Globus: Making It Easy to Move, Share and Publish Your Data

Vas Vasiliadis
vas@uchicago.edu

Thank you to our sponsors!

U.S. DEPARTMENT OF
ENERGY

THE UNIVERSITY OF
CHICAGO

NIST
National Institute of Standards and Technology
U.S. Department of Commerce

Argonne
NATIONAL LABORATORY

...and Thank YOU!

5 major services	**190 PB** transferred	**30 Bn** files processed	**45,000** registered users
13 national labs use Globus	**10,000** active endpoints	**10,000** active users/year	**99.9%** uptime
60+ institutional subscribers	**1 PB** largest single transfer to date	**3 months** longest continuously managed transfer	**130** federated campus identities

Agenda

- **Overview for those new to Globus**
- **Survey of “recent” enhancements**
- **Brief demonstration**
- **Questions**

Globus delivers…

Big data transfer, sharing, publication, and discovery…

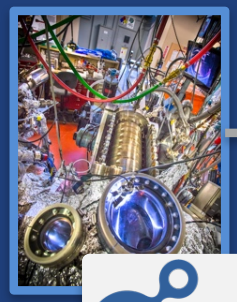
…directly from your own storage systems…

…via software-as-a-service

Data management today

- **Move?**
- **Share?**
- **Discover?**
- **Reproduce?**

Globus and the research data lifecycle



Instrument

Globus transfers files reliably, securely

2

Compute Facility

4 Globus controls access to shared files on existing storage; no need to move files to cloud storage!

7 Curator reviews and approves; data set published on campus or other system

Publication Repository

globus connect

Transfer

1

Researcher initiates transfer request; or requested automatically by script, science gateway

3

Share

Researcher selects files to share, selects user or group, and sets access permissions

6

Researcher assembles data set; describes it using metadata (Datacite & domain-specific)

Collaborator logs in to Globus and accesses shared files; no local account required; download via Globus

5

Publish

6

8

Personal Computer

Peers, collaborators search and discover datasets; transfer and share using Globus

Discover

- **Only a Web browser required**
- **Use storage system of your choice**
- **Access using your campus credentials**

Another perspective...

Campus Bridging

…within and beyond campus boundaries

Bridge to national cyberinfrastructure

Move datasets to supercomputer, national facility

Move results to campus (…)

Bridge to instruments

Analysis store

- Clear staging store
- Data cleansing
- Pre-processing
- ...

Raw Source Data

High durability, low cost store

Bridge to your collaborators

globus connect

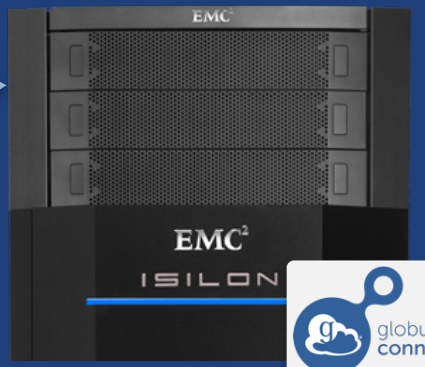
XSEDE
Extreme Science and Engineering Discovery Environment

EC2

Jetstream

NERSC

Local/Campus stores



globus connect

- Group management
- Share creation
- Permissions configuration

Cloud stores

ceph

openstack

Jetstream

Google Drive

amazon web services S3

globus connect

Bridge to community/public

(Re)format…

- PDF/A
- HDF
- …

Describe…

- Datacite metadata
- Domain metadata
- Provenance info
- ...

Replicated, public repositories

Persistent identifier

- DOI

Multi-step curation workflow

- Review → Update → Resubmit cycle

Why use Globus?

- **Ease of use: consistent UI across systems**
- **“Fire-and-forget” file transfer**
- **Maximized WAN throughput**
- **Secure access w/multi-tier security model**
- **No/low-overhead external collaboration**
- **Rapid deployment w/standard packages**
- **Highly automatable: CLI, RESTful API**

What's new?

