

Introduction to Globus

SGX3's FacultyHack24
September 19, 2024

Lev Gorenstein
lev@globus.org
levg@uchicago.edu



THE UNIVERSITY OF
CHICAGO





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.



Our team comprises ...

professional software developers
and business operators with
extensive experience in industry
and academia



Our freemium sustainability model

- Basic capabilities are available free of charge to anyone engaged in non-profit research
- Subscriptions enable multiple enhanced features for both end-users and system administrators
- Subscription pricing model based on fairness and equity

globus.org/subscriptions



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



542,000
REGISTERED
USERS/APPS

1,880+
IDENTITY
PROVIDERS

LOCAL
STORAGE

RCC
INSTITUTIONAL
STORAGE



USERS IN
80+
COUNTRIES



60,000+
ACTIVE
ENDPOINTS

TAPE
ARCHIVES

HIGH
PERFORMANCE
COMPUTING

RELIABLE
TRANSFER
1.8PB
PER DAY

2,600+
CONNECTED
INSTITUTIONS

15,000+
ACTIVE SHARED
ENDPOINTS

COMMERCIAL
CLOUD
STORAGE

2,504,099,232,162 MB TRANSFERRED





Globus is a data management platform that delivers fast and reliable data transfer, collaboration, and automation services, directly from **your own storage systems** via **software-as-a-service** using **existing identities**.

“your own storage” can mean different things



Research Computing HPC



ACCESS
NICIS
NATIONAL INTEGRATED
CYBERINFRASTRUCTURE SYSTEM
aarnet
Australia's Academic
and Research Network

National Resources



Personal Systems

Globus unifies data across
disparate systems [®]



