



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.



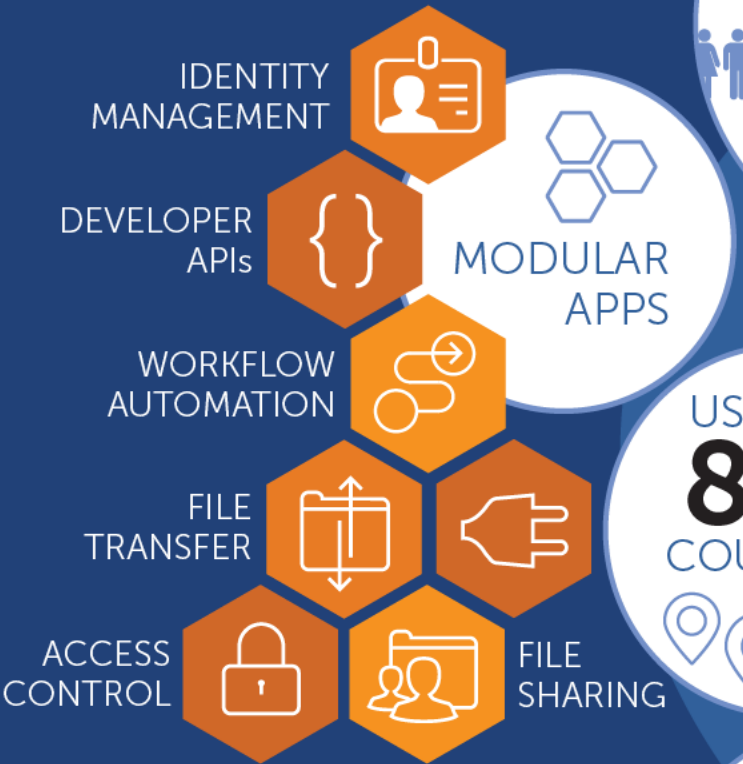
...or “We enable researchers to outsource the mundane—yet critical—data management tasks that impede their work.”

300,000
REGISTERED
USERS

1,400+
IDENTITY
PROVIDERS

LOCAL
STORAGE

RCC
INSTITUTIONAL
STORAGE



MODULAR
APPS

USERS IN
80+
COUNTRIES

SOFTWARE AS A SERVICE
PLATFORM AS A SERVICE



globus

39,000
ACTIVE
ENDPOINTS

TAPE
ARCHIVES

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



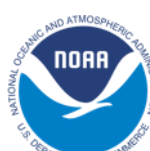
THE UNIVERSITY OF
CHICAGO



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



Transfer

3
Researcher selects files to share, selects user or group, and sets access permissions

Share

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

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

Search

Automate

1
Researcher initiates transfer request; or requested automatically by script, science gateway



5
Collaborator logs in to Globus and accesses shared files; no local account required; download via Globus

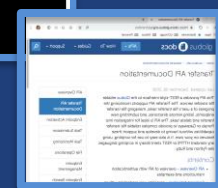
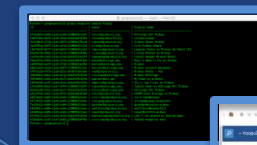
7
... for building science gateways, portals and automations.

