

Introduction to Globus for ALCF Users

Research Data Transfer

2022 ALCF Computational Performance Workshop
May 26, 2022

Greg Nawrocki

ggreg@globus.org

nawrocki@uchicago.edu

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



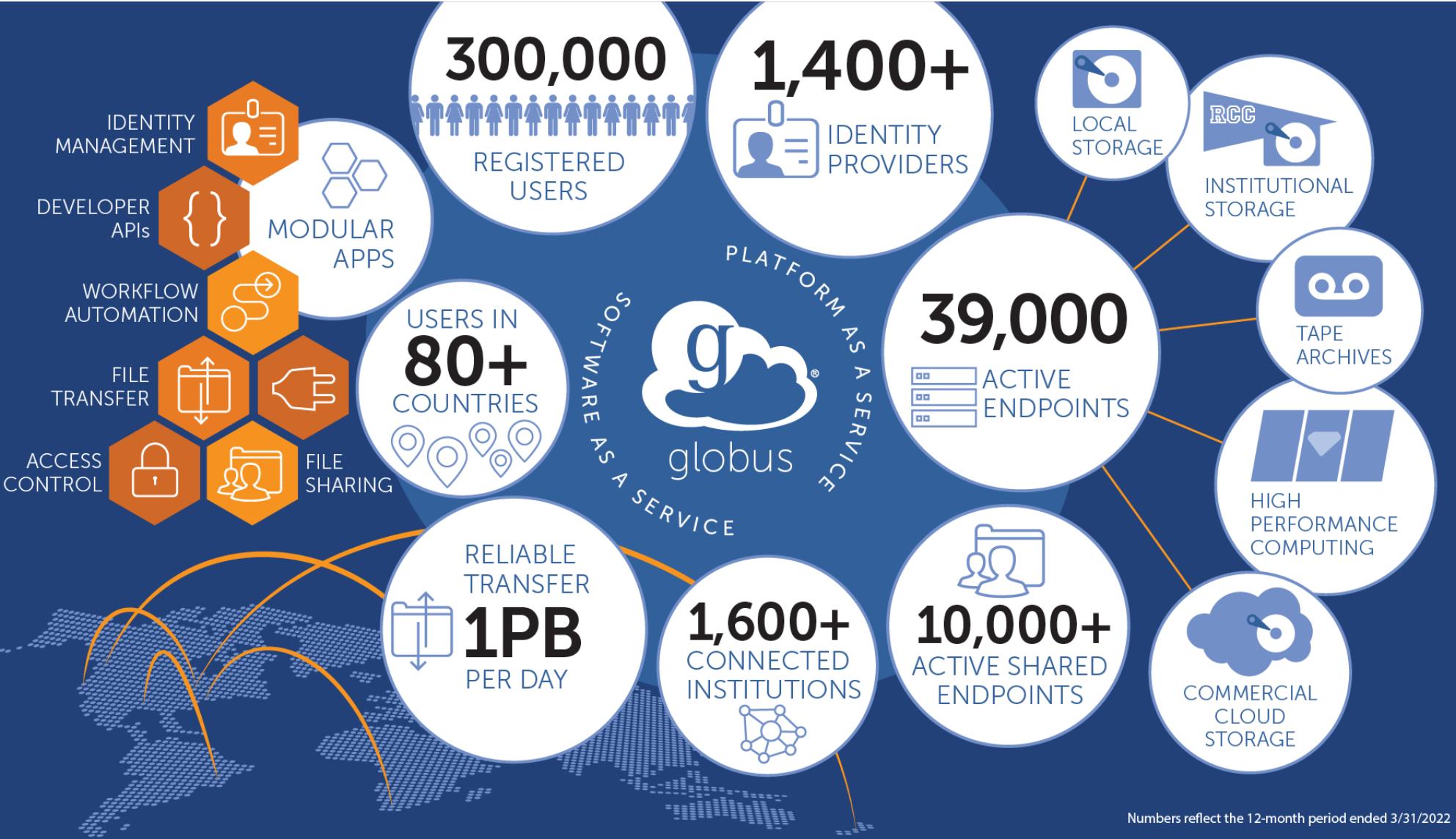
NIST
National Institute of
Standards and Technology
U.S. Department of Commerce

