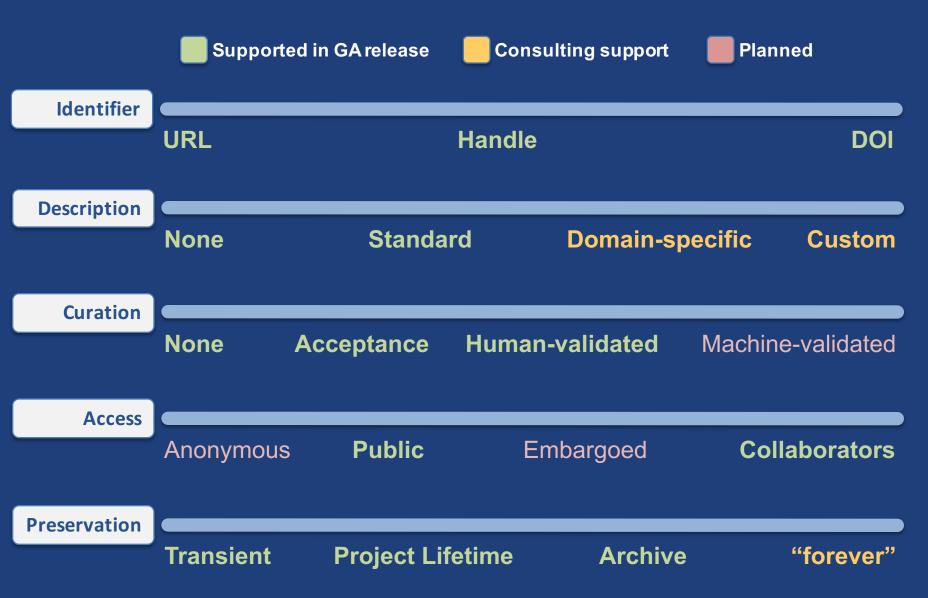
## Typical publication workflow



16



## Publication GA release





## Globus Platform-as-a-Service



**KBase** 

Globus APIs

• • •

**Data Publication & Discovery** 

File Sharing

File Transfer & Replication

Identity, Group, and Profile Management

**Globus Connect** 













**Globus Toolkit** 



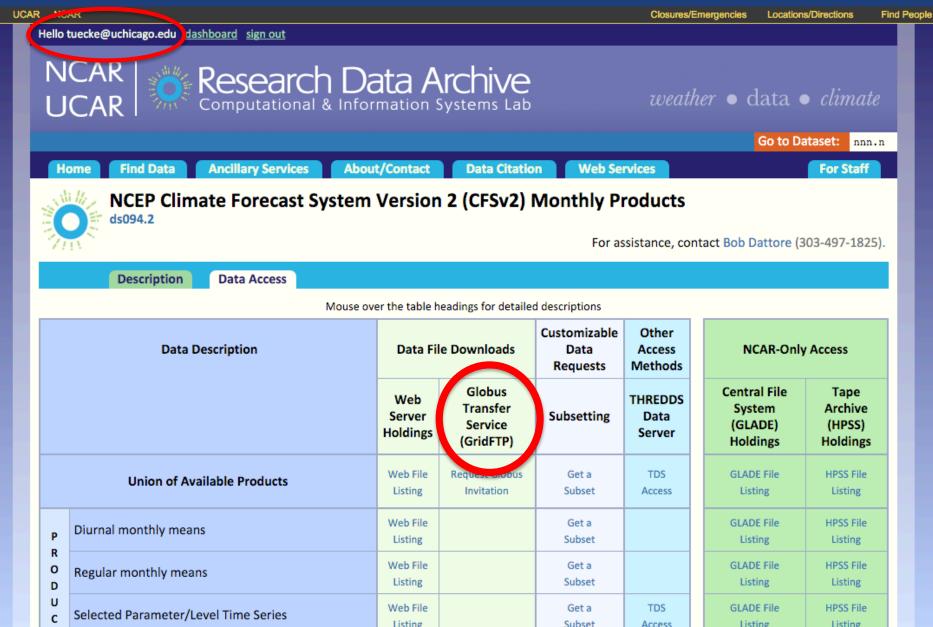
## Platform Questions

- How do you leverage Globus services in your own applications?
- How do you extend Globus with your own services?

