



globus

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



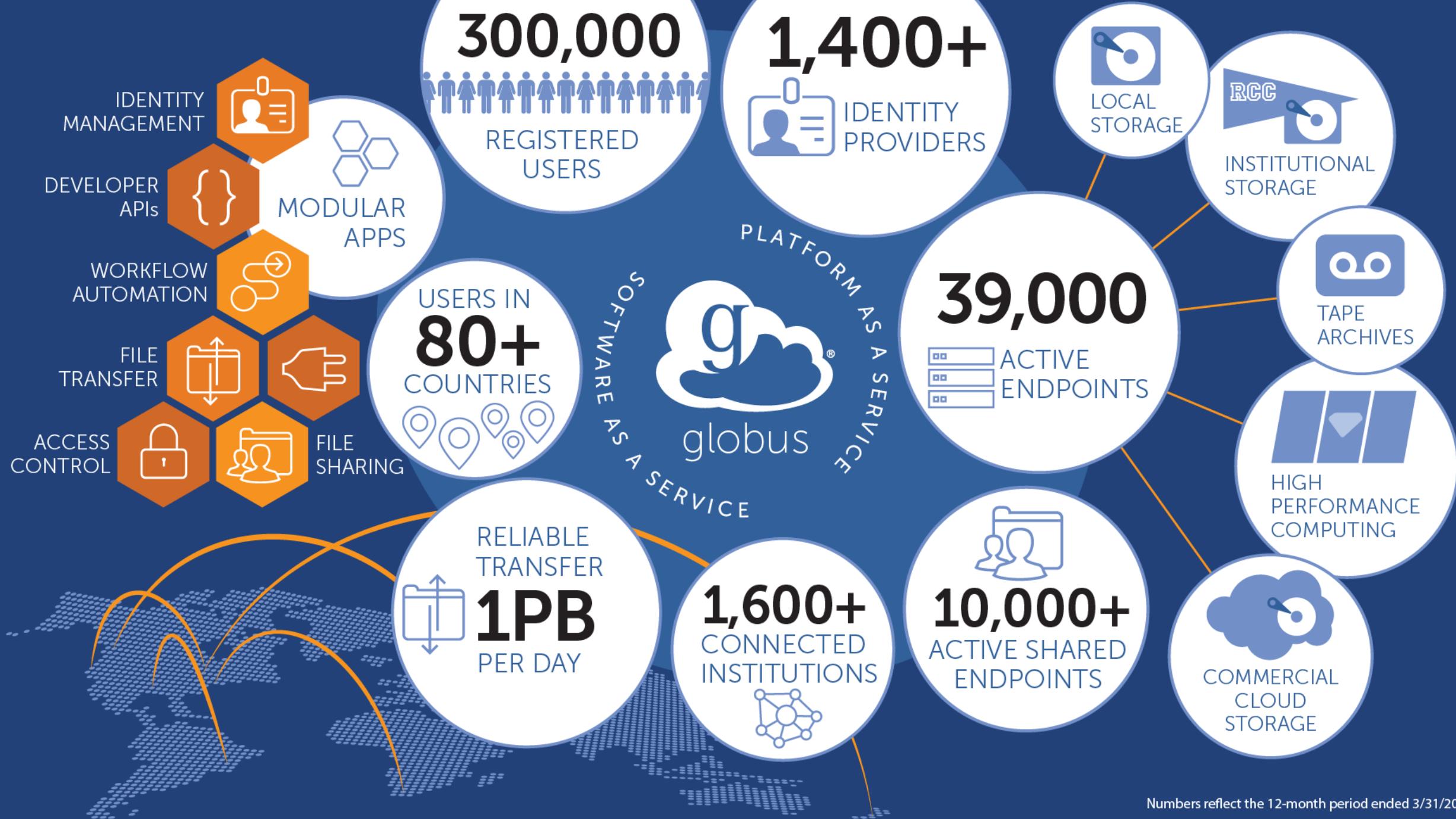
THE UNIVERSITY OF
CHICAGO



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.”





Development is partly funded by...



