

# Globus – A Platform for Research Data Management Challenges

ARGONNE  
**ATPESCC2023**  
EXTREME - SCALE COMPUTING

**Greg Nawrocki**  
[greg@globus.org](mailto:greg@globus.org)  
[nawrocki@uchicago.edu](mailto:nawrocki@uchicago.edu)  
[nawrocki@anl.gov](mailto:nawrocki@anl.gov)



THE UNIVERSITY OF  
**CHICAGO**





Globus is ...

a non-profit service  
developed and operated by



THE UNIVERSITY OF  
CHICAGO



## Our Mission

Increase the efficiency and effectiveness of researchers engaged in data-driven science and scholarship through **sustainable** software

 Development is funded by...



U.S. DEPARTMENT OF  
**ENERGY**



THE UNIVERSITY OF  
**CHICAGO**



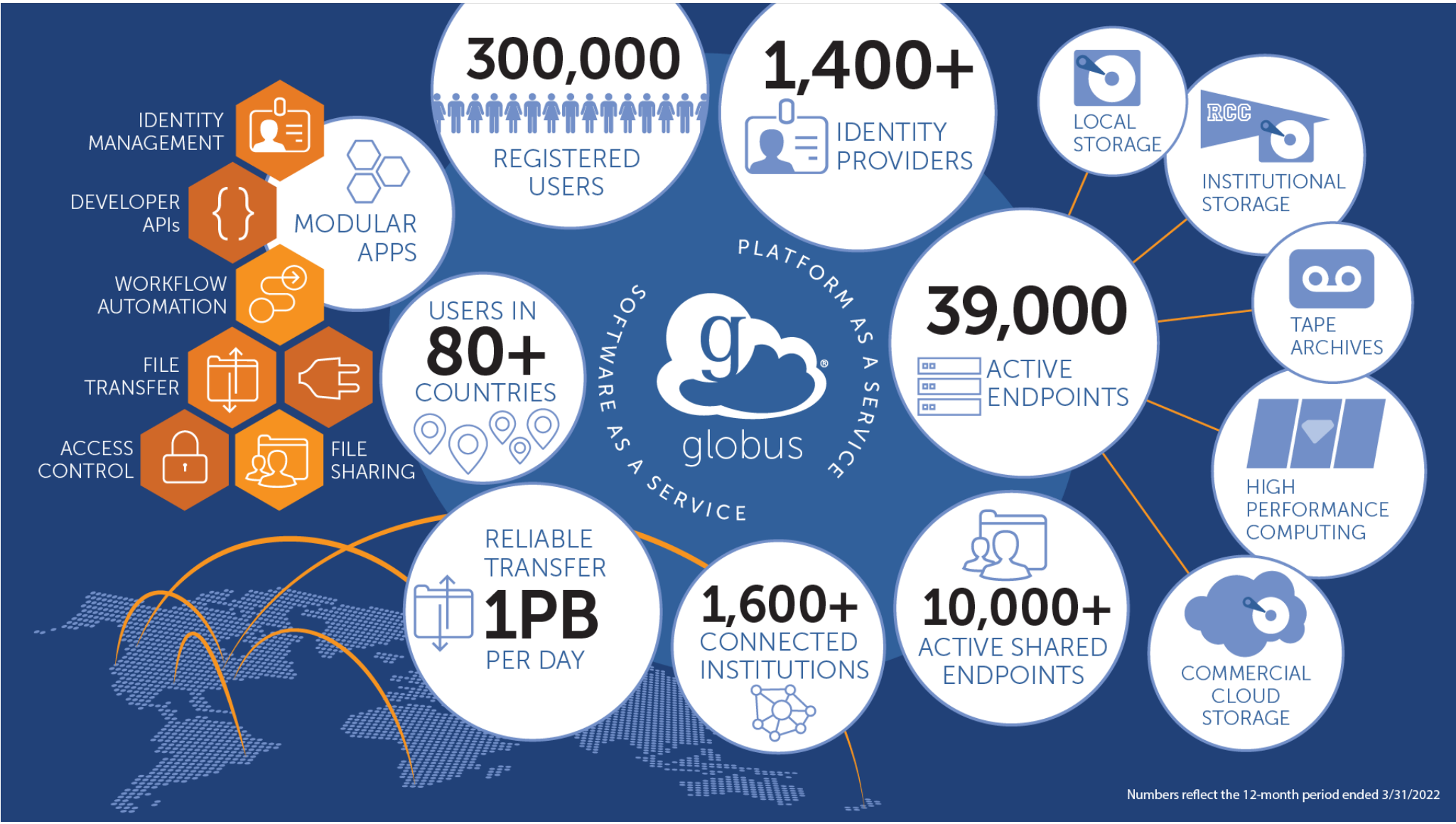
**NIST**  
National Institute of  
Standards and Technology  
U.S. Department of Commerce

