

# Globus: Platform for Data Driven Research

Rachana Ananthakrishnan  
Executive Director, Globus  
[ranantha@uchicago.edu](mailto:ranantha@uchicago.edu)



THE UNIVERSITY OF  
CHICAGO



IN-PERSON WORKSHOP

**Empowering Secure Data-Driven Research:  
Leveraging Science DMZ, Globus, and National  
Cyberinfrastructure Resources**



Globus is ...

a non-profit service  
developed and operated by



THE UNIVERSITY OF  
CHICAGO



Our mission is to...

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



# Deliver Platform for Research IT



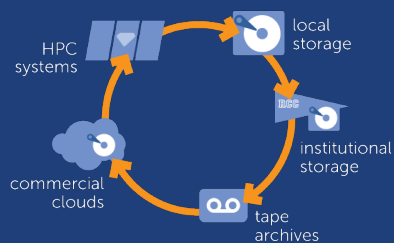
**Managed transfer & sync**



**Publication & discovery**



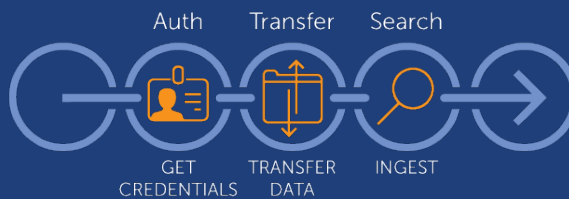
**Collaborative data sharing**



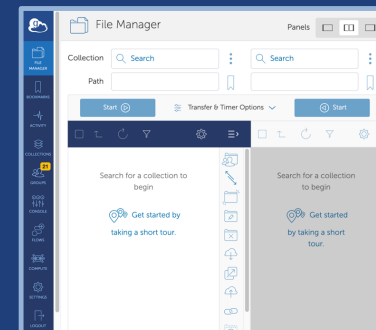
**Unified data access**



**Managed remote execution**



**Reliable automation**



**Software-as-a-Service**



**Platform-as-a-Service**





## Our freemium sustainability model

- **Basic capabilities are available free of charge to anyone engaged in non-profit research**
- **Subscriptions enable enhanced features for both researchers and system administrators**
- **Subscription pricing based on level of research supported**

[\*\*globus.org/subscriptions\*\*](https://globus.org/subscriptions)



**Which best describes your primary role at your institution?**



- ☐ Research computing manager/leader
- ☐ Research computing staff/System administrator
- ☐ Compliance/security staff/leader
- ☐ Researcher/postdoc
- ☐ Student
- ☐ Infrastructure/Service provider



# In service for science and scholarship...



## Digital agriculture – University of Winnipeg

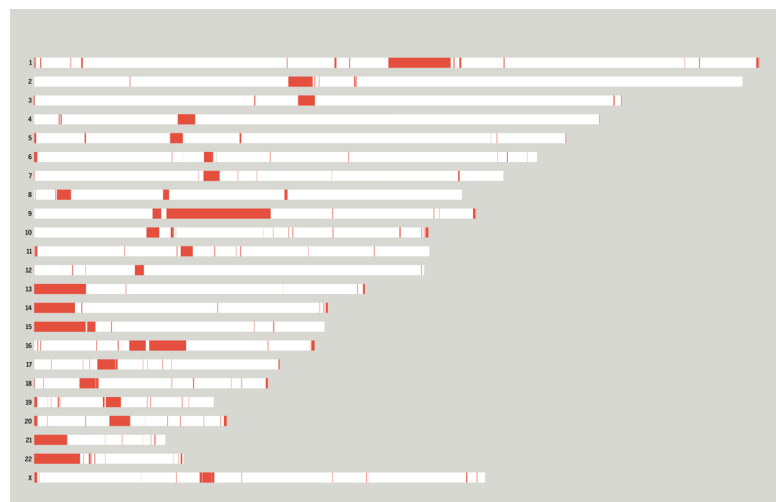
- Increasing crop yields using machine learning models
- Building training data sets
  - 40K images per day, tagged with metadata
  - Move data from diverse sources to campus storage, then onto Compute Canada HPC to run models
- Orchestrate data transfer using Globus CLI



Credit: Dilbarjot and Michael Beck,  
Physics and Applied Computer Science , University of Winnipeg



# Secure data sharing for international collaboration



Resolved sequences by the T2T-CHM13v2.0 reference genome.  
Resource: T2T consortium

[globus.org/user-stories/globus-enables-multi-institutional-data-sharing](https://globus.org/user-stories/globus-enables-multi-institutional-data-sharing)



# Instrument data delivery at scale



BIOMEDICAL RESEARCH CORE FACILITIES  
**ADVANCED GENOMICS CORE**  
UNIVERSITY OF MICHIGAN

**Use Globus to deliver  
100s of TB of genomic  
data to researchers**

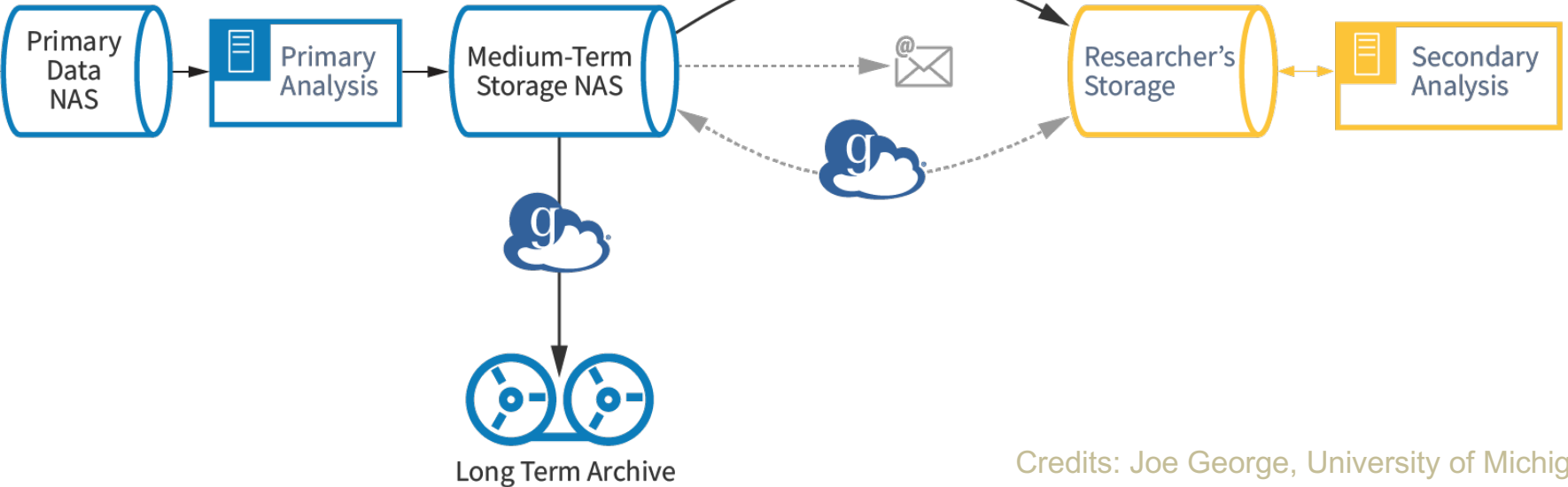
Sequencer



Sequencer



Sequencer

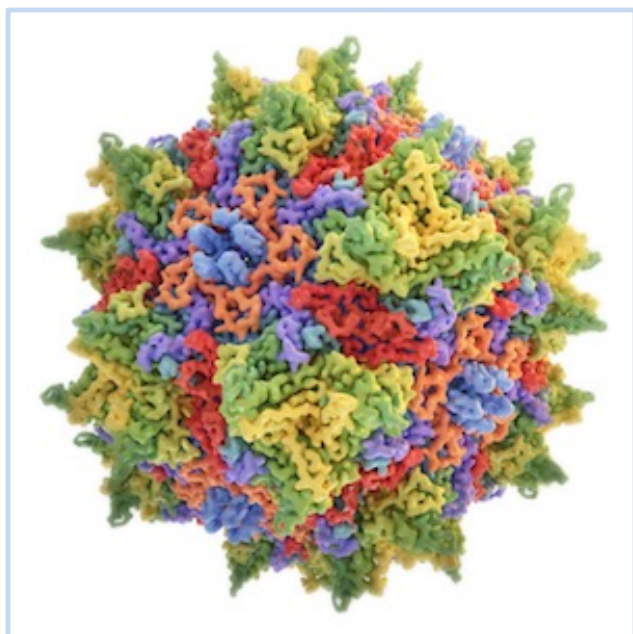


Credits: Joe George, University of Michigan



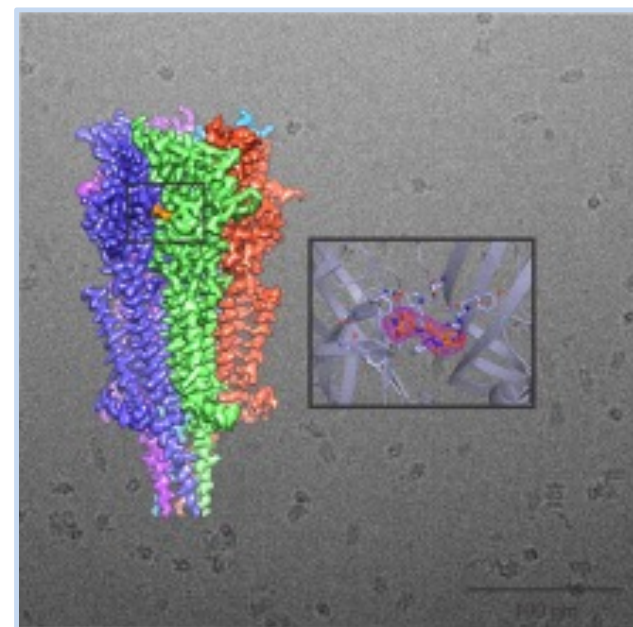
# Data Management at Cryo-EM Facilities

**Pacific Northwest Cryo-EM Processing Center  
(PNNL and Oregon Health Sciences University)**



Credit: <https://pncc.labworks.org/about-us>

**Case Western Reserve – Cryo-EM Core**



Credit: <https://case.edu/medicine/research/som-core-facilities/cryo-electron-microscopy-core>