Desktop Workstations



Archives



Instruments



Microsoft Azure
Blob Storage 
aws | **S3**
Google Cloud 
red cloud

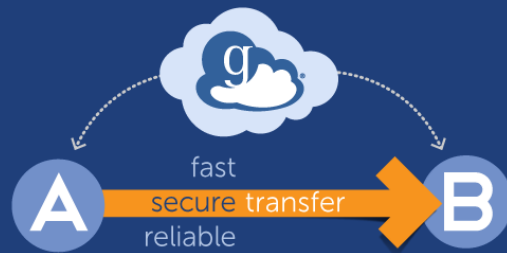
Public Cloud Storage

Globus Connectors support diverse systems

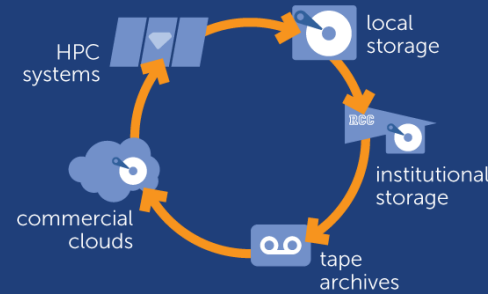




The Globus Web App and file transfer made simple



Managed transfer & sync

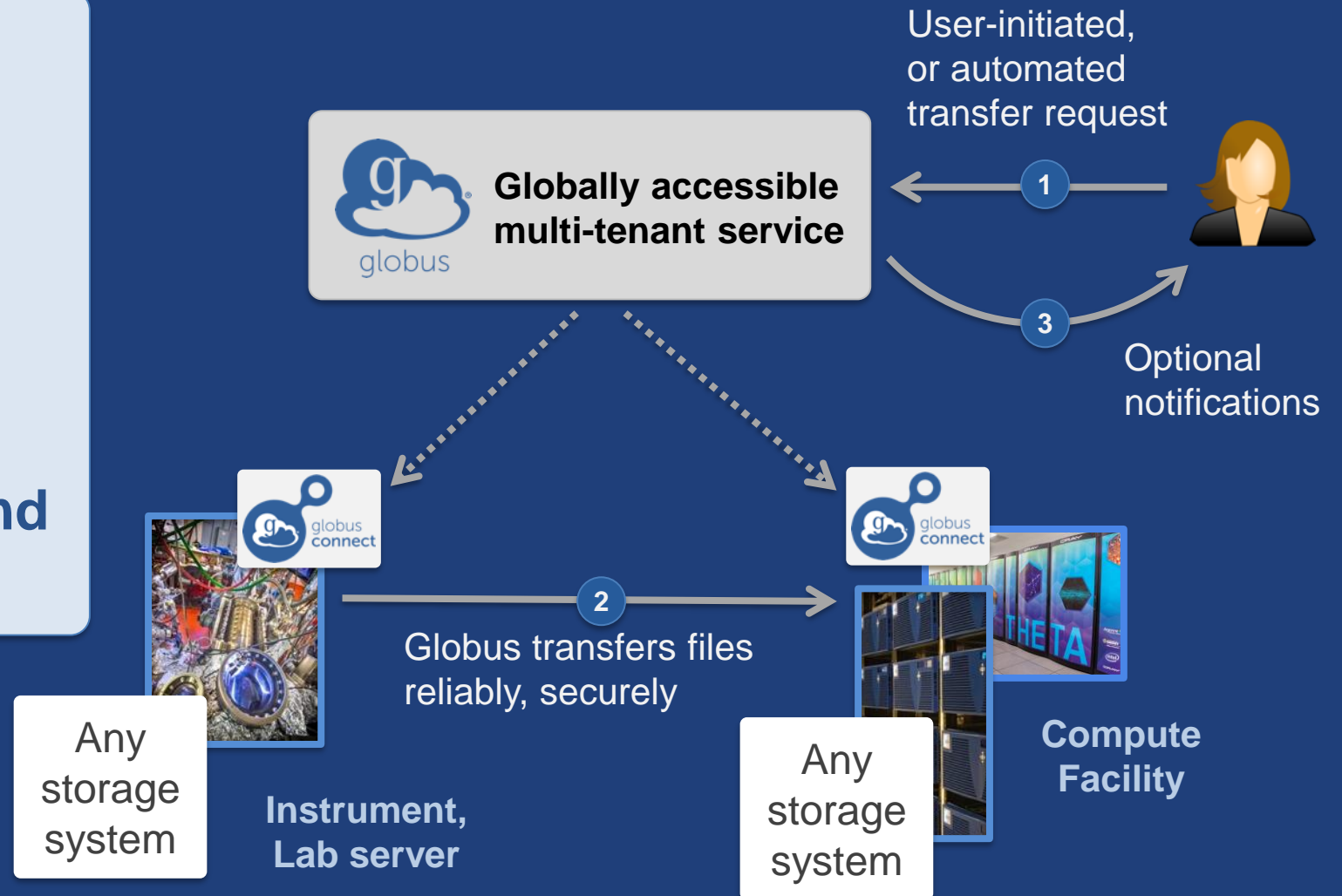


Unified data access



Assured data transfer ...no babysitter required.

- Fire-and-forget file transfer and synchronization
- Automated resolution of transient errors
- Browser-based access to “small” data
- Support for on-premises and cloud storage systems



Endpoints, Collections and Globus Connect

- Globus Connect is installed to instantiate an Endpoint
- Endpoints host Collections
 - Collections abstract storage
 - Users interact with Collections
 - There are two types of Collections
 - Mapped Collections – Accessed by a user that has a local account on the storage system that collection abstracts.
 - Guest Collections – A collection that uses an existing mapped collection and adds the ability of a user to share access to their data on that collection.

Globus Connect

- **Globus Connect Server (GCS)**

- Multi user Linux systems
- May have different storage types
- Installed by an administrator
- Multiple mapped collections



- **Globus Connect Personal (GCP)**

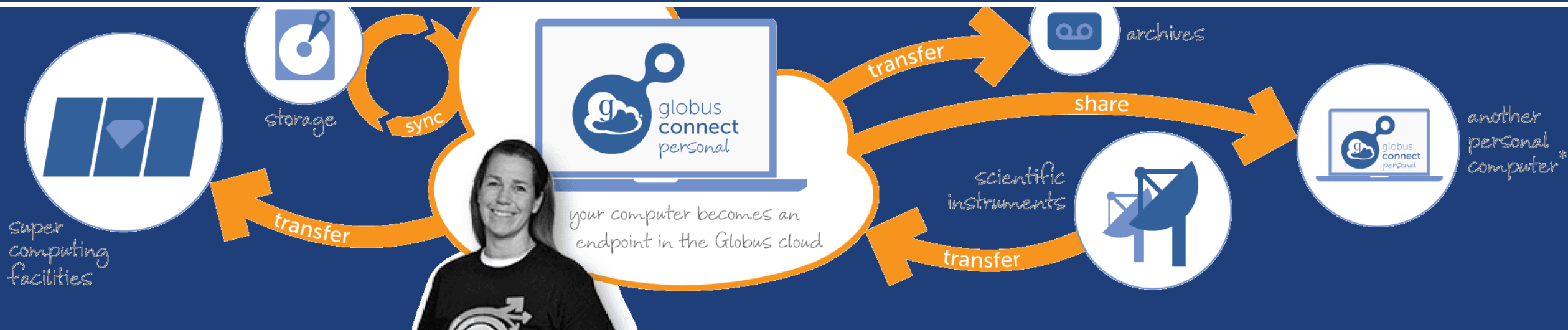
- Personal workstations and laptops (single user)
- Installed by the user
- Single mapped collection





