Globus: Platform for Data Driven Research

IN-PERSON WORKSHOP

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

Rachana Ananthakrishnan
Executive Director, Globus
ranantha@uchicago.edu

THE UNIVERSITY OF CHICAGO

globus

Globus is …

a non-profit service developed and operated by

THE UNIVERSITY OF CHICAGO

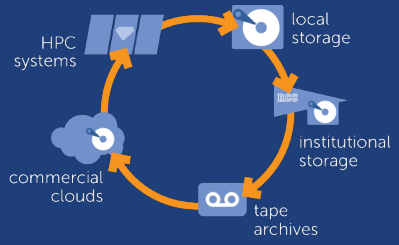
Our mission is to…

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

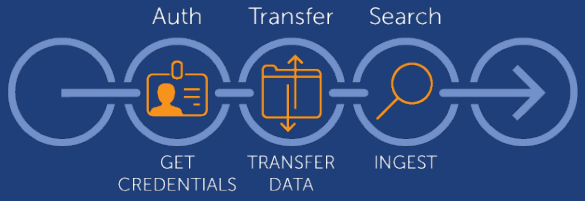
Collaborative data sharing



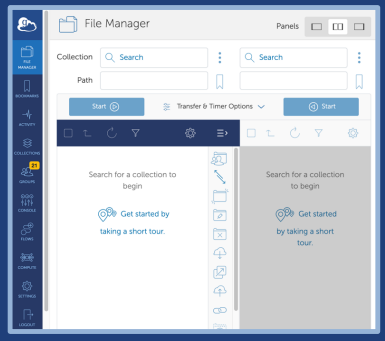
Unified data access

Publication & discovery

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

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…

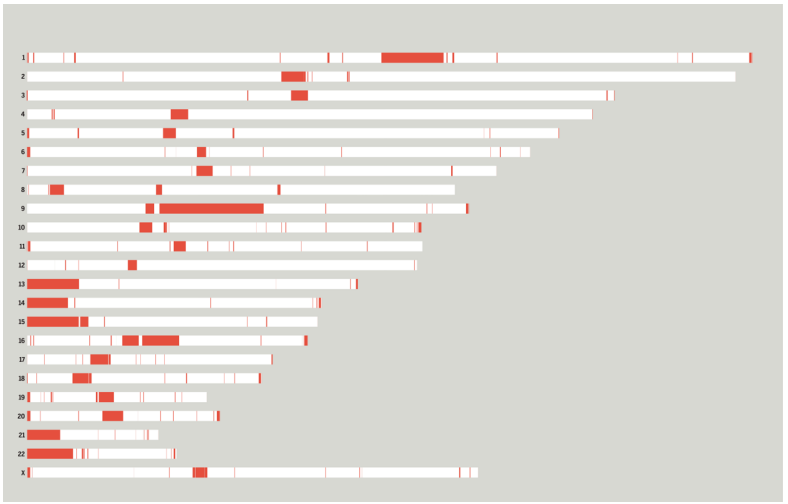
Digital agriculture – University of Winnipeg

- 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

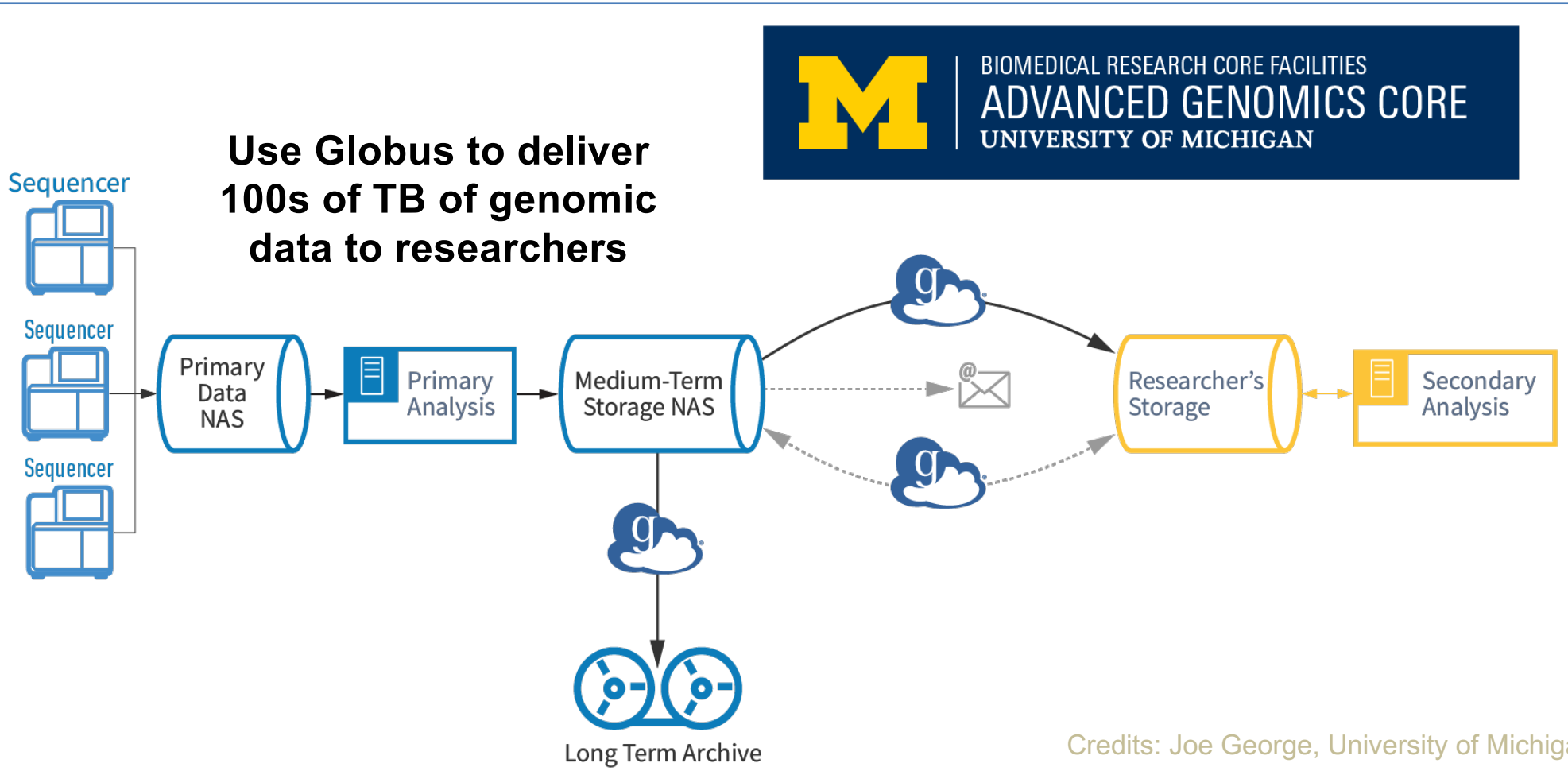
T2T Consortium
TELOMERE-TO-TELOMERE CONSORTIUM



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

globus.org/user-stories/globus-enables-multi-institutional-data-sharing

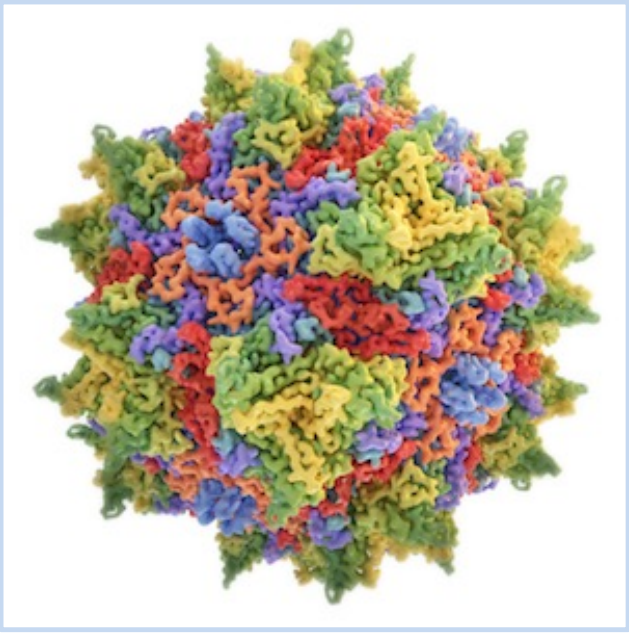
Instrument data delivery at scale



Credits: Joe George, University of Michigan

Data Management at Cyro-EM Facilities

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



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

Globus for

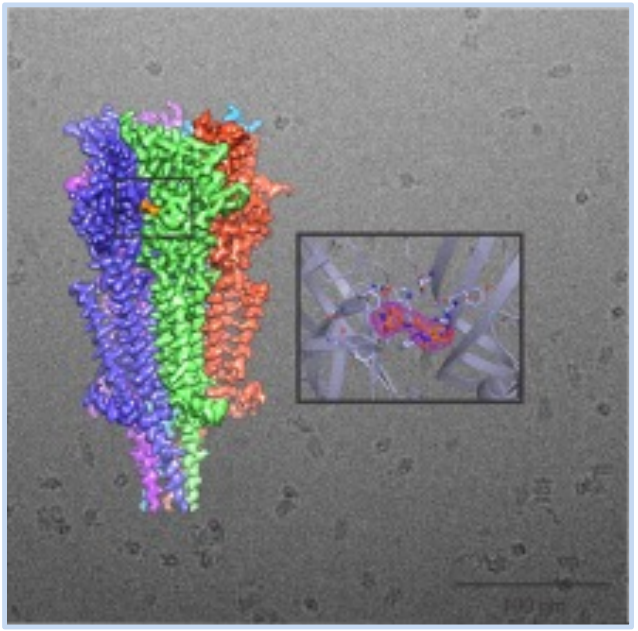
automated data sync as new data is collected

