iRODS and Globus Deployment at the VSC

Ingrid Barcena-Roig, KU Leuven
Vas Vasiliadis, University of Chicago

iRODS UGM – July 7, 2022

KU LEUVEN

THE UNIVERSITY OF CHICAGO

Globus is …

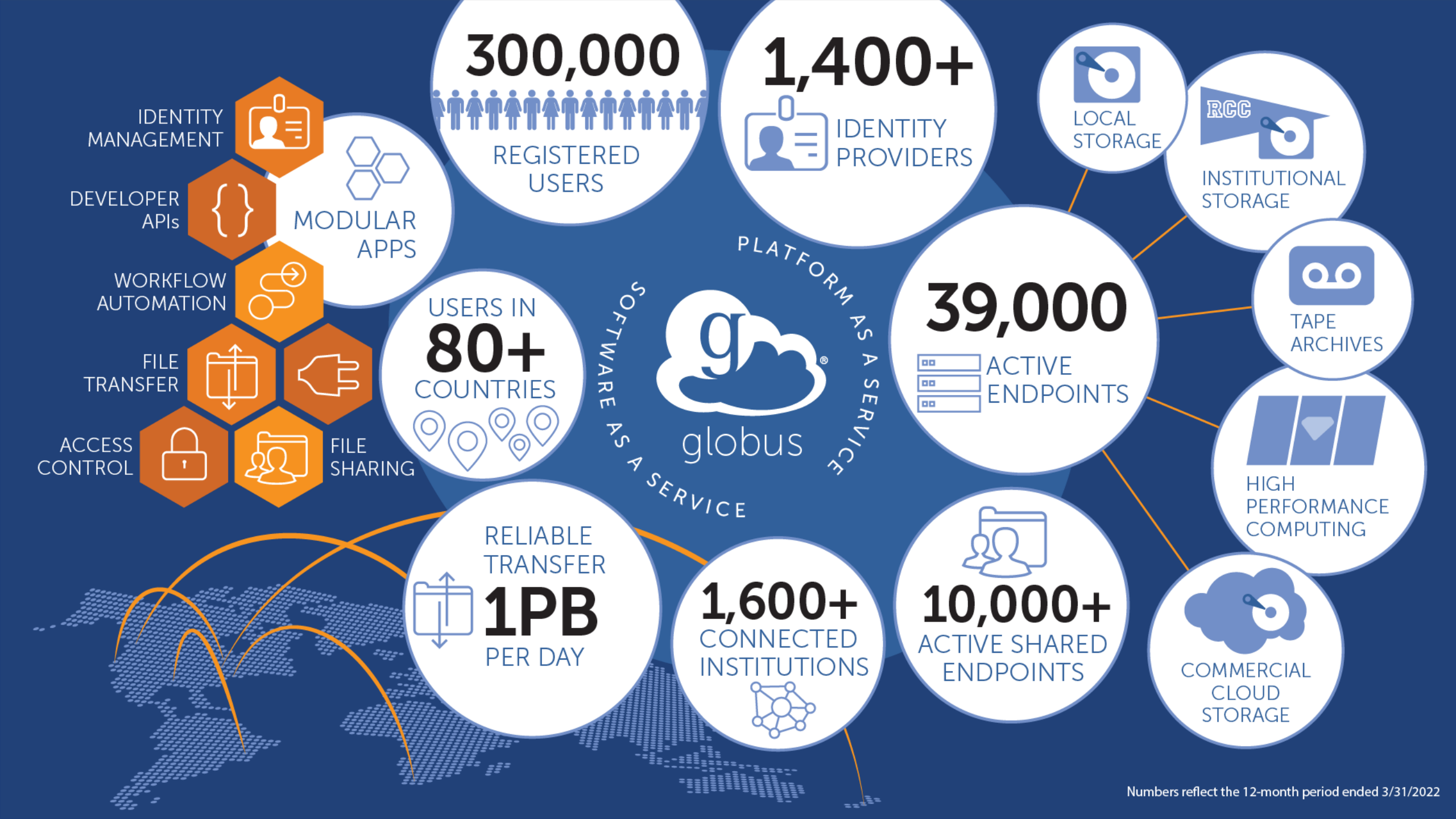
a non-profit service developed and operated by

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.

…or “We enable researchers to outsource the mundane–yet critical–data management tasks that impede their work.”


300,000
REGISTERED USERS
1,400+
IDENTITY PROVIDERS
IDENTITY MANAGEMENT
DEVELOPER APIs
WORKFLOW AUTOMATION
FILE TRANSFER
ACCESS CONTROL
FILE SHARING
MODULAR APPS
USERS IN 80+ COUNTRIES
SOFTWARE AS A SERVICE
PLATFORM AS A SERVICE
globus
LOCAL STORAGE
RCC
INSTITUTIONAL STORAGE
TAPE ARCHIVES
39,000
ACTIVE ENDPOINTS
HIGH PERFORMANCE COMPUTING
RELIABLE TRANSFER
1PB
PER DAY
1,600+
CONNECTED INSTITUTIONS
10,000+
ACTIVE SHARED ENDPOINTS
COMMERCIAL CLOUD STORAGE
Numbers reflect the 12-month period ended 3/31/2022

Development is partly 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

Globus and the research data lifecycle

Instrument

Globus transfers files reliably, securely

2

Compute Facility

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

Transfer

3

Share

1

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

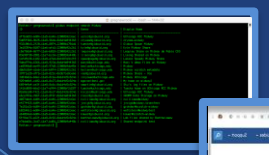
Researcher selects files to share, selects user or group, and sets access permissions

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

5

Personal Computer

6 The Globus Command Line Interface, API sets, Python SDK and Action Providers give you a platform…



Build

Search

Automate

7 … for building science gateways, portals and automations.

8

Streamlining research workflows and ensuring those that need access to the data have it.

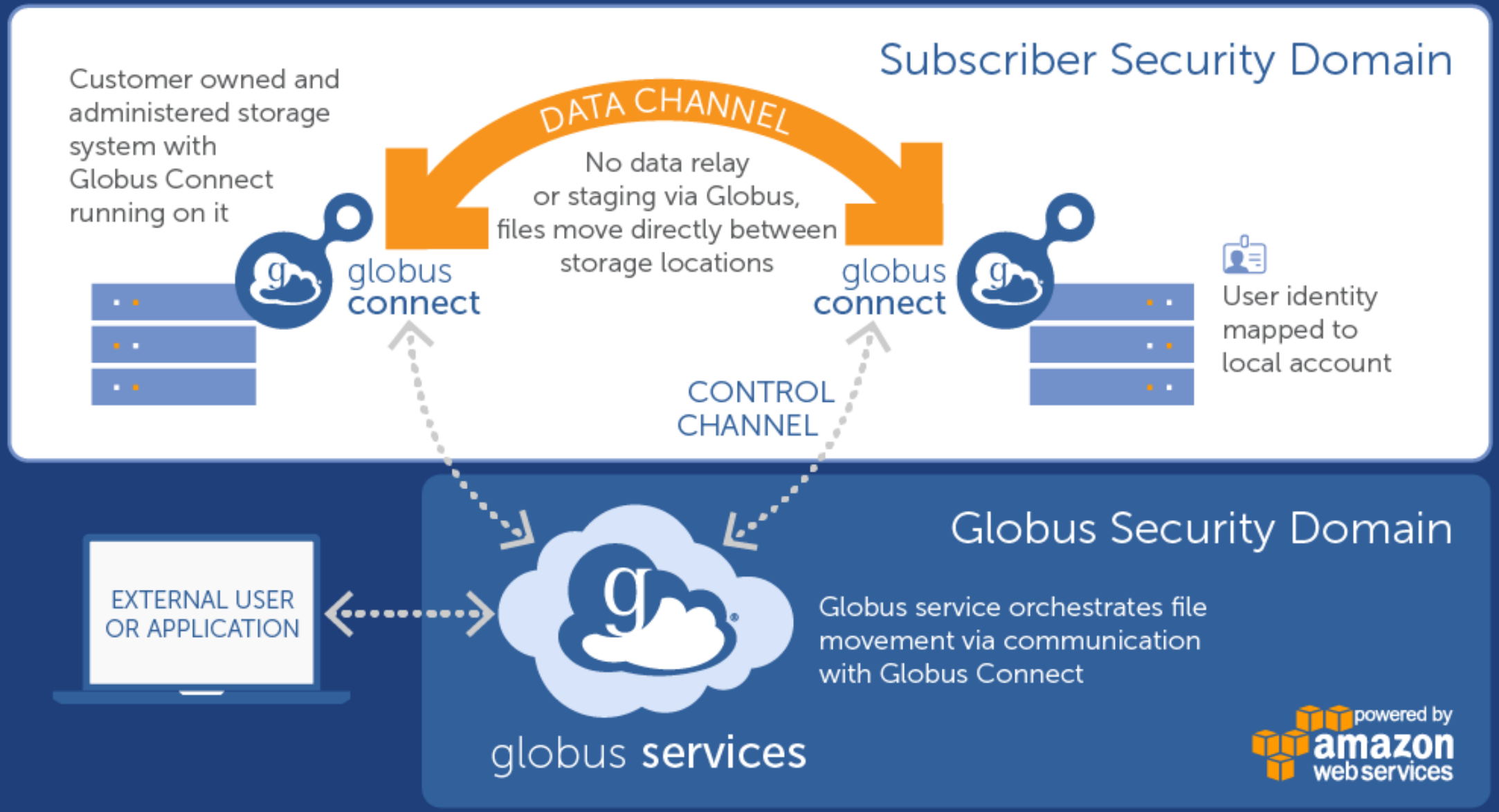
- **Use a Web browser, CLI or platform services**
- **Access any storage**
- **Use an existing identity**