Argonne
NATIONAL LABORATORY

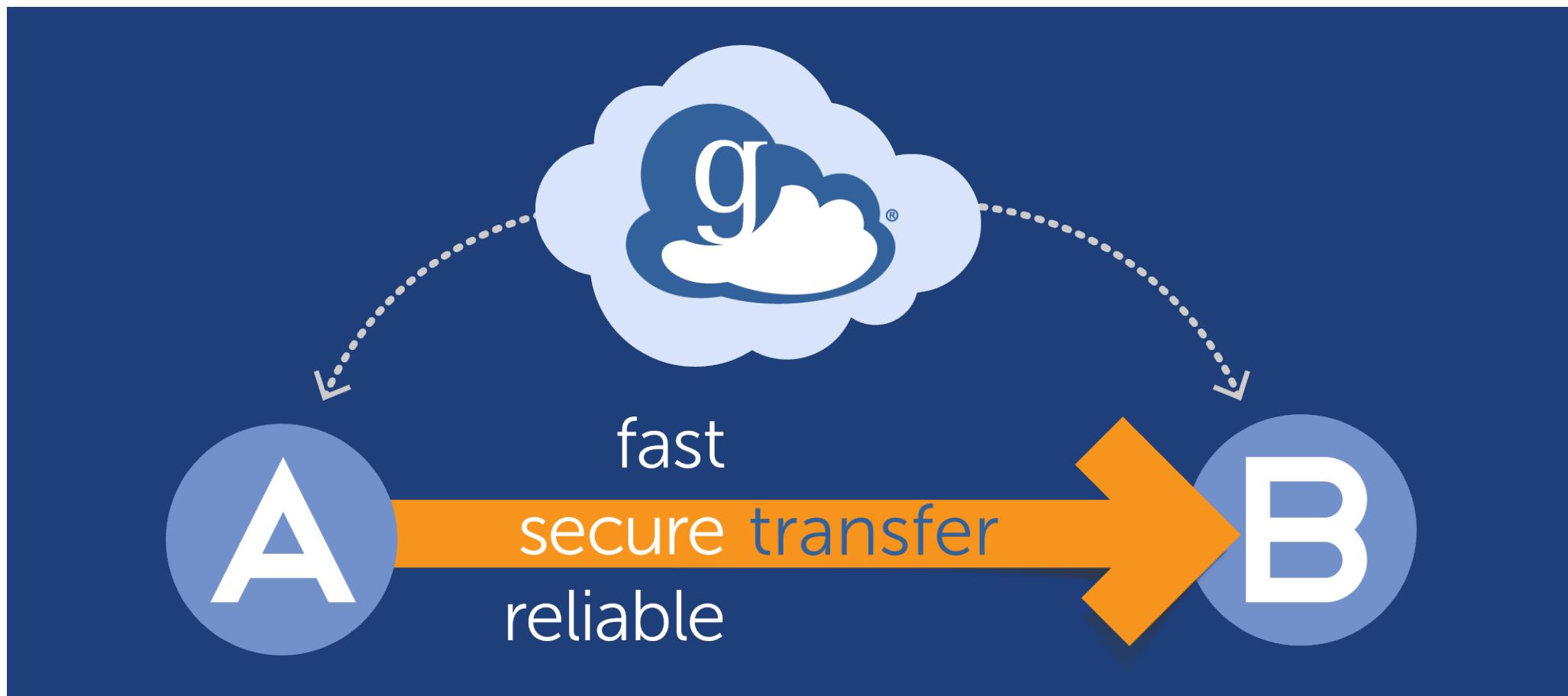


Operations are funded by subscribers





1,587,616,487,880 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



Google Cloud

Microsoft Azure
Blob Storage



IBM Cloud



Google Drive