**Globus for**

**automated data sync as new  
data is collected**

**provisioning of data access  
for researchers**

**reliable, secure data access  
for users**

**monitoring and management  
via console**





## Streamlining processing of field data

*Having a Globus Flow developed in collaboration with our Research Computing Colleagues and maintained in a library of flows allows high-speed computing to be **available to a larger number of potential users**. In my case, the Globus flow structure will allow me to incorporate collaborators and volunteers more easily into my research, **which increases community impact and engagement**.*

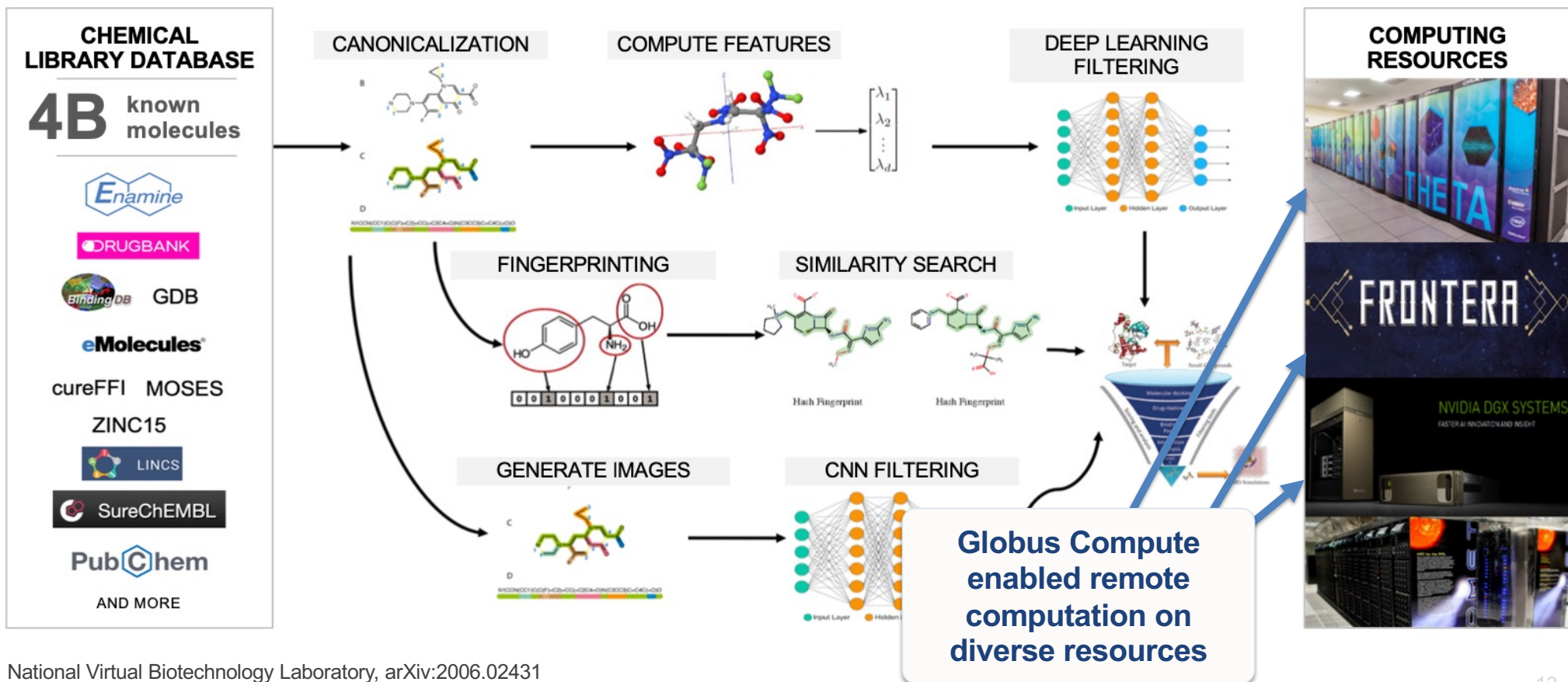
-Dan Ardia, Charles A. Dana Professor of Biology, F&M College

FRANKLIN & MARSHALL  
COLLEGE





# Exemplar Use Case: ML-based drug screening



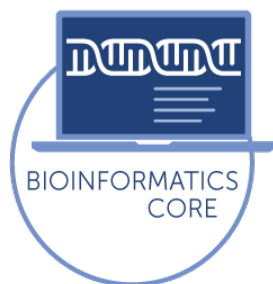
National Virtual Biotechnology Laboratory, arXiv:2006.02431



# Need to navigate...

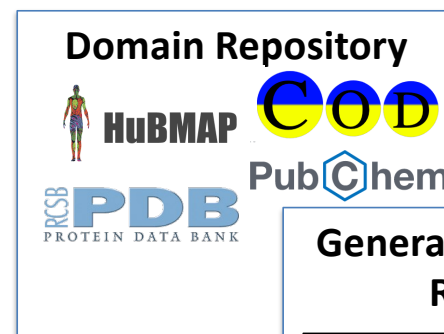


## Local/remote compute systems



## Experimental and core facilities

## Discovery and repository services





# Stakeholder perspectives or wishlist

## Researchers

No/low-friction path for access to data and computing systems by me and my collaborators

## Core Facility Staff

Reliable data egress maximize instrument utilization, and user support for analysis

## System Administrators

Managed, centralized visibility into system utilization, access policies, etc.

## Librarians/ Repository Managers

Simplified creation and implementation of DMPs, meet FAIR data standards

## Cybersecurity and Privacy Staff

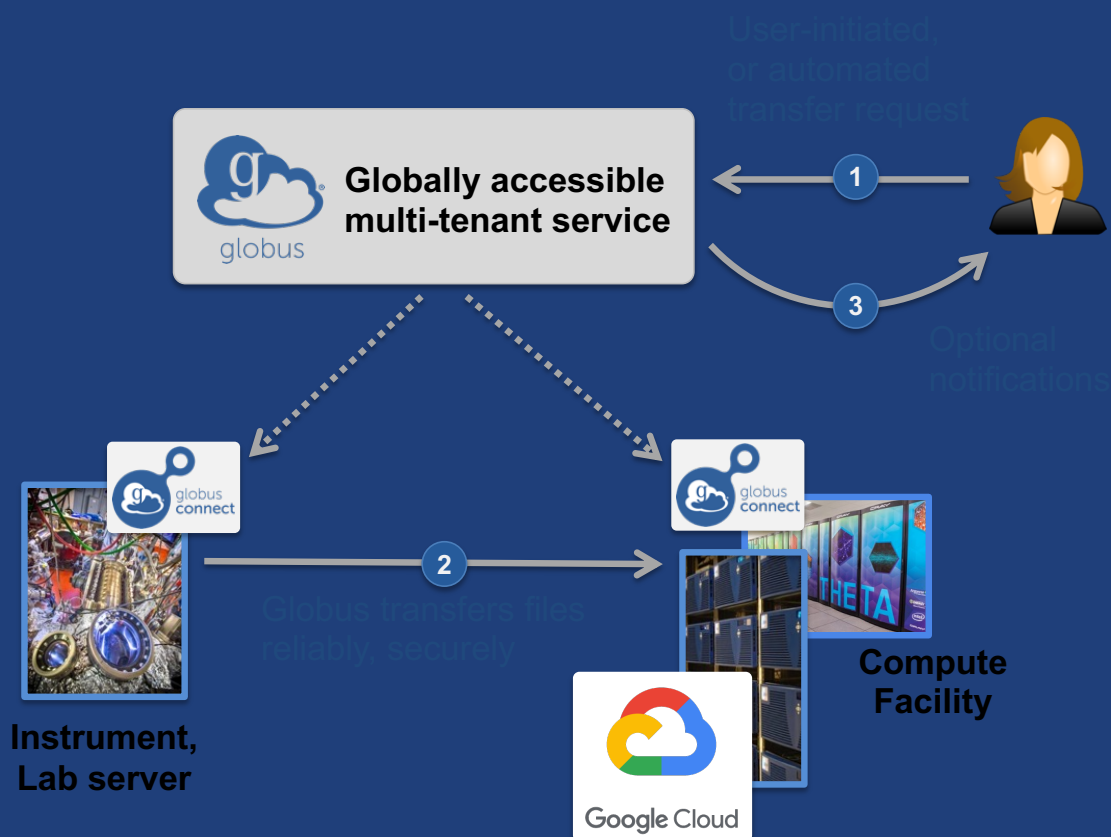
Auditable enforcement of variety of compliance requirements



# Unified research data access and sharing

# Fast, reliable file transfer ...from any to any system

- Fire-and-forget transfers/sync
- Optimized speed
- Assured reliability
- Unified view of storage
- HTTP/S access to data





# Intuitive web application interface

FILE MANAGER

BOOKMARKS

ACTIVITY

COLLECTIONS

21 GROUPS

CONSOLE

FLOW

COMPUTE

SETTINGS

LOGOUT

HELP & SITEMAP

File Manager

Panels

Collection

Path

Start

Transfer & Timer Options

Start

NAME	LAST MODIFIED	SIZE
E099_HPPG_100_55_025C_att06_...	3/17/2023, 11:2...	110.45 KB
E099_HPPG_100_55_025C_att06_...	3/17/2023, 11:2...	113.91 KB
esgf_demo	3/11/2023, 12:1...	—
GW_Demo	4/18/2023, 02:...	—
TestFolder	9/30/2022, 12:...	—
TestUser1	3/20/2023, 05:...	—

view

ALCF Username

Password

Cryptocard or Mobile token password

SIGN IN

This is a Federal computer system and is the property of the United States Government. It is for authorized use only. Users (authorized or unauthorized) have no explicit or implicit expectation of privacy.

Any or all uses of this system and all files on this system may be intercepted, monitored, recorded, copied, audited, inspected, and disclosed to authorized site, Department of Energy, and law enforcement personnel, as well as authorized officials of other agencies, both domestic and foreign. By using this system, the user consents to such interception, monitoring, recording, copying, auditing, inspection, and disclosure at the discretion of authorized site or Department of Energy personnel.