Argonne   
NATIONAL LABORATORY



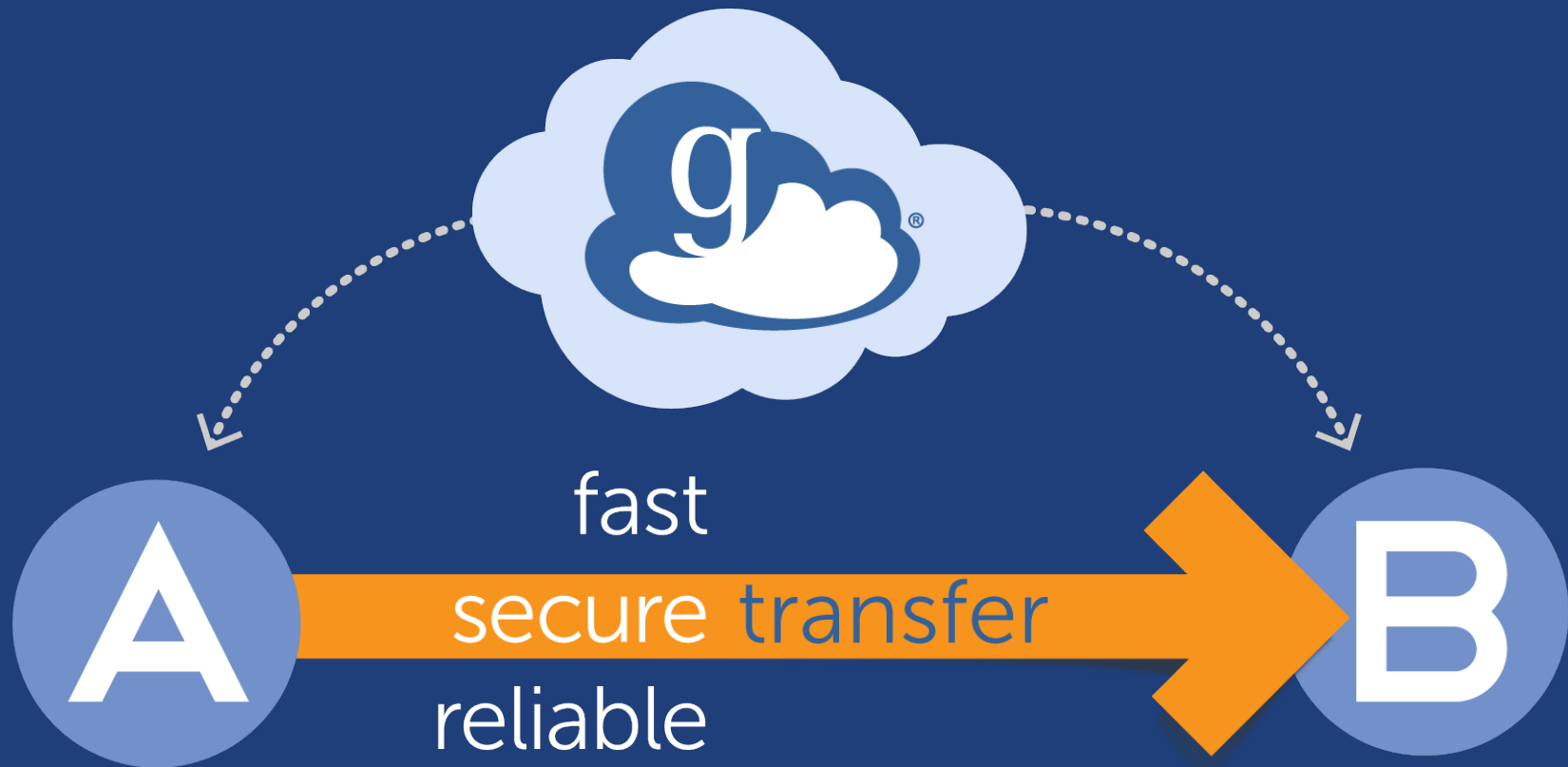
# Operations are funded by subscribers





Numbers reflect the 12-month period ended 3/31/2022

2,094,846,905,883 MB TRANSFERRED





Globus delivers...

Fast and reliable big data transfer,  
sharing, and platform services...

...directly from your own storage  
systems...