Globus is SaaS and PaaS

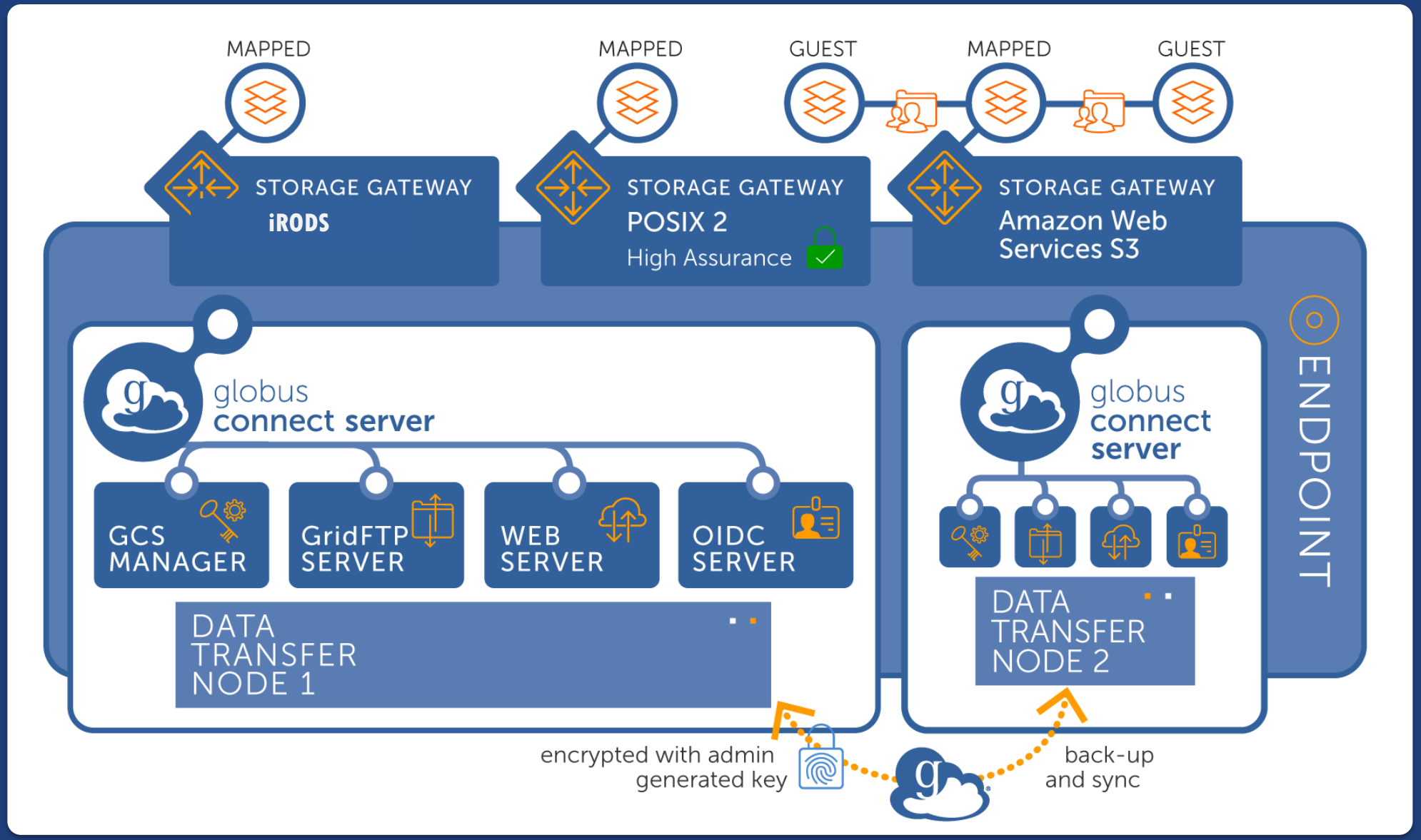
Web app simplifies *ad hoc* data management

Platform services simplify creation of automated data management apps for diverse instrument types