- **New authentication service**
- **Endpoint search and bookmarks**
- **Management console**
- **Data publication and discovery**
- **Storage connectors**
- **Globus Platform-as-a-Service**

Demonstration

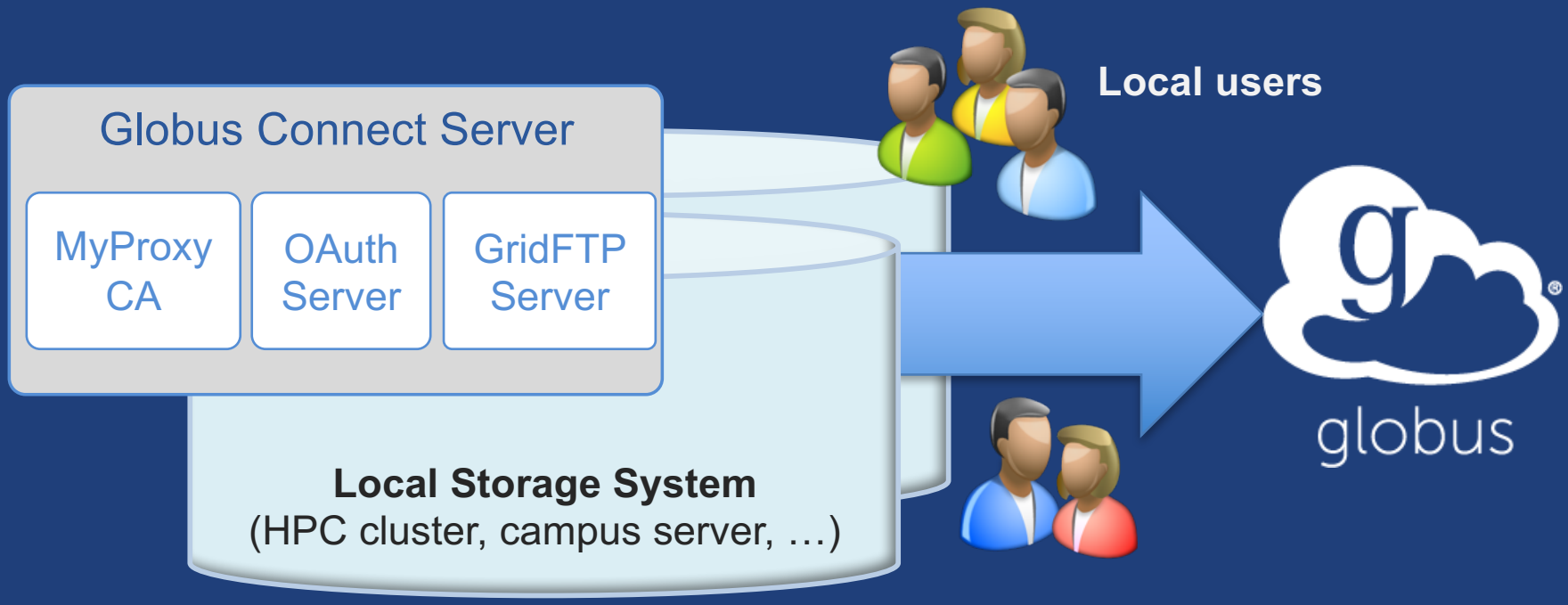
Globus Connect…

Makes your storage system a Globus endpoint

Globus Connect Personal

- **Installers do not require admin access**
- **Zero configuration; auto updating**
- **Handles NATs**