provisioning of data access for researchers

reliable, secure data access for users

monitoring and management via console

Case Western Reserve – Cryo-EM Core



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

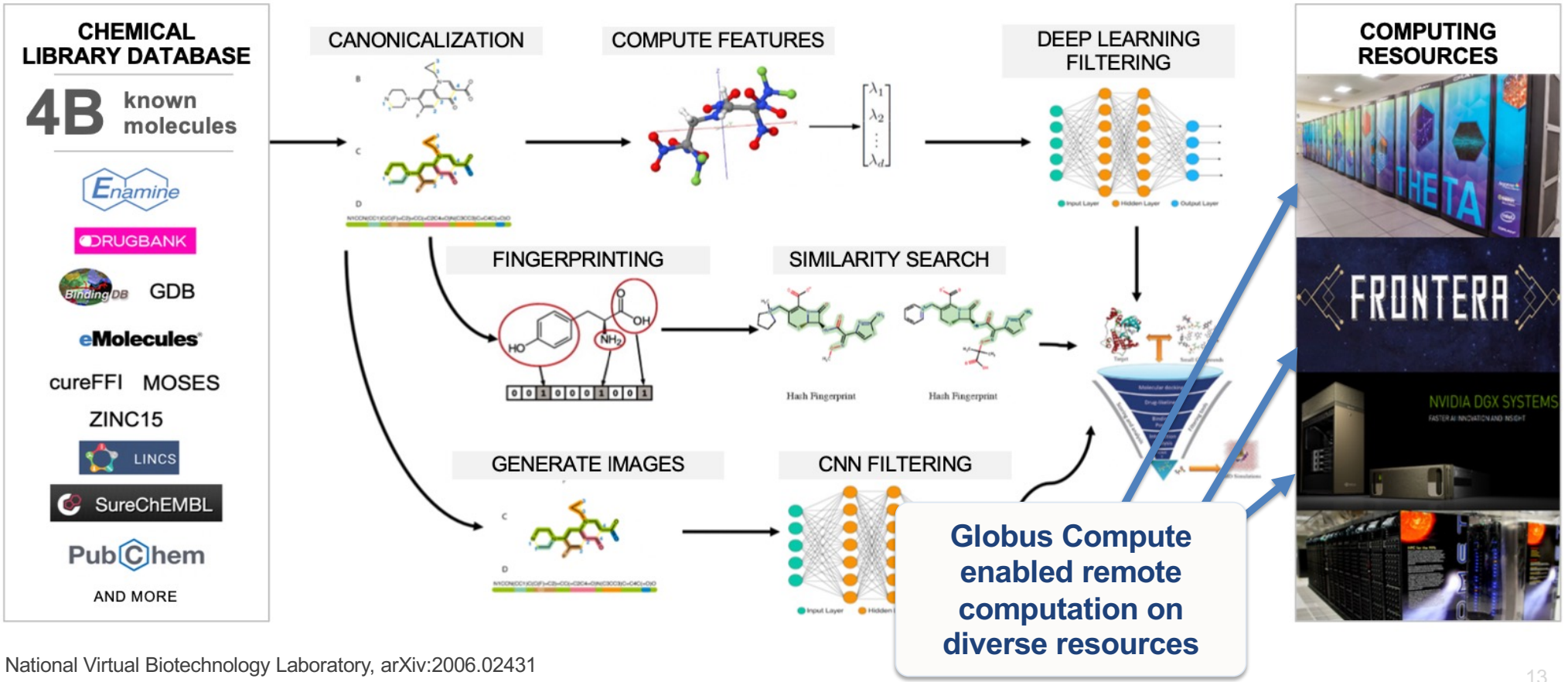
Streamlining processing of field data

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

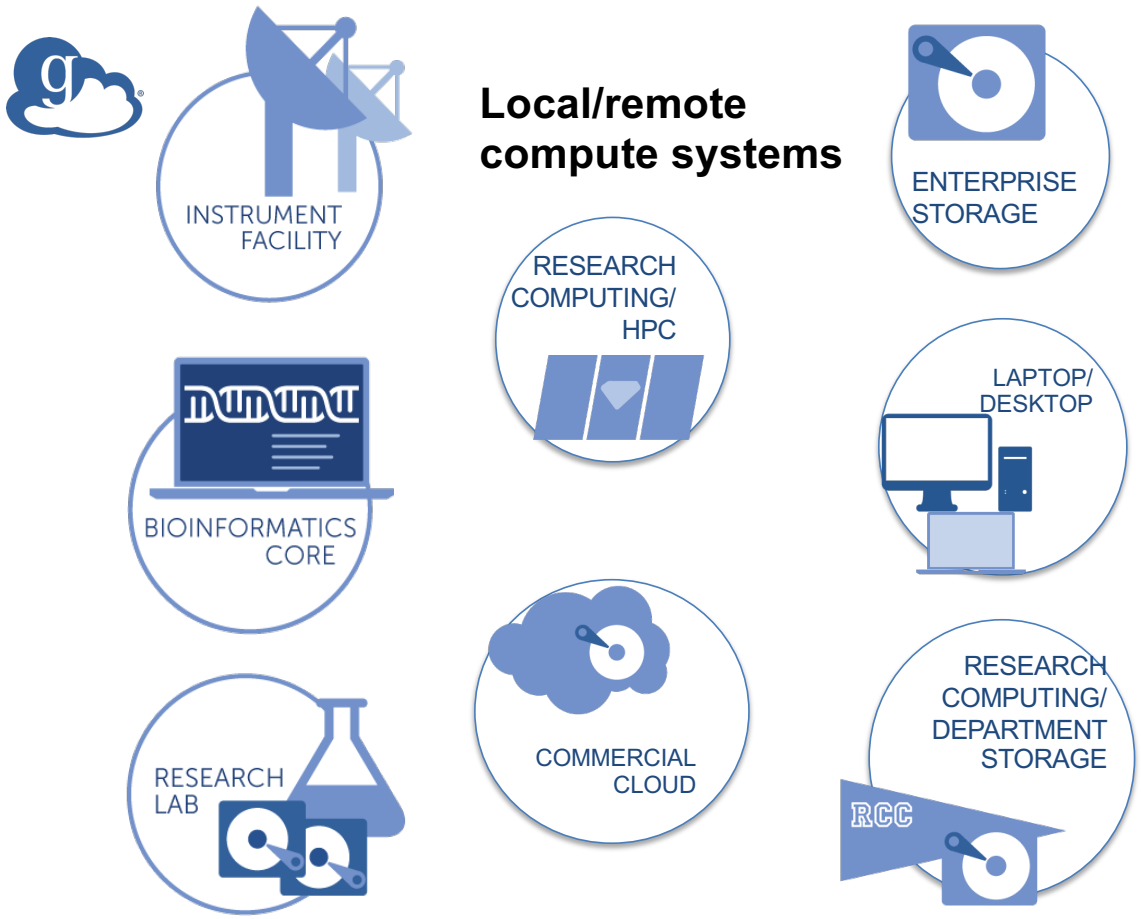
FRANKLIN & MARSHALL COLLEGE

Exemplar Use Case: ML-based drug screening



National Virtual Biotechnology Laboratory, arXiv:2006.02431

Need to navigate…



Discovery and repository services

Domain Repository

HuBMAP

COD

PubChem

RCSB PDB PROTEIN DATA BANK

Generalist/Institutional Repository

GREI generalist repository ecosystem initiative

DataCite

PIDs@ OSTI.GOV

Research Resource IDentification

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

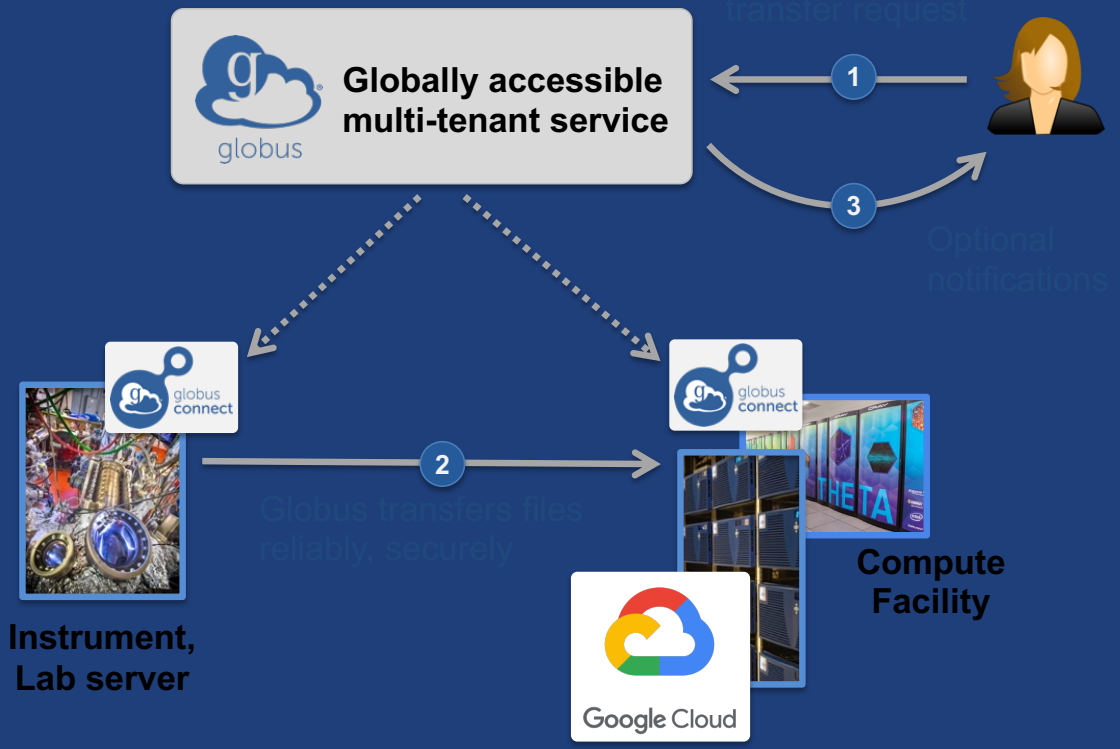
Cybersecurity and Privacy Staff

Auditable enforcement of variety of compliance requirements

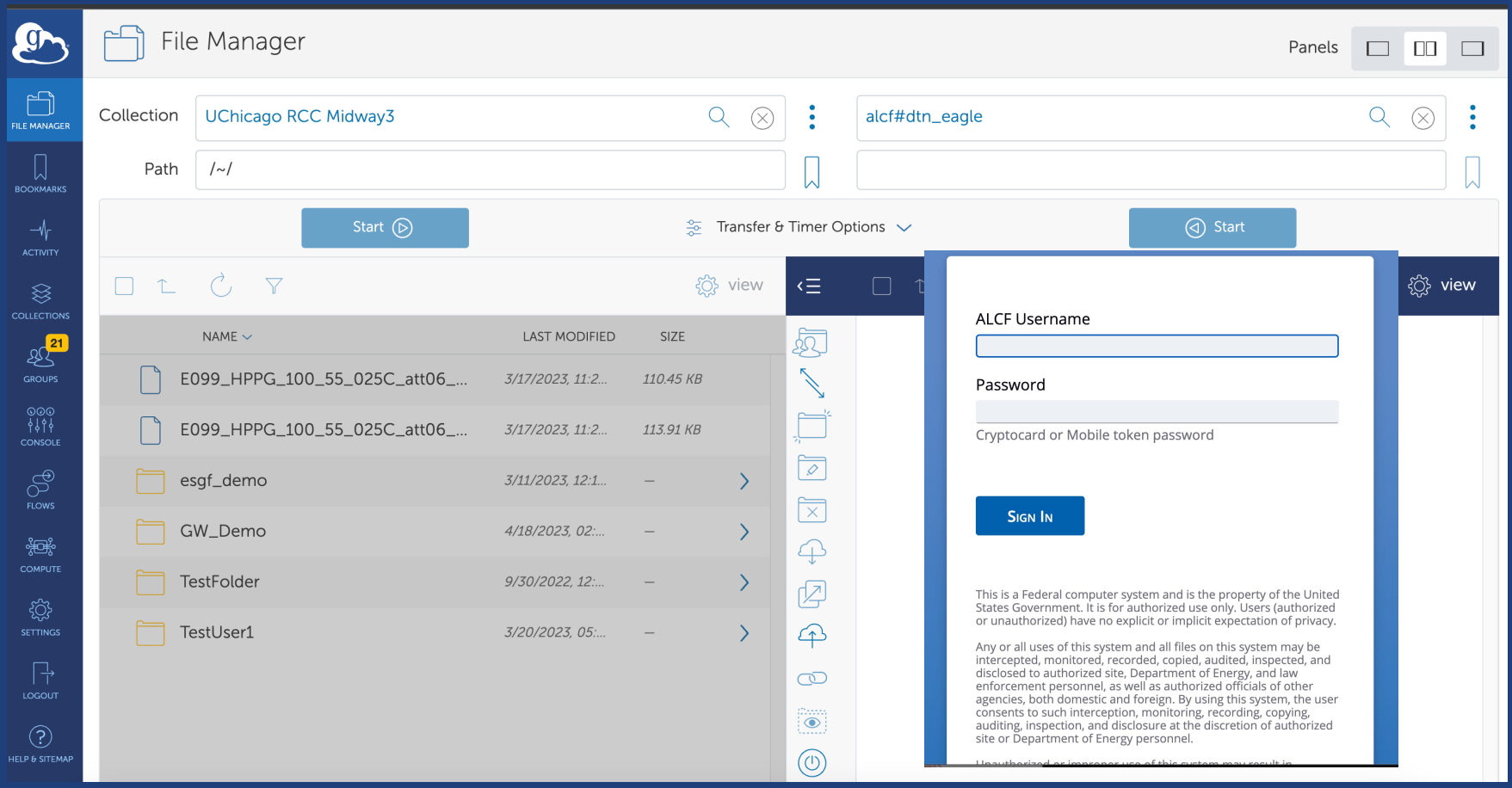
Unified research data access and sharing