...via software-as-a-service using  
existing identities with the overarching  
goal of...





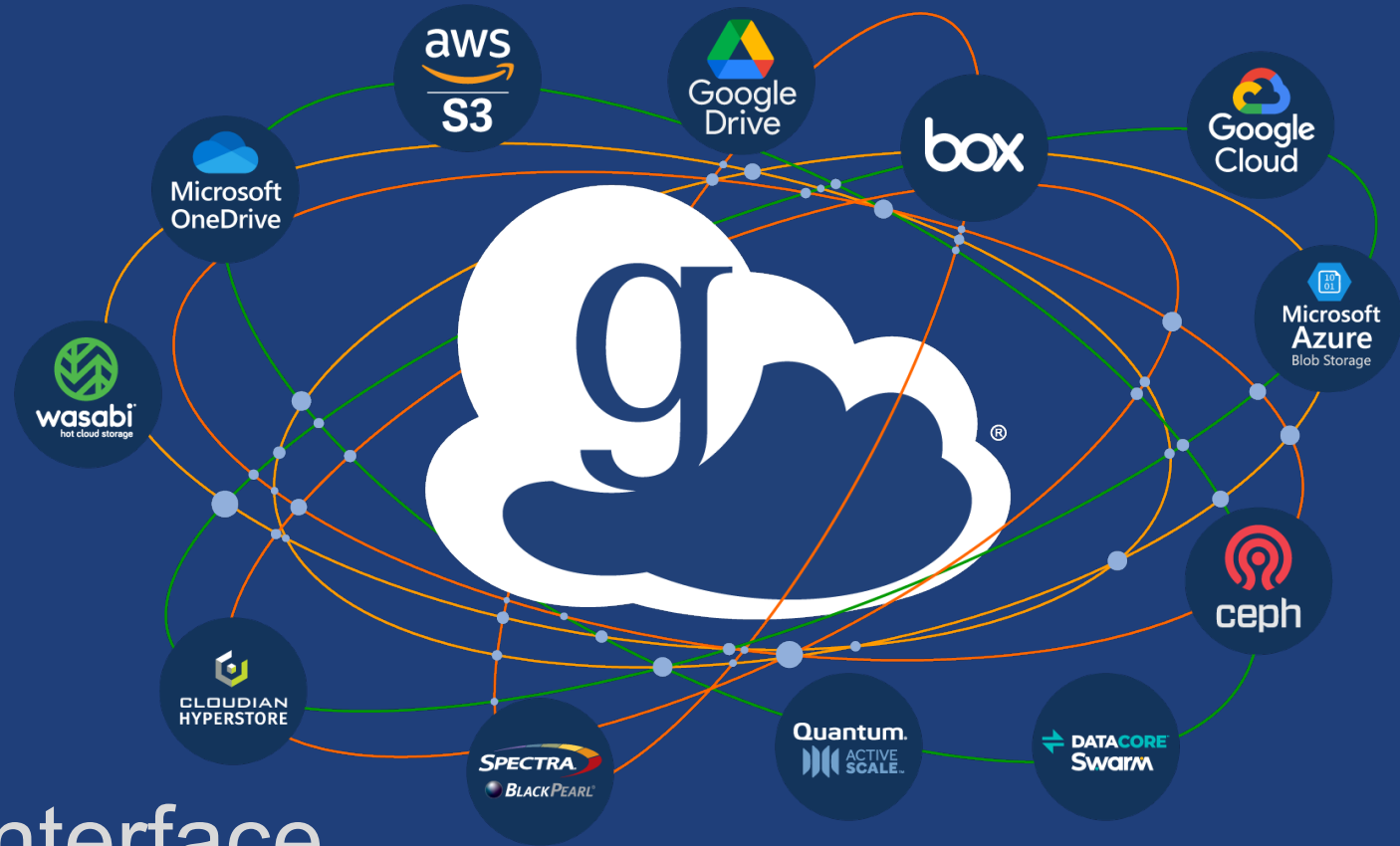
# ...unifying access to data across tiers





# PREMIUM globus connectors





Uniform interface  
Consistent user experience

# Globus SaaS / PaaS: Research data lifecycle

Instrument



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

2 Globus transfers files reliably, securely

Transfer

Compute Facility



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

Share

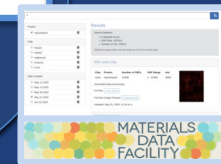
4 Globus controls access to shared files on existing storage; no need to move files to cloud storage!

6 The Timer and Automation Services, the Command Line Interface, API sets, and Python SDK provide the tools...

Build



7 ... for building science gateways, portals, publication services...