# Transfer/sync options

Start 1 Transfer & Timer Options Start

Label This Transfer

Transfer Settings

NOTE: These settings will persist during this session unless changed.

☒ sync - only transfer new or changed files ⓘ

where the modification time is newer

Files which are newer on the source will be overwritten by this option.

☐ delete files on destination that do not exist on source ⓘ

☐ preserve source file modification times ⓘ

☐ do NOT verify file integrity after transfer ⓘ

☐ encrypt transfer ⓘ

☐ Skip files on source with errors ⓘ

☐ Fail on quota errors ⓘ

Notification Settings

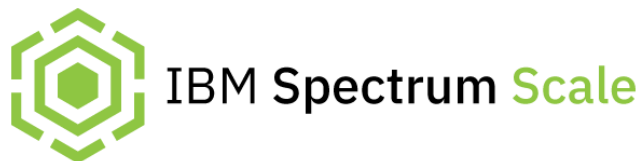
☐ Disable success notification ⓘ

☐ Disable failure notification ⓘ

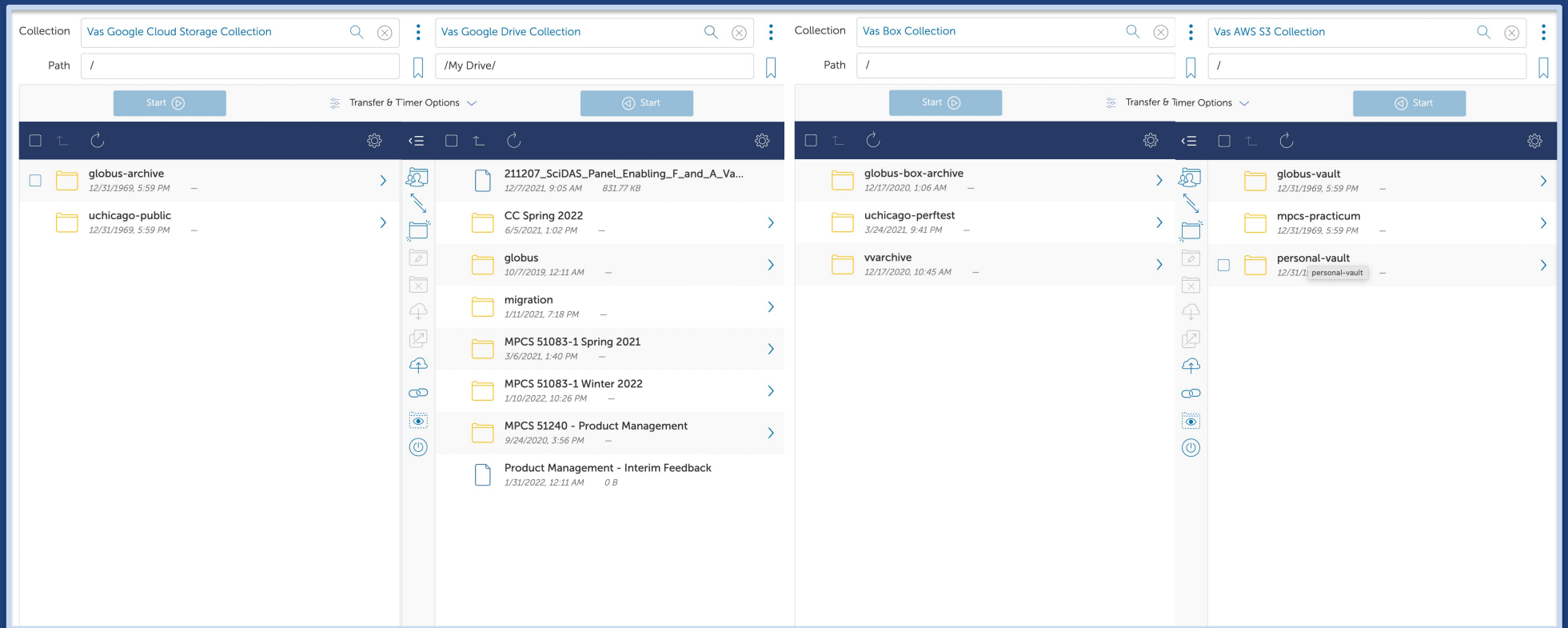
☐ Disable inactive notification ⓘ



# Globus Connectors support diverse systems



# Uniform interface, consistent user experience



The screenshot displays the Globus interface with four collections, each showing a list of files and folders. The interface is consistent across all collections, featuring a search bar, a path field, and a 'Start' button. The 'Transfer & Timer Options' dropdown is also present in each collection's header.

Collection	Path	File/Folder Name	Size	Created
Vas Google Cloud Storage Collection	/	globus-archive	-	12/31/1969, 5:59 PM
		uchicago-public	-	12/31/1969, 5:59 PM
		211207_SciDAS_Panel_Enabling_F_and_A_Va...	831.77 KB	12/7/2021, 9:05 AM
		CC Spring 2022	-	6/5/2021, 1:02 PM
		globus	-	10/7/2019, 12:11 AM
		migration	-	1/11/2021, 7:18 PM
		MPCS 51083-1 Spring 2021	-	3/6/2021, 1:40 PM
		MPCS 51083-1 Winter 2022	-	1/10/2022, 10:26 PM
		MPCS 51240 - Product Management	-	9/24/2020, 3:56 PM
		Product Management - Interim Feedback	0 B	1/31/2022, 12:11 AM
Vas Box Collection	/	globus-box-archive	-	12/17/2020, 1:06 AM
		uchicago-perftest	-	3/24/2021, 9:41 PM
		vvarchive	-	12/17/2020, 10:45 AM
Vas AWS S3 Collection	/	globus-vault	-	12/31/1969, 5:59 PM
		mpcs-practicum	-	12/31/1969, 5:59 PM
		personal-vault	-	12/31/1969, 5:59 PM

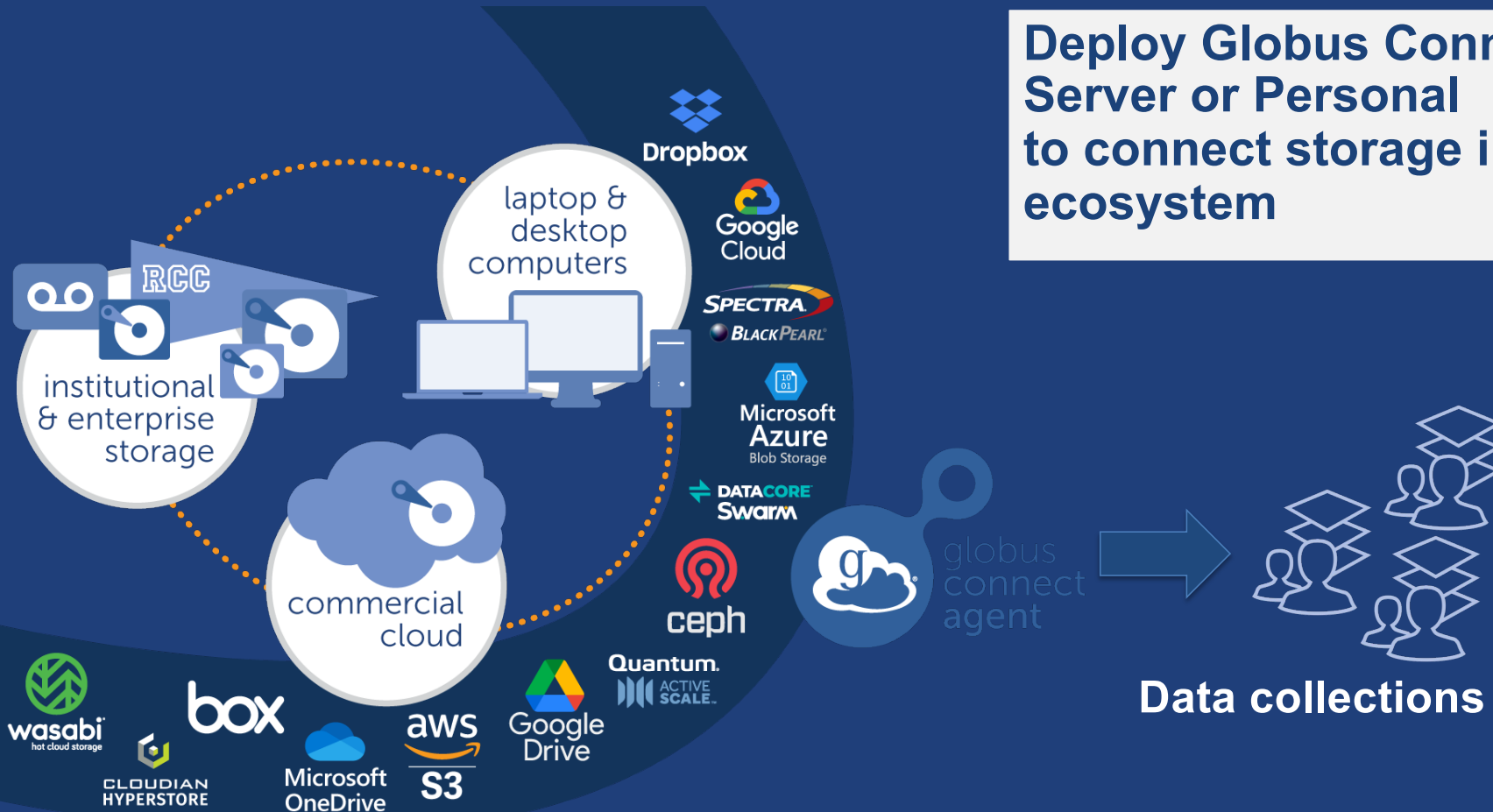
## What types of storage systems do you use/offer?