box iRODS®

Quantum. ACTIVE SCALE™





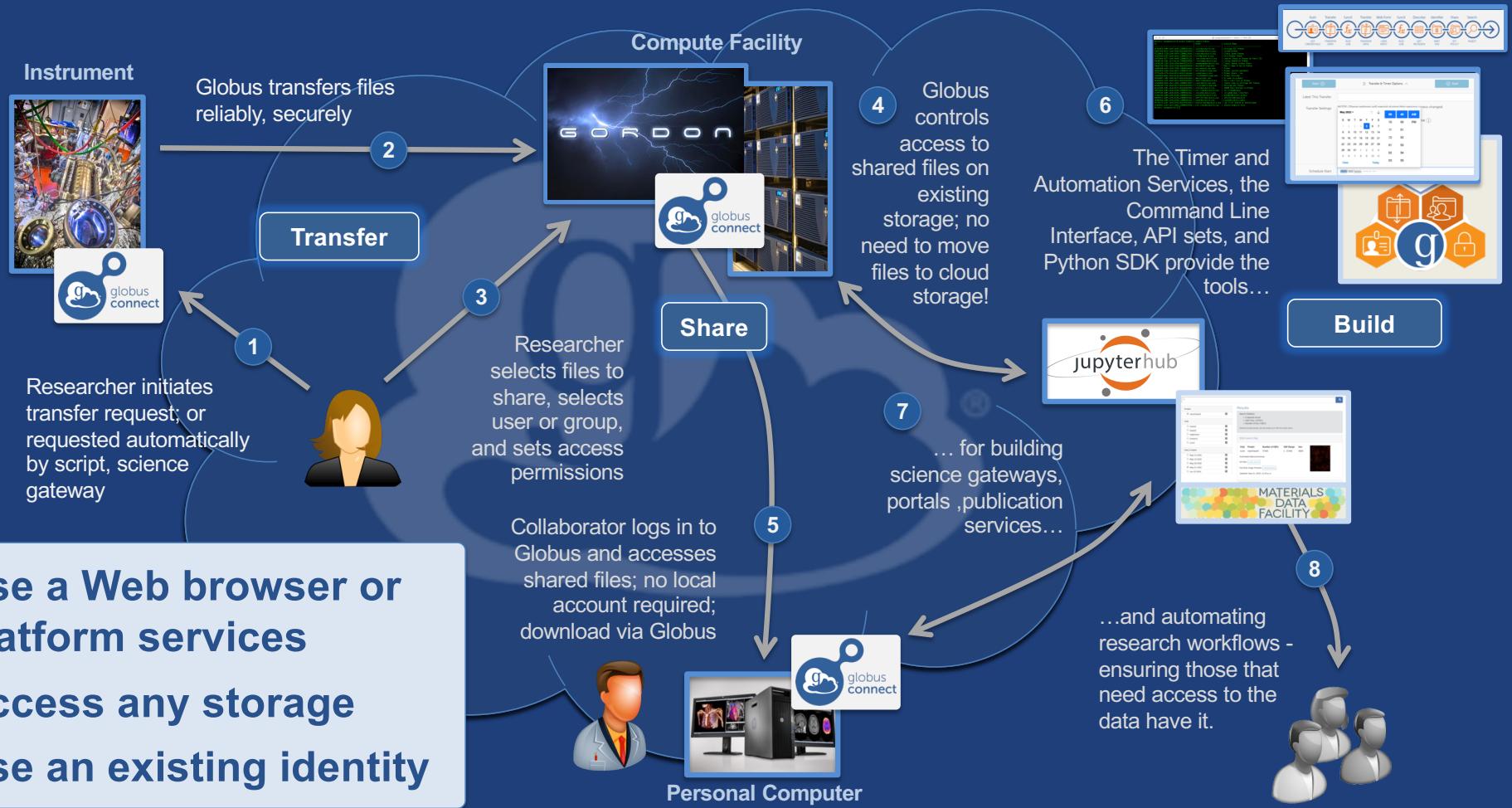
HPSS





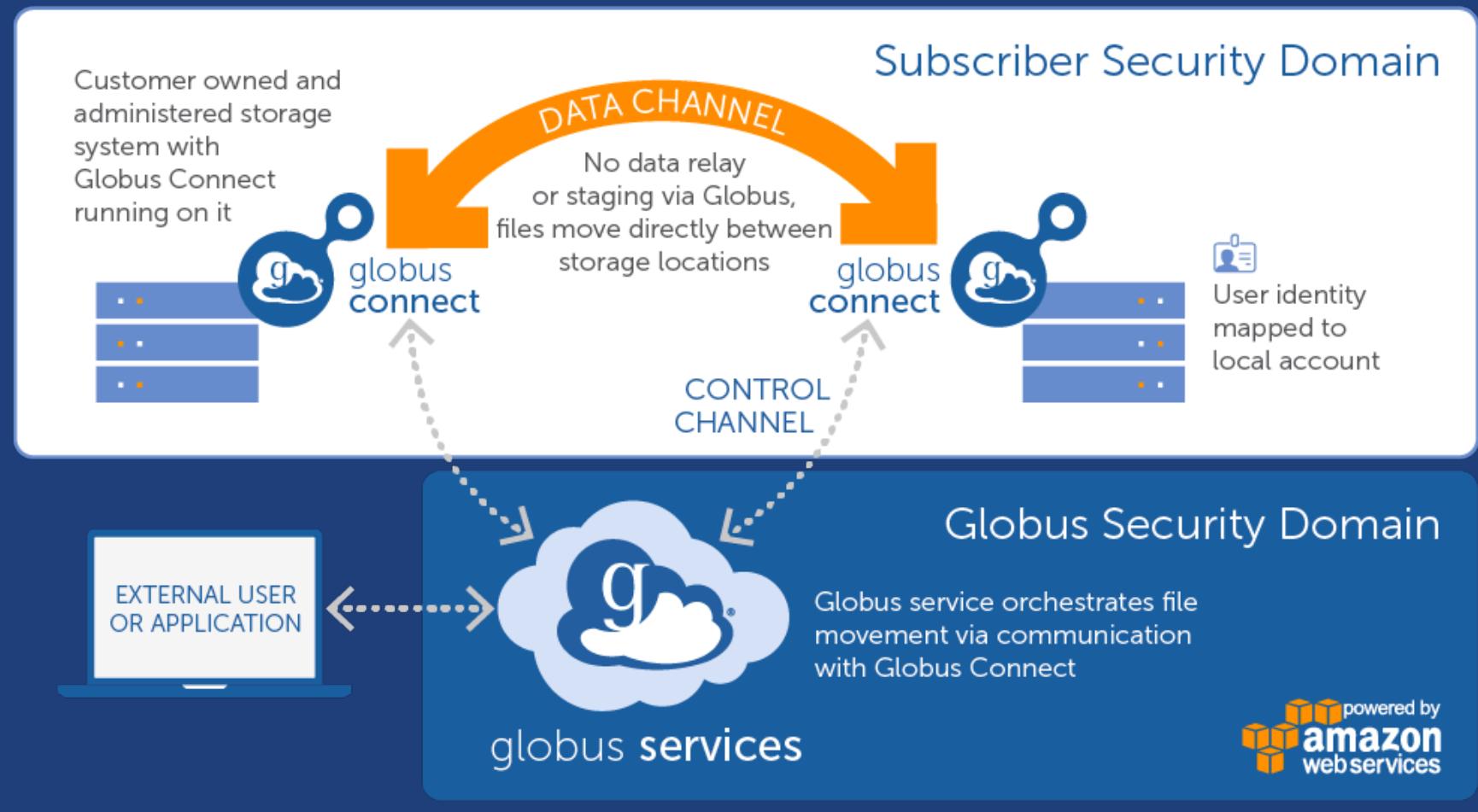


Globus SaaS / PaaS: Research data lifecycle



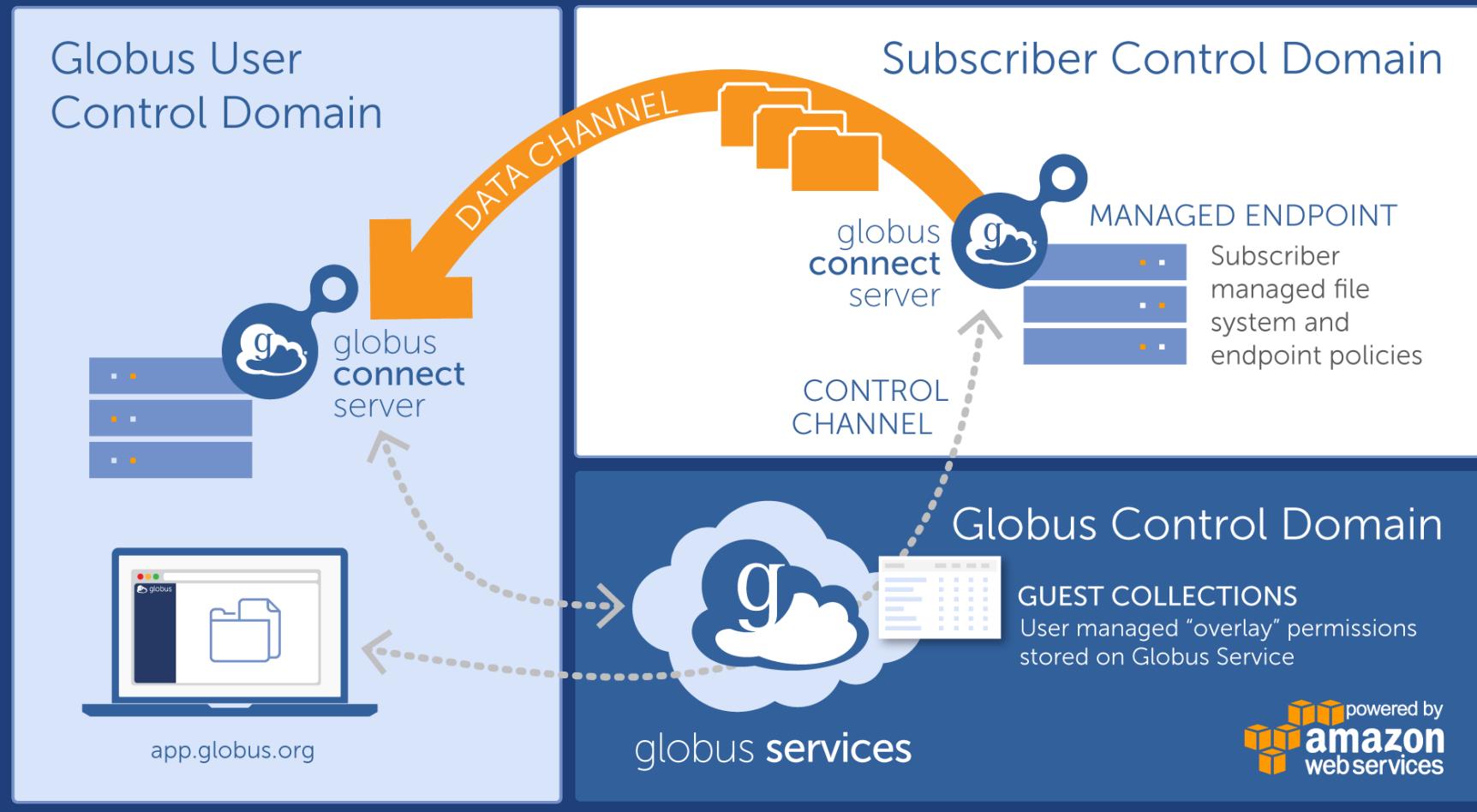


Hybrid SaaS – Transfer – Mapped Collections





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**





Collections and Globus Connect

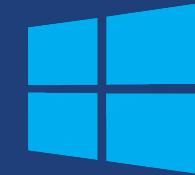
- **Globus Connect Server**

- 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/>





Globus Transfer Demo

- Authenticating to Globus
 - Linking in other identities
- The hamburger menu – file / folder management
- Transfer to and from ALCF resources
 - theta, cooley, thetagpu, grand
 - University of Chicago RCC Midway
- Globus Connect Personal
 - Installation
 - Transfer