8 ...and automating research workflows - ensuring those that need access to the data have it.



- Use a Web browser or platform services
- Access any storage
- Use an existing identity

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



Personal Computer





# Endpoints, Collections and Globus Connect



- **Globus Connect Server**



- Multi user Linux Systems

- <https://docs.globus.org/globus-connect-server/>

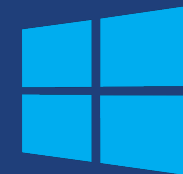
- **Globus Connect Personal**

- Personal Workstations and Laptops

- <https://www.globus.org/globus-connect-personal>

- OS specific instructions

- <https://docs.globus.org/how-to/>





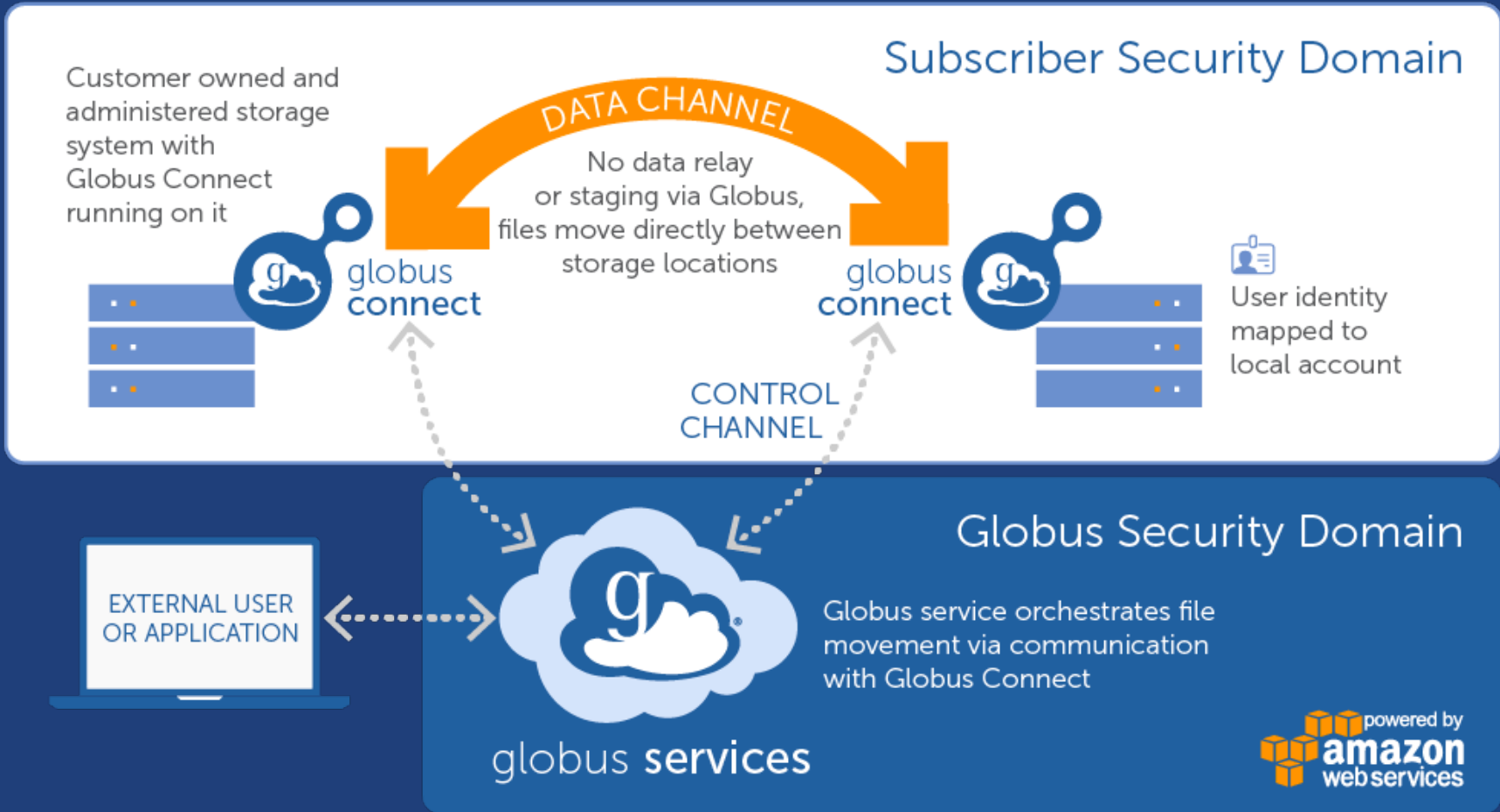
**GCSv4 has been  
deprecated  
July 2023 and will be  
discontinued December  
2023**