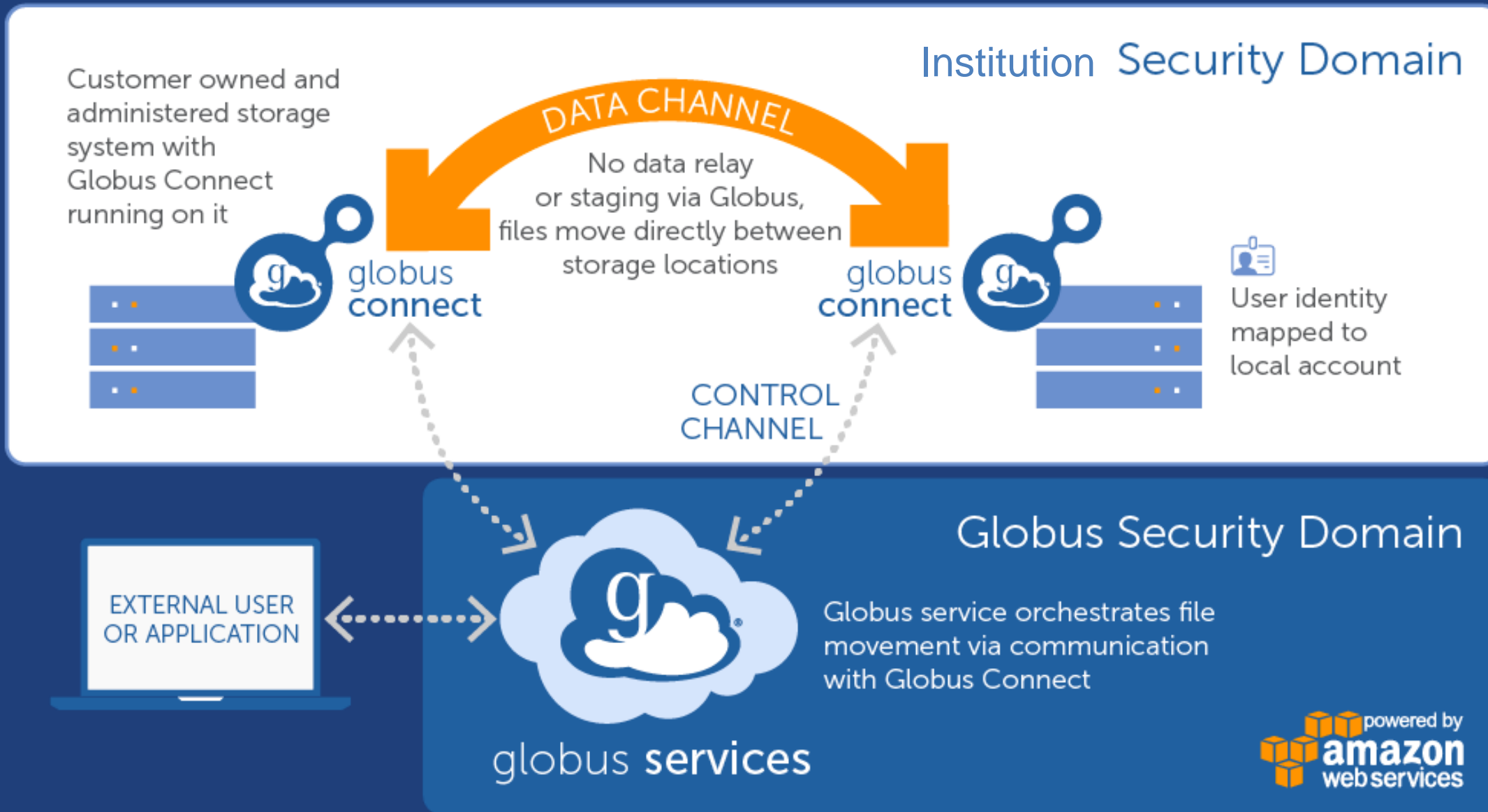
Installing Globus Connect Personal

globus.org/globus-connect-personal





Quick review





Globus data movement, a tale of two protocols

- Globus Transfer
 - GridFTP
 - Fast
 - Reliable (network back off, fire and forget, retries on failures)
 - A remote control for data
- Globus Downloads
 - HTTPS
 - Intended for situations where users don't have the option of a destination endpoint
 - Direct to the machine doing the pulling
 - Likely to bring great sadness for large collections of files