Conceptual Architecture: Hybrid SaaS

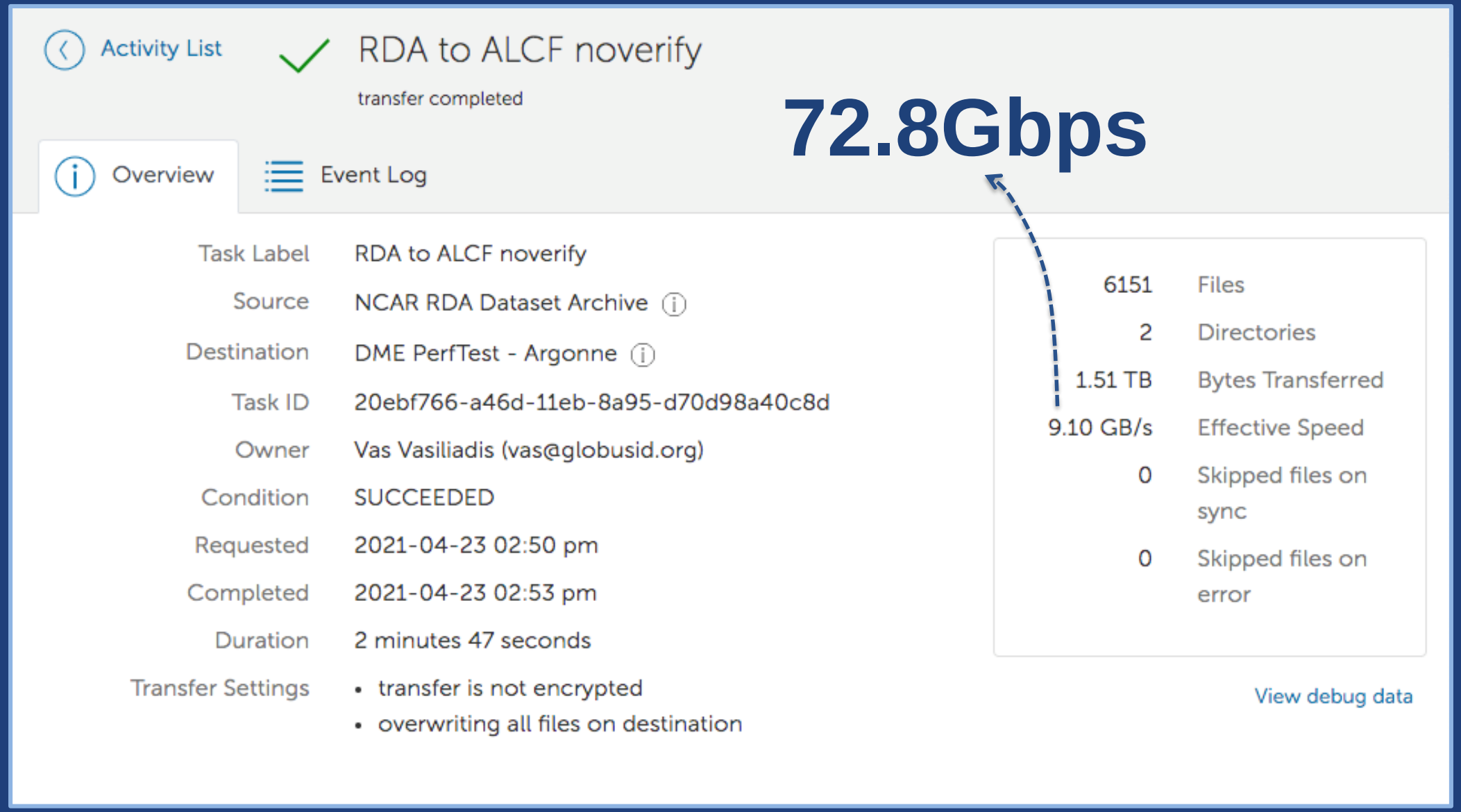


Globus Connect Server Architecture



Globus supports a broad set of storage systems

Globus is fast …and, more importantly, reliable



Globus is secure

- **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**

Managing protected data with Globus

Security controls

→ NIST 800-53

→ 800-171 Low+

Restricted data handling

→ PHI, PII, CUI

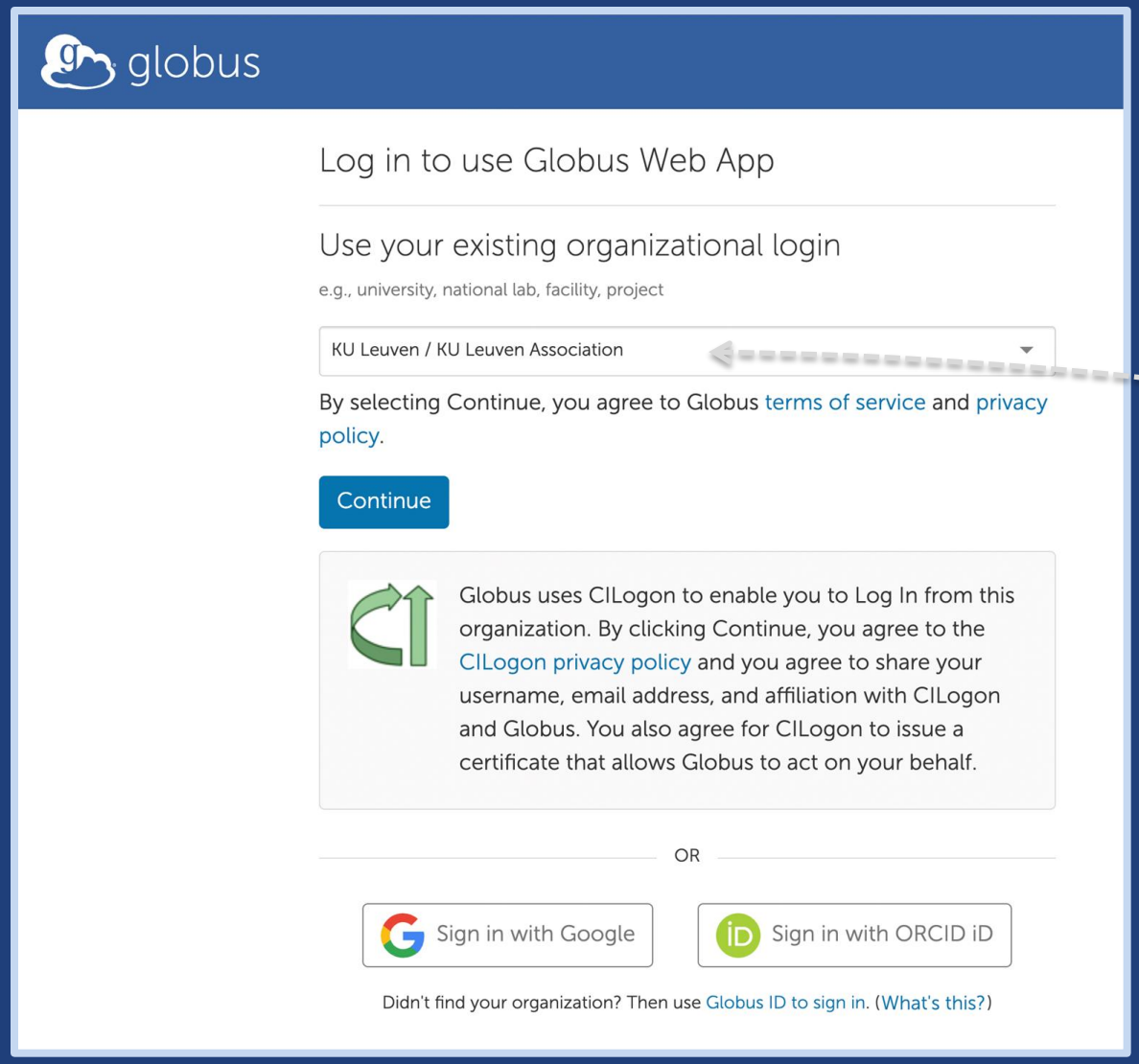
→ Compliant data sharing

BAA w/Uchicago

→ UChicago BAA with Amazon

A quick peek…

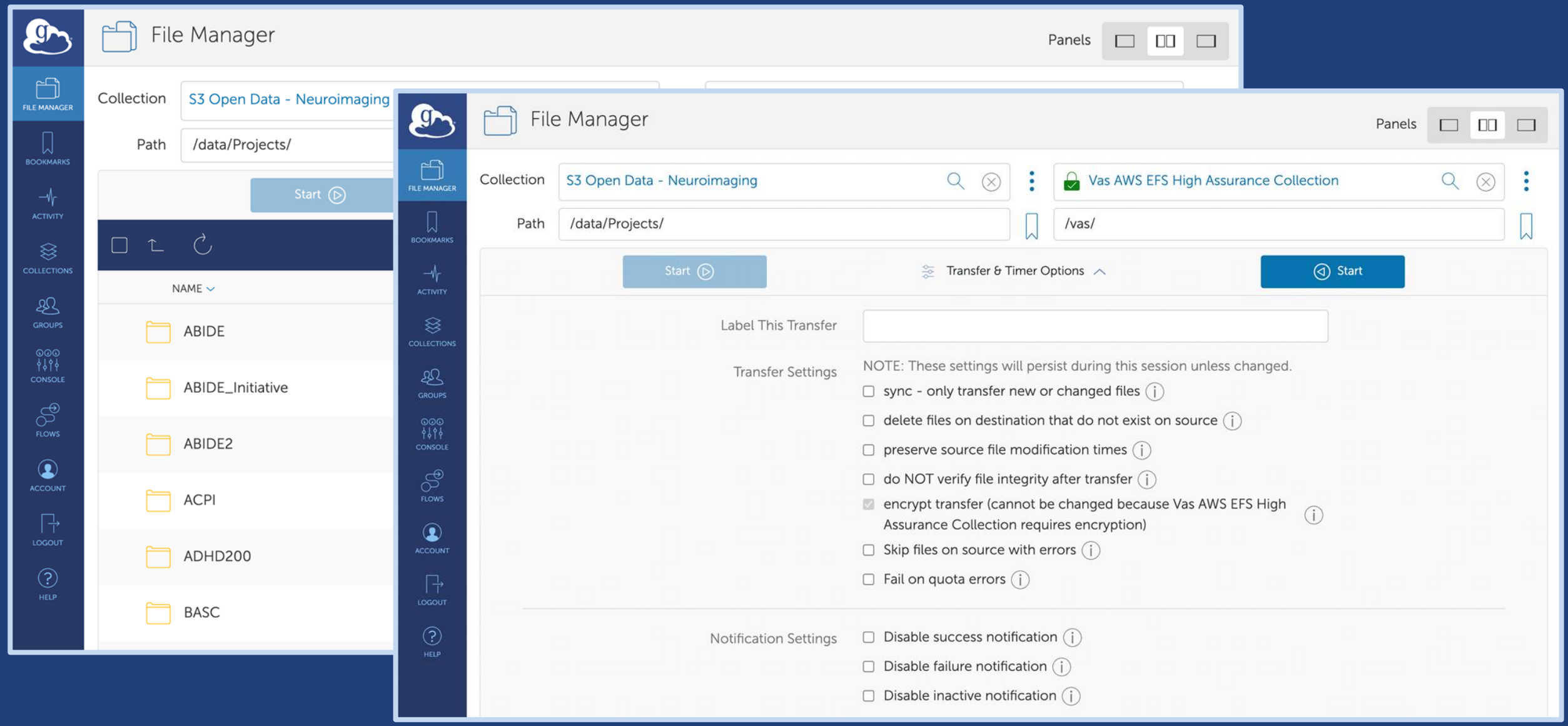
Federated login (via Globus Auth: OAuth2/OIDC)



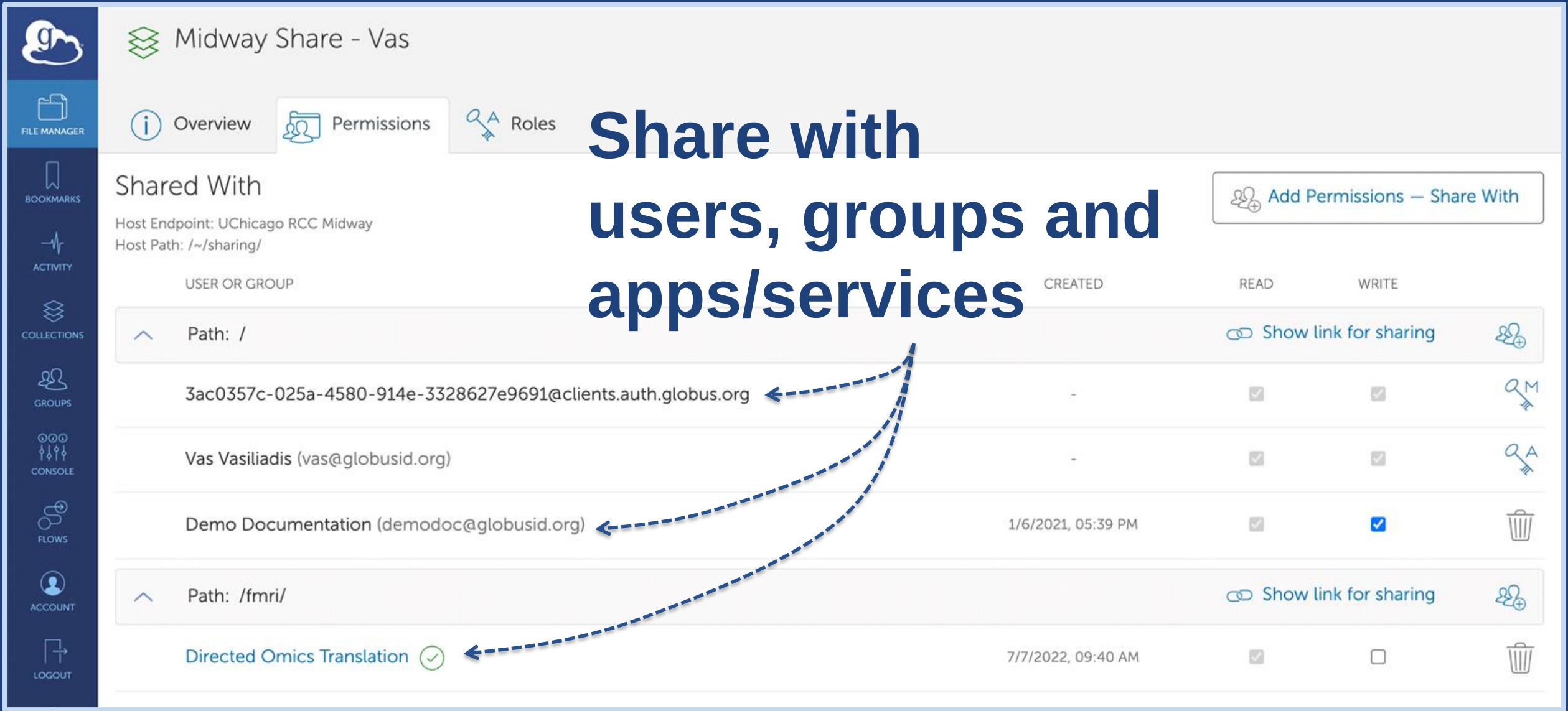
1,400+ IdPs

…via released R&S attributes or custom configuration

Transfer options and Globus “cron”