U. S. DEPARTMENT OF
ENERGY



THE UNIVERSITY OF
CHICAGO



powered by
amazon
web services



NIST
National Institute of
Standards and Technology
U.S. Department of Commerce

Argonne
NATIONAL LABORATORY



...Operations are funded by subscribers

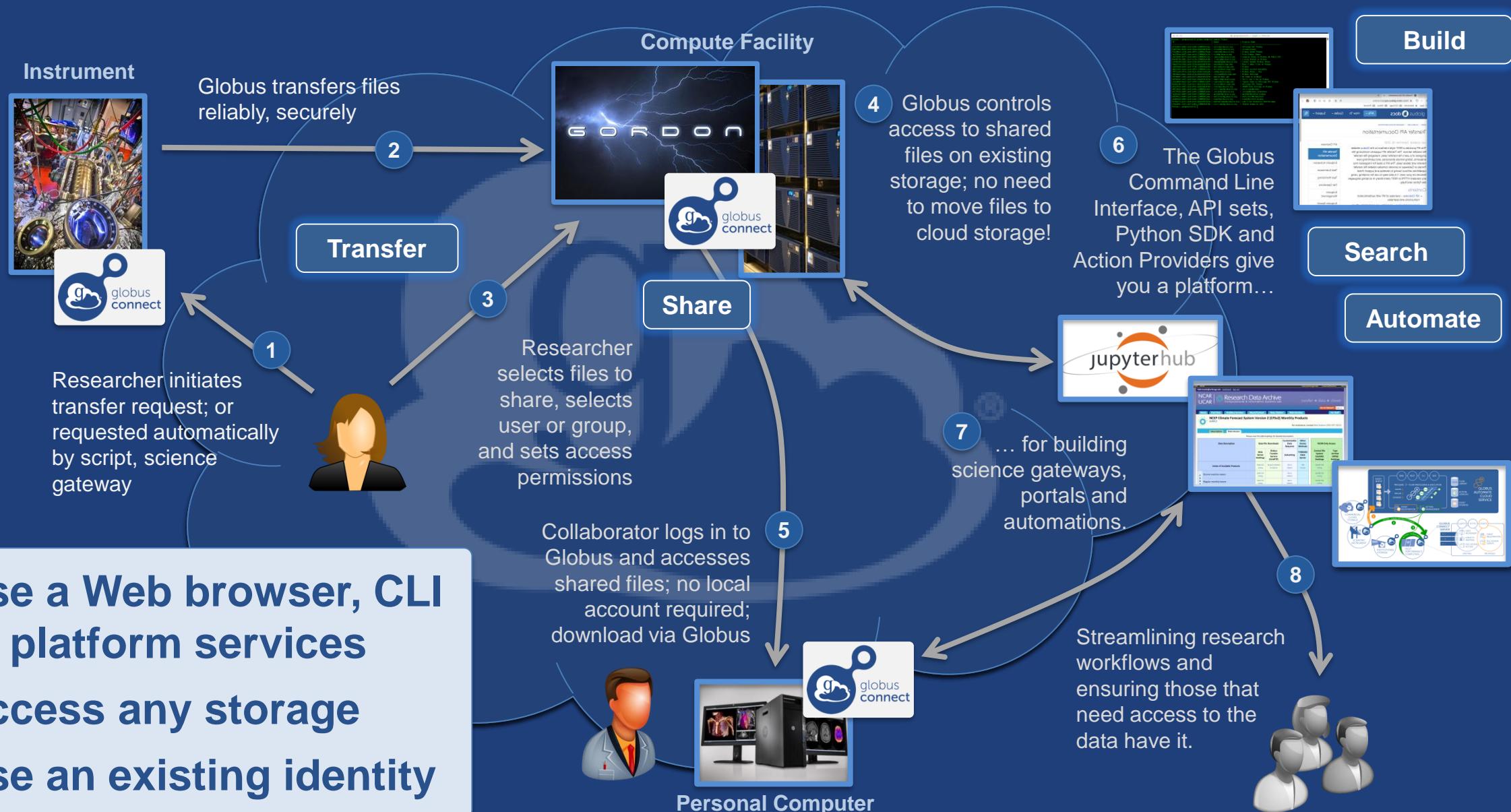


THE ROCKEFELLER UNIVERSITY





Globus and the research data lifecycle



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



Globus is SaaS and PaaS

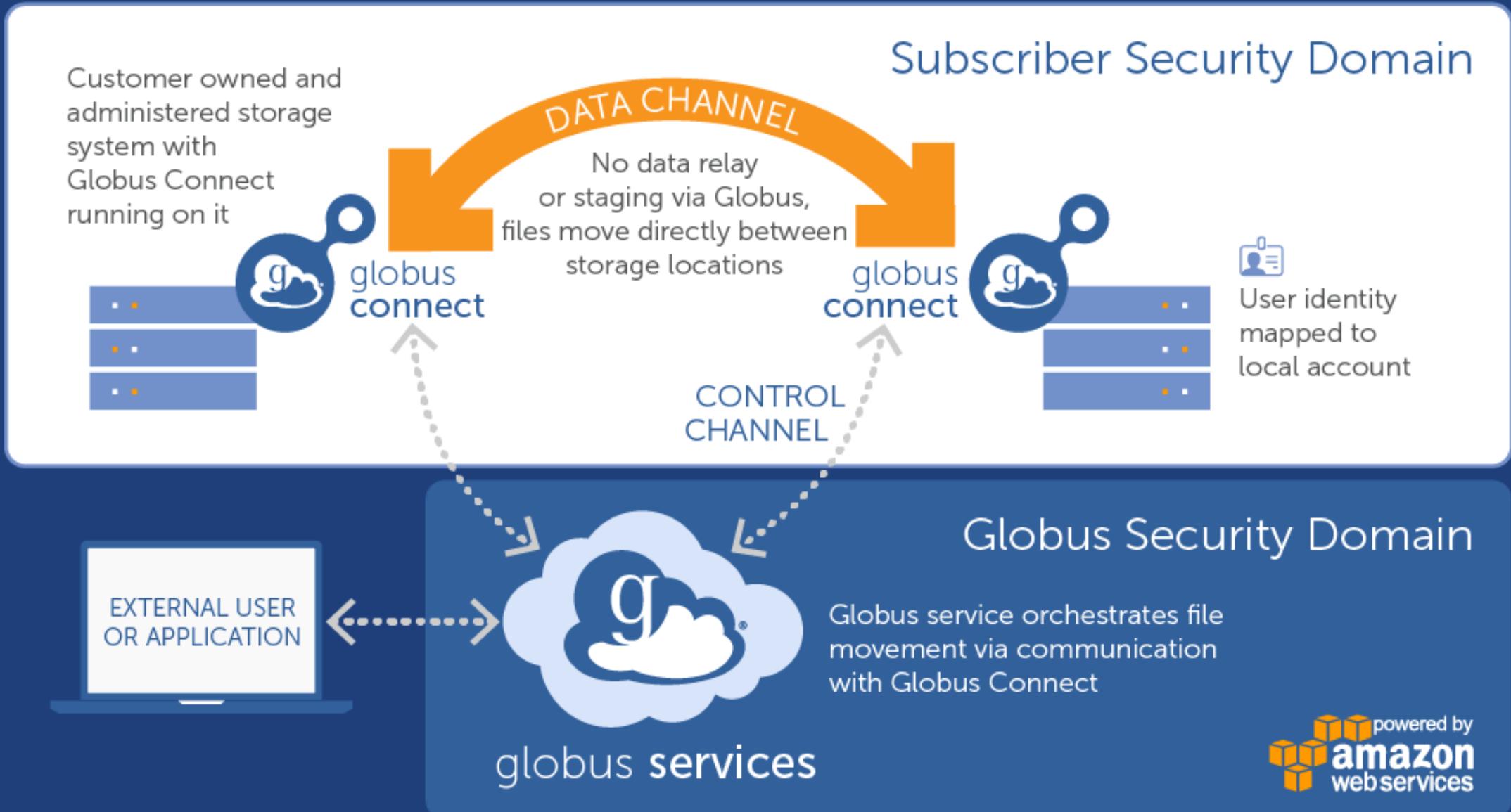
Web app simplifies *ad hoc* data management