GridFTP is reliable, secure ...and fast!

Activity List



RDA to ALCF noverify

transfer completed

72.8Gbps



Overview



Event Log

Task Label	RDA to ALCF noverify
Source	NCAR RDA Dataset Archive ⓘ
Destination	DME PerfTest - Argonne ⓘ
Task ID	20ebf766-a46d-11eb-8a95-d70d98a40c8d
Owner	Vas Vasiliadis (vas@globusid.org)
Condition	SUCCEEDED
Requested	2021-04-23 02:50 pm
Completed	2021-04-23 02:53 pm
Duration	2 minutes 47 seconds

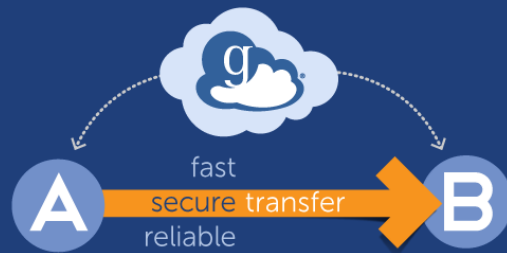
- Transfer Settings
- transfer is not encrypted
 - overwriting all files on destination

6151	Files
2	Directories
1.51 TB	Bytes Transferred
9.10 GB/s	Effective Speed
0	Skipped files on sync
0	Skipped files on error

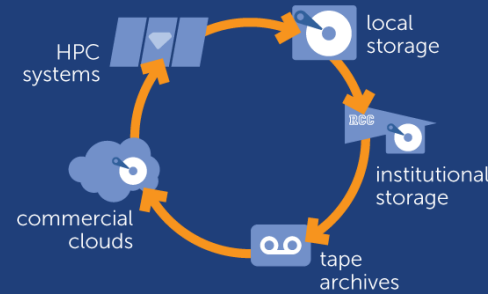
[View debug data](#)