**Migration tools are available, please  
update!**

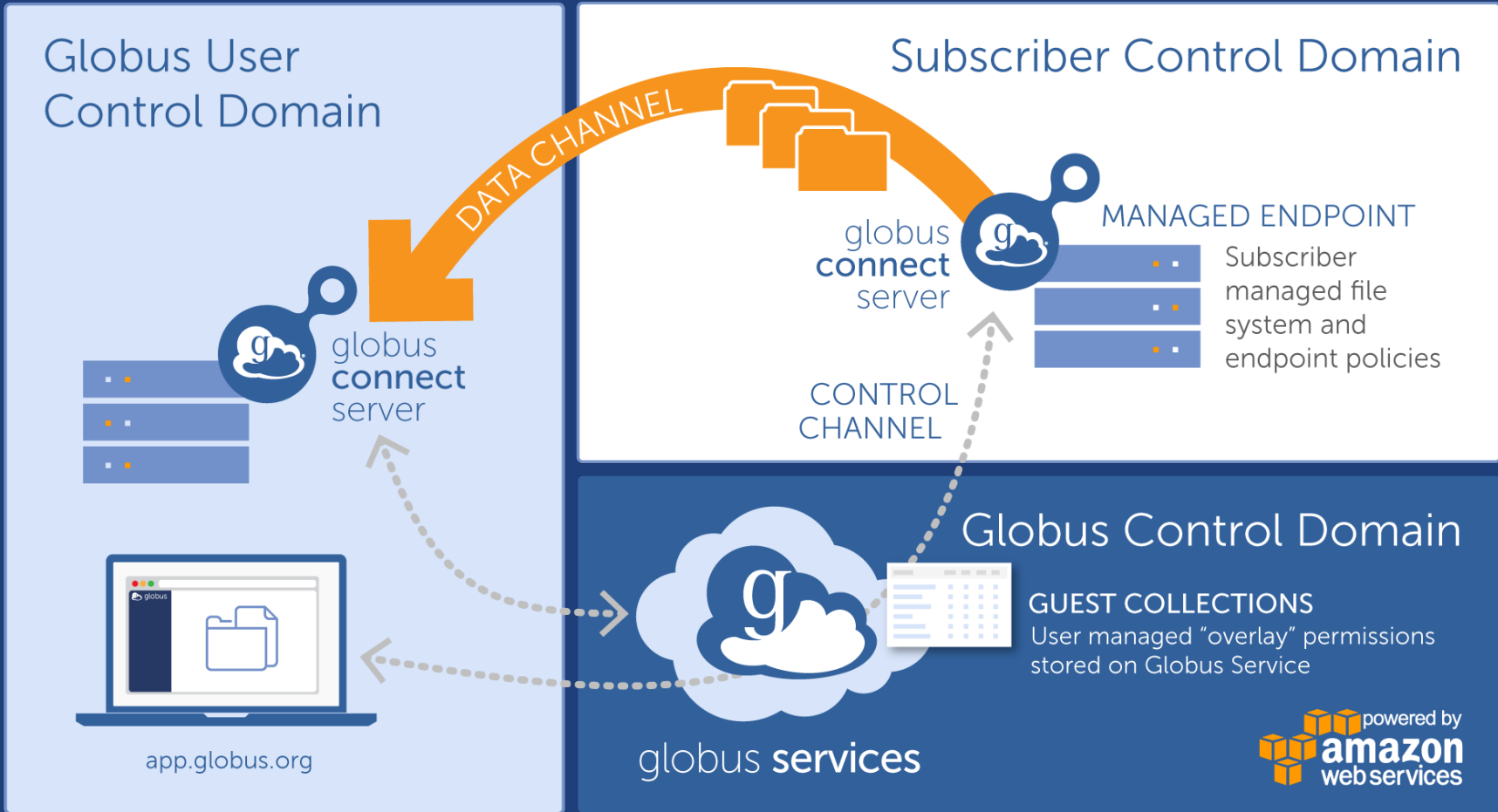


# Hybrid SaaS – Transfer – Mapped Collections





# Hybrid SaaS – Sharing – Guest Collections







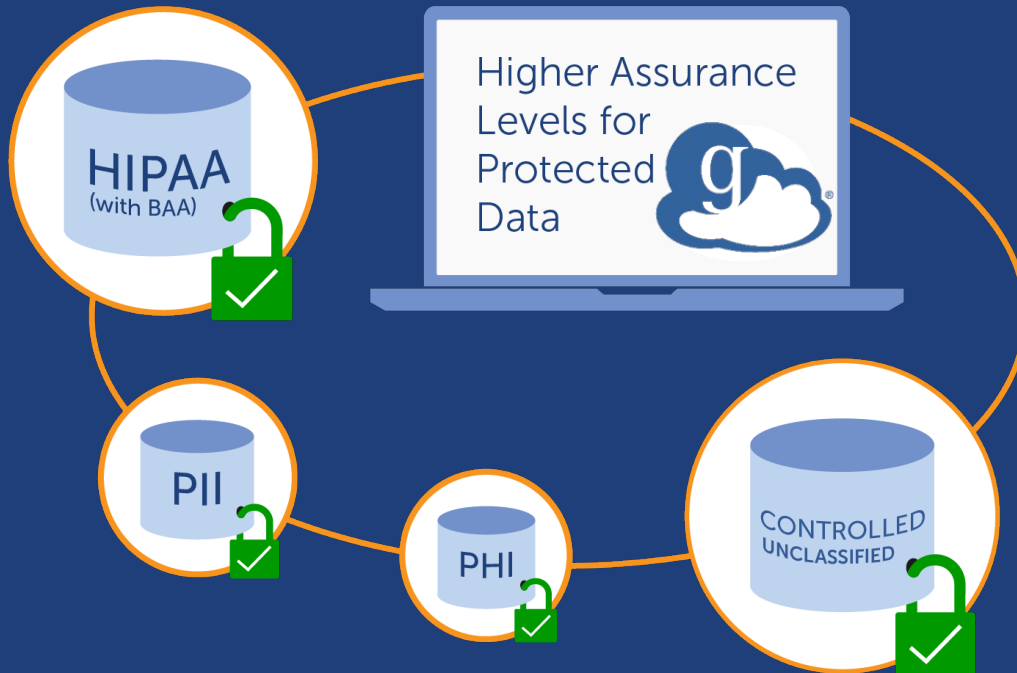
# Globus core security features



- **Access Control**
  - Identities provided and managed by institution
  - Institution controls all access policies
  - Globus is identity broker; no access to/storage of user credentials
  - Fine grained access control on the collections
- **Data remain at institutions, not stored by Globus**
- **Data does not flow through the Globus Service but directly between Endpoints and their Collections**
- **Integrity checks of transferred data**
- **High availability and redundancy**
- **Encryption of user files and Globus control data**

# Manage Protected Data – Subscription Level

## Higher assurance levels for HIPAA and other regulated data



- Support for protected data such as health related information
- Share data with collaborators while meeting compliance requirements
- Includes BAA option
  - Business Associate Agreements (BAA) will be between University of Chicago and our subscribers



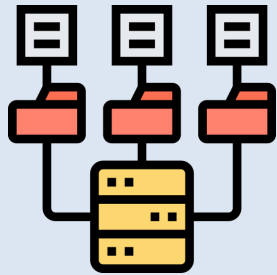
## High Assurance features for PHI, CUI

- **Additional authentication assurance**
  - Reauthentication after specified time period