- ☐ Object store – on prem
- ☐ Tape archives
- ☐ AWS S3
- ☐ Google – Cloud/Drive
- ☐ MS – Blob/OneDrive
- ☐ Dropbox
- ☐ Box





# Enabling data ecosystem



**Deploy Globus Connect Server or Personal to connect storage into the ecosystem**

**Data collections**

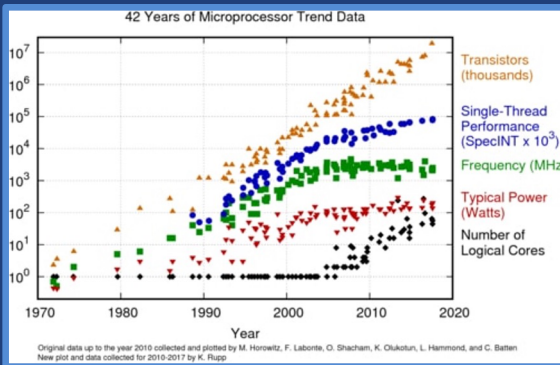


# Unified access to compute resources

# Why do we need to rethink research computing?

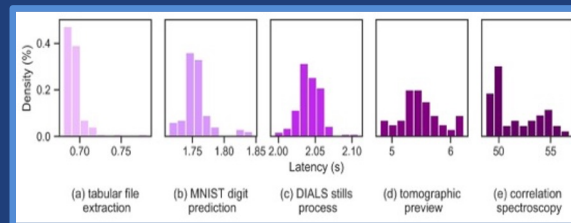
## Resources

- Hardware specialization
- Specialization leads to distribution



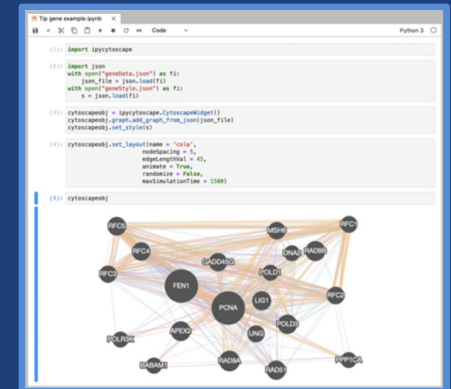
## Workloads

- Interactive, real-time workloads
- Machine learning training and inference
- Components may best be executed in different places



## Users

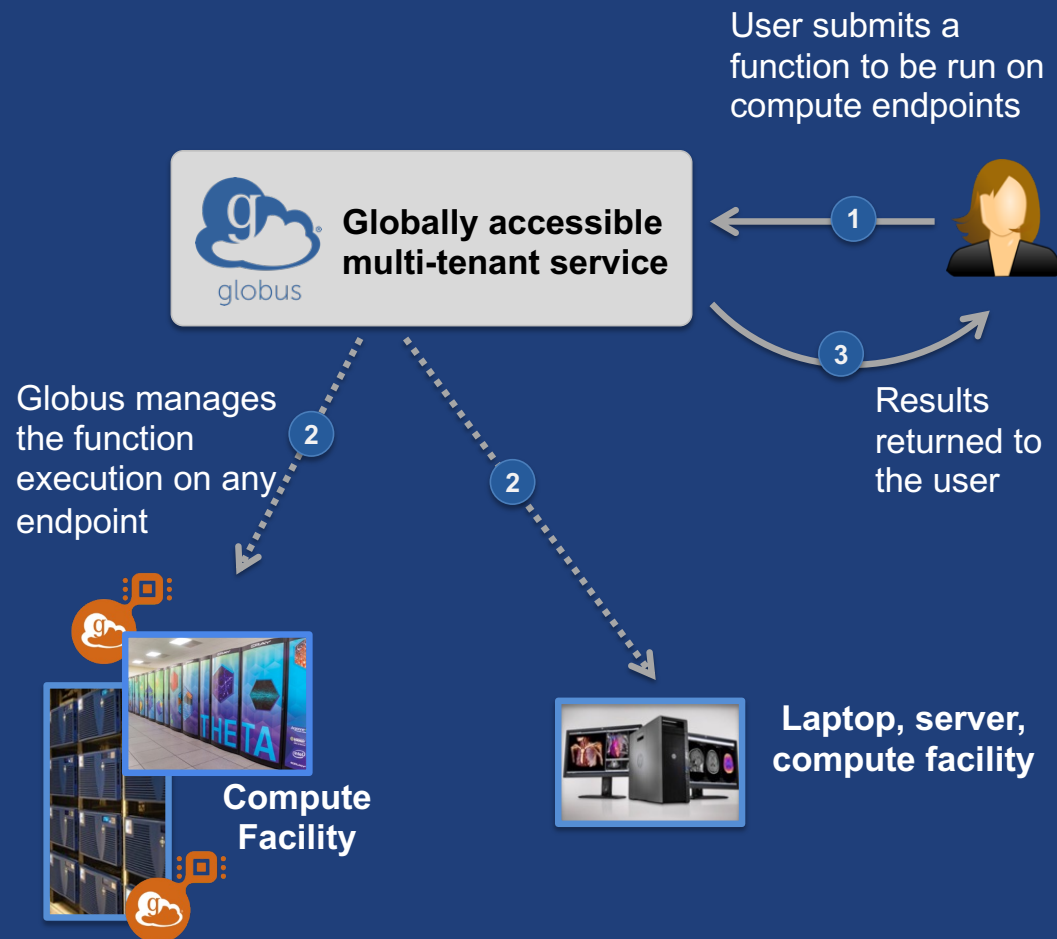
- Diverse backgrounds and expertise
- Different user interfaces (e.g., notebooks)





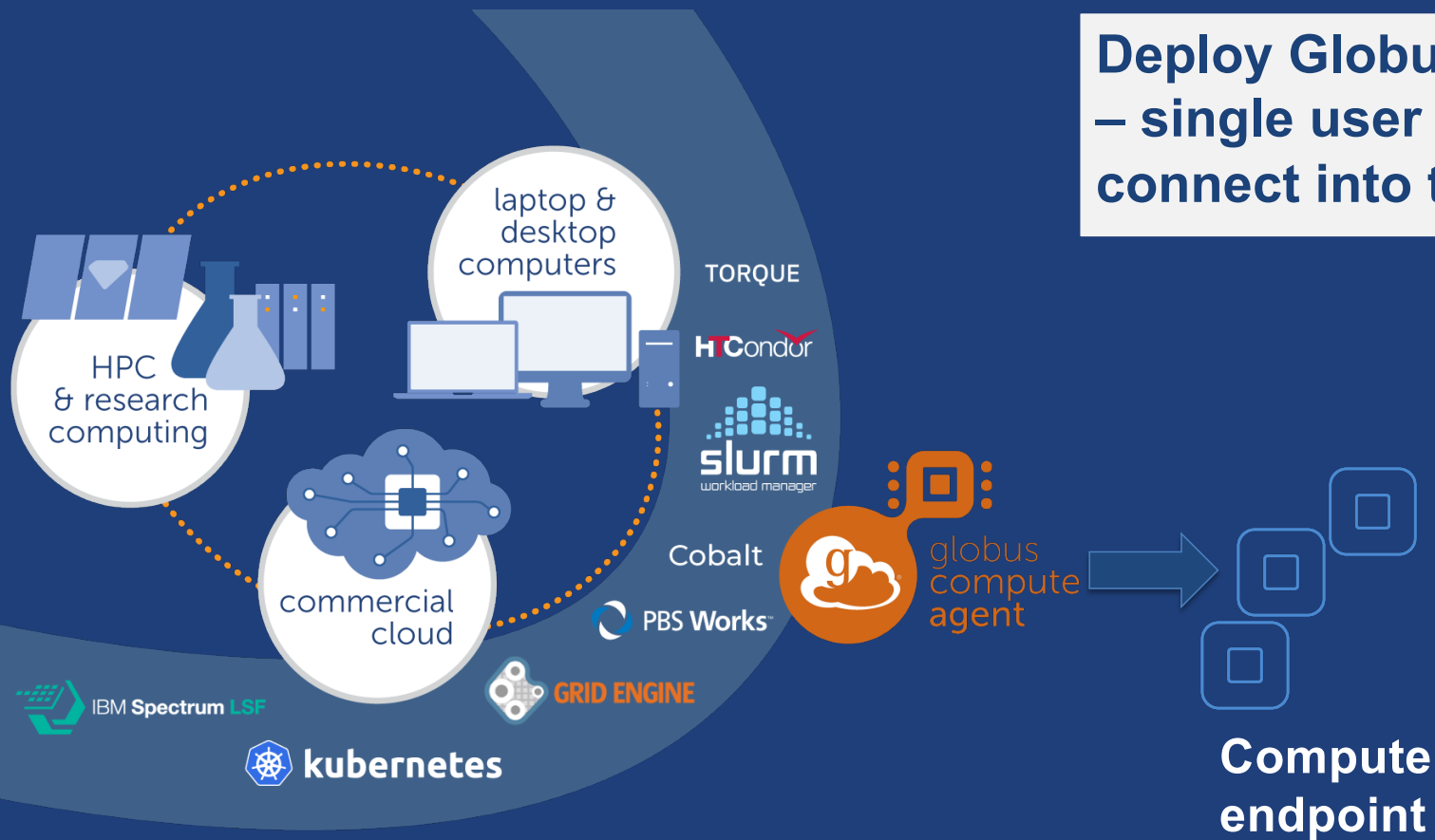
# Managed compute ...on any system

- **Programmatic access** compute resources
- **Consistent user interface**
- **“Fire and forget” function execution**
- **Federated authentication, and local access control**
- **Support use of Python for functions**