Fast, reliable file transfer …from any to any system

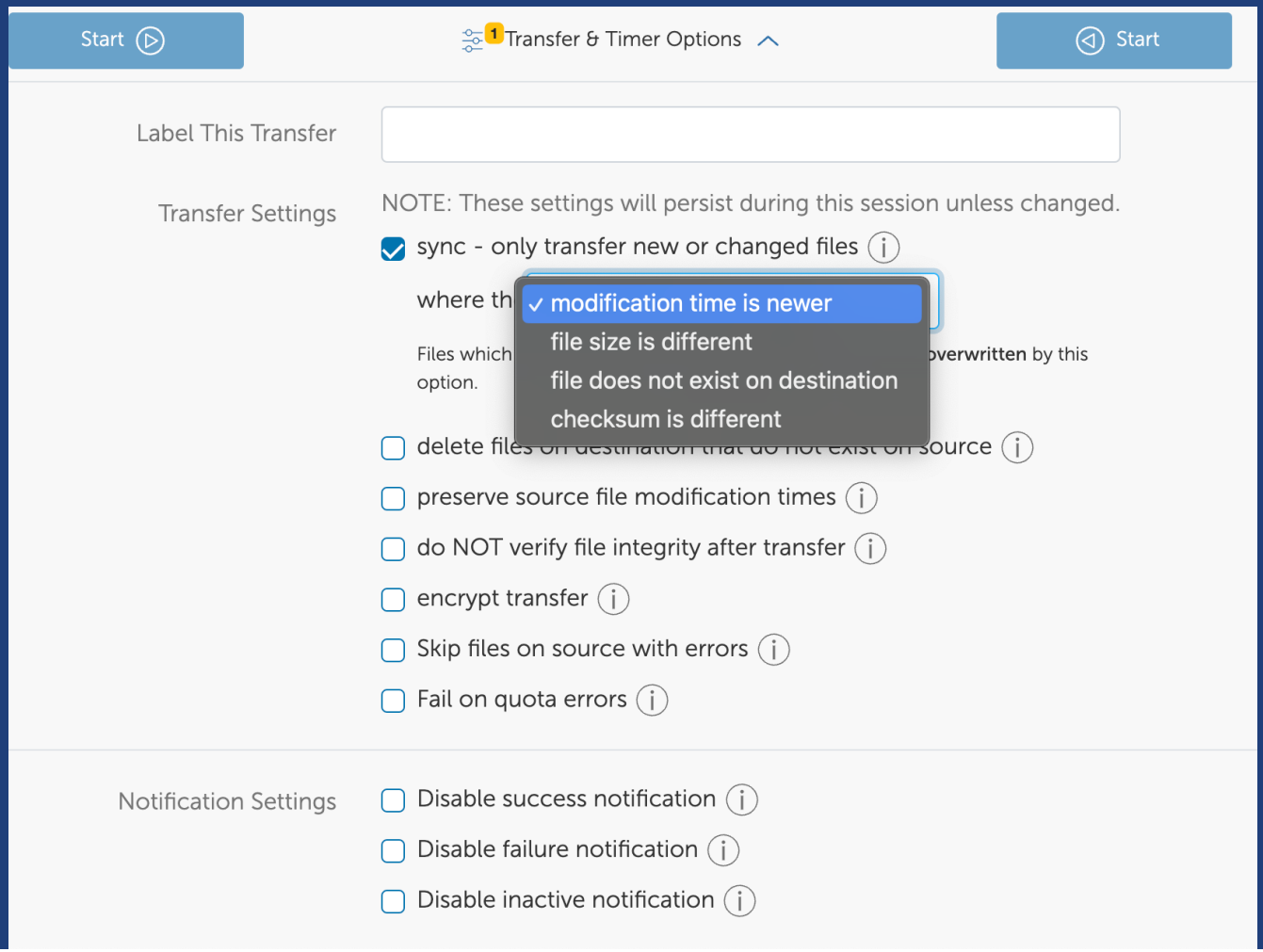
- **Fire-and-forget transfers/sync**
- **Optimized speed**
- **Assured reliability**
- **Unified view of storage**
- **HTTP/S access to data**