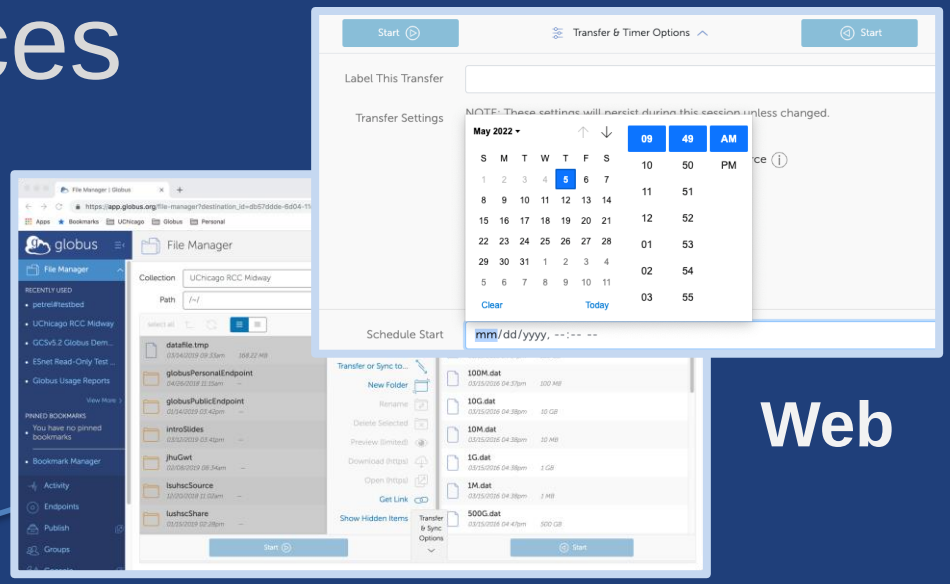
Sharing: Globus maintains a virtual ACL



One service, many interfaces

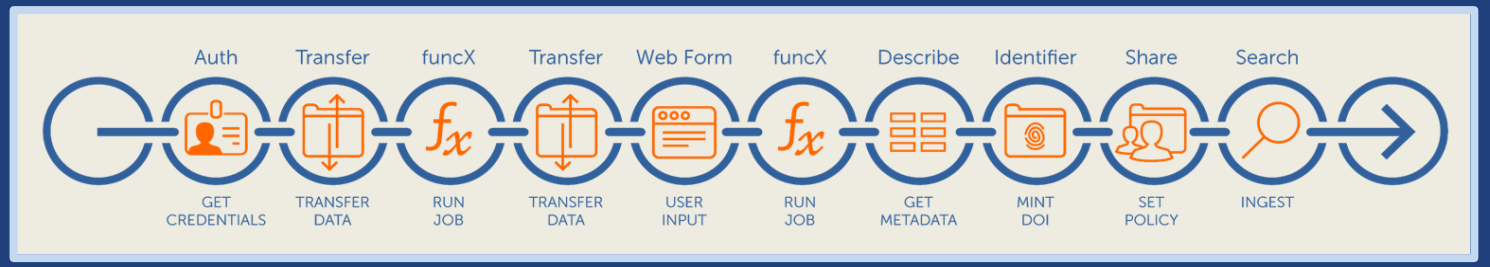
```
(globus-cli) jupiter:~ vas$ globus
Usage: globus [OPTIONS] COMMAND [ARGS]...

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

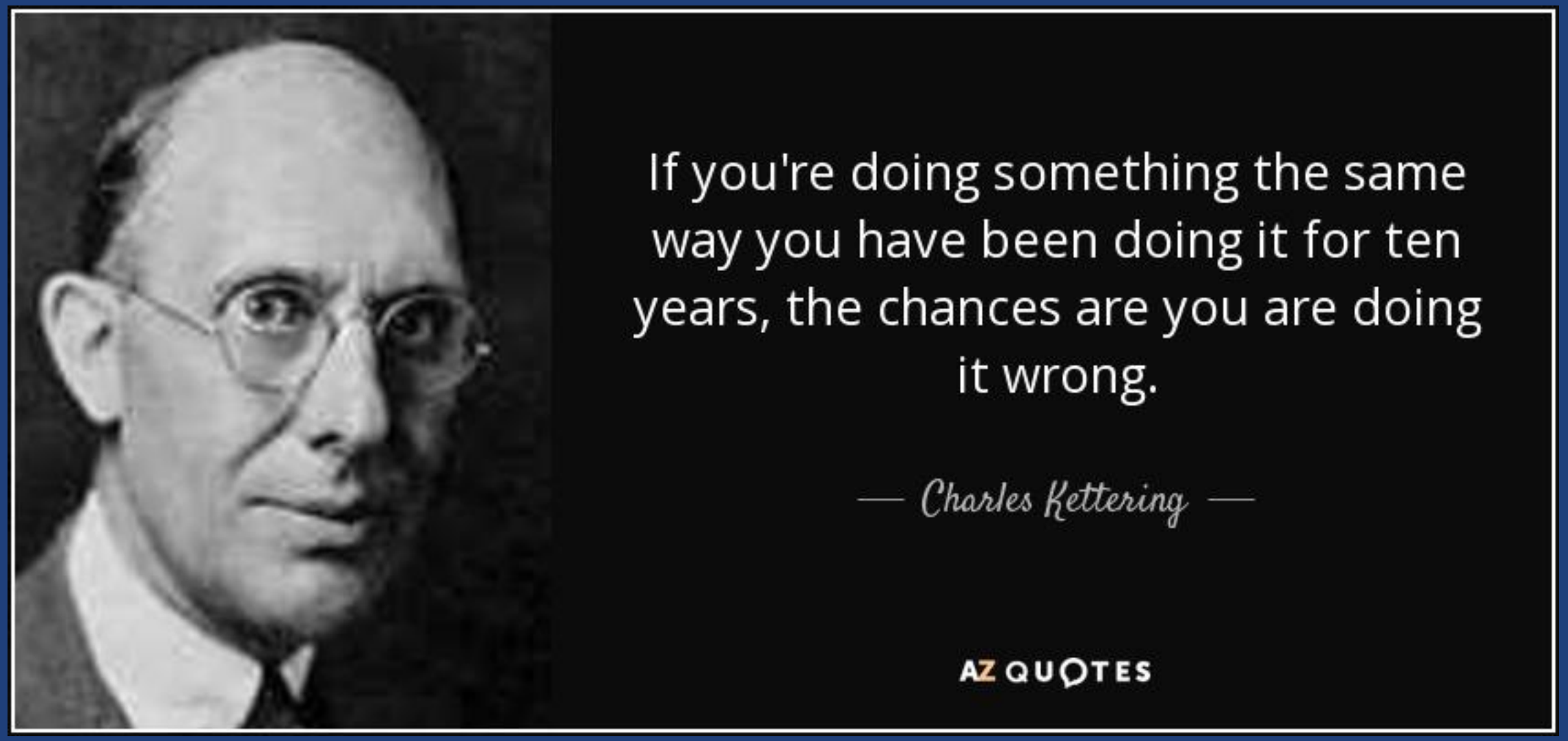
```
GET /endpoint/go%23ep1
PUT /endpoint/vas#my_endpt
200 OK
X-Transfer-API-Version: 0.10
Content-Type: application/json
…
```

Rest API

What's driving the development agenda?

Compliance with diverse data privacy and other legal protections

“I love scp/sftp/…; it works great in my perl scripts!”



Instrument innovations, research processes are evolving at breakneck speed.