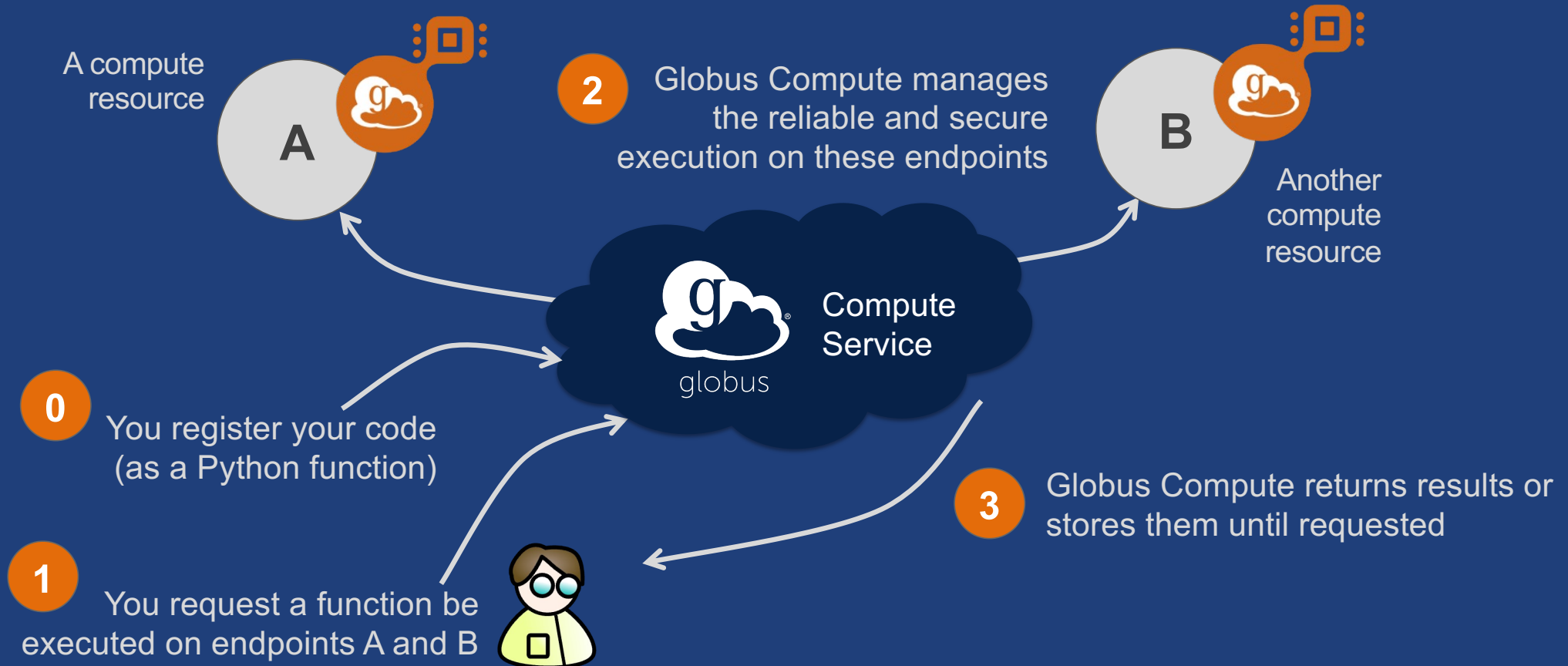
# Compute ecosystem



**Deploy Globus Compute agent  
– single user or multi user to  
connect into the ecosystem**



# How does it look from the researcher's PoV?





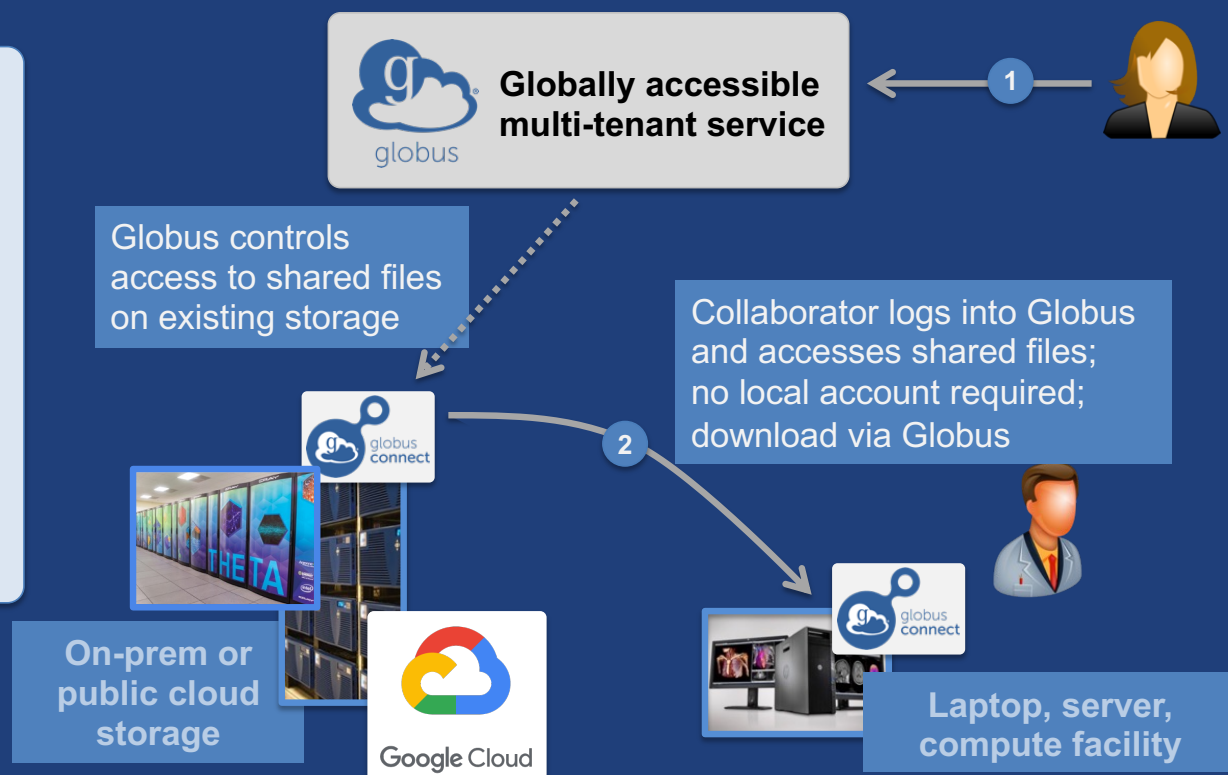
# Compliant and secure collaboration



# Secure data sharing ...from any storage

- Fine-grained access control “overlay” on storage system
- Share with any identity, email, group
- No need to stage data just for sharing

Select files to share, select user or group, and set access permissions





# Data sharing – permissions & roles

**Presentation Materials - RA**

Overview Permissions Roles

USER OR GROUP	CREATED	EXPIRATION	READ	WRITE
Permissions granted by role				
Path: /				
Brigitte Raumann (braumann@uchicago.edu)	3/8/2024, 04:33 PM	never expires	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Identity 6afa2dc5-d219-4078-9a06-ba37aa32c739 (6afa2dc5-d219-4078-9a06-ba37aa32c739@clients.auth.globus.org)	12/11/2023, 06:29 PM	never expires	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Path: /RPI/				
Public	2/27/2024, 03:00 PM			
Identity a7a2bd06-919b-478e-9477-447f14198a63 (foster@anl.gov)	5/3/2024, 04:27 PM			
Path: /Stanford/				
All Users	12/11/2023, 06:29 PM			

Overview Permissions Roles

Assigned Roles

[Assign New Role](#)

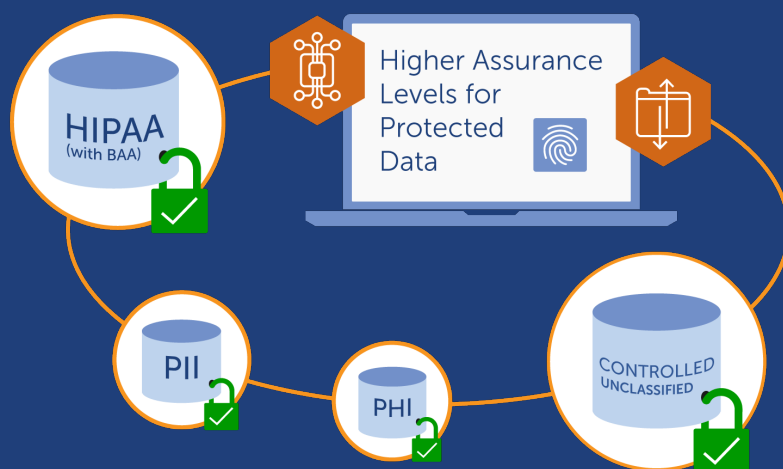
	Rachana Ananthakrishnan	ranantha@uchicago.edu	Owner	
	Rachana Ananthakrishnan	ranantha@uchicago.edu	Administrator	
	7d69a95e-48a8-491e-8e72-ed1631e81954	7d69a95e-48a8-491e-8e72-ed1631e81954@clients.auth.globus.org	Access Manager	

## Common sharing scenarios enabled by Globus

- **Disable sharing; not very FAIR, but may be necessary**
- **Allow sharing only by specific users/groups**
- **Allow sharing only from specific systems/directories**
- **Limit sharing to specific institution(s)**
- **Limit access based on time**

# Globus High Assurance for managing protected data

## Security controls



## Restricted data handling

→ PHI, PII, CUI  
→ Compliant data sharing

## BAA w/UChicago



## What are some of your compliance requirements?



- ☐ NIST 800-53
- ☐ NIST 800-171
- ☐ SOC-2
- ☐ HIPAA
- ☐ CMMC
- ☐ GDPR/State Privacy laws
- ☐ Others



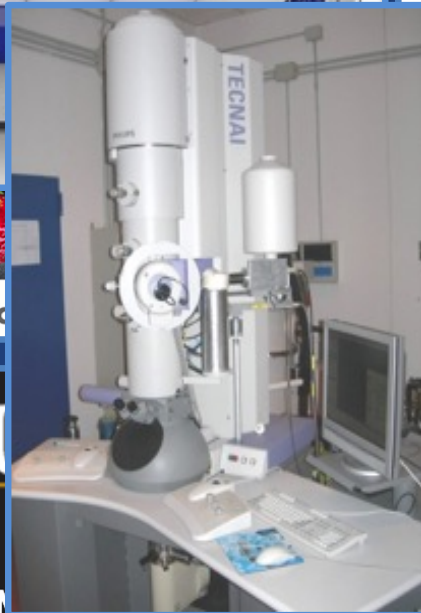
# Core Facility data automation





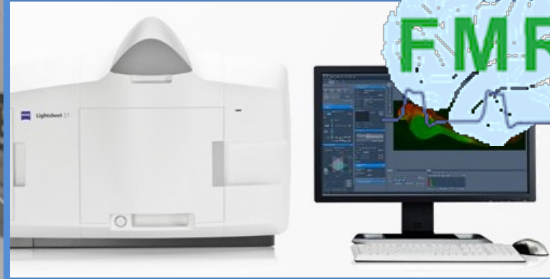
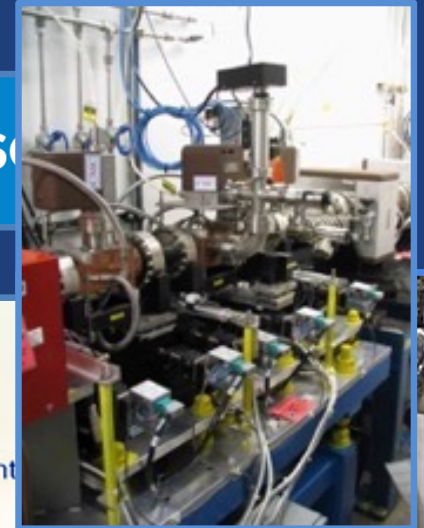
High Resolution  
Cryo-EM