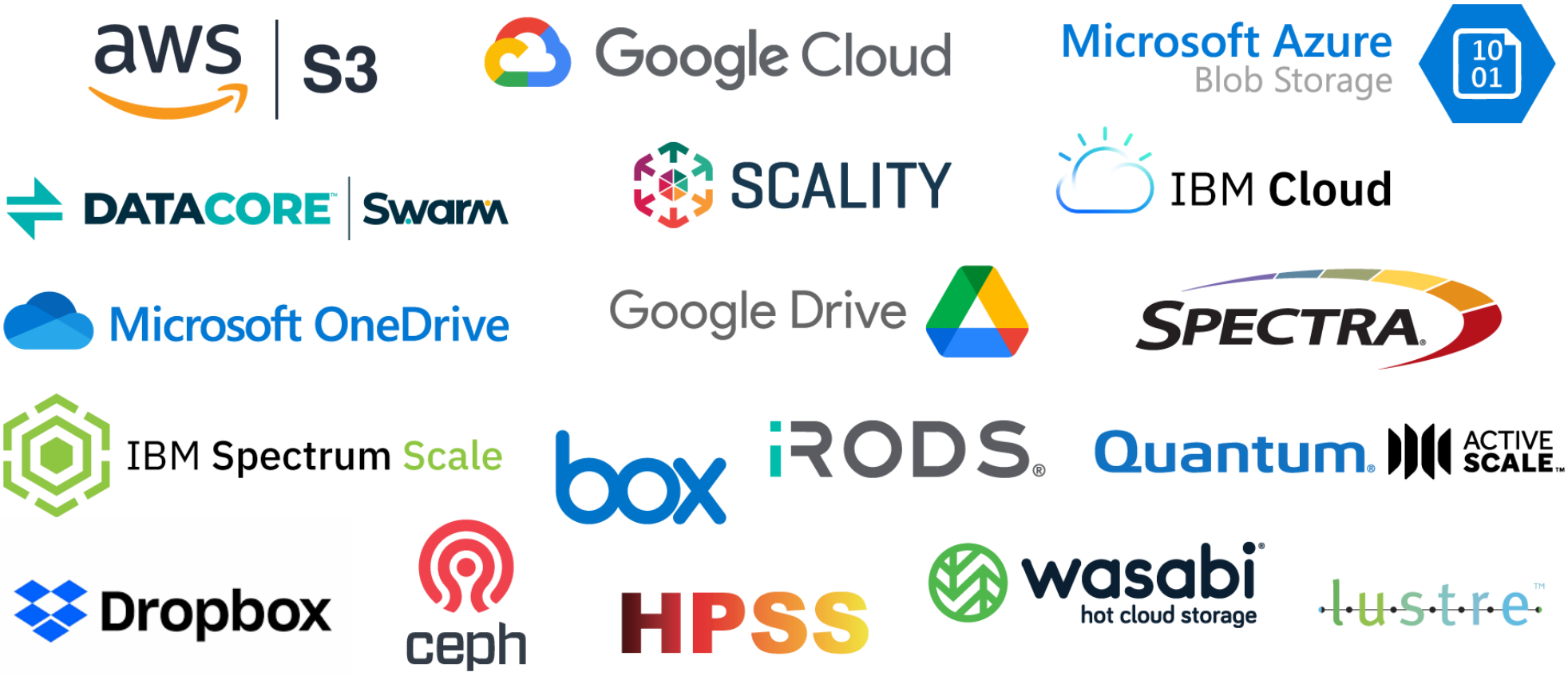
Intuitive web application interface



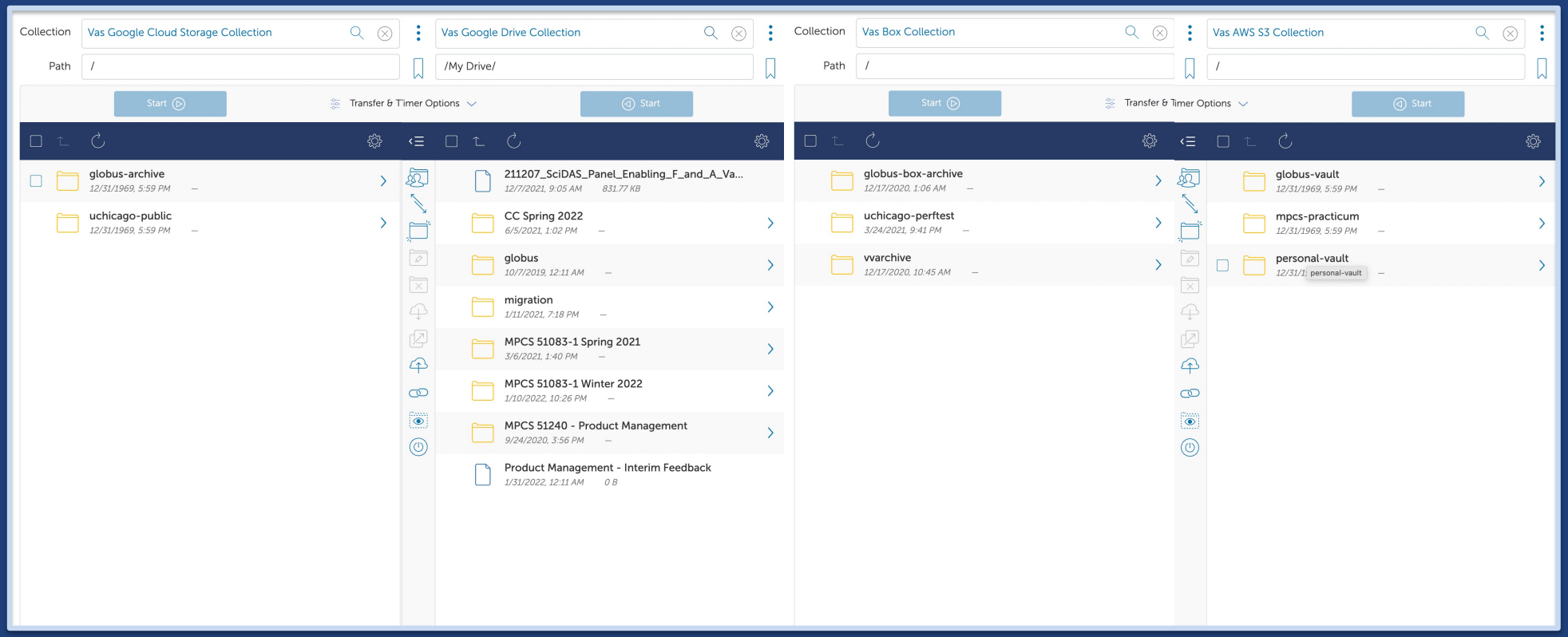
Transfer/sync options



Globus Connectors support diverse systems



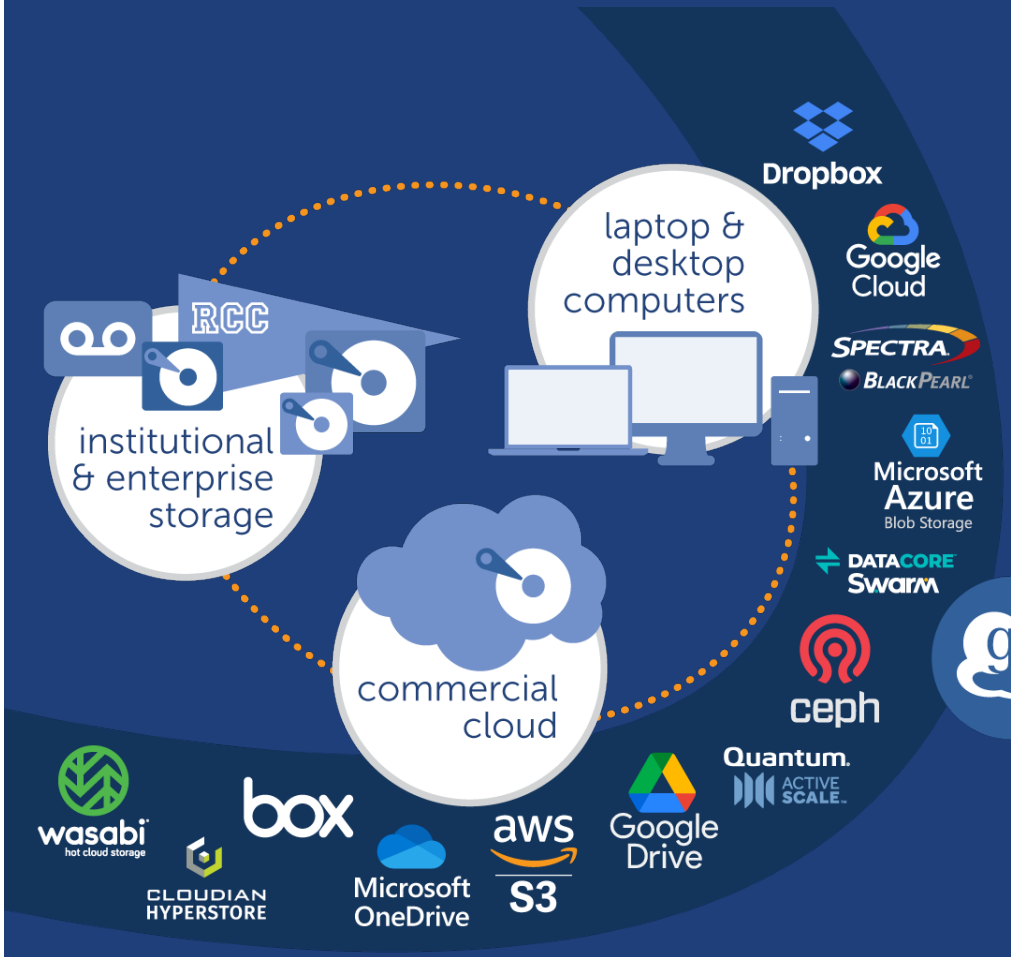
Uniform interface, consistent user experience



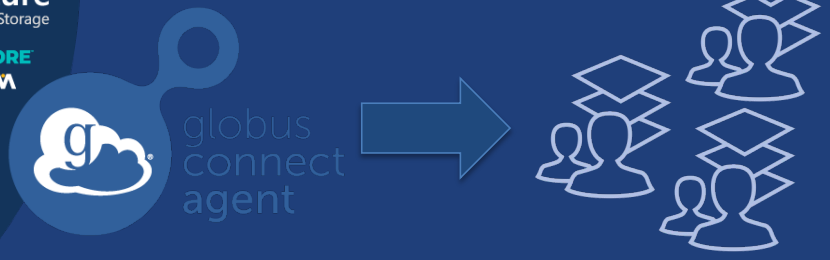
What types of storage systems do you use/offer?

- ❑ 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



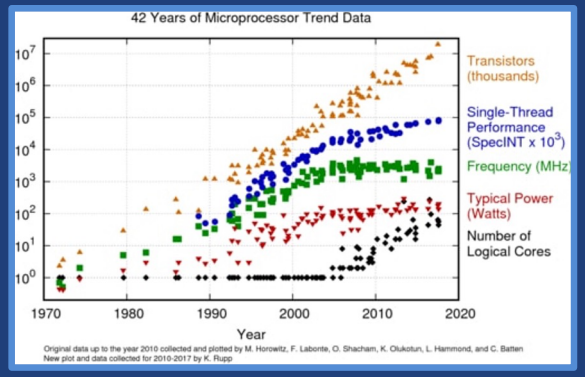
Data collections

Unified access to compute resources

Why do we need to rethink research computing?