Globus Connect Server

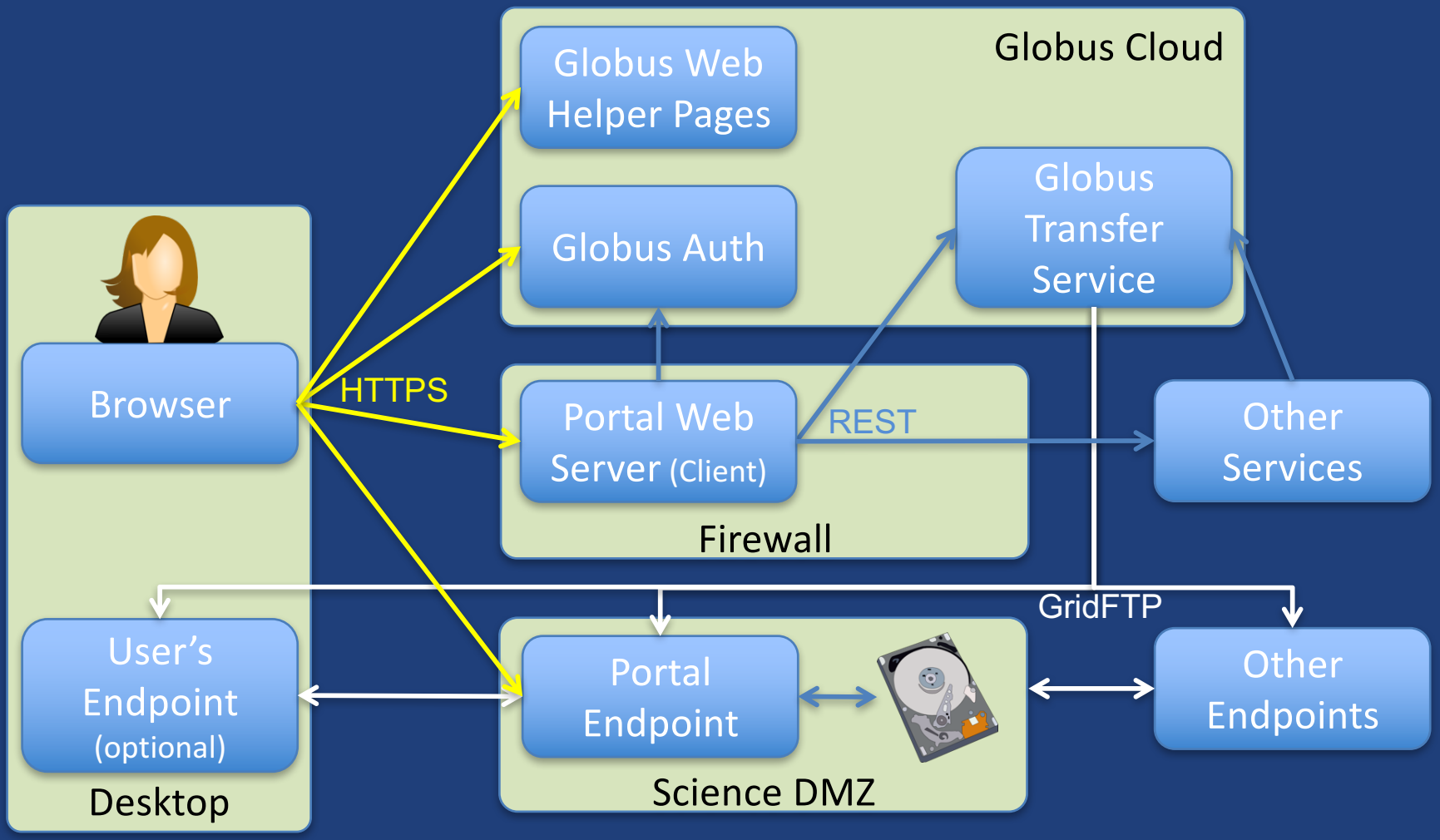


- **Enable access for all users with local accounts**
- **Create endpoint on practically any filesystem**
- **Native packages: RPMs and DEBs**