 How do we empower the research community to create an integrated ecosystem of services and applications?

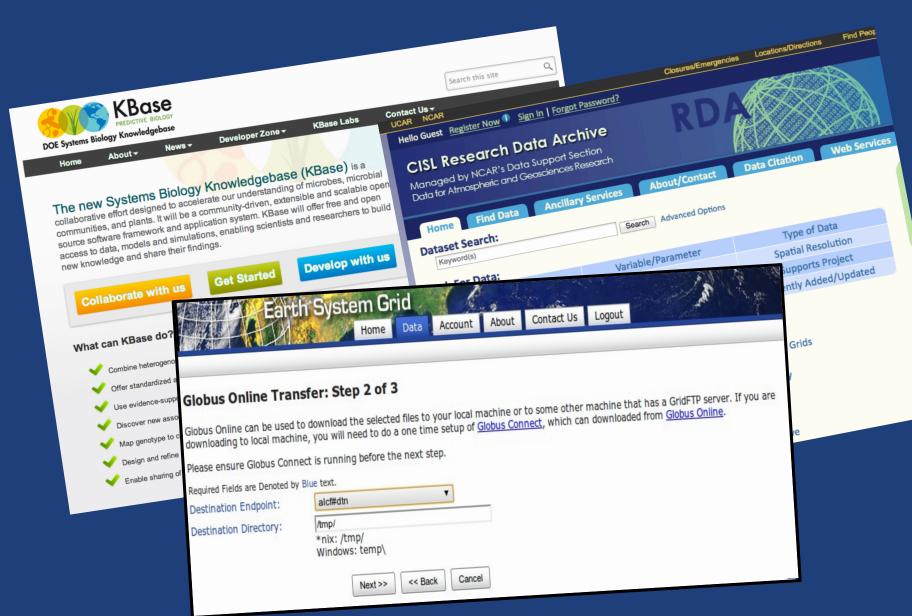


## Research data portal





## Platform adopters





## Globus at a glance

major services

100 PB transferred

20 billion

files processed

25,000

registered users

13 national labs use Globus services

8,000 active endpoints

>300

users are active daily

99.95% uptime over the past

two years

>30 institutions subscribed

1 PB largest single transfer to date

3 months

longest continuously managed transfer

**157** federated campus identities



## Globus endpoints worldwide



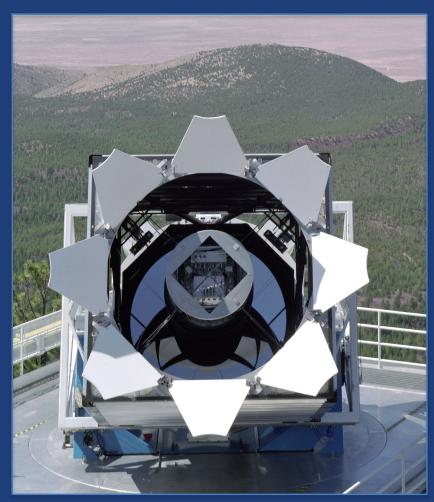


## Globus is enabling...

Study of the structure and evolution of galaxies, the nature of dark energy, and cosmological history of the universe

Joel Brownstein University of Utah





Sloan Digital Sky Survey Source: University of Utah

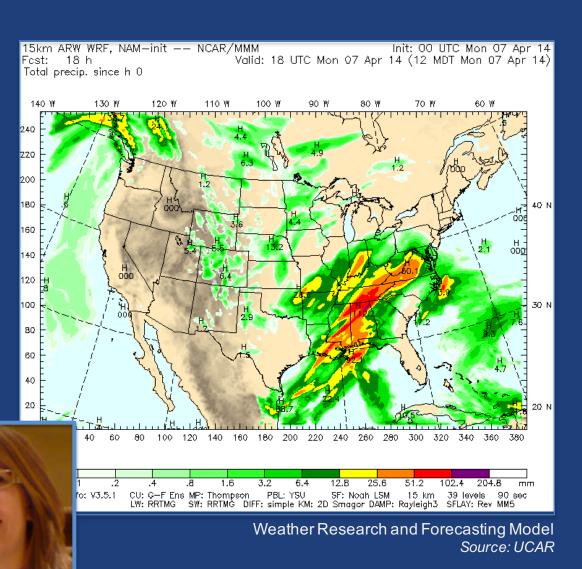


## Globus is enabling...

Development of numerical simulations of severe storms for improved responsiveness to weather