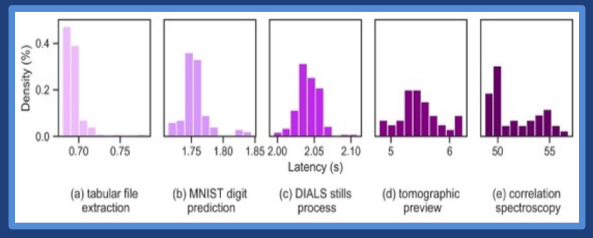
Resources

- Hardware specialization
- Specialization leads to distribution



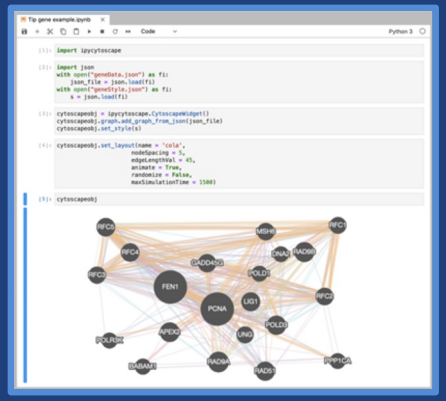
Workloads

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



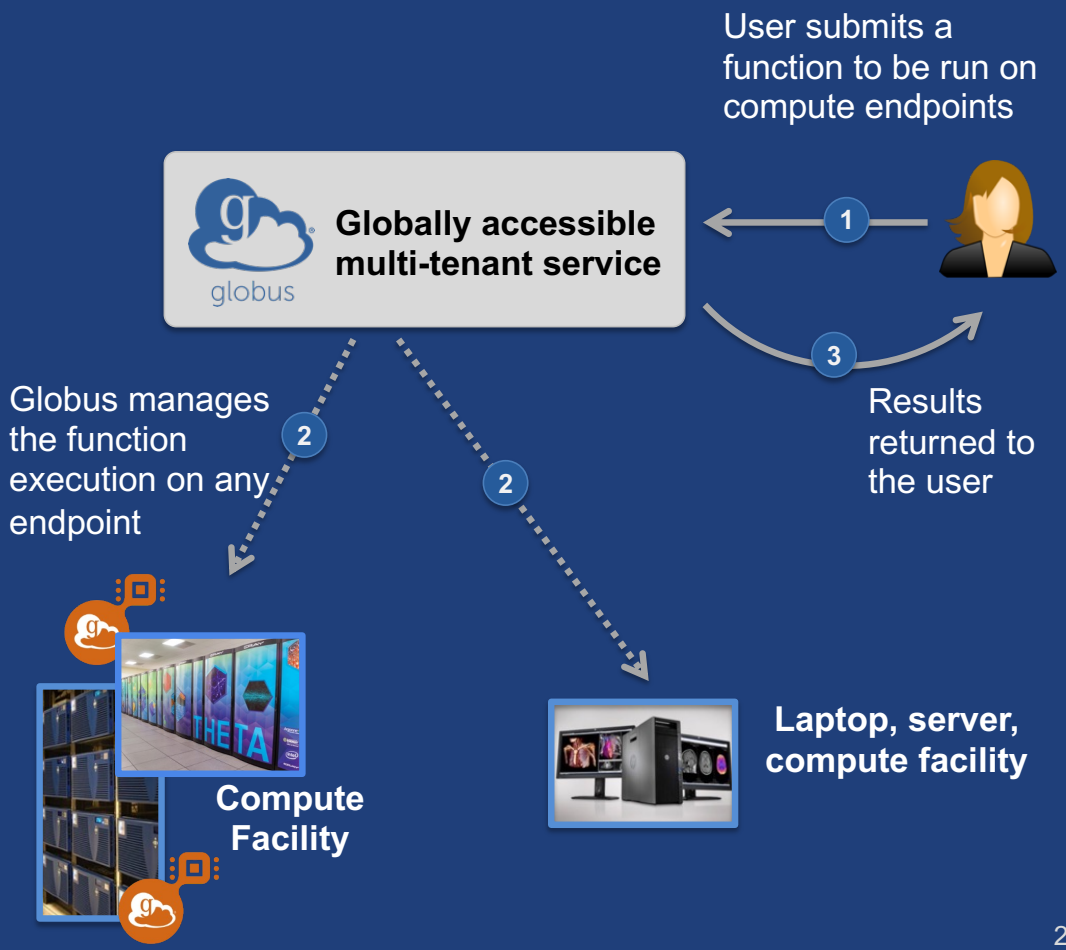
Users

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

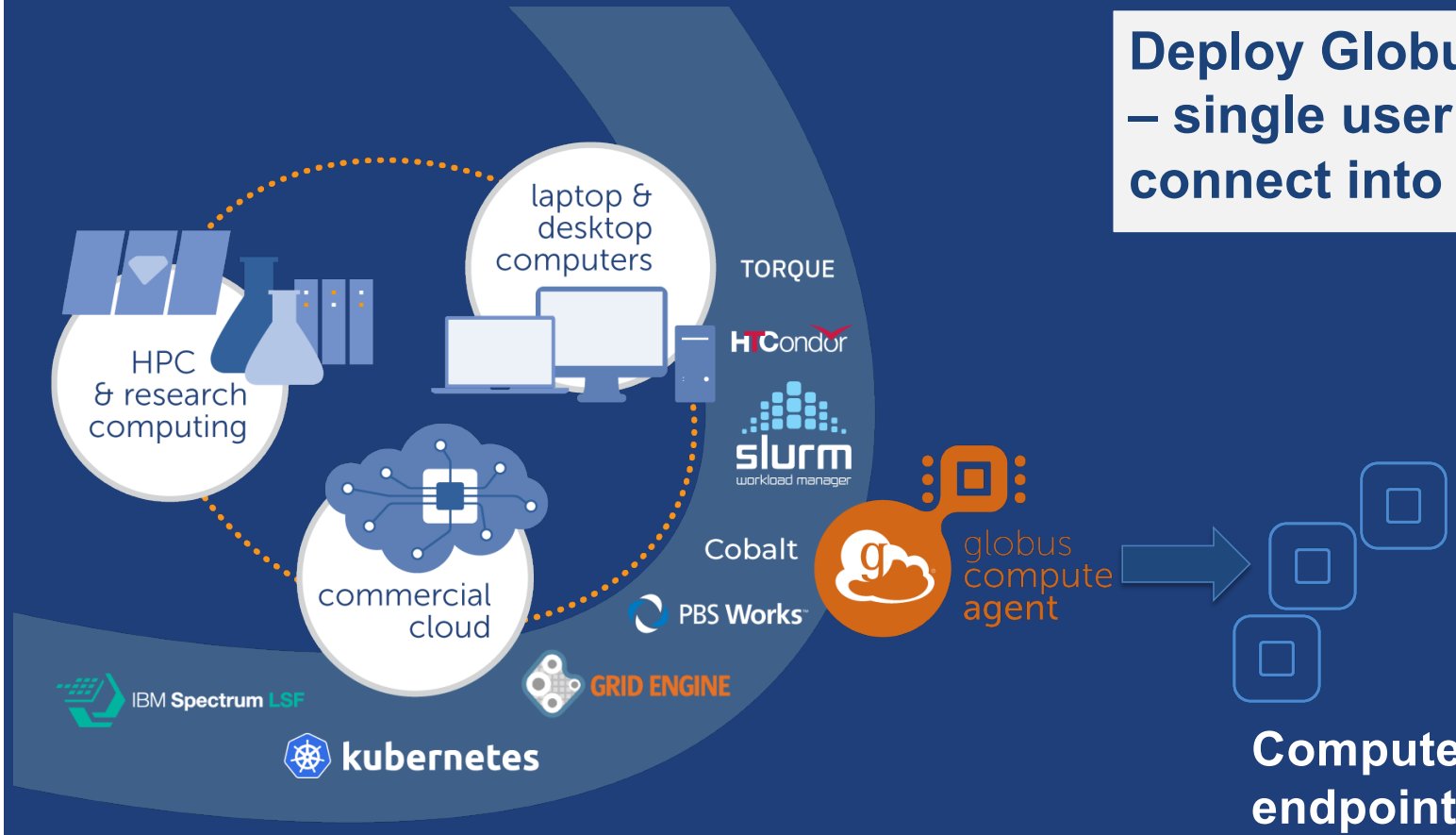


Managed compute …on any system

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



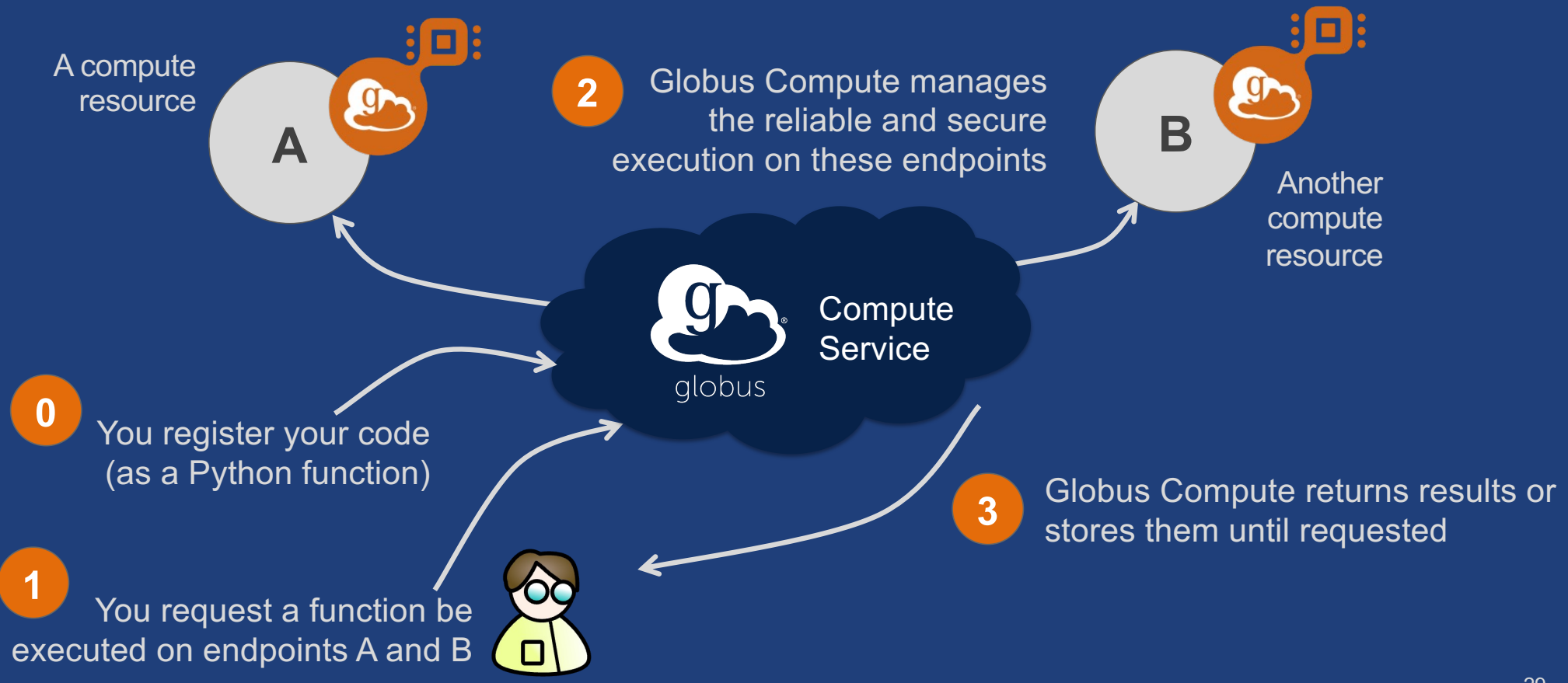
Compute ecosystem



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

Compute endpoint

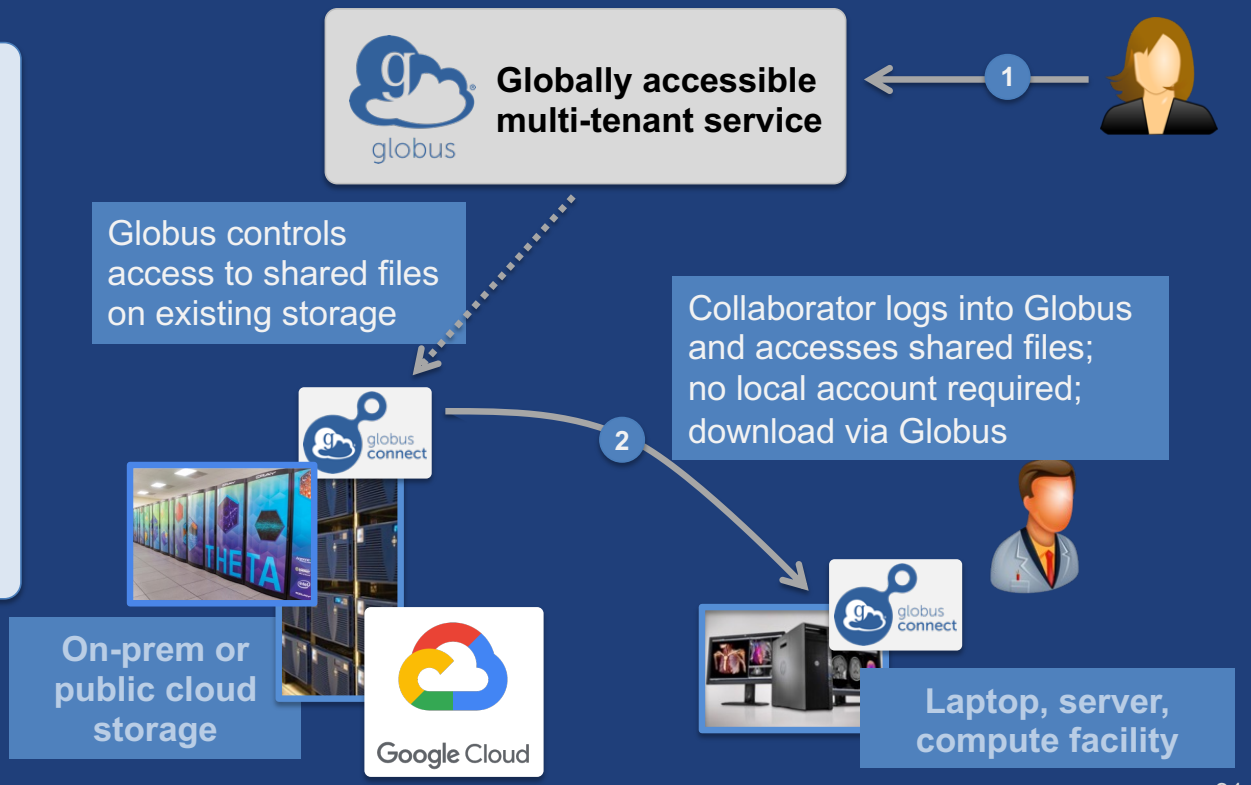
How does it look from the researcher's PoV?