8
Streamlining research workflows and ensuring those that need access to the data have it.

Personal Computer



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



Build



Globus is SaaS and PaaS

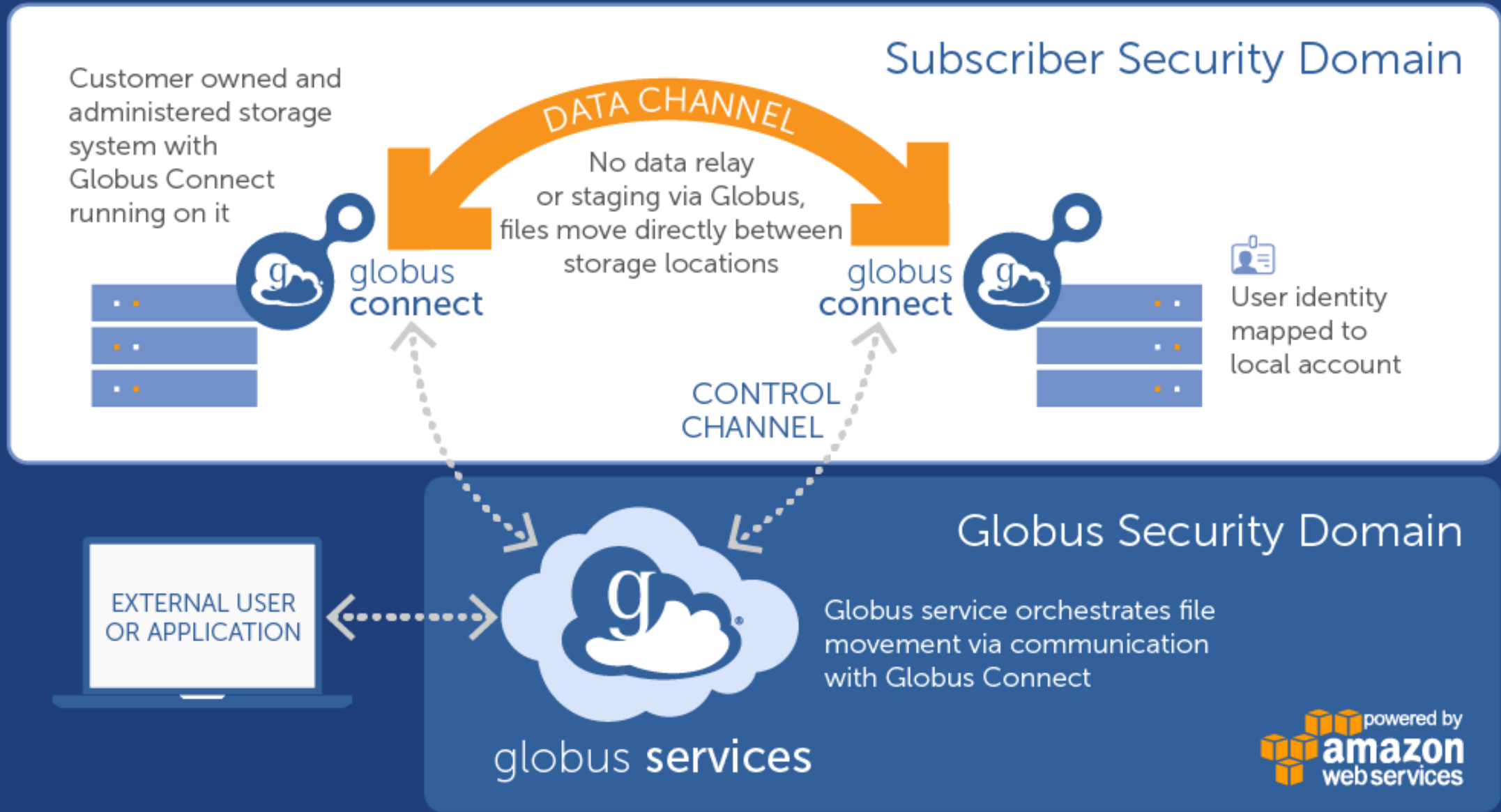
Web app simplifies *ad hoc* data management