Ann Syrowski University of Illinois

events



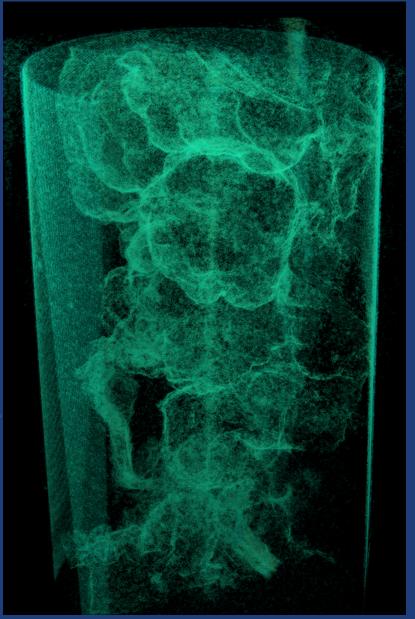




Dan Kozak (Caltech) replicates 1 PB LIGO astronomy data for resilience



Erin Miller (PNNL) collects data at **Advanced Photon** Source, renders at PNNL, and views at ANL





## Globus Subscription

## Globus Subscriptions

Globus file transfer is free to use for non-profit research and educational purposes. Subscriptions offer researchers access to enhanced features such as file sharing and data publication, and provide system administrators with a comprehensive view into how storage systems are being utilized. Your subscription helps us make Globus self-sustaining so we can continue to offer high-quality data management infrastructure to all researchers.

### Click here to start your 90-day free trial of Globus premium features

### Subscriptions for Non-Profit Research and Education

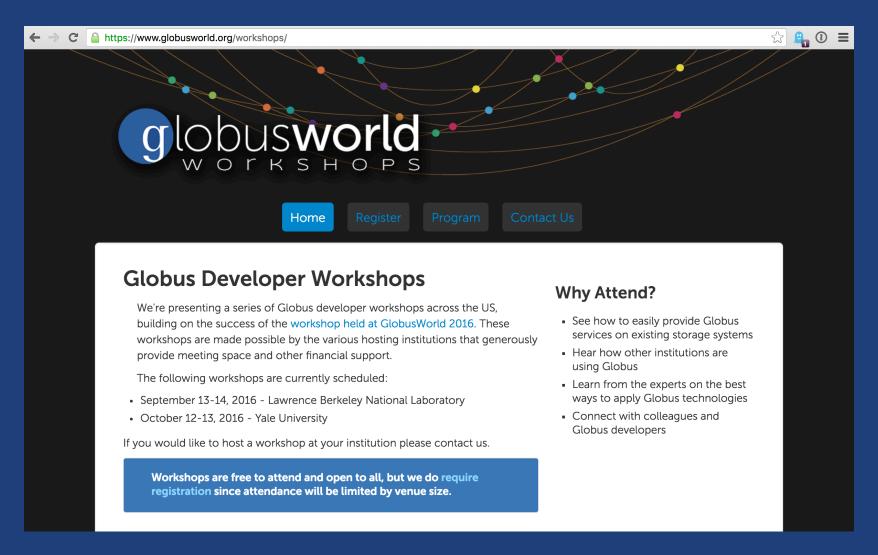
If you wish to use Globus in a commercial setting, you must have a commercial subscription.