National Cryo-Electron Microscopy



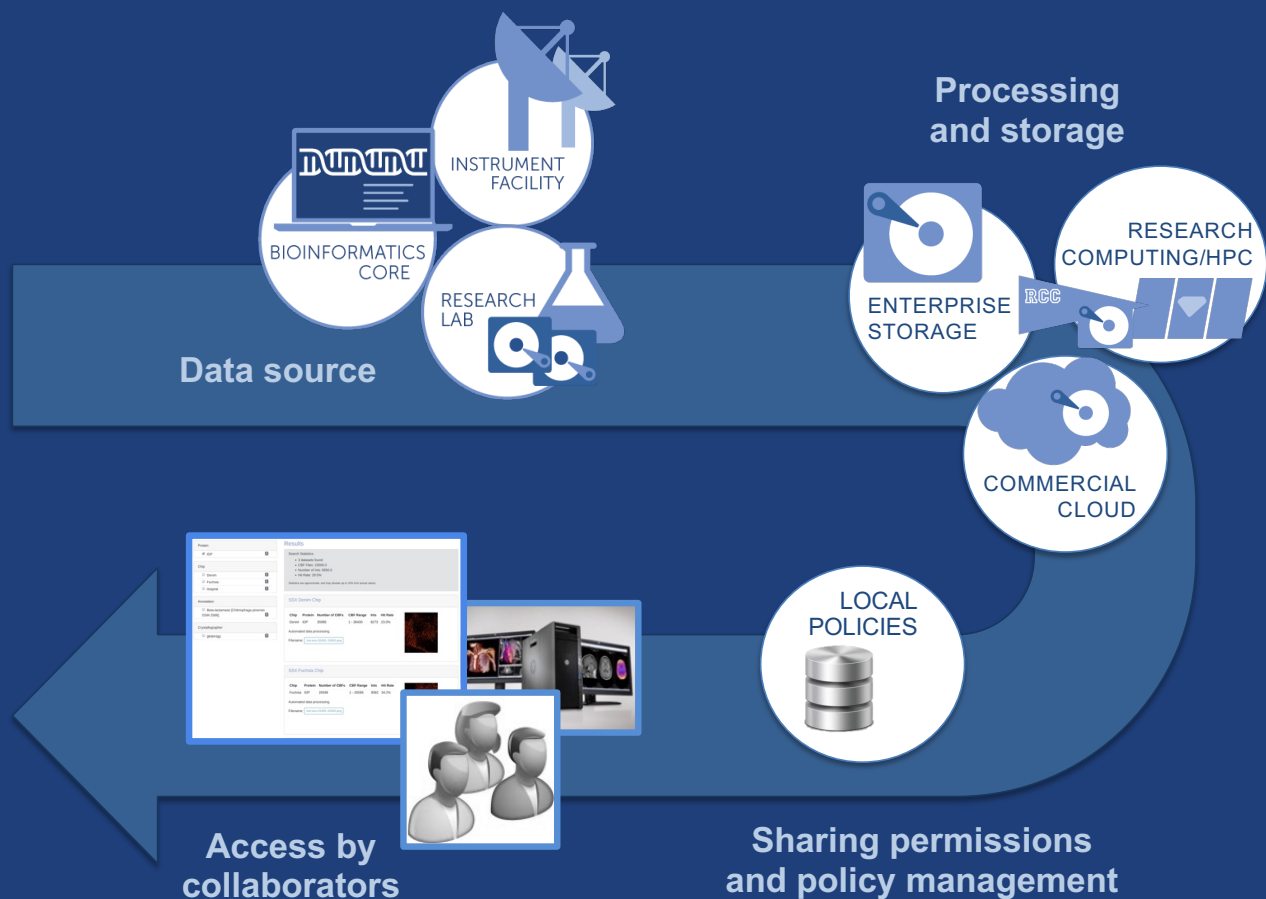
 **ADVANCED LIGHT SOURCE**

**NISC**  
NIH Intramural Sequencing Center





# Instrument data processing pattern



At minimum, make data available to others in your lab, campus colleagues and/or external collaborators



## How do you interact with core/instrument facilities?

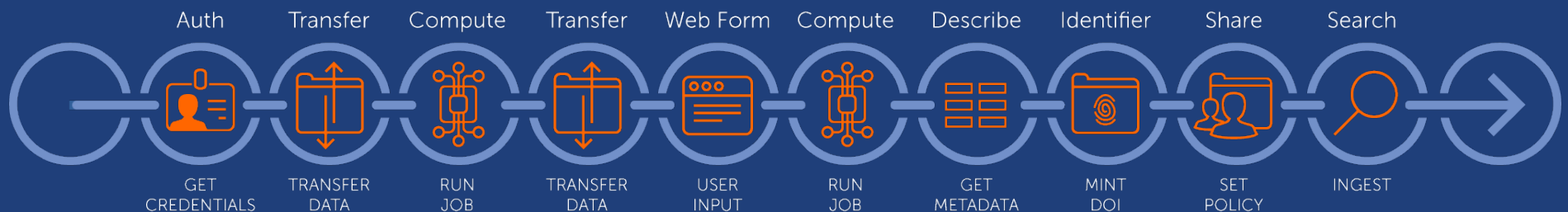


- ☐ User of a local facility
- ☐ User of remote facility
- ☐ Administrator/manager for facility
- ☐ Provide services to facility  
(storage/compute/etc)
- ☐ Other interactions



# Globus Flows for managed automation

- **Managed, secure, reliable task orchestration—at scale**
- **Define, run, and share distributed research pipelines**
- **Event driven execution model**





# Automating cryoEM

Globus  
Flows



Transfer



Transfer  
raw files

Compute



Launch  
analysis job

Carbon!



Correct,  
classify, ...

Compute



Extract  
metadata

Share



Set access  
controls

Transfer

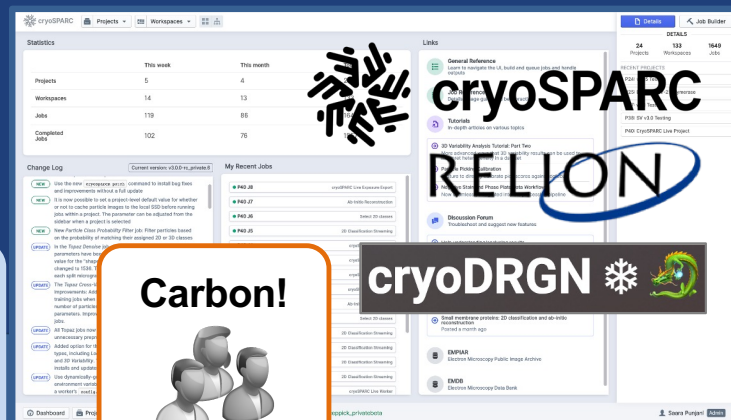
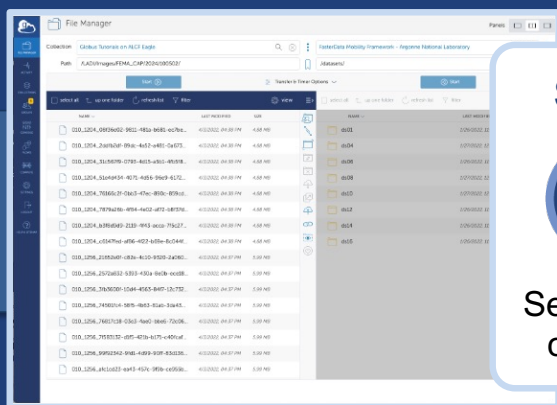