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



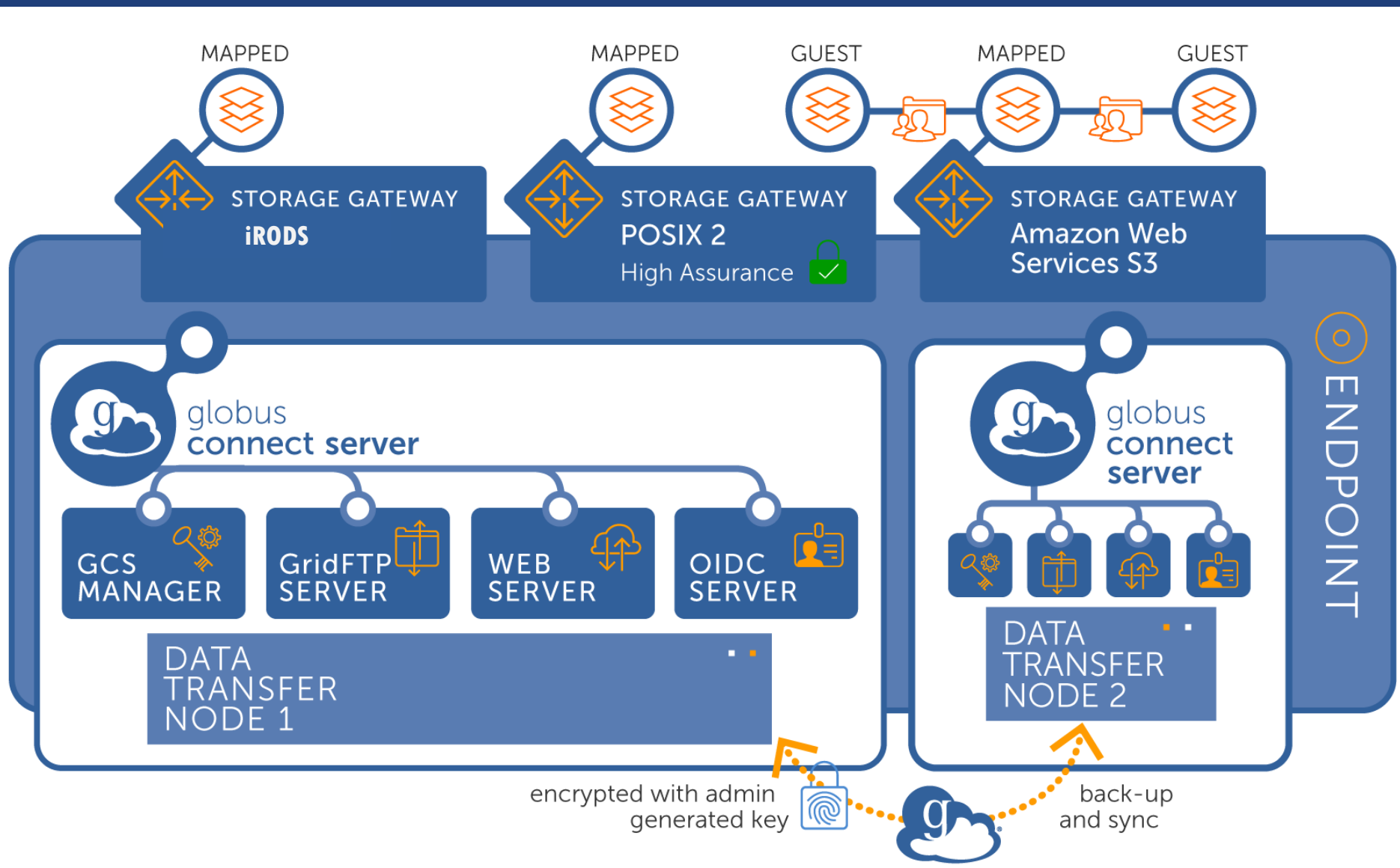


Conceptual Architecture: Hybrid SaaS

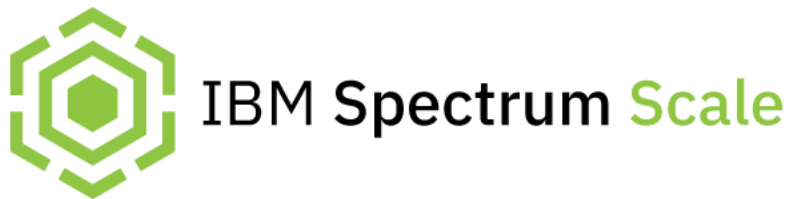




Globus Connect Server Architecture



Globus supports a broad set of storage systems





Globus is fast ...and, more importantly, reliable

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](#)



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



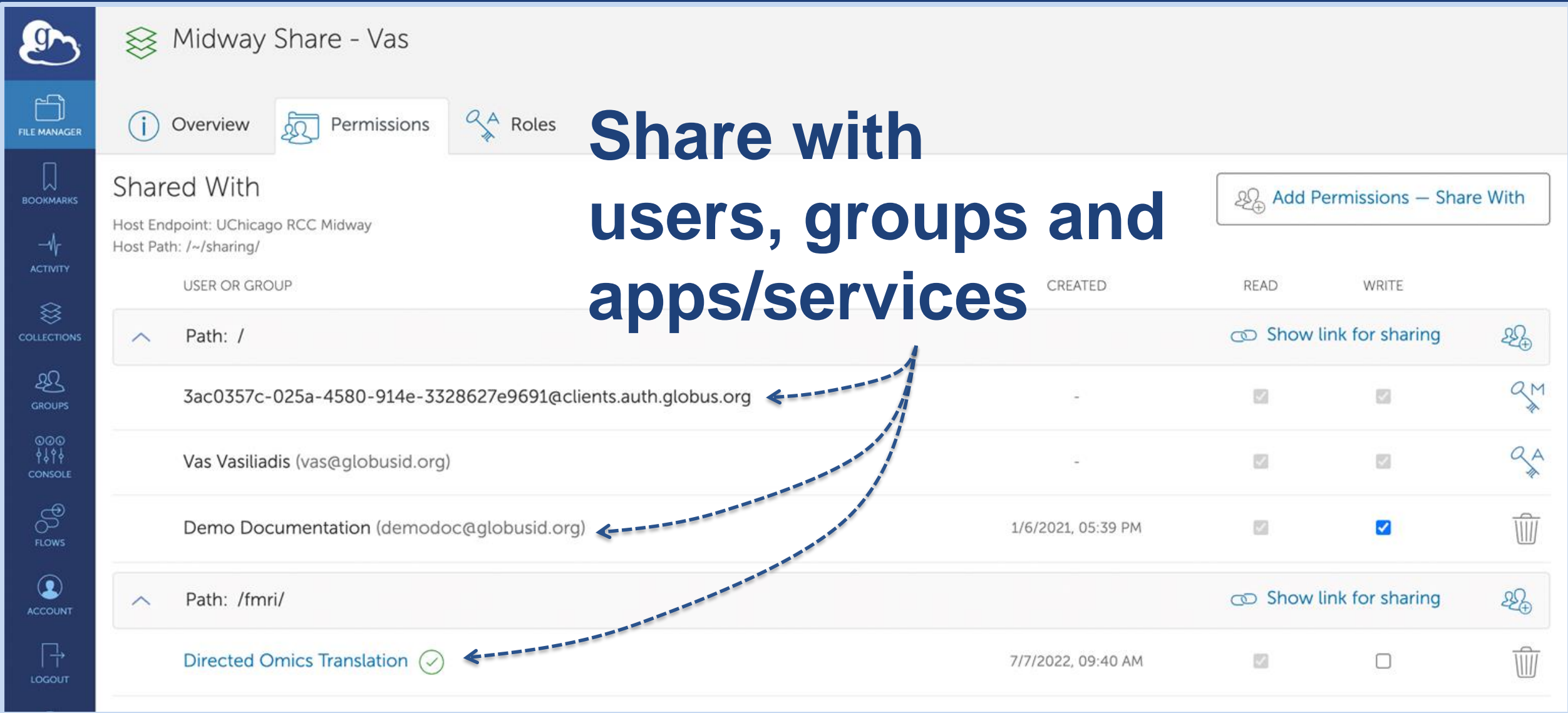
Transfer options and Globus "cron"