The Globus Web App and a common data transfer use case

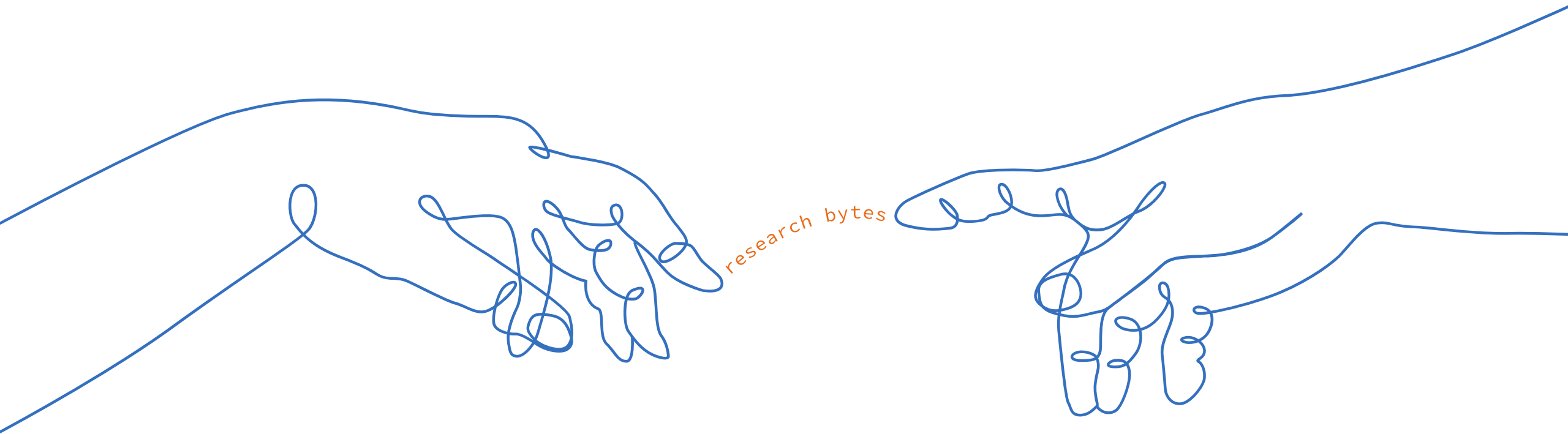


Managed transfer & sync



Unified data access

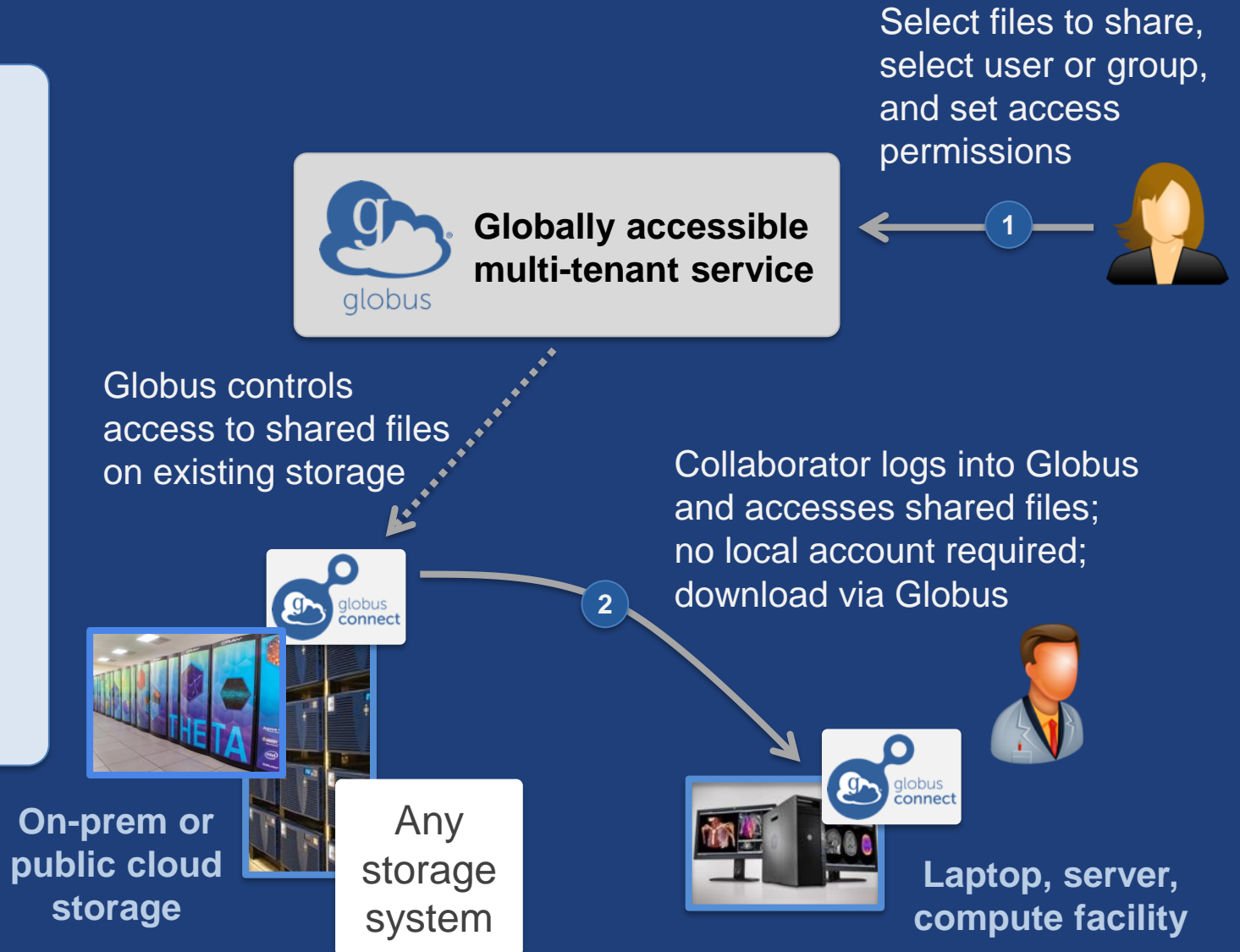




Scholarship is inherently collaborative ...but how do we share, distribute data easily yet securely?