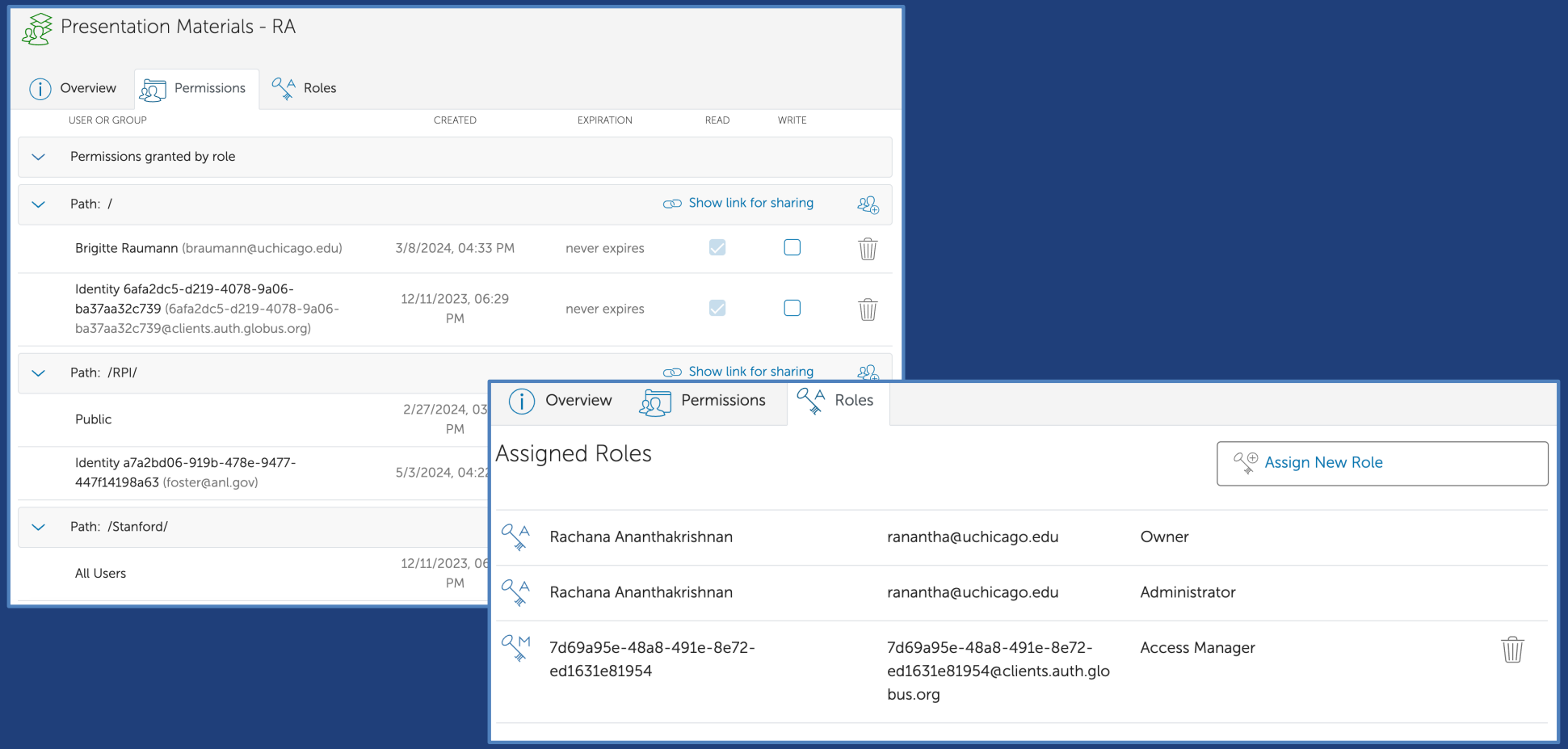
Compliant and secure collaboration

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**



Data sharing – permissions & roles

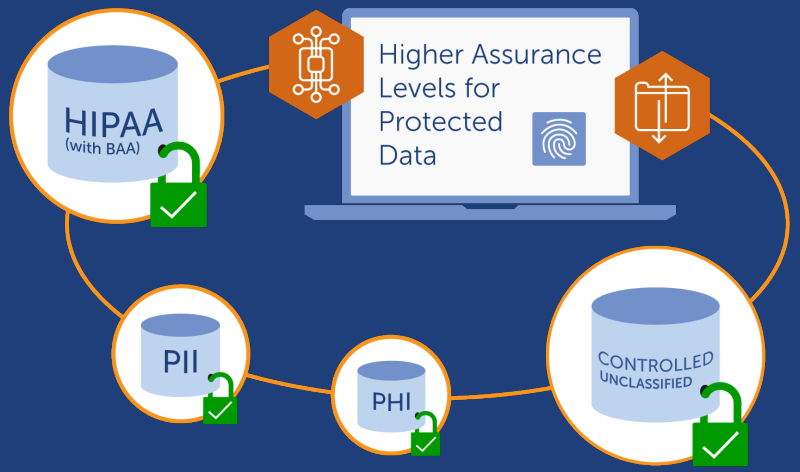


Common sharing scenarios enabled by Globus

- **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

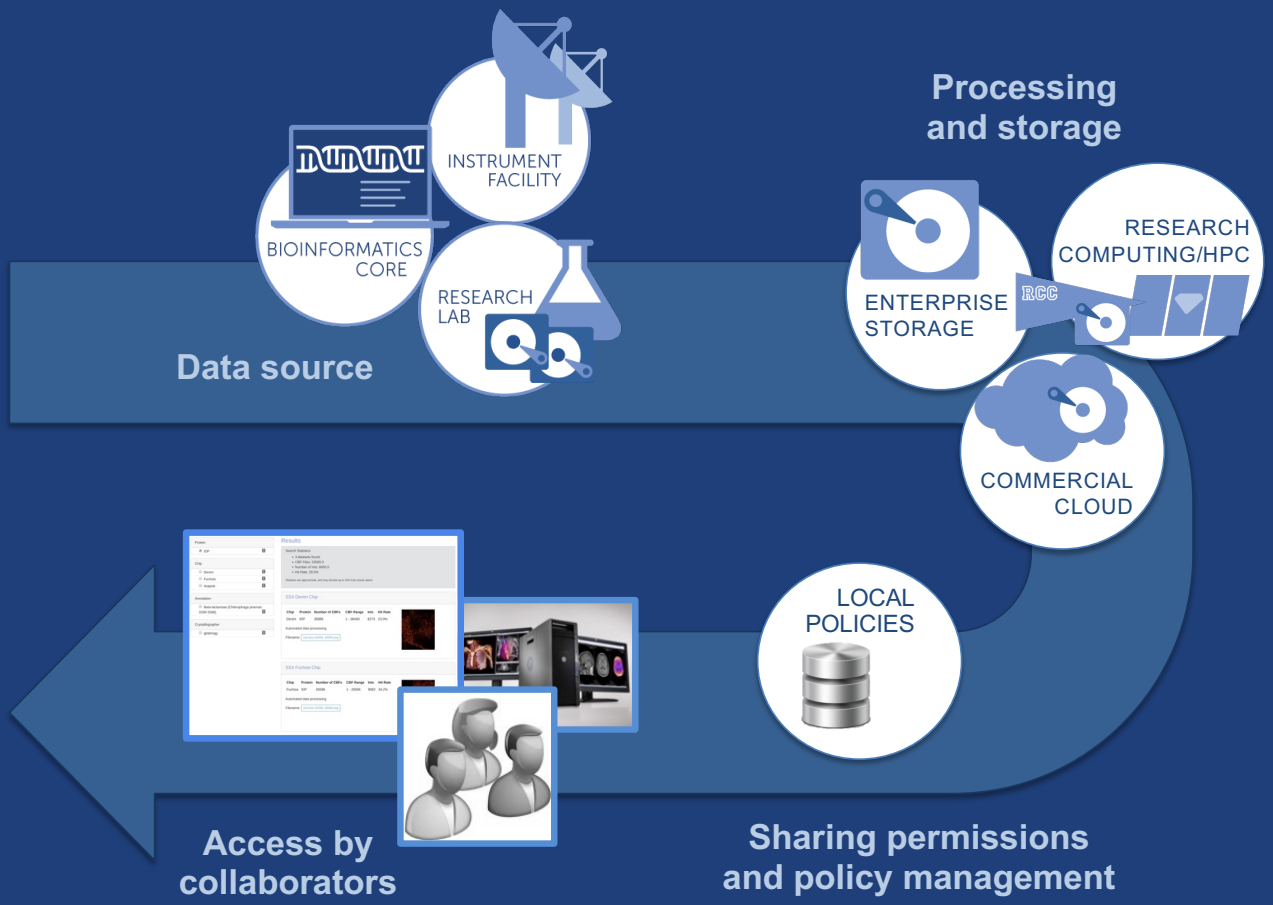
What are some of your compliance requirements?

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

Core Facility data automation


ion torrent
ADVANCED LIGHT S
NEW YO
GENOME CEN
High Resolution
Cryo-EM
National Cryo-Electron Microsc
NIH Intramural Sequencing Cent
TECNAI
PURDU
S C I E N
Purdue Cryo-EM Facility
FMRIF

Instrument data processing pattern



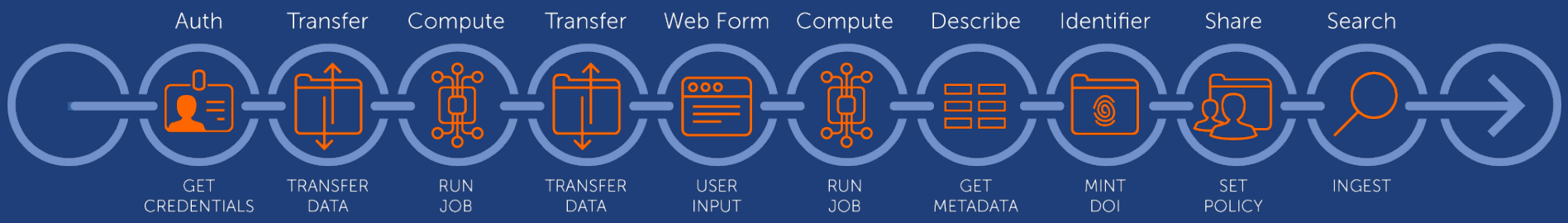
At minimum, make data available to others in your lab, campus colleagues and/or external collaborators

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

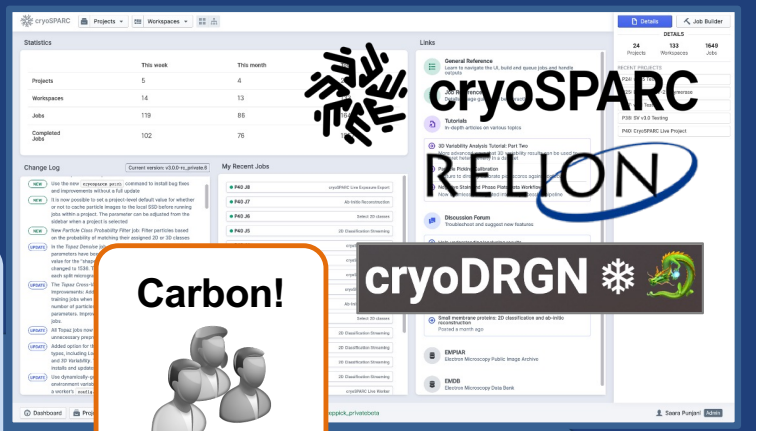
Globus Flows

Transfer

Transfer raw files

Compute

Launch analysis job

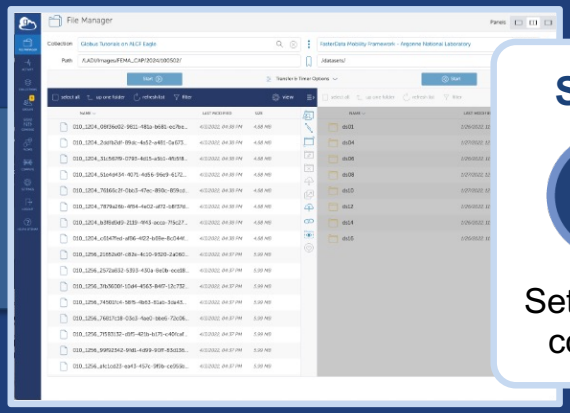


Carbon!