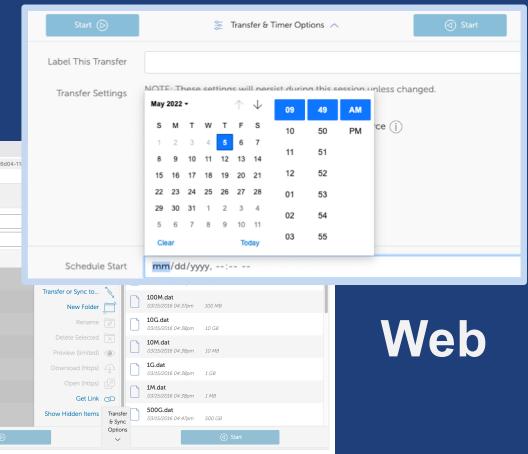
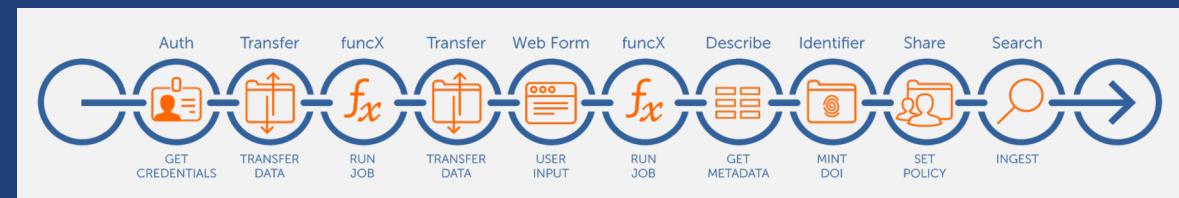


One service, many interfaces



CLI

```
(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" for the attack.
Commands:
  bookmark               Manage Endpoint Bookmarks
  config                 Modify, view, and manage your Globus CLI config.
```



Web

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

Rest
API



Globus Automation Capabilities



Timer Service

Scheduled and recurring transfers
(*a.k.a. Globus cron*)

Command Line Interface

Ad hoc scripting and integration



Globus Flows service

Comprehensive task (data and compute) orchestration with human in the loop interactions



Globus APIs

- Auth
 - Groups
 - Transfer
 - Search
 - Timer
 - Flows
 - GCS Manager
- **Globus Web App consumes public Transfer API**
 - **Resource named by URL (standard REST approach)**
 - **Globus APIs use JSON for documents**
- docs.globus.org/api/transfer



Globus Python SDK

- **Python client library for the Globus REST APIs**
- **Largely direct mapping to REST API**
- **globus_sdk.TransferClient class handles connection management, security, framing, marshaling**

globus-sdk-python.readthedocs.io/en/stable/
[globus.github.io/globus-sdk-python](https://github.com/globus-sdk/python)



Support Resources

- **Globus Documentation:** docs.globus.org
- **Globus Timer:** globus.org/blog/scheduled-and-recurring-transfers-now-available-globus-web-app
- **Globus CLI:** docs.globus.org/cli/
- **Globus Automation Services and Flows:** docs.globus.org/globus-automation-services/
- **YouTube Channel:** youtube.com/user/GlobusOnline



Developer References

- Globus API / SDK Documentation
 - Transfer API : docs.globus.org/api/transfer/
 - SDK: globus-sdk-python.readthedocs.io/en/stable/
- Globus GitHub: github.com/globus/
 - Jupyter Notebooks
 - Stand alone notebooks and hub integrations that walk through much of the functionality of our SDK
 - <https://github.com/globus/globus-jupyter-notebooks>
 - Automation Examples
 - Shell scripted CLI and Python module examples of common research data management use cases
 - <https://github.com/globus/automation-examples>

Introduction to Globus for ALCF Users

Research Data Transfer

2022 ALCF Computational Performance Workshop
May 26, 2022

Greg Nawrocki

ggreg@globus.org

nawrocki@uchicago.edu

nawrocki@anl.gov



THE UNIVERSITY OF
CHICAGO

Argonne
NATIONAL LABORATORY