How can we adapt our data management to keep pace?

Current Focus:
Automating data management at scale

3 Degrees of Automation

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

Understanding SARS-CoV-2 Protein Structure

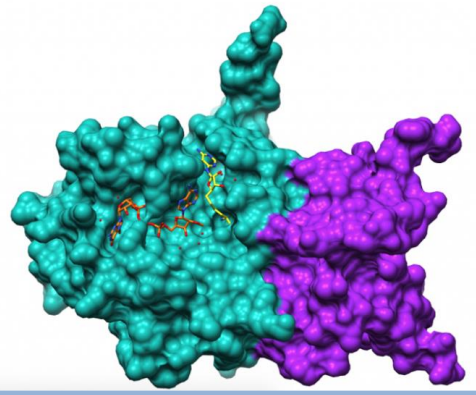
"These data services have taken the time to solve a structure from weeks to days and now to hours"

Darren Sherrell, SBC beamline scientist APS Sector 19

SCIENCE

Argonne researchers use Theta for real-time analysis of COVID-19 proteins

AUTHOR NILS HEINONEN
PUBLISHED 07/28/2020
DOMAIN BIOLOGICAL SCIENCES
SYSTEMS THETA

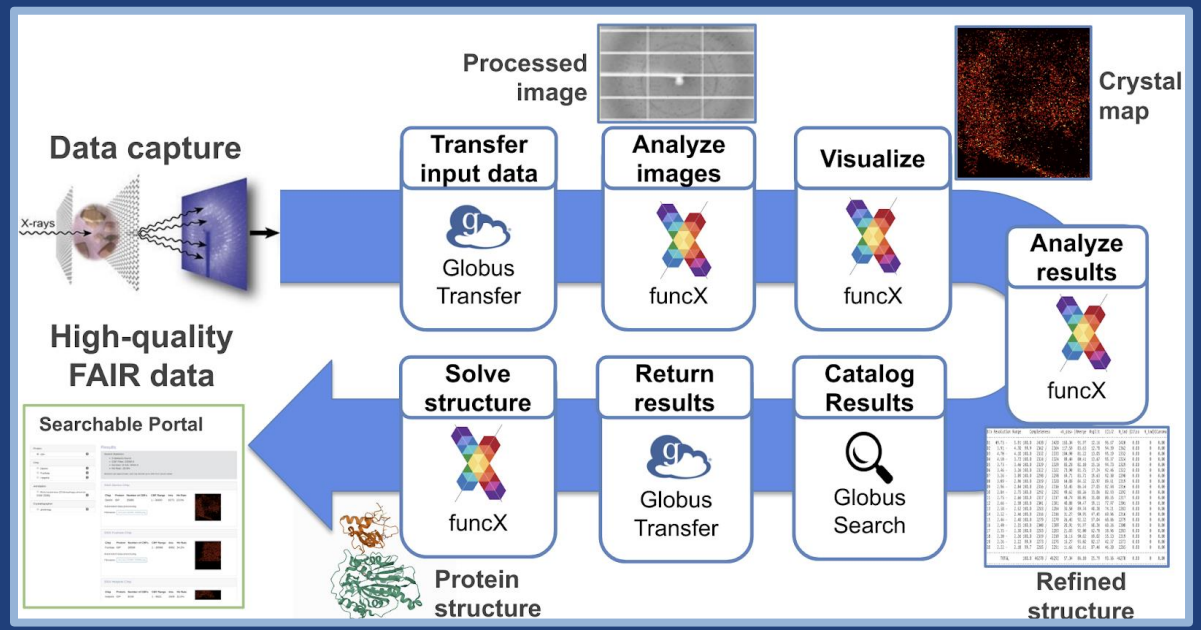


HPC wire

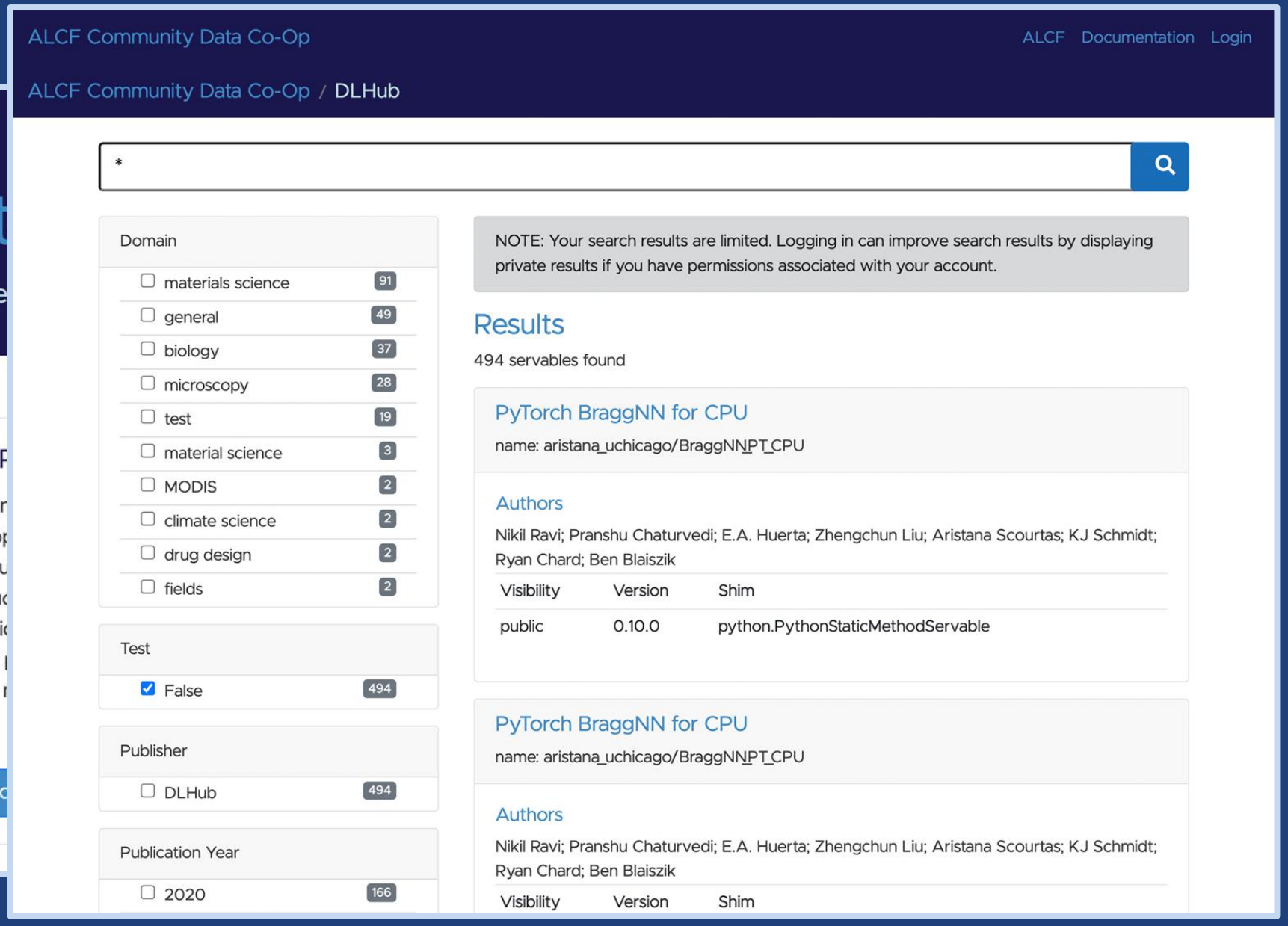