Easy data sharing ...no “special” treatment required

- Share with collaborators at any institution, company
- Share from current storage - no data staging required
- Storage system administrators can enforce compliance policies
- Fine-grained access control





The Globus Web App and a data sharing use case



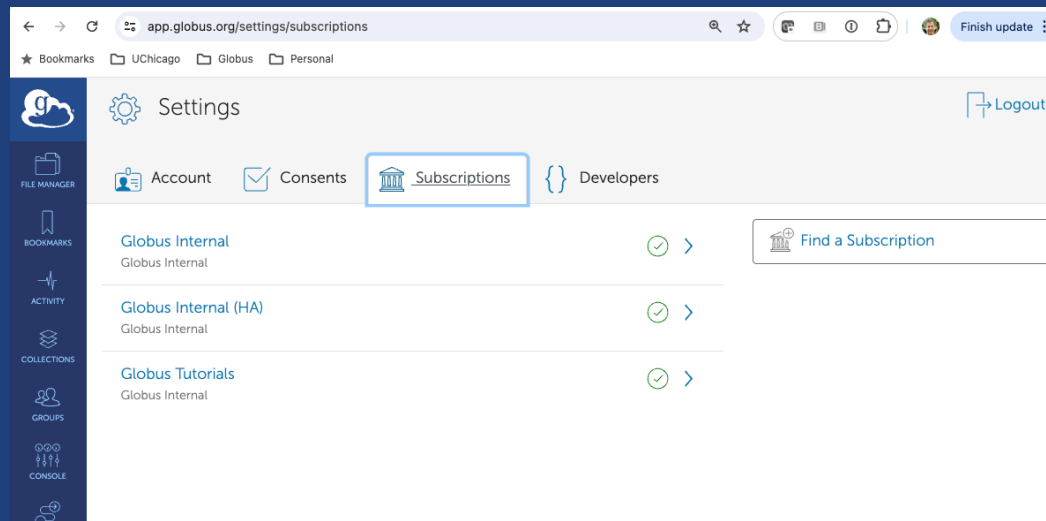


Globus Subscriptions

- By having a role in a Subscription Management Group and / or subscribing resources, Subscribers can enable
 - File Sharing: The ability to create Guest Collections
 - Premium Connectors: beyond POSIX
 - Administrative Oversight: The Globus Console
 - Usage Reports: Monthly and “since the beginning” stats
 - Data Automation Services: The ability to deploy more than one Flow
 - HTTPS: Add upload / download to collections
 - Manage protected data: Subscription options for protected data
 - High Assurance: PII, CUI
 - BAA: PHI
 - ... and more!
 - globus.org/why-subscribe



How can I find out if my institution is a Globus Subscriber and how can I join our Subscription Group?





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
- **Data remain at institutions, not stored by Globus**
- **Integrity checks of transferred data**
- **High availability and redundancy**
- **Encryption of user files and Globus control data**

Globus for protected data management

Security controls

- NIST 800-53
- 800-171 Low+



Restricted data handling

- PHI, PII, CUI
- Compliant data sharing

BAA w/Uchicago

- UChicago BAA with Amazon

Globus High Assurance features

- **Additional authentication assurance**
 - Authenticate with a *specific identity* within session
 - Reauthenticate after specified time period
- **Session/device isolation**
 - Authentication context is per application, per session
- **Enforces encryption of all user data in transit**
- **Audit logging**





With Globus, it's not a tradeoff you need to make...

Access
Sharing
Collaboration



Security
Privacy
Compliance



TIMTOWTDI

- **Globus webapp (interactive): app.globus.org**
- **Globus CLI (scriptable!): docs.globus.org/cli/**
- **Platform tools (developer-friendly): globus.org/platform/services/tools/**
 - Python SDK
 - JavaScript SDK
 - Web Helper pages

Data portals

- Either static, or backed by Globus Search service (an indexed database hosted by us)
- Example: acdc.alcf.anl.gov
- You can have one too
 - Django framework (you host)
github.com/globus/django-globus-portal-framework
 - Or zero infrastructure one (based on GitHub Pages)
github.com/globus/template-data-portal

Support resources

- Globus documentation: docs.globus.org
- YouTube channel: youtube.com/GlobusOnline
- Helpdesk: support@globus.org
- Mailing Lists: globus.org/mailing-lists
- Customer engagement team (office hours)
- Professional services team (advisory, custom work)