  - Authenticates with the specific identity within session
- **Isolation of applications**
  - Authentication context is per application, per session
- **Enforced encryption of data in transit**
- **Local audit logging**
- **Option to require MFA**
- **Prevention of anonymous or open sharing**
- **Follow HIPAA, NIST SP 800-171, NIST SP 800-53 standards**





# Subscription growth drivers



## Premium Storage Connectors

9.5% → 32.0%

2021-22

2022-23



28.6% → 45.8%

## Protected Data Management



# Transfer and Sharing Demo

Sharing

Responsive  
Interface

Identities and  
Accounts

Transfer

Bookmarks

Groups

The Activity Monitor

Roles

The Console

Transfer Details

The Hamburger  
Menu



# Globus Mapped Collections at the ALCF

- **Swift**
  - /home via alcf#dtn\_home
- **Theta**
- **Grand**
- **HPSS**
- **<https://docs.alcf.anl.gov/data-management/data-transfer/using-globus/>**



## Globus Sharing (Guest Collections) at the ALCF

- Eagle
- Sharing on Eagle Using Globus
- <https://www.alcf.anl.gov/support-center/theta-and-thetagpu/eagle-data-sharing#Accessing%20your%20Eagle%20Project%20Directory>
- Just Google “ALCF Globus Eagle”



# Globus is SaaS and PaaS

Web app facilitates  
*ad hoc* data management

Platform services simplify  
creation of portals,  
gateways and applications  
for automating data  
management in instrument  
cores and other facilities