The screenshot shows the Globus File Manager interface. The top bar includes the Globus logo and the title 'File Manager'. Below the title, there are fields for 'Collection' (S3 Open Data - Neuroimaging) and 'Path' (/data/Projects/). A 'Start' button is visible. The main area displays a list of folders with the following names: ABIDE, ABIDE_Initiative, ABIDE2, ACPI, ADHD200, and BASC. A vertical sidebar on the left contains navigation icons for FILE MANAGER, BOOKMARKS, ACTIVITY, COLLECTIONS, GROUPS, CONSOLE, FLOWS, ACCOUNT, LOGOUT, and HELP.

The screenshot shows the Globus File Manager interface with transfer settings. The top bar includes the Globus logo and the title 'File Manager'. Below the title, there are fields for 'Collection' (S3 Open Data - Neuroimaging) and 'Path' (/data/Projects/). A second 'Collection' field shows 'Vas AWS EFS High Assurance Collection' with a lock icon. A 'Path' field shows '/vas/'. A 'Start' button is visible. The main area displays transfer settings under the heading 'Transfer & Timer Options'. The settings include:

- Label This Transfer:
- Transfer Settings: NOTE: These settings will persist during this session unless changed.
 - sync - only transfer new or changed files ⓘ
 - delete files on destination that do not exist on source ⓘ
 - preserve source file modification times ⓘ
 - do NOT verify file integrity after transfer ⓘ
 - encrypt transfer (cannot be changed because Vas AWS EFS High Assurance Collection requires encryption) ⓘ
 - Skip files on source with errors ⓘ
 - Fail on quota errors ⓘ
- Notification Settings:
 - Disable success notification ⓘ
 - Disable failure notification ⓘ
 - Disable inactive notification ⓘ

Sharing: Globus maintains a virtual ACL



Midway Share - Vas

Overview Permissions Roles

Share with users, groups and apps/services

Shared With

Host Endpoint: UChicago RCC Midway
Host Path: /~/sharing/

Add Permissions — Share With

USER OR GROUP	CREATED	READ	WRITE	
Path: /				Show link for sharing
3ac0357c-025a-4580-914e-3328627e9691@clients.auth.globus.org	-	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
Vas Vasiliadis (vas@globusid.org)	-	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
Demo Documentation (demodoc@globusid.org)	1/6/2021, 05:39 PM	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
Path: /fmri/				Show link for sharing
Directed Omics Translation	7/7/2022, 09:40 AM	<input checked="" type="checkbox"/>	<input type="checkbox"/>	



One service, many interfaces

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

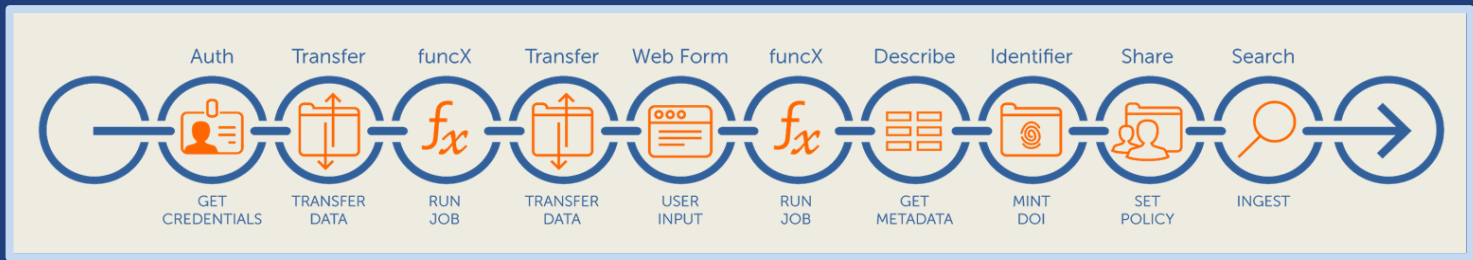
Options:
  -v, --verbose          Control level of output
  -h, --help             Show this message and exit.
  -F, --format [json|text] Output format for stdout. Defaults to text
  --map-http-status TEXT Map HTTP statuses to any of these exit codes:
                        0,1,50-99. e.g. "404=50,403=51"

Commands:
  bookmark      Manage Endpoint Bookmarks
  config        Modify, view, and manage your Globus CLI config.
```

CLI

The screenshots show the Globus web interface. The top screenshot displays a calendar for scheduling a transfer, with a note: "NOTE: These settings will persist during this session unless changed." The bottom screenshot shows a file manager view with a list of files and folders, including endpoints like 'globusPersonalEndpoint' and 'globusPublicEndpoint'.

Web



Flows



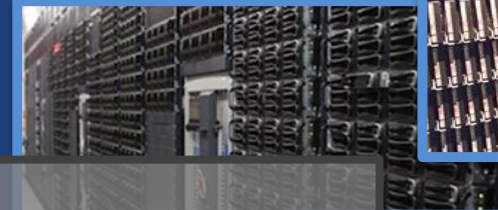
Globus service

```
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!”