Features (click 1) for description)	Basic - Free	Starter Subscription	Standard Subscription
Transfer level 1	Unlimited	Unlimited	Unlimited
User level <b>1</b>	Unlimited	Unlimited	Unlimited
Managed endpoints 1	None	1	Unlimited
Management console 1	_	•	•
Usage reports 1	_	•	•
Support for Globus Connect, Web, CLI 1	_	•	•
Shared endpoints 1	_	•	•
Globus Plus users (1)	_	_	•

https://www.globus.org/pricing



## GlobusWorld Developer Workshops



https://www.globusworld.org/workshops/



## Sample python data portal

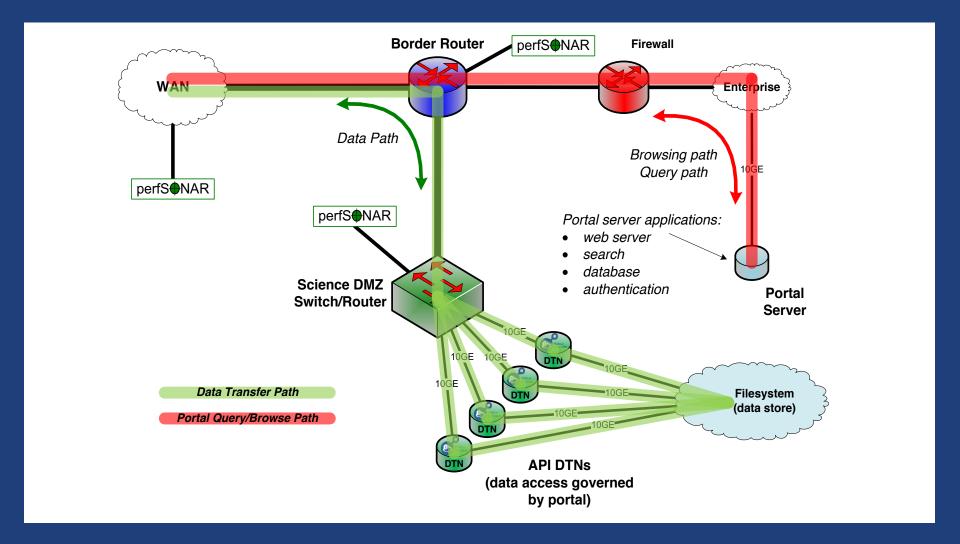


https://github.com/globus/globus-sample-data-portal

When all else fails...

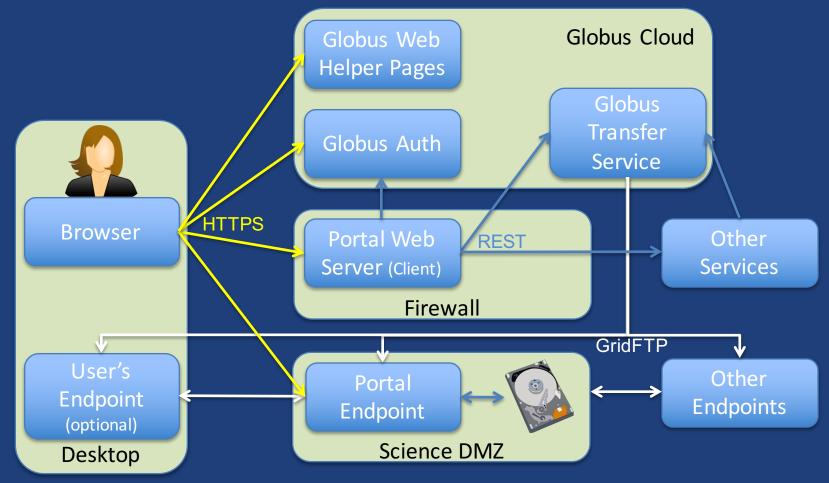


## Next-Generation Portal Leverages Science DMZ





## Prototypical research data portal



- Move portal storage into Science DMZ, with Globus endpoint
- Leave Portal Web server behind firewall
- Globus handles the security and data heavy lifting



## Developer Workshops scheduled

- September 13-14, 2016 LBNL, Berkeley, CA
- October 12-13, 2016 Yale University, New Haven, CT
- Date TBD NCAR, Boulder, CO



More information:

www.globus.org

www.globus.org/data-publication

Contact:

support@globus.org



## http://tinyurl.com/globusatpesc