Correct, classify, …

Compute

Extract metadata



Share

Set access controls

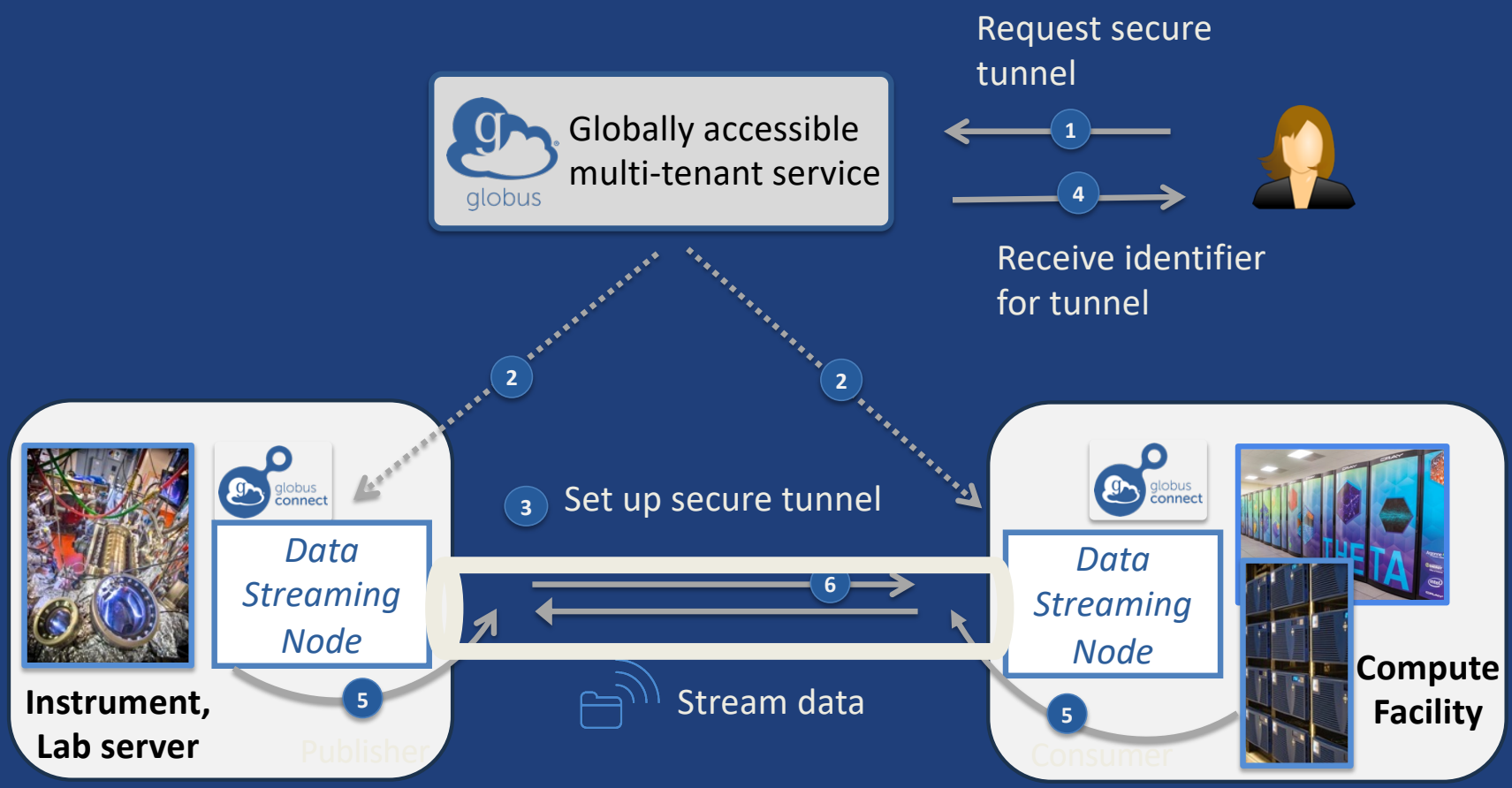
Transfer

Move final files

Search

Ingest to index

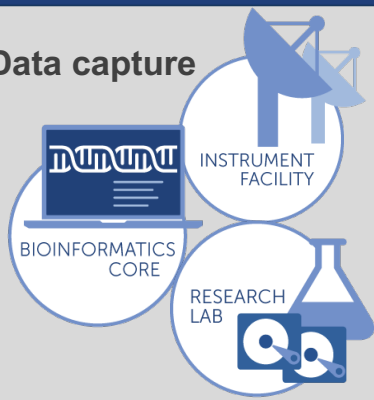
Streaming data using Globus

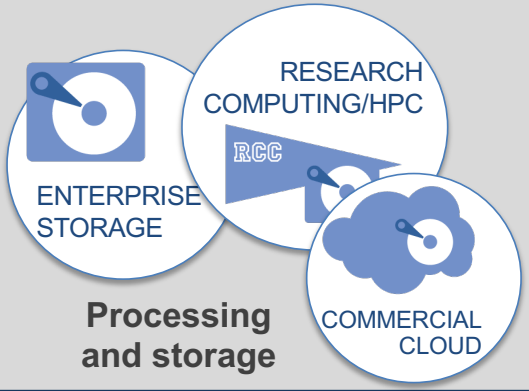


Beyond automation

Flows to ensure “FAIRness”

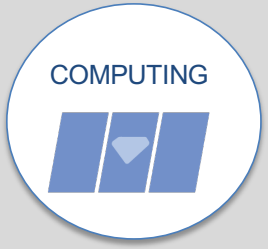
Data capture



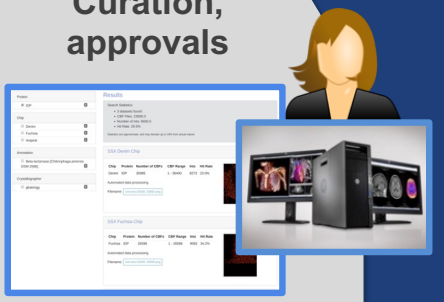


Processing and storage

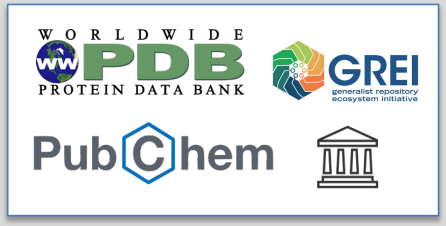
Metadata extraction, index ingest



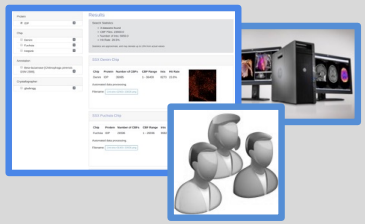
Curation, approvals



Community access via portal, science gateway or data commons



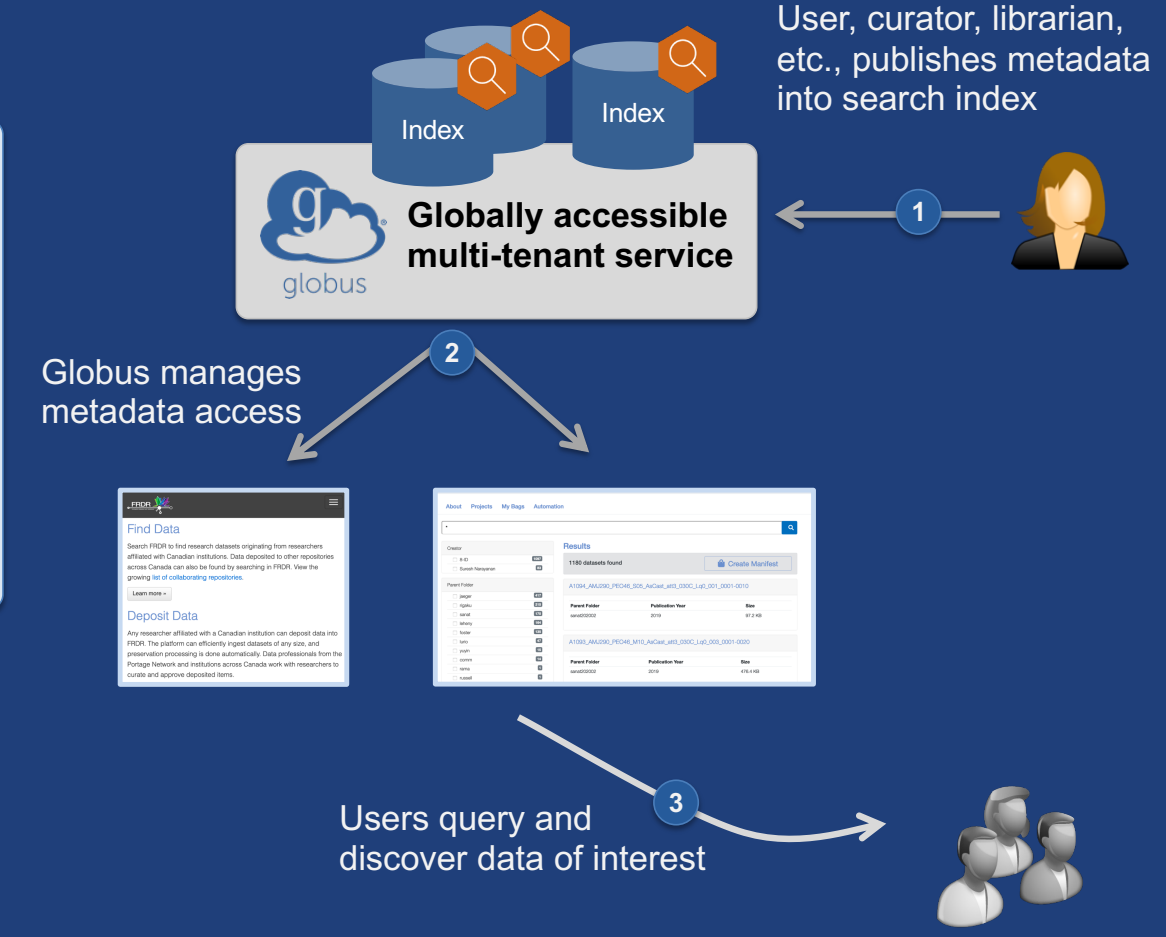
Access control/ publication for discovery