If you're doing something the same way you have been doing it for ten years, the chances are you are doing it wrong.

— *Charles Kettering* —

AZ QUOTES



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

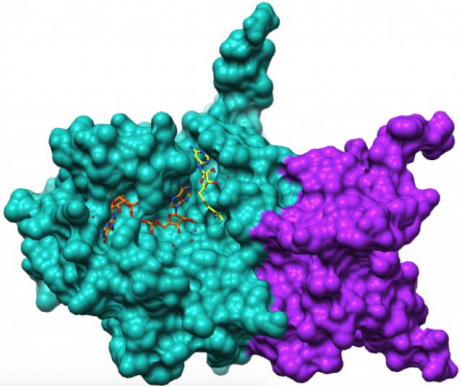
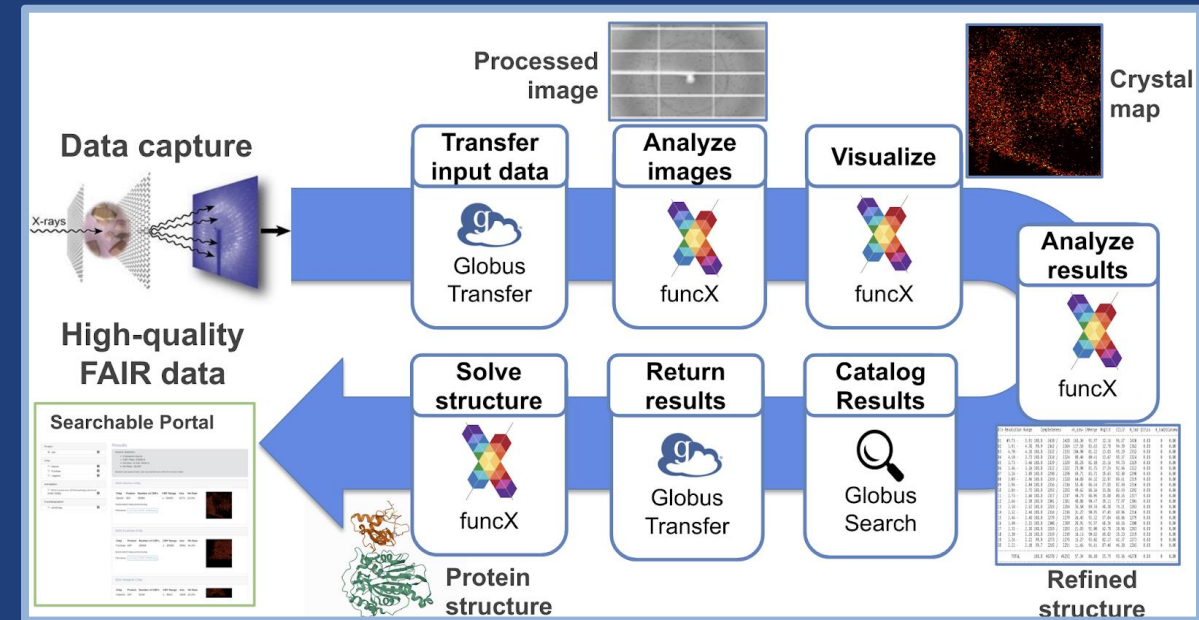
"These data services have taken the time to solve a structure from weeks to days and now to hours"

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

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

June 12, 2020

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

Displaying 1 to 4 of 4 Structures Page 1 of 1 Display 25 per page

7JHE [Download File](#) [View File](#)

Room Temperature Structure of SARS-CoV-2 Nsp10/Nsp16 Methyltransferase in a Complex with 2'-O-methylated m7GpppA Cap-1 and SAH Determined by Fixed-Target Serial Crystallography

Wilamowski, M., Sherrell, D.A., Minasov, G., Kim, Y., Shuvalova, L., Lavens, A., Chard, R., Rosas-Lemus, M., Maltseva, N., Jedrzejczak, R., Michalska, K., Satchell, K.J.F., Joachimiak, A., Center for Structural Genomics of Infectious Diseases (CSGID)

To be published

Released 2020-08-26
Method X-RAY DIFFRACTION 2.25 Å
Organisms Severe acute respiratory syndrome coronavirus 2
Macromolecule 2'-O-methyltransferase (protein)
Non-structural protein 10 (protein)
Unique Ligands CL, MGP, SAH, V9G, ZN

6XKM [Download File](#) [View File](#)

Room Temperature Structure of SARS-CoV-2 NSP10/NSP16 Methyltransferase in a Complex with SAM Determined by Fixed-Target Serial Crystallography

Wilamowski, M., Sherrell, D.A., Minasov, G., Kim, Y., Shuvalova, L., Lavens, A., Chard, R., Rosas-Lemus, M., Maltseva, N., Jedrzejczak, R., Michalska, K., Satchell, K.J.F., Joachimiak, A., Center for Structural Genomics of Infectious Diseases (CSGID)

To be published

Released 2020-07-08
Method X-RAY DIFFRACTION 2.25 Å
Organisms Severe acute respiratory syndrome coronavirus 2
Macromolecule 2'-O-methyltransferase (protein)
Non-structural protein 10 (protein)
Unique Ligands CL, SAM, ZN