Move final  
files

Search

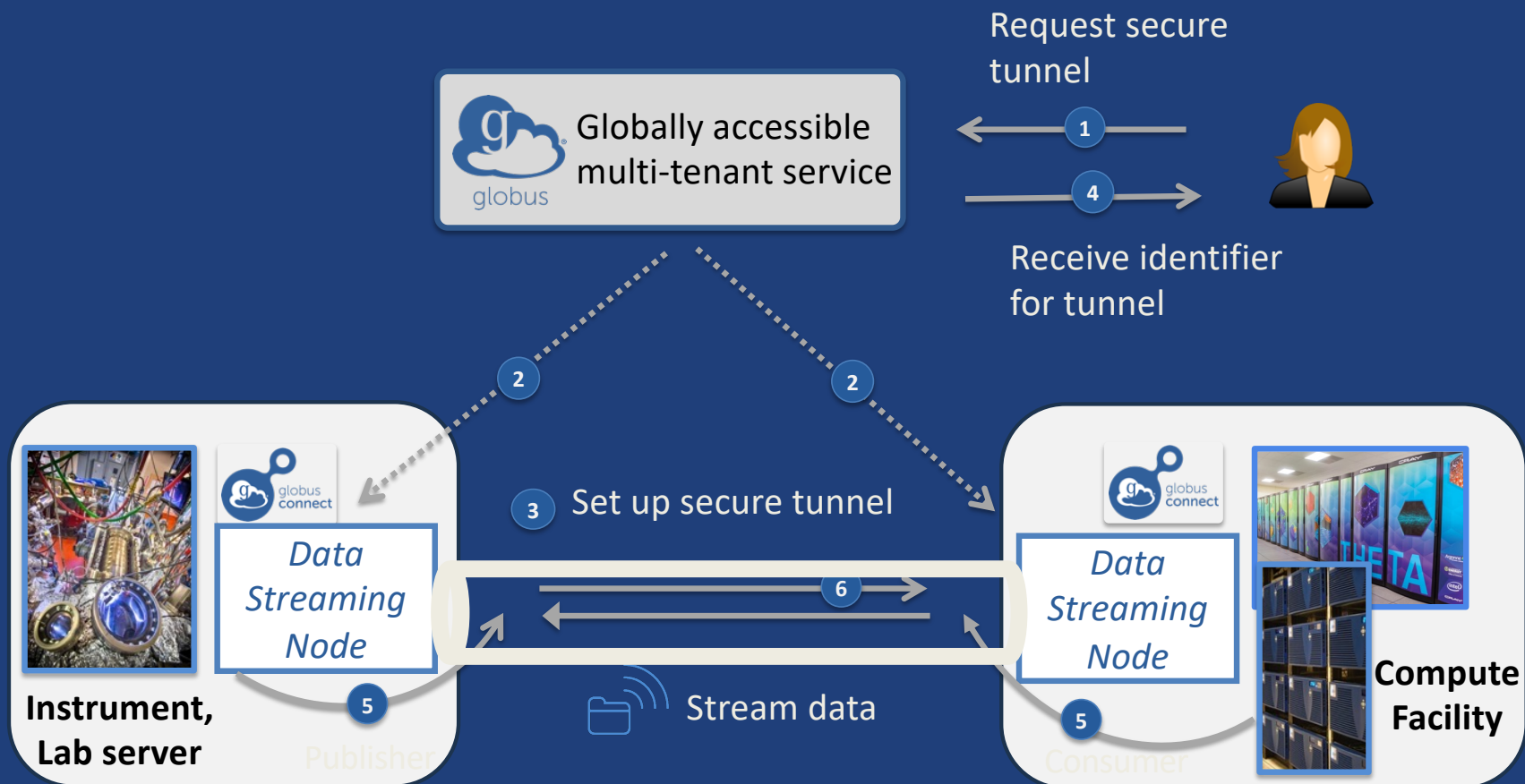


Ingest to  
index





# Streaming data using Globus

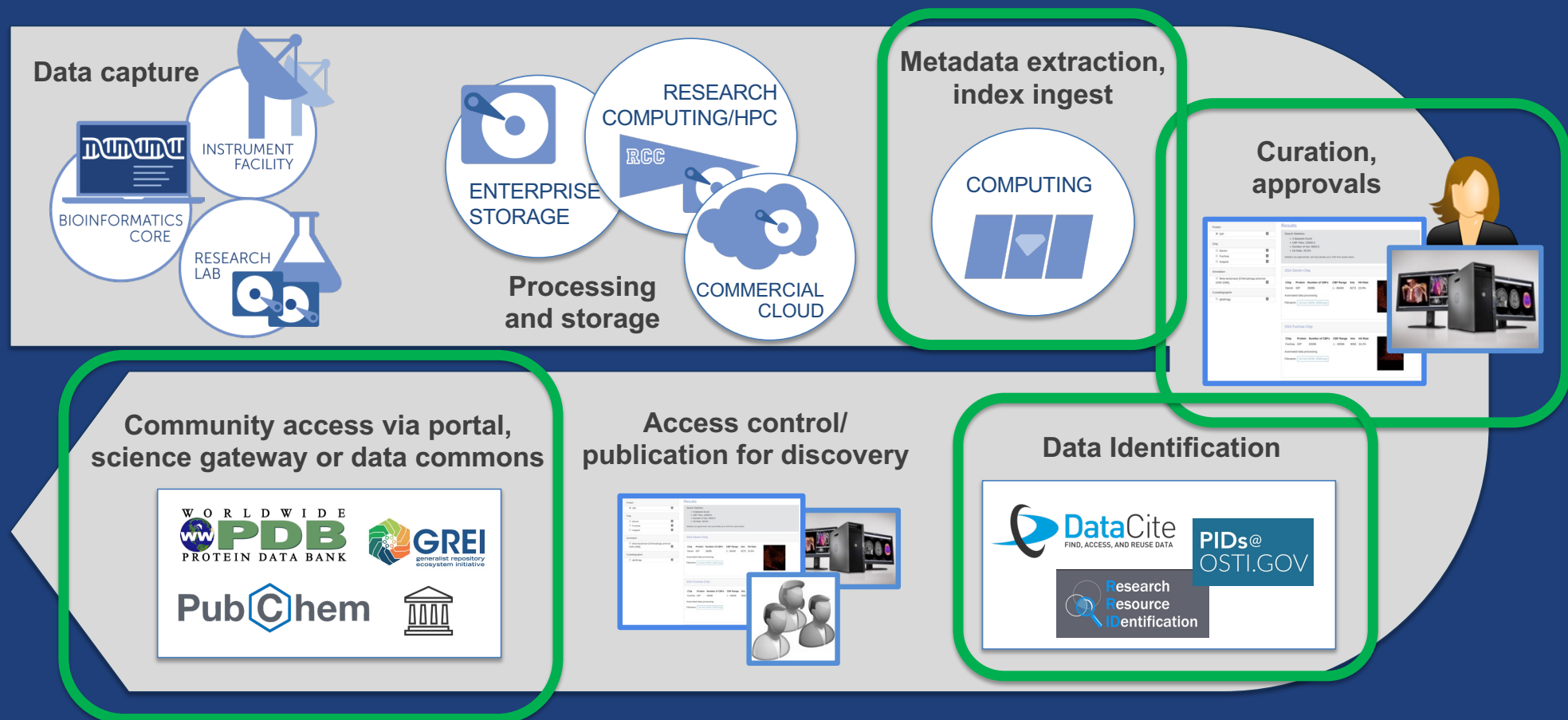




# Beyond automation



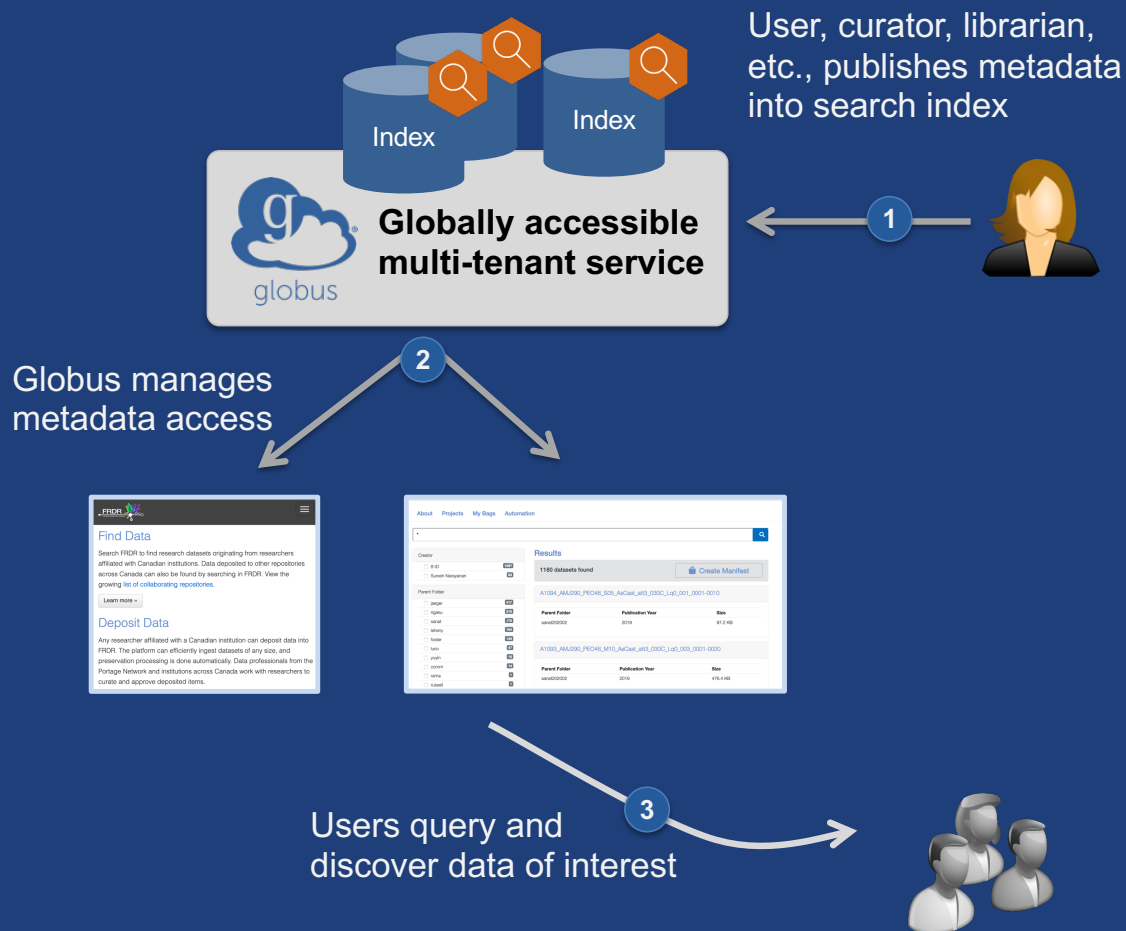
# Flows to ensure “FAIRness”