Since 1987 - Covering the Fastest Computers in the World and the People Who

Argonne's User Facilities Continue to Enable Critical Work Combating and Addressing the Impacts of the COVID-19 Epidemic

June 12, 2020



Portals for data publication and faceted search



globus.org

app.globus.org

docs.globus.org

globus.org/**connectors**

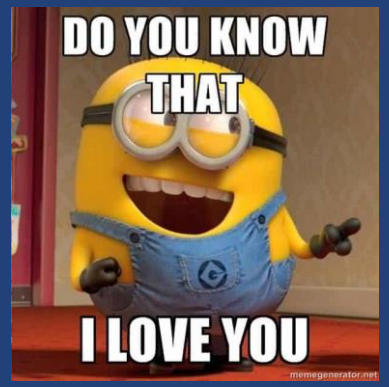
globus.org/**subscriptions**

support@globus.org

VSC iRODS and Globus use cases

KU LEUVEN

iRODS and Globus?



But:

- A lot of windows users: WSL/iCommands? PRC?...
- Reliability: network issues, server problems …..
- Repetitive transfers: scripting for everyone?
- Sharing with externals: create externals accounts, itickets?

KU LEUVEN

iRODS and Globus!

iRODS

Store
Describe
Discover
Automate

Globus

Move
Share

KU LEUVEN

VSC use cases

Transfer

Ingest

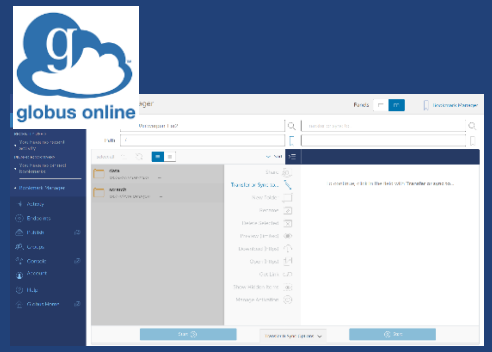
Share

KU LEUVEN

Transfer

Authentication
institutional account

ORCID



Laptop

Department

KU LEUVEN

OneDrive iRODS HPC–Tier-2

GHENT UNIVERSITY

OneDrive HPC–Tier-2

Universiteit Antwerpen

OneDrive HPC–Tier-2

VUB

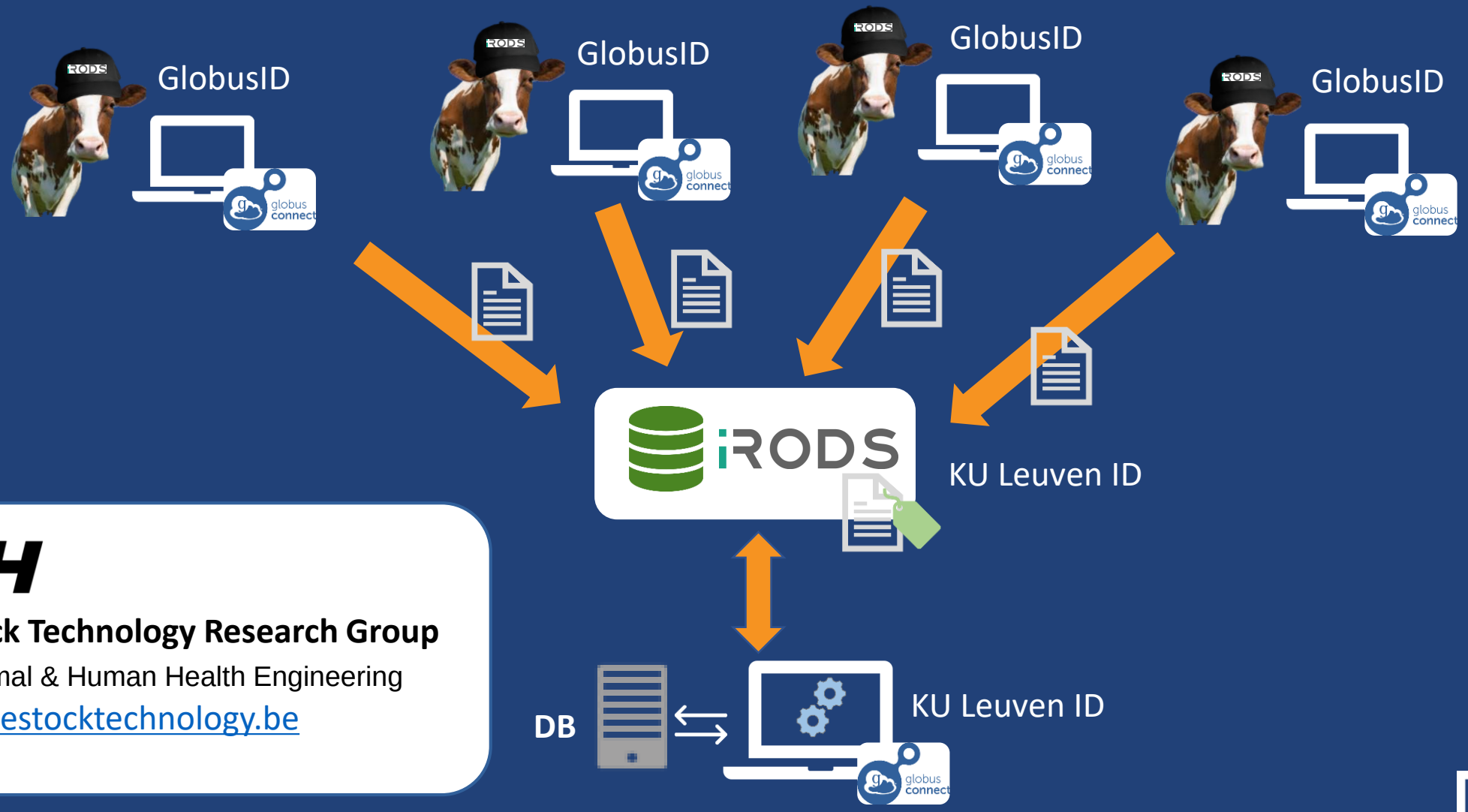
OneDrive HPC–Tier-2

VSC

HPC-Tier-1b HPC–Tier-1c iRODS

KU LEUVEN

Ingest (PoC): milk quality control at dairy farms



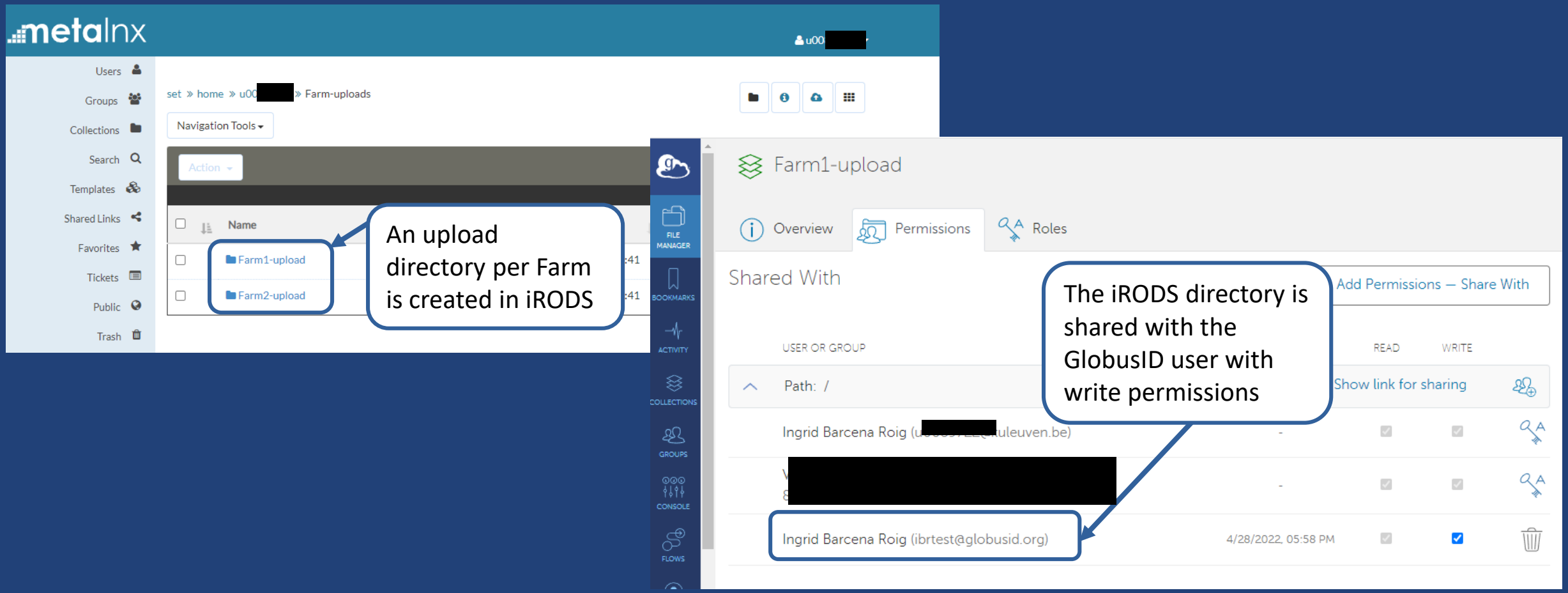
Livestock Technology Research Group

Dept Animal & Human Health Engineering

www.livestocktechnology.be

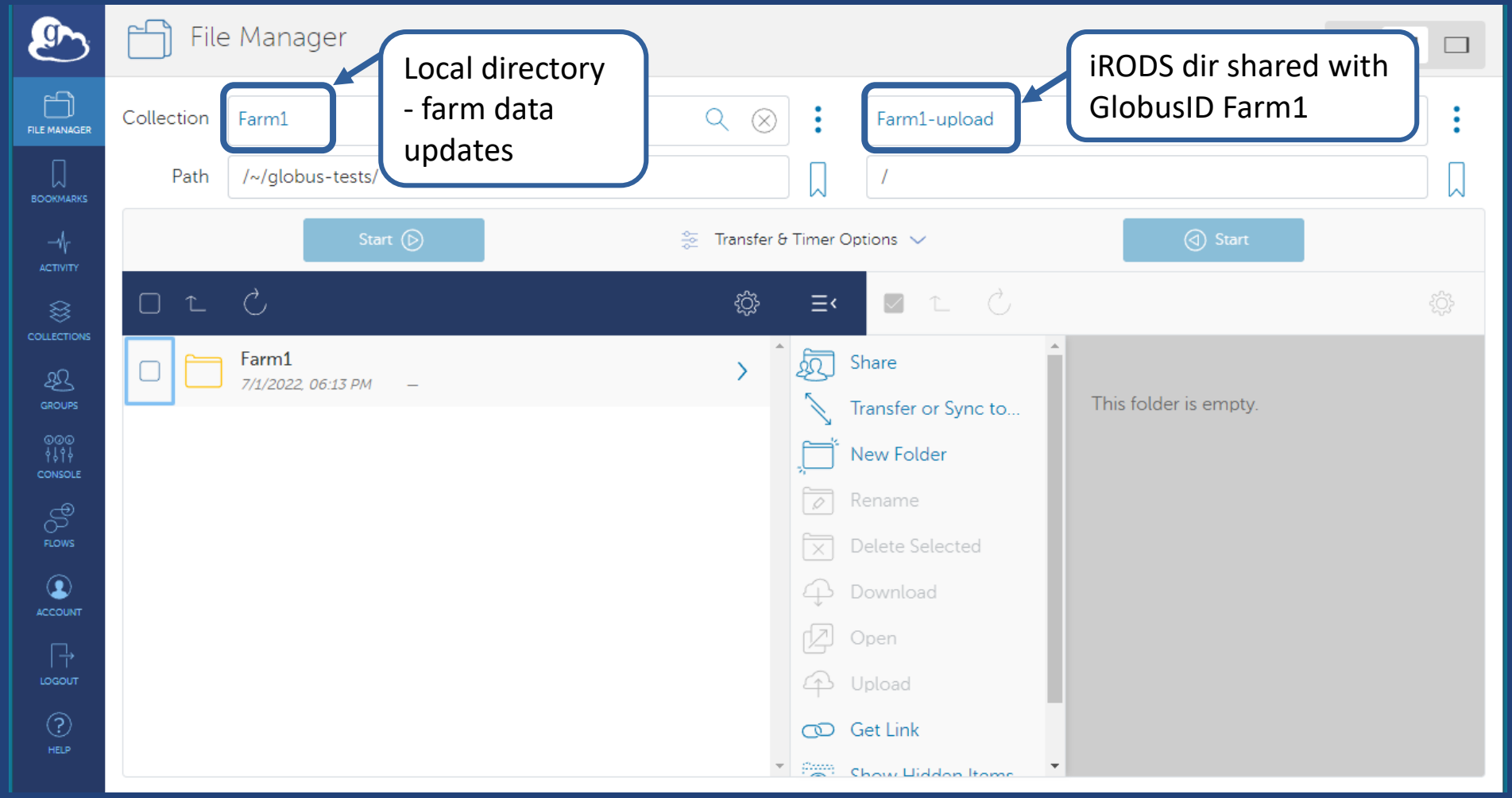