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



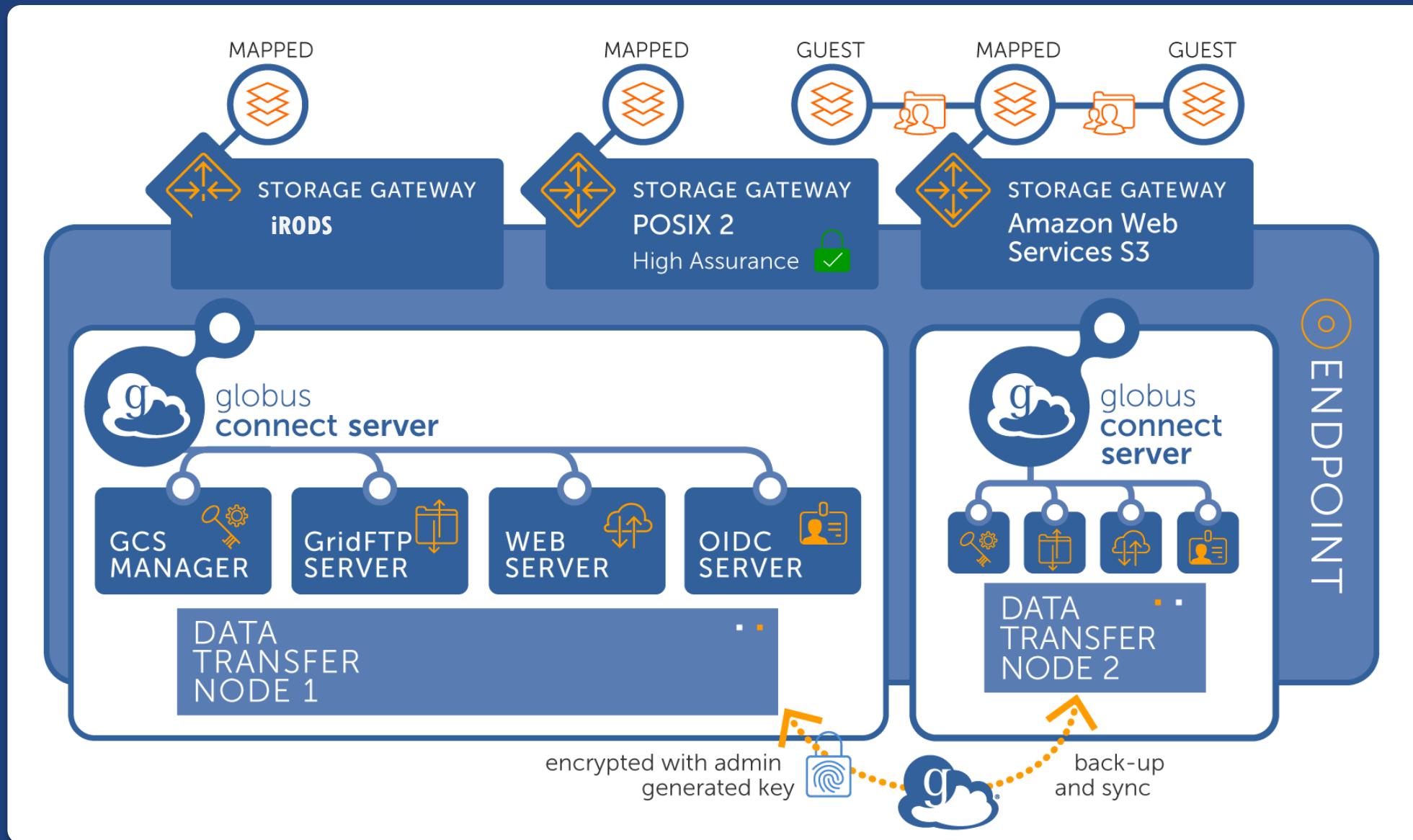


Conceptual Architecture: Hybrid SaaS





Globus Connect Server Architecture





Globus supports a broad set of storage systems





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

Activity List ✓ RDA to ALCF noverify
transfer completed

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	<ul style="list-style-type: none">transfer is not encryptedoverwriting all files on destination

72.8Gbps

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)

The screenshot shows the Globus Federated login page. At the top, there's a blue header bar with the Globus logo and the text "globus". Below it, the main content area has a white background. It starts with the heading "Log in to use Globus Web App". Underneath, there's a sub-instruction "Use your existing organizational login" followed by a note "e.g., university, national lab, facility, project". A dropdown menu is open, showing "KU Leuven / KU Leuven Association" as the selected option. Below the dropdown, a note states: "By selecting Continue, you agree to Globus [terms of service](#) and [privacy policy](#)". A large blue "Continue" button is centered below this note. To the left of the "Continue" button is a green CILogon logo icon. To the right of the "Continue" button is a callout box containing text about CILogon's role in enabling logins from organizations. At the bottom of the page, there are two additional sign-in options: "Sign in with Google" (with a red "G" icon) and "Sign in with ORCID iD" (with a green "id" icon). A small note at the very bottom says: "Didn't find your organization? Then use [Globus ID](#) to sign in. (What's this?)".