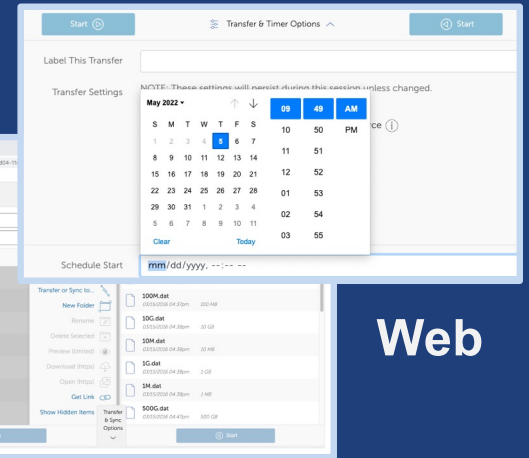
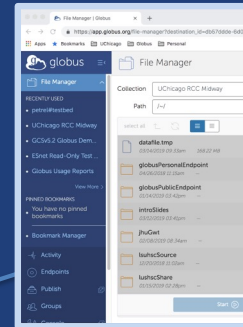
# One service, many interfaces

```
(globus-cli) jupiter:~ vas$ globus
Usage: globus [OPTIONS] COMMAND [ARGS]...

Options:
  -v, --verbose          Control level of output
  -h, --help             Show this message and exit.
  -F, --format [json|text] Output format for stdout. Defaults to text
  --map-http-status TEXT Map HTTP statuses to any of these exit codes:
                        0,1,50-99. e.g. "404=50,403=51"

Commands:
  bookmark      Manage Endpoint Bookmarks
  config        Modify, view, and manage your Globus CLI config.
```

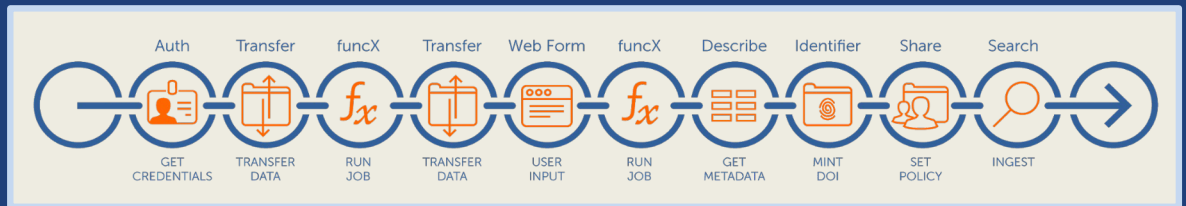
CLI



Web



Globus service



Flows

```
GET /endpoint/go%23ep1
PUT /endpoint/vas#my_endpt
200 OK
X-Transfer-API-Version: 0.10
Content-Type: application/json
...
```

Rest API

# Globus Platform and Automation Capabilities



## Timer Service

The Globus WebApp supports recurring and scheduled transfers. (*a.k.a. Globus cron*)



## Command Line Interface

The CLI provides an interface to Globus services from the shell and is suited to both interactive and scripting use cases.



## Globus API / SDK

Our open REST APIs and Python SDK empower you to create an integrated ecosystem of research data services and applications. Harness the power of the Globus platform so you can focus on building your application.



# Automation using Globus Flows

## Available to all Globus Subscribers\*

- **Managed, secure (Globus Auth), reliable, step function based task orchestration**
- **Support for heterogenous resources**
- **Extensible and authorable event driven execution model**
  - Flow Definition (JSON)
  - Input Schema (JSON)
  - Deployment
- **Extensible via custom actions**

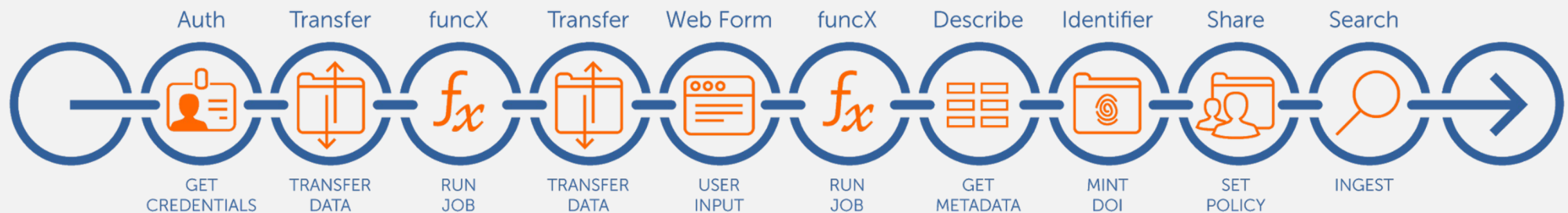


\* Non-subscribers can author one flow per user

# Managed automation of tasks

- **Flows:** A platform service for defining, applying, and sharing distributed research automation flows
- Flows comprise **Actions**
- **Action Providers:** Called by Flows to perform tasks
- **Triggers\***: Start flows based on events