Portals for data publication and faceted search

ALCF Community

Providing Researchers Data Sharing Capabilities

APS 2ID XRF

X-Ray fluorescence imaging for a variety of applications

Private

[Search This Project](#)

High-Resolution Microscopy

Microscopy techniques for the structural analysis of individual molecules and bulk materials

[Search](#)

ALCF Community Data Co-Op ALCF Documentation Login

ALCF Community Data Co-Op / DLHub

Domain

- materials science 91
- general 49
- biology 37
- microscopy 28
- test 19
- material science 3
- MODIS 2
- climate science 2
- drug design 2
- fields 2

Test

- False 494

Publisher

- DLHub 494

Publication Year

- 2020 166

NOTE: Your search results are limited. Logging in can improve search results by displaying private results if you have permissions associated with your account.

Results

494 servables found

PyTorch BraggNN for CPU

name: aristana_uchicago/BraggNNPT_CPU

Authors

Nikil Ravi; Pranshu Chaturvedi; E.A. Huerta; Zhengchun Liu; Aristana Scourtas; KJ Schmidt; Ryan Chard; Ben Blaiszik

Visibility	Version	Shim
public	0.10.0	python.PythonStaticMethodServable

PyTorch BraggNN for CPU

name: aristana_uchicago/BraggNNPT_CPU

Authors

Nikil Ravi; Pranshu Chaturvedi; E.A. Huerta; Zhengchun Liu; Aristana Scourtas; KJ Schmidt; Ryan Chard; Ben Blaiszik

Visibility	Version	Shim



globus.org

app.globus.org

docs.globus.org

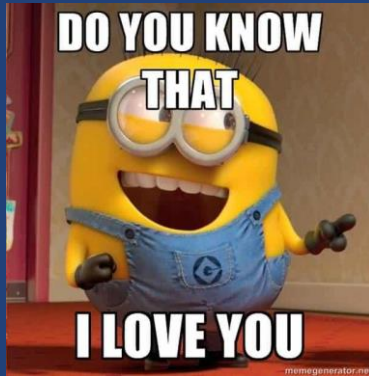
globus.org/connectors

globus.org/subscriptions

support@globus.org

VSC iRODS and Globus use cases

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?

iRODS and Globus!



iRODS

Store
Describe
Discover
Automate



Globus

Move
Share



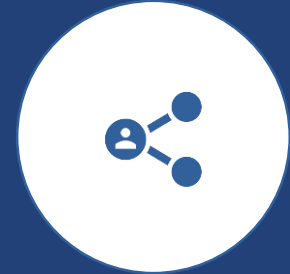
VSC use cases



Transfer



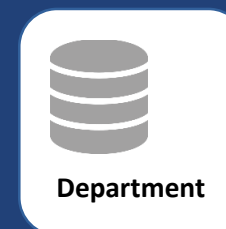
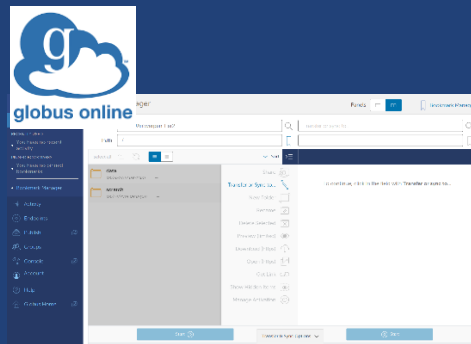
Ingest