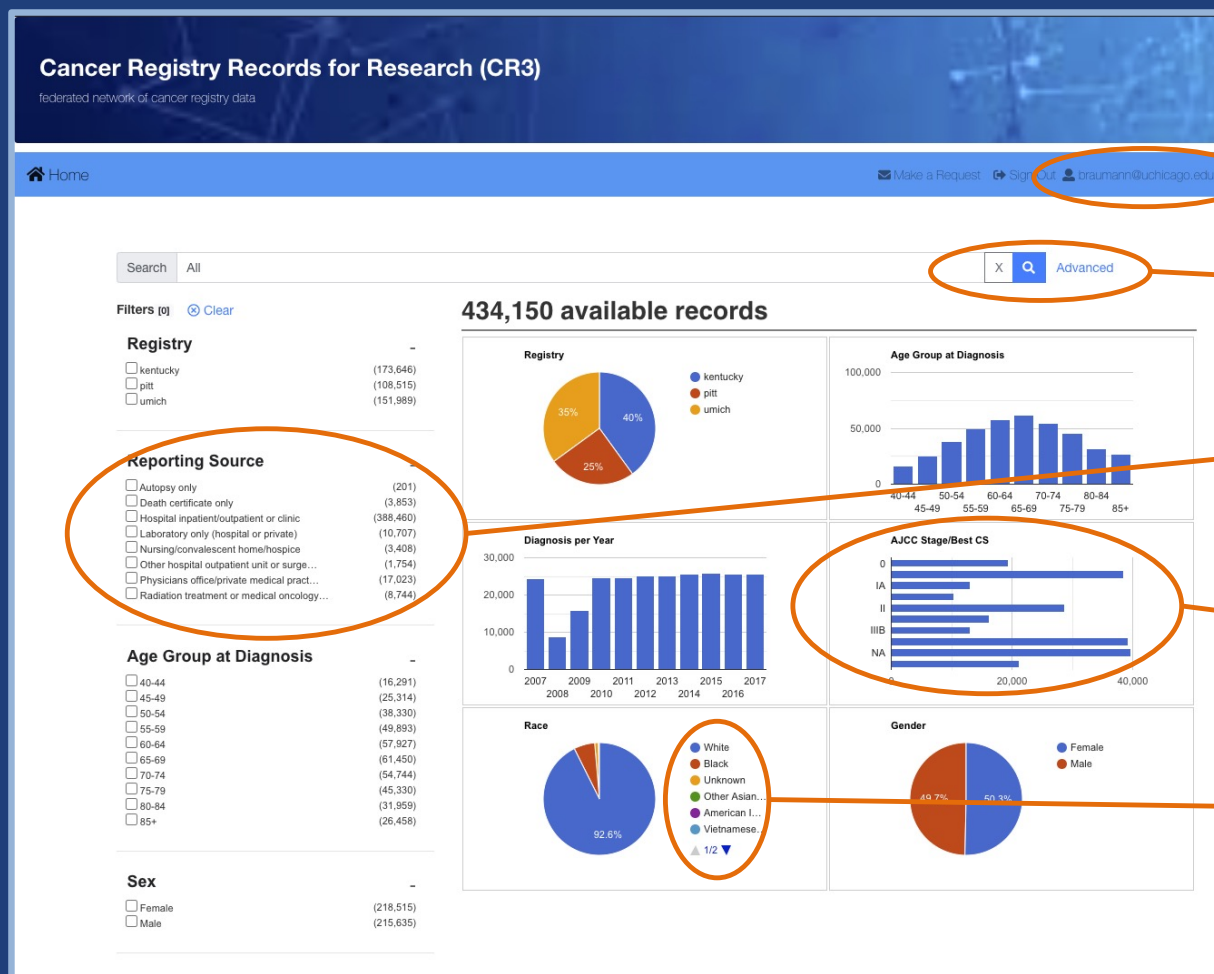


# Globus Search: Data discovery for any domain

- Scalable metadata store
- Fine grained visibility controls
- Schema agnostic
- Federated auth integration
- Queried via API with facets



# An example of how we can get to FAIR data



Federated login → instant access using your institutional credentials

Google-like text search with facets for filtering

Variable facets based on source registry index

Dynamically updating charts as facets change

Restricted visibility to sensitive data



# Flows and Transfer for data lifecycle management

Tar and transfer

Secure data  
egress

Extract metadata  
and push to an  
index

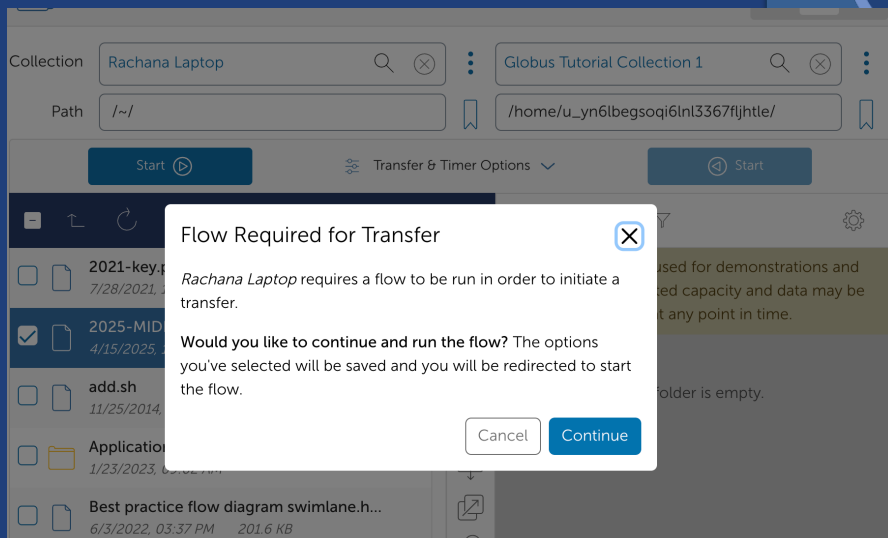
Automated  
processing on data  
deposit

***Configure a flow to run when user transfers data to or from a collection***



# Secure egress

Globus  
Flows



User initiates a transfer, and policy triggers a flow to process request

Transfer



Transfer  
files staging  
area

Compute



Launch  
screening  
job

Notify



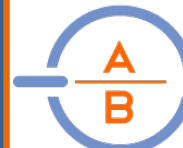
Notify  
admin

Transfer



Move final  
files

Choice



Admin  
approval

# Using Globus Platform



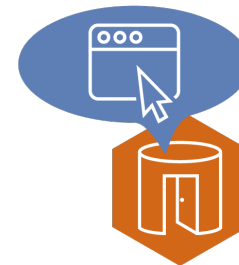
Globus Python  
SDK



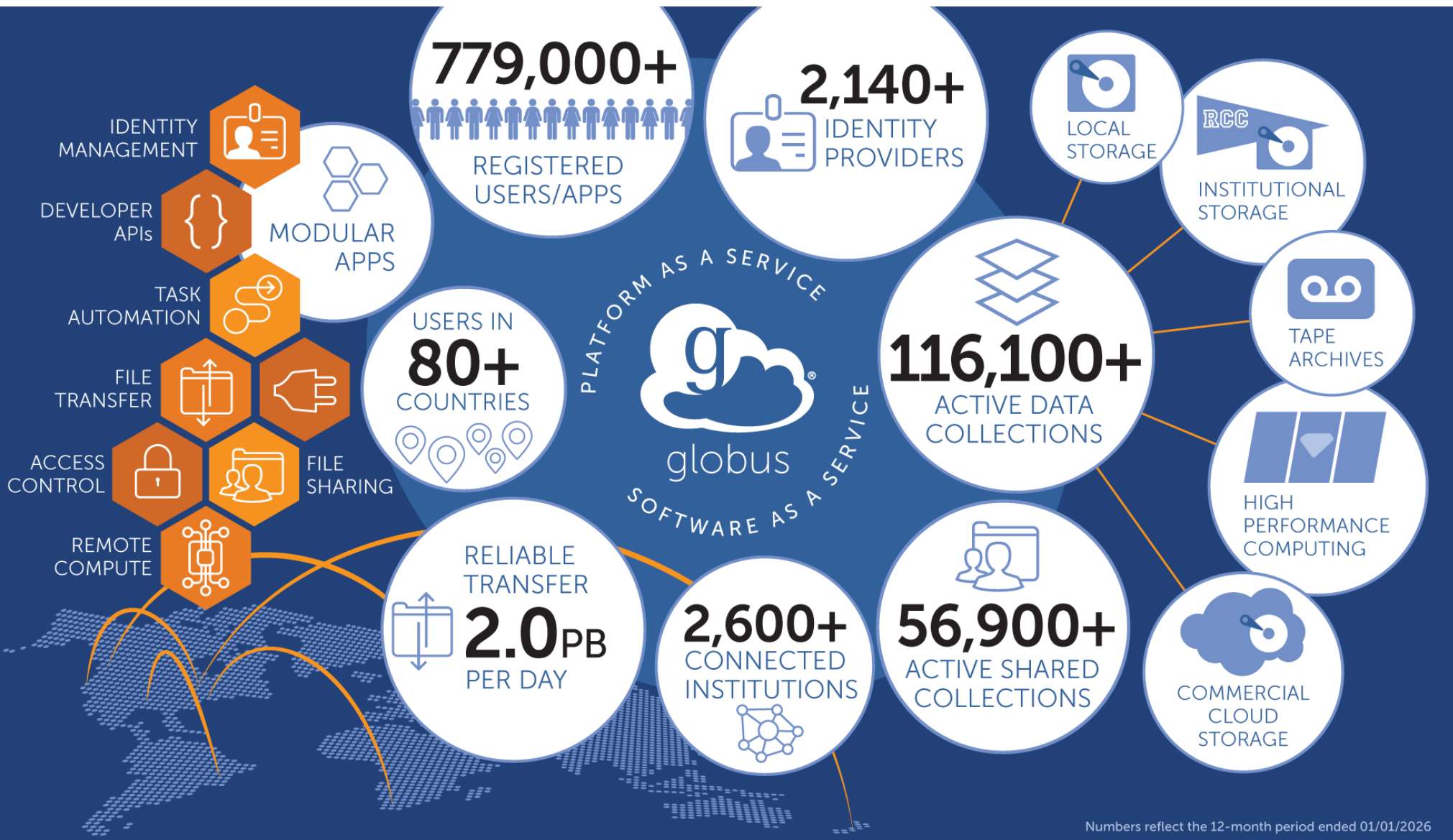
Globus Django  
Portal Framework



Globus JavaScript  
SDK



Globus Serverless  
Portal



Numbers reflect the 12-month period ended 01/01/2026

 **Thank you!**

- **Engage: [ranantha@uchicago.edu](mailto:ranantha@uchicago.edu)**
- **Website: [globus.org](https://globus.org)**
- **Documentation: [docs.globus.org](https://docs.globus.org)**
- **Support: [support@globus.org](mailto:support@globus.org)**