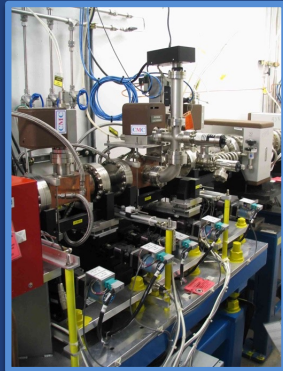
\* In development



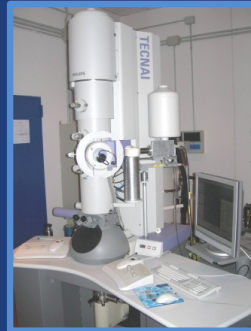
# Automation use case, manage data from instruments



Next-Gen Sequencer



Advanced Light Source



Cryo-EM



MRI



Light Sheet Microscope



Analysis store



High-durability, low-cost store



Remote visualization



Personal system





# Globus Flows Demo



# Globus Compute

**Managed, federated  
Function-as-a-Service for  
reliably, scaleably and  
securely executing functions  
on remote endpoints from  
laptops to supercomputers**





# Evolution of funcX





# Globus Compute – Function as a Service

- **Compute service** — Highly available cloud-hosted service; managed fire-and-forget function execution
- **Compute endpoint** — Abstracts access to compute resources; edge device to supercomputer
- **SDK** — Python interface for interacting with the service; familiar Globus look and feel
- **Security** — Leverages Globus Auth; user authentication and identity mapped to local account



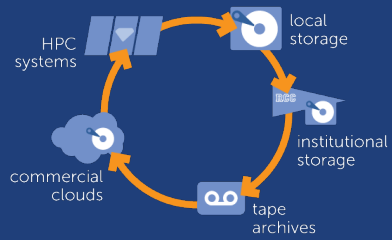


## Compute service will evolve rapidly

- **Multi-user compute endpoints**
- **Native integration with transfer for stage in and stage out of data for compute tasks**
- **Expanding compute service interfaces in the webapp for administrators and users**



# Research IT. Reimagined.



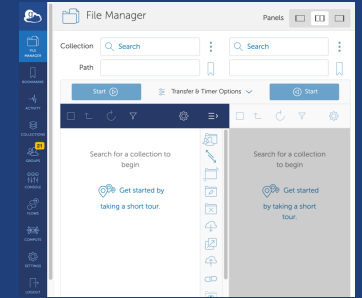
**Unified data access**



**Data sharing**



**Publication & discovery**



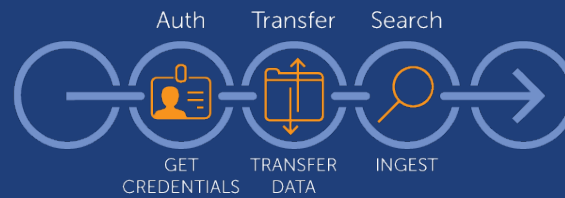
**Software-as-a-Service**



**Managed transfer & sync**



**Managed remote execution**



**Reliable automation**



**Platform-as-a-Service**



## Support Resources

- **Globus Documentation: [docs.globus.org](https://docs.globus.org)**
- **Globus Automation Services and Flows: [docs.globus.org/globus-automation-services/](https://docs.globus.org/globus-automation-services/)**
- **YouTube Channel: [youtube.com/user/GlobusOnline](https://youtube.com/user/GlobusOnline)**
- **Helpdesk: [support@globus.org](mailto:support@globus.org)**
- **Mailing Lists: [globus.org/mailing-lists](https://globus.org/mailing-lists)**
- **Customer engagement team (office hours)**
- **Professional services team (advisory, custom work)**

# Globus – A Platform for Research Data Management Challenges

ARGONNE  
**ATPESC2023**  
EXTREME - SCALE COMPUTING

**Greg Nawrocki**  
[greg@globus.org](mailto:greg@globus.org)  
[nawrocki@uchicago.edu](mailto:nawrocki@uchicago.edu)  
[nawrocki@anl.gov](mailto:nawrocki@anl.gov)



THE UNIVERSITY OF  
**CHICAGO**