1,400+ IdPs
...via released
R&S attributes
or custom
configuration



Transfer options and Globus “cron”

The screenshot displays two instances of the Globus File Manager interface. The left instance shows a file list for the collection "S3 Open Data - Neuroimaging" at the path "/data/Projects/". The right instance shows the same collection and path, but with a "Transfer & Timer Options" panel open.

Left Instance (File Manager View):

- Collection: S3 Open Data - Neuroimaging
- Path: /data/Projects/
- Start button
- File list: ABIDE, ABIDE_Initiative, ABIDE2, ACPI, ADHD200, BASC

Right Instance (Transfer & Timer Options Panel):

- Collection: S3 Open Data - Neuroimaging
- Path: /data/Projects/
- Transfer & Timer Options panel:
 - Label This Transfer: (empty input field)
 - Transfer Settings:
 - NOTE: These settings will persist during this session unless changed.
 - sync - only transfer new or changed files i
 - 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 (cannot be changed because Vas AWS EFS High Assurance Collection requires encryption) 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
- Start button

Sharing: Globus maintains a virtual ACL

Midway Share - Vas

FILE MANAGER

BOOKMARKS

ACTIVITY

COLLECTIONS

GROUPS

CONSOLE

FLOWS

ACCOUNT

LOGOUT

Overview Permissions Roles

Shared With

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

USER OR GROUP

Path: /

3ac0357c-025a-4580-914e-3328627e9691@clients.auth.globus.org

Vas Vasiliadis (vas@globusid.org)

Demo Documentation (demodoc@globusid.org)

Path: /fmri/

Directed Omics Translation (checkmark)

Add Permissions – Share With

Show link for sharing

CREATED	READ	WRITE
-	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
-	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
1/6/2021, 05:39 PM	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
7/7/2022, 09:40 AM	<input checked="" type="checkbox"/>	<input type="checkbox"/>

Share with users, groups and apps/services

The diagram illustrates the sharing interface for a folder named 'Midway Share - Vas'. It shows three entries under 'Shared With': 1) A Globus user ID (3ac0357c-025a-4580-914e-3328627e9691@clients.auth.globus.org). 2) A Globus group (Vas Vasiliadis (vas@globusid.org)). 3) A service (Demo Documentation (demodoc@globusid.org)). Each entry has a 'Show link for sharing' button and a 'Permissions' section with 'READ' and 'WRITE' checkboxes. Dashed blue arrows point from each entry to its corresponding 'WRITE' checkbox, demonstrating how Globus maintains a virtual Access Control List (ACL) for each share.