Share

Transfer

Authentication
institutional account



KU LEUVEN

GHENT
UNIVERSITY

Universiteit
Antwerpen

VUB

VSC

OneDrive iRODS HPC-Tier-2

OneDrive HPC-Tier-2

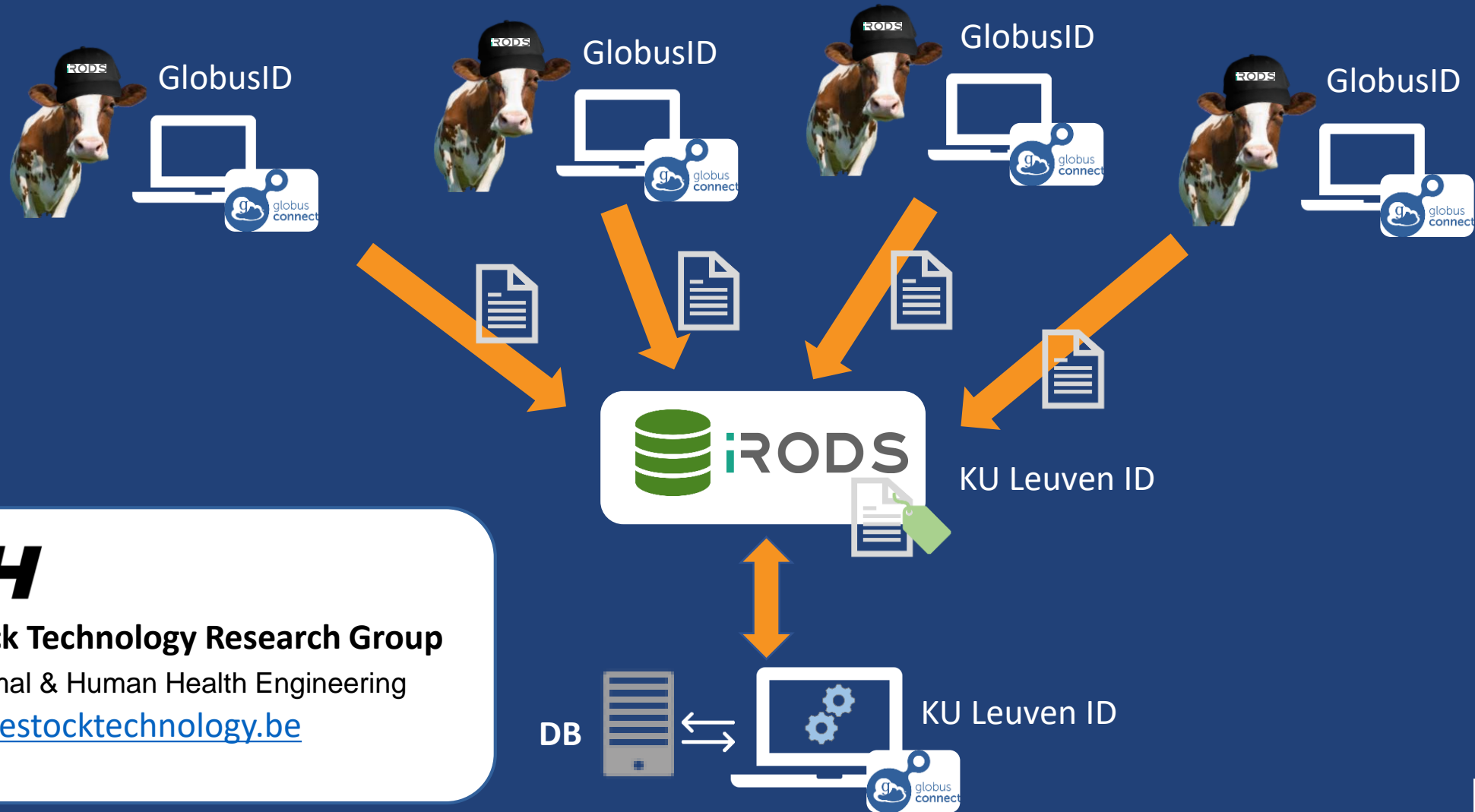
OneDrive HPC-Tier-2

OneDrive HPC-Tier-2

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

Configuration: iRODS

metalnx u00 [redacted]

set » home » u00 [redacted] » Farm-uploads

Navigation Tools

Action

<input type="checkbox"/>	Name	Size
<input type="checkbox"/>	Farm1-upload	41
<input type="checkbox"/>	Farm2-upload	41

An upload directory per Farm is created in iRODS

Farm1-upload

Overview Permissions Roles

Shared With

Add Permissions – Share With

USER OR GROUP	READ	WRITE	
Path: /			
Ingrid Barcena Roig (u00 [redacted]@kuleuven.be)	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
[redacted]	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
Ingrid Barcena Roig (ibrtest@globusid.org)	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	

The iRODS directory is shared with the GlobusID user with write permissions

Configuration: farm (1/2)

File Manager

Collection **Farm1**

Path /~/globus-tests/

Farm1 7/1/2022, 06:13 PM

- Share
- Transfer or Sync to...
- New Folder
- Rename
- Delete Selected
- Download
- Open
- Upload
- Get Link
- Show Hidden Items

Local directory - farm data updates

iRODS dir shared with GlobusID Farm1

Configuration: farm (2/2)

The screenshot displays the File Manager configuration interface for a farm. The interface is divided into several sections:

- Collection:** Farm1
- Path:** /~/globus-tests/
- Transfer & Timer Options:** Includes a "Start" button and a "Transfer & Timer Options" button.
- Transfer Settings:**
 - sync - only transfer new or changed files (where the checksum is different)
 - delete files on destination that do not exist on source
 - preserve source file modification times
 - do NOT verify file integrity after transfer
 - encrypt transfer (cannot be changed because of farm1-u)
 - Skip files on source with errors
 - Fail on quota errors
- Notification Settings:**
 - Disable success notification
 - Disable failure notification
 - Disable inactive notification
- Schedule:**
 - Schedule Start: 07/02/2022 06 00 AM
 - Repeat: hours
 - every: 6
 - end: never

Select sync option

Transfer is forced to be encrypted (defined by the sharing config)

A sync process is started every 6 hours

Configuration: department

The screenshot shows the File Manager configuration page for a sync process. The interface includes a sidebar with navigation options like FILE MANAGER, BOOKMARKS, ACTIVITY, COLLECTIONS, GROUPS, CONSOLE, FLOWS, ACCOUNT, LOGOUT, and HELP. The main area displays the following configuration details:

- Collection:** VSC iRODS set.irods.icts.kuleuven.be
- Path:** /set/home/u0089722/
- Transfer & Timer Options:** Transfer & Timer Options
- Label This Transfer:** Farm-upload-iRODStoDept
- Transfer Settings:**
 - sync - only transfer new or changed files (i)
 - where the checksum is different
 - Files which were part of an incomplete transfer will not be overwritten by this option.
 - delete files on destination that do not exist on source (i)
 - preserve source file modification times (i)
 - do NOT verify file integrity after transfer (i)
 - encrypt transfer (i)
 - Skip files on source with errors (i)
 - Fail on quota errors (i)
- Notification Settings:**
 - Disable success notification (i)
 - Disable failure notification (i)
 - Disable inactive notification (i)
- Schedule Start:** 07/05/202 06 30 AM
- Repeat:** days
- every:** 1 days

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

Monitoring

Activity

Tasks Timers

Timers are scheduled and repeating transfer tasks. They can be created while setting up a transfer in [File Manager](#) using settings under the [Transfer & Timer Options](#) tab

QUICK FILTERS ACTIVE COMPLETED filter by timer name

Sort

LATEST REQUEST	NEXT REQUEST	EXECUTIONS	FREQUENCY	ENDS ON
-	7/1/2022, 07:50 PM	0 - no limit set	every 10 minutes	n/a

Farm1-upload-4h
loaded - next request on 7/1/2022, 08:10 PM

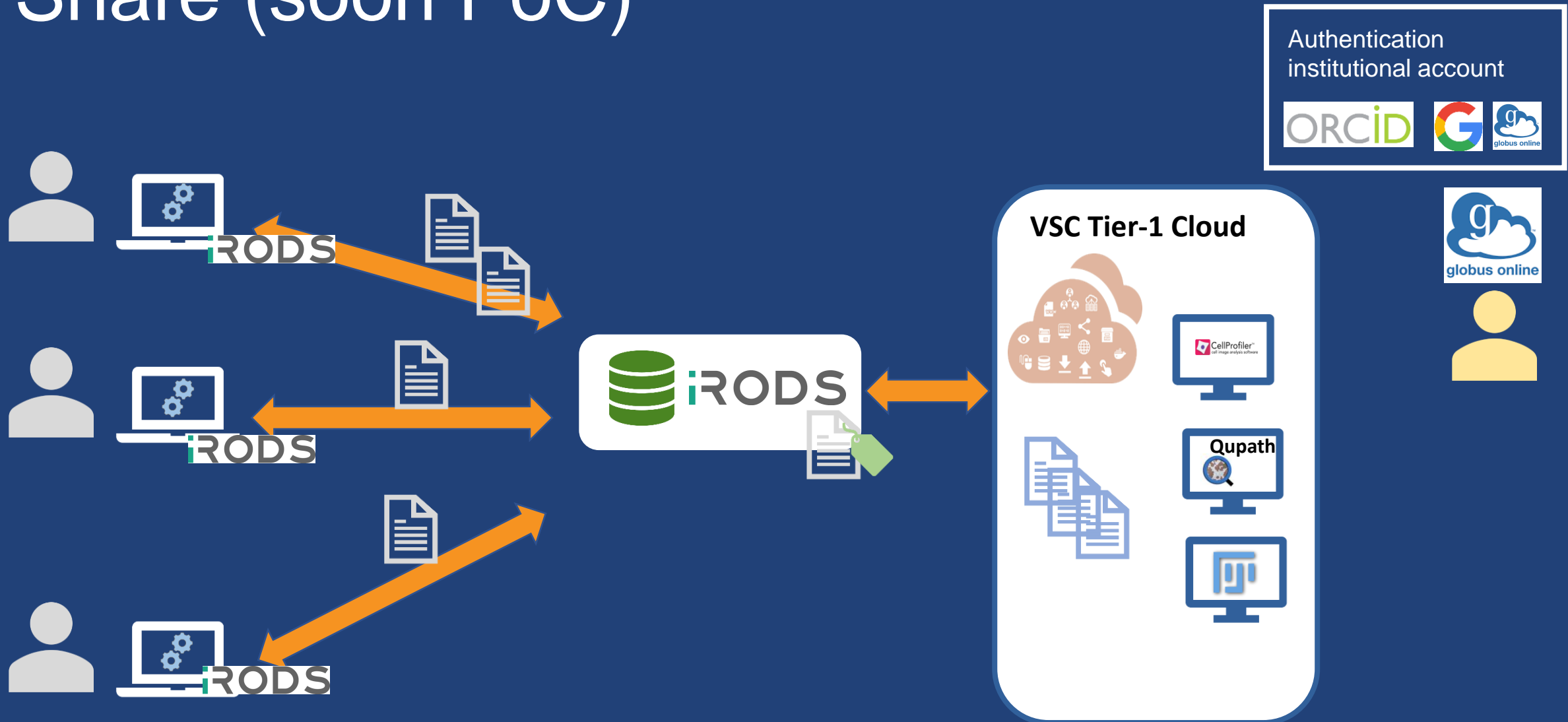
Overview Timer Log

All Requests Requests with Faults

< First Next >

7/1/2022, 08:00 PM - Timer Farm1-upload-4h, run 2	view task
7/1/2022, 07:50 PM - Timer Farm1-upload-4h, run 1	view task

Share (soon PoC)





Questions?