Storage connectors

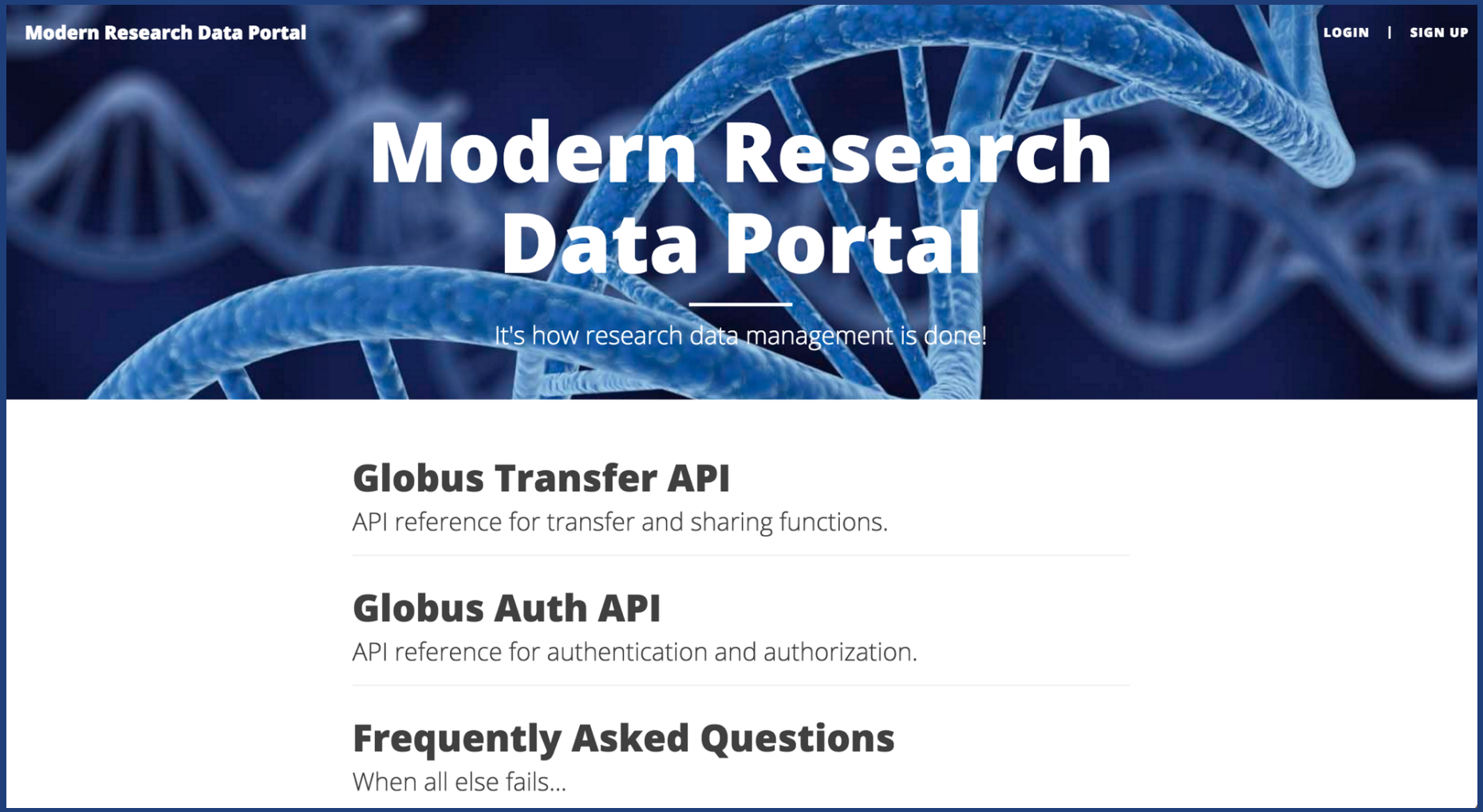
- **Standard storage connectors (Posix)**
  - Linux, Windows, MacOS
  - Lustre, GPFS, OrangeFS, etc.
- **Premium storage connectors**
  - HPSS
  - HDFS
  - S3
  - Ceph RadosGW (S3 API)
  - Spectra Logic BlackPearl
  - Google Drive (coming soon)