One service, many interfaces

```
(globus-cli) jupyter:~ 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" for the attack

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

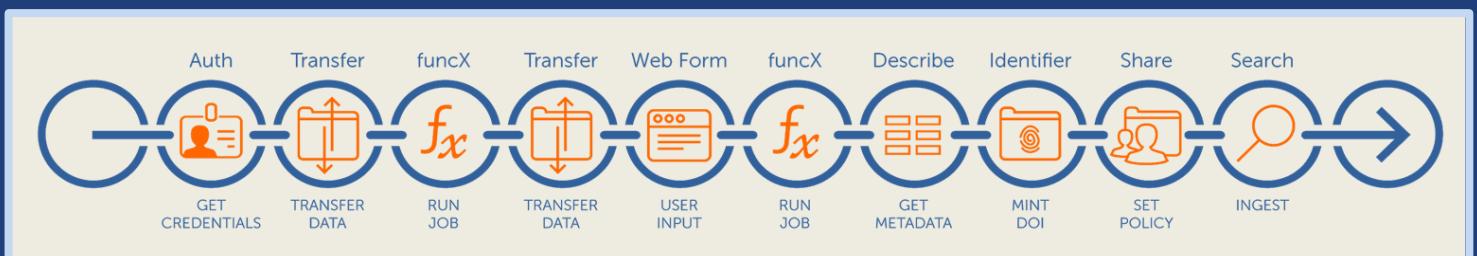
CLI



Globus service

The screenshot shows the Globus Web interface. On the left, there's a sidebar with 'File Manager' and 'Transfer Settings'. In the center, there's a calendar for May 2022 with a scheduled transfer set for 09:49 AM. Below the calendar is a list of files and folders, with some having 'Transfer Options' dropdown menus. A 'Transfer & Sync to...' button is visible at the bottom right of the main content area.

Web



Flows

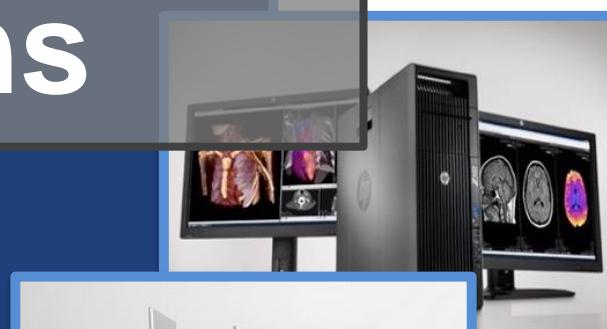
```
GET /endpoint/go%23ep1
PUT /endpoint/vas#my_endpt
200 OK
X-Transfer-API-Version: 0.10
Content-Type: application/json
...
```

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!”



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

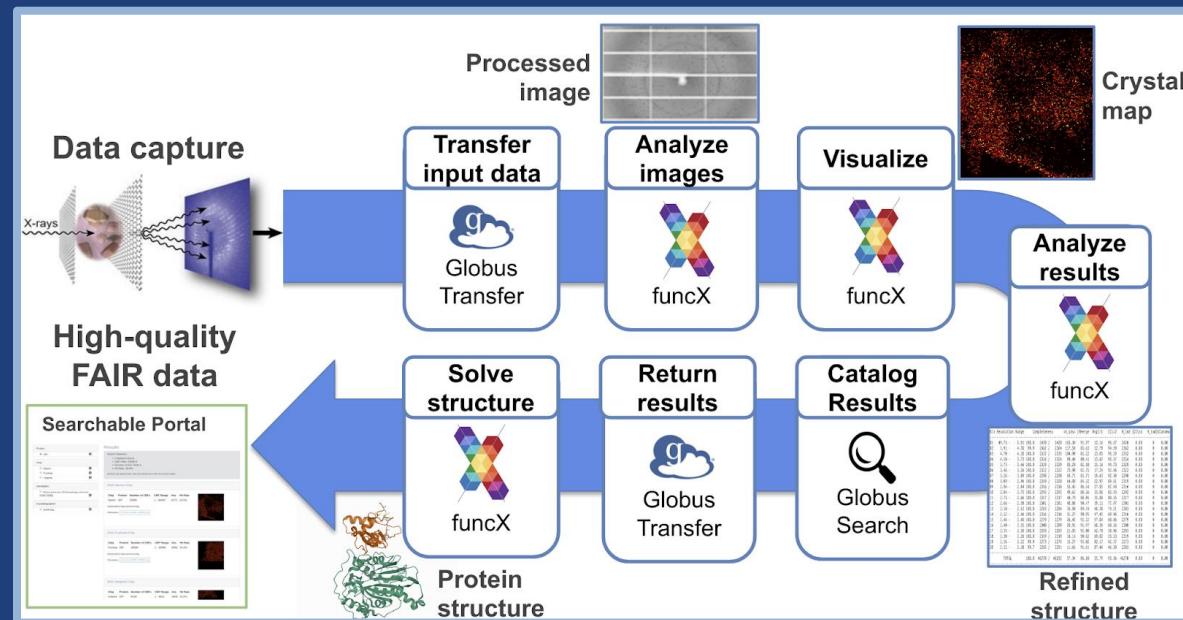
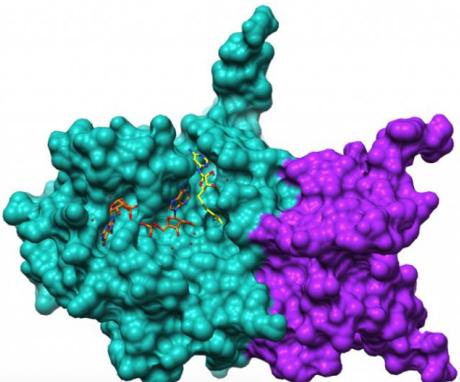
"These data services have taken the time to solve a structure from weeks to days and now to hours"

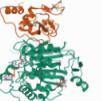
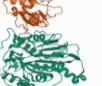
Darren Sherrell, SBC beamline scientist APS Sector 19

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



Displaying 1 to 4 of 4 Structures		Page 1 of 1	Display 25 <input type="button" value="per page"/>
 3D View	 7JHE	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	<input type="button" value="Download File"/> <input type="button" value="View File"/>