KU LEUVEN

Configuration: iRODS



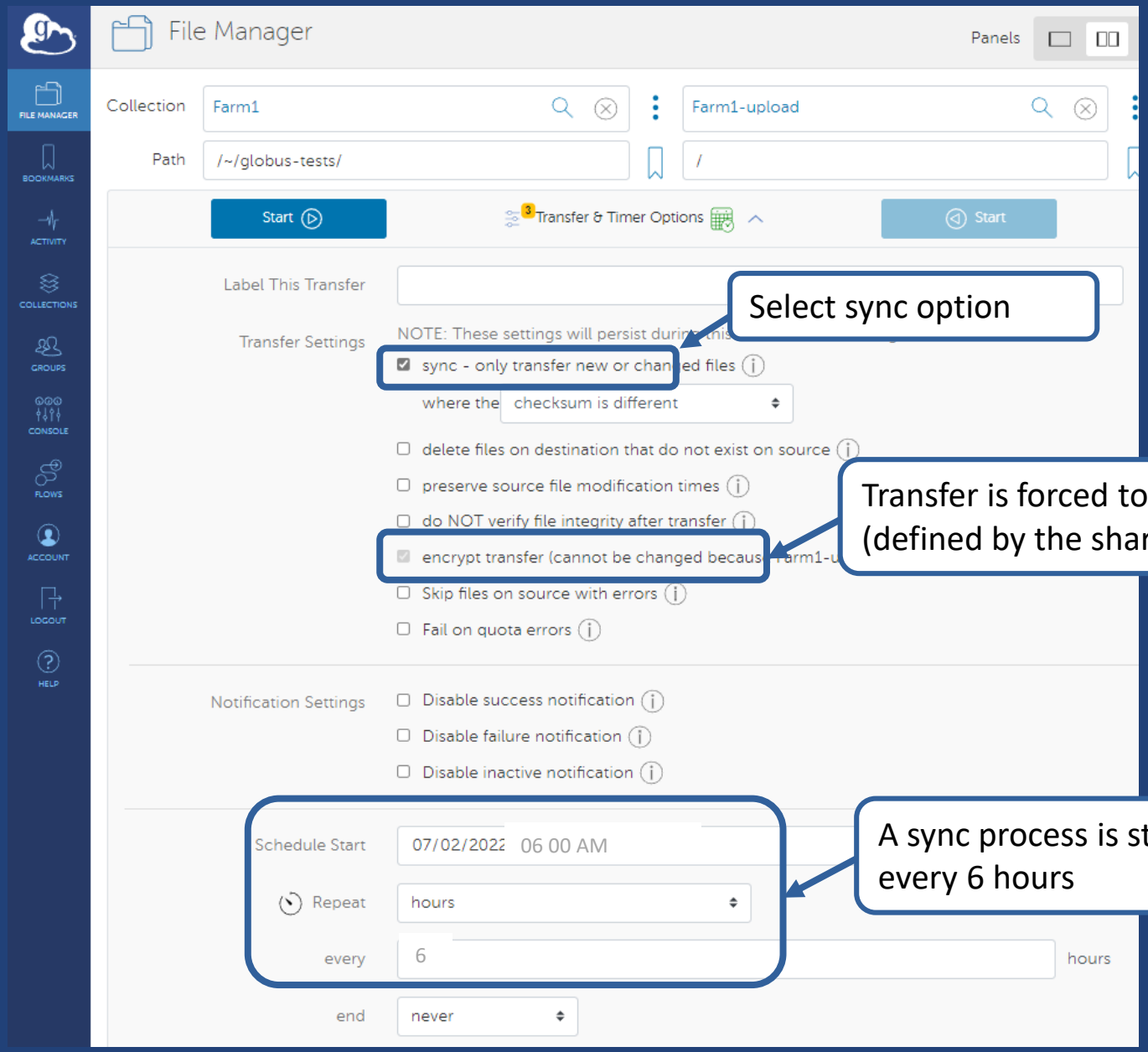
KU LEUVEN

Configuration: farm (1/2)



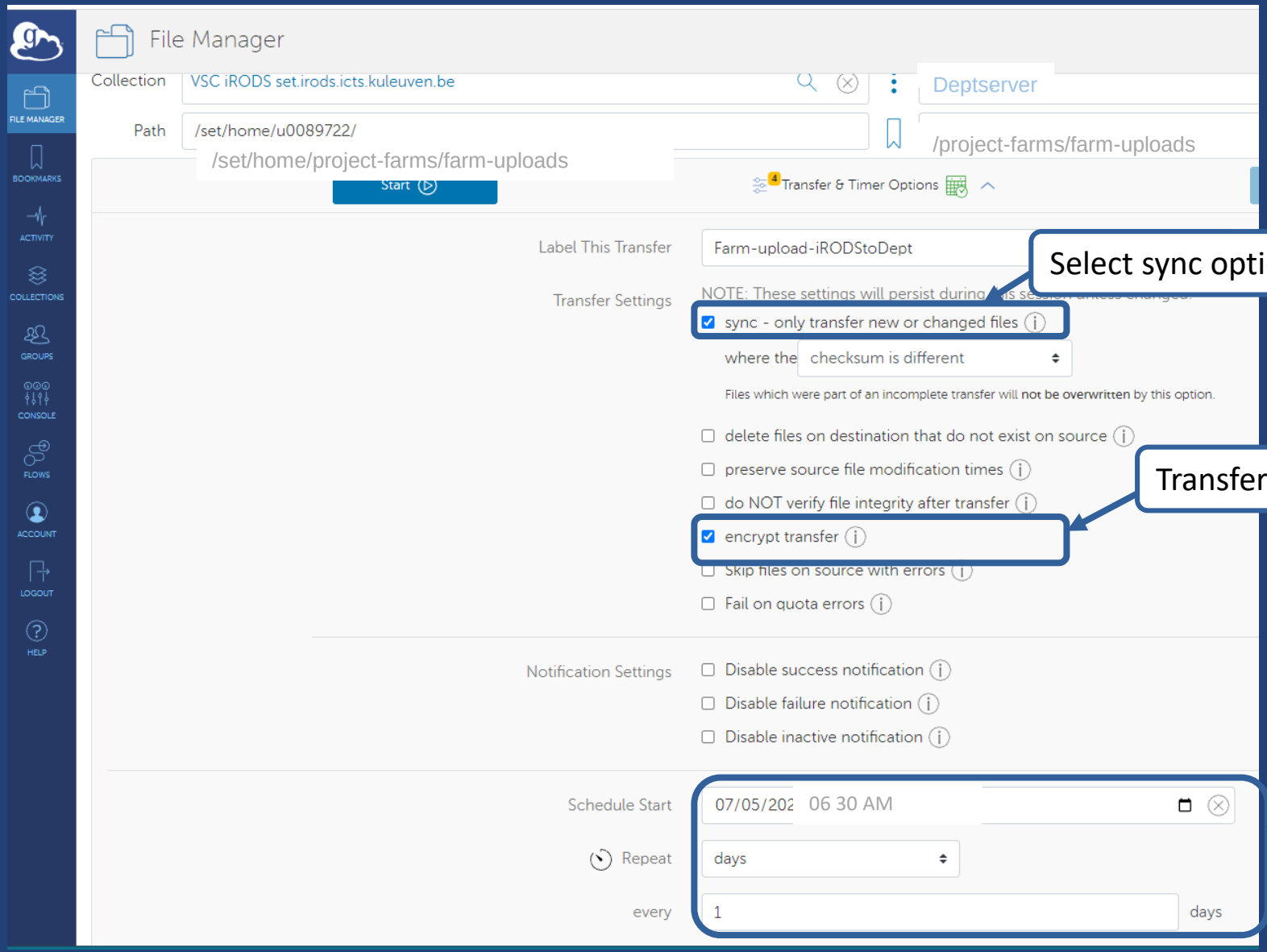
KU LEUVEN

Configuration: farm (2/2)



KU LEUVEN

Configuration: department



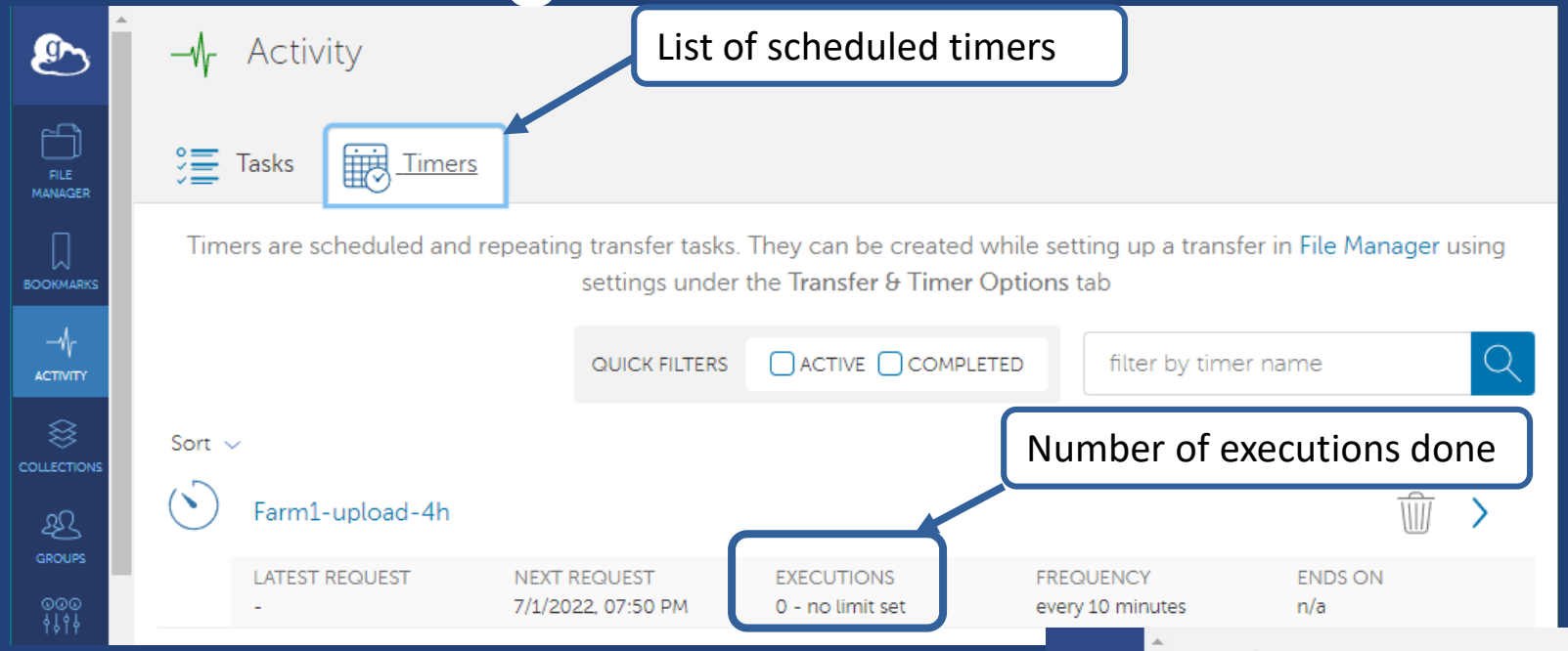
Select sync option

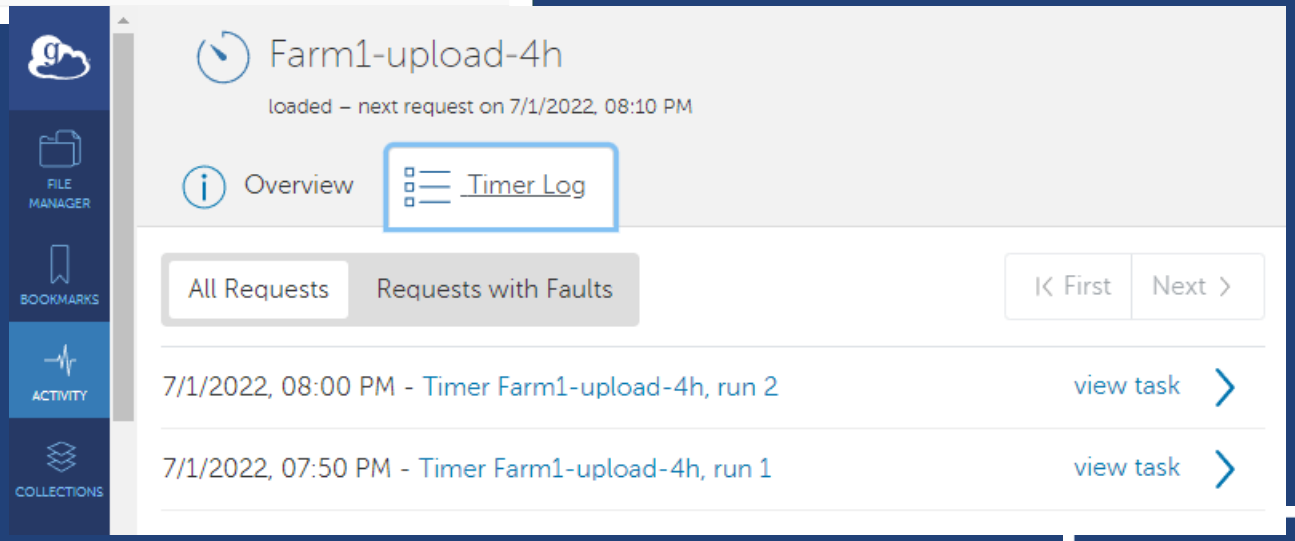
Transfer is forced to be encrypted

A sync process is started 30m after farm dta sync to keep nearly real time processing of the data

KU LEUVEN

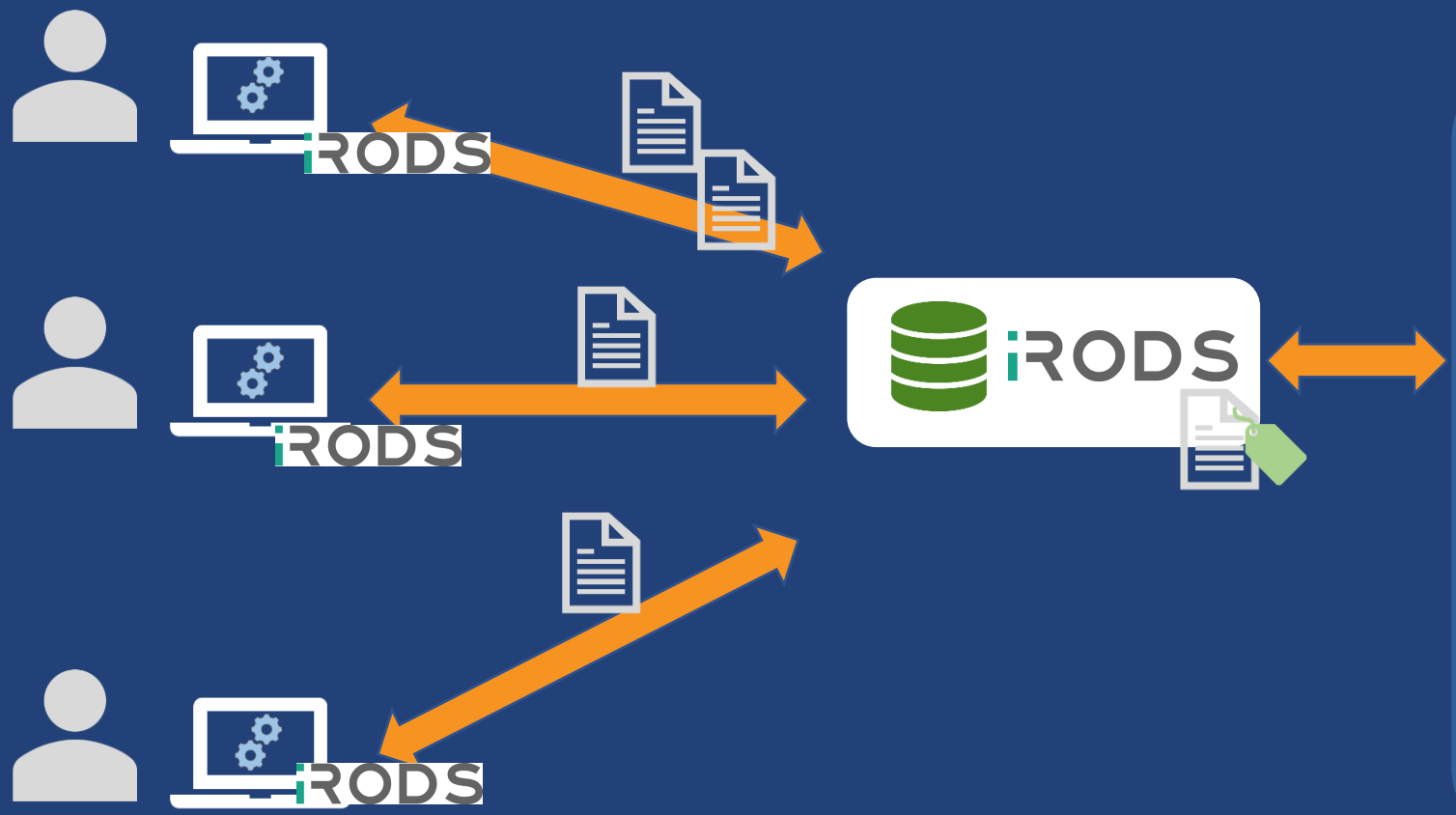
Monitoring





KU LEUVEN

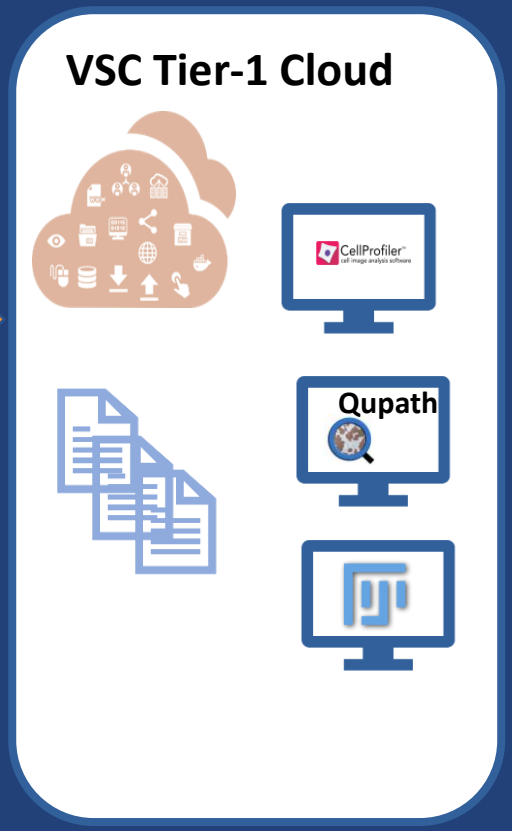
Share (soon PoC)



Authentication
institutional account

ORCID

globus online



KU LEUVEN

Questions?

KU LEUVEN