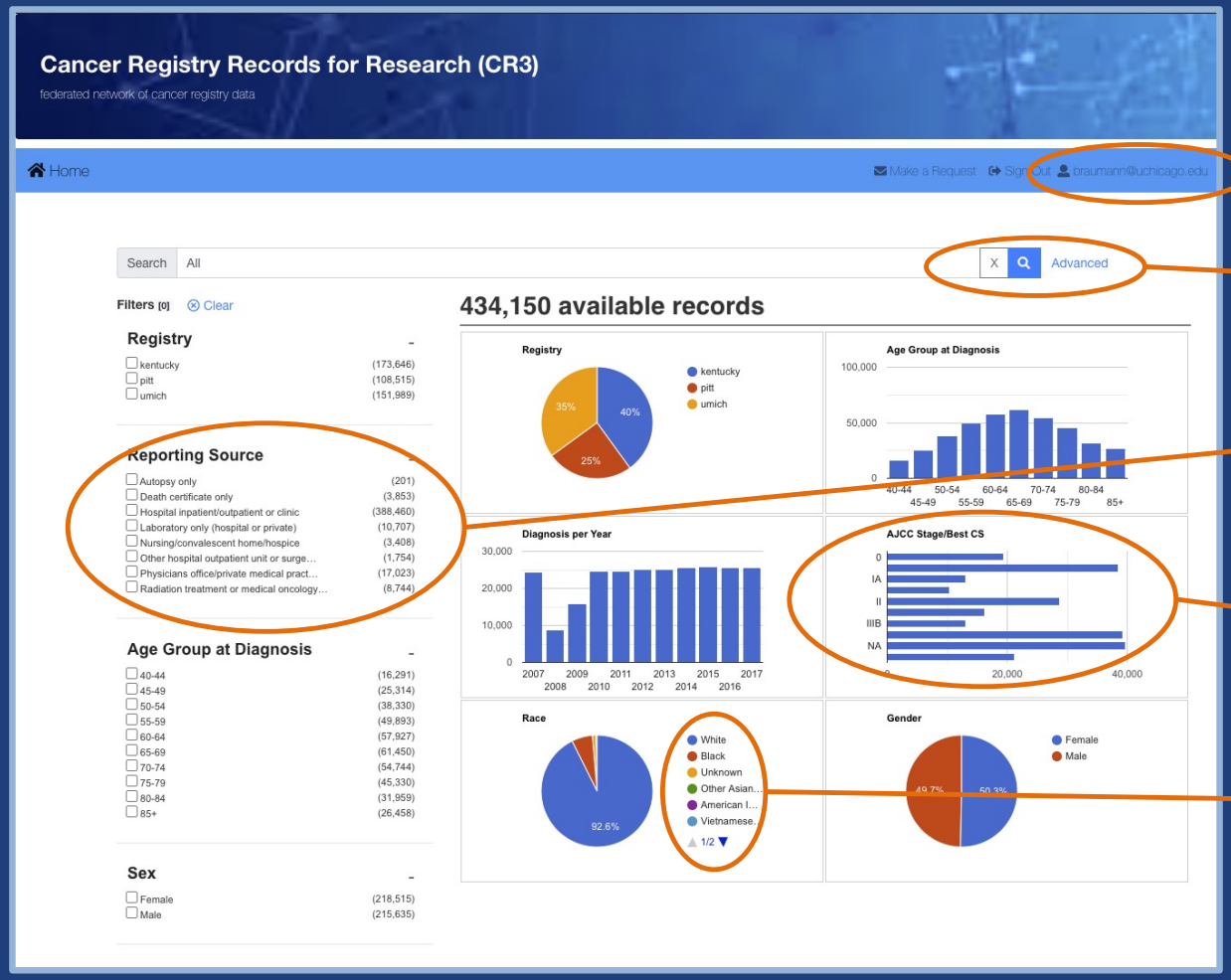
Data Identification

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

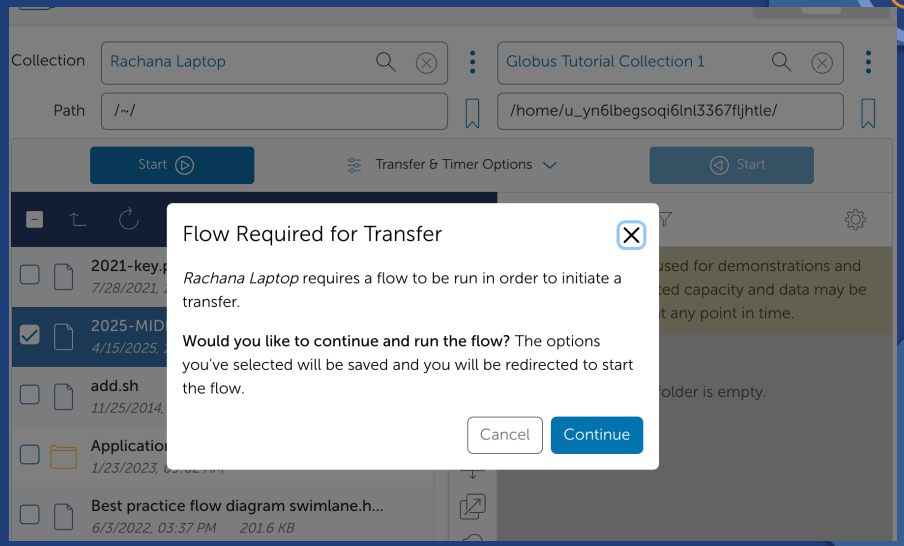
Restricted visibility to sensitive data

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



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

Globus Flows

Transfer

Transfer files staging area

Compute

Launch screening job

Notify

Notify admin

Choice

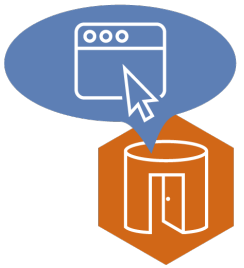
A / B

Admin appoval

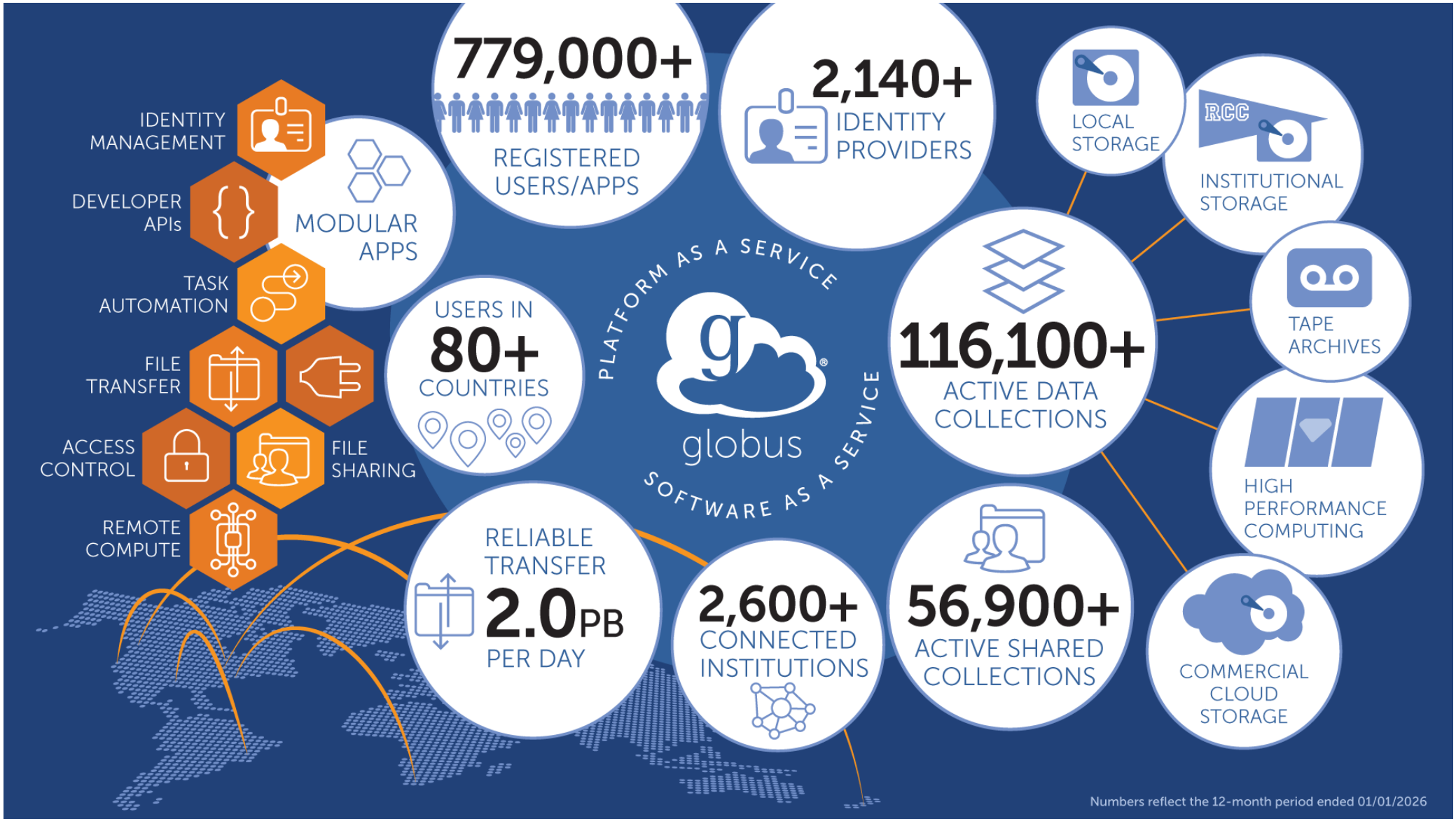
Transfer

Move final files

Using Globus Platform



Globus Python SDK	Globus Django Portal Framework	Globus JavaScript SDK	Globus Serverless Portal


779,000+
REGISTERED USERS/APPS
2,140+
IDENTITY PROVIDERS
IDENTITY MANAGEMENT
DEVELOPER APIs
TASK AUTOMATION
FILE TRANSFER
ACCESS CONTROL
FILE SHARING
REMOTE COMPUTE
MODULAR APPS
USERS IN
80+
COUNTRIES
PLATFORM AS A SERVICE
globus
SOFTWARE AS A SERVICE
RELIABLE TRANSFER
2.0PB
PER DAY
2,600+
CONNECTED INSTITUTIONS
116,100+
ACTIVE DATA COLLECTIONS
56,900+
ACTIVE SHARED COLLECTIONS
LOCAL STORAGE
RCC
INSTITUTIONAL STORAGE
TAPE ARCHIVES
HIGH PERFORMANCE COMPUTING
COMMERCIAL CLOUD STORAGE
Numbers reflect the 12-month period ended 01/01/2026

Thank you!

- **Engage: ranantha@uchicago.edu**
- **Website: globus.org**
- **Documentation: docs.globus.org**
- **Support: support@globus.org**