Globus Platform-as-a-Service



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

Sample python data portal



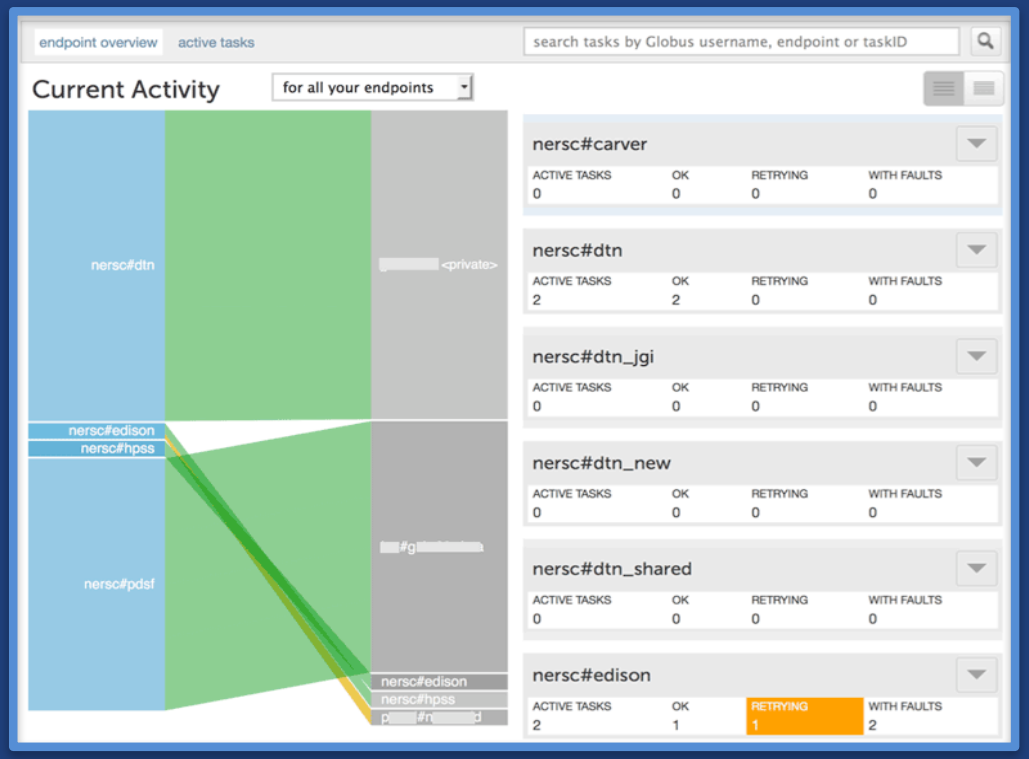
https://github.com/globus/globus-sample-data-portal

Coming Soon

- **HTTPS support**
  - Synchronous alternative GridFTP
  - Implements same fine-grained access control model
  - Greatly simplifies sharing/transfer for “small” datasets
  - Supports standard browser behaviors such as viewing known file types in the browser
  - Allows integration with clients and web apps to further leverage existing research storage systems
- **New command line interface**
  - Locally installed Python client
  - Similar to AWS command line tools

Globus Subscriptions

- **Standard Subscription**
  - Shared endpoints
  - Data publication
  - HTTPS support*
  - Management console
  - Usage reporting
  - Priority support
  - Application integration
- **Branded Web Site**
- **Premium Storage Connectors**
  - Amazon S3, Ceph, HPSS, Spectra, Google Drive*, …
- **Alternate Identity Provider (InCommon is standard)**



*Available in early 2017

New pricing model

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 ⓘ for description)	**Basic - Free**	**Starter Subscription**	**Standard Subscription**
Transfer level ⓘ	Unlimited	Unlimited	Unlimited
User level ⓘ	Unlimited	Unlimited	Unlimited
Managed endpoints ⓘ	None	1	Unlimited
Management console ⓘ	–	✓	✓
Usage reports ⓘ	–	✓	✓
Support for Globus Connect, Web, CLI ⓘ	–	✓	✓
Shared endpoints ⓘ	–	✓	✓
Globus Plus users ⓘ	–	–	✓

https://www.globus.org/subscriptions

Representative subscribers

Learn more on the GlobusWorld Tour

- **www.globusworld.org/tour/**
- **Webinar: October 6th, 1pm CDT**
- **SC16: Full-day developer tutorial (Booth #436)**

Join the Globus community

- Access the service: **globus.org/login**
- Create a personal endpoint: **globus.org/app/endpoints/create-gcp**
- Create a server endpoint: **docs.globus.org/resource-provider-guide**
- Need help? **support@globus.org**
- Subscribe: **globus.org/subscriptions**
- Follow us: **@globusonline**