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, VBG, ZN	
To be published			
 3D View	 6XKM	Room Temperature Structure of SARS-CoV-2 NSP10/NSP16 Methyltransferase in a Complex with SAM Determined by Fixed-Target Serial Crystallography	<input type="button" value="Download File"/> <input type="button" value="View File"/>
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	
To be published			



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



Portals for data publication and faceted search

ALCF Community Data Co-Op

Providing Researchers Data Sharing Capabilities

APS 2ID XRF
X-Ray fluorescence imaging for a variety of applications

APL High-Energy Microscopy
Microscopic techniques for the structure of individual bulk particles

Private

Search This Project

ALCF Community Data Co-Op

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

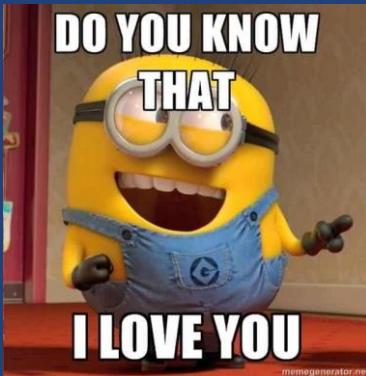
globus.org/connectors

globus.org/subscriptions

support@globus.org

VSC iRODS and Globus use cases

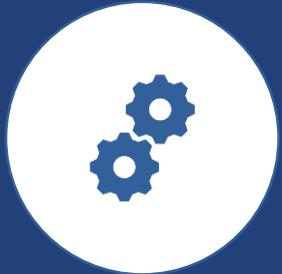
iRODS and Globus?



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

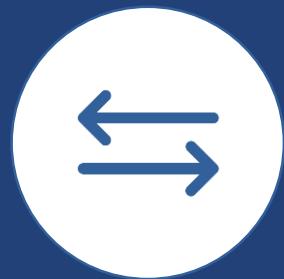


Globus

Move
Share



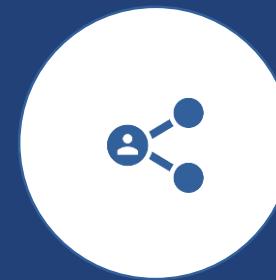
VSC use cases



Transfer



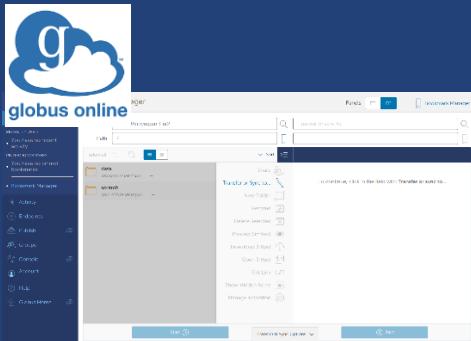
Ingest



Share

Transfer

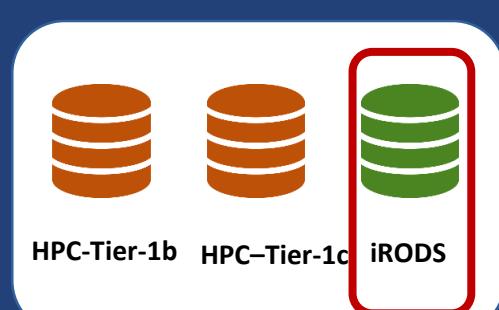
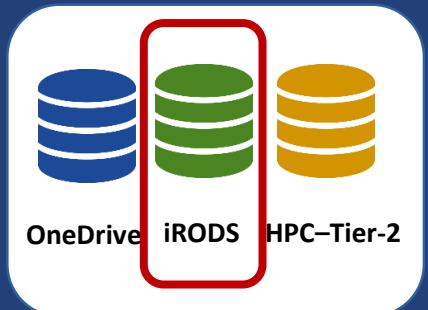
Authentication
institutional account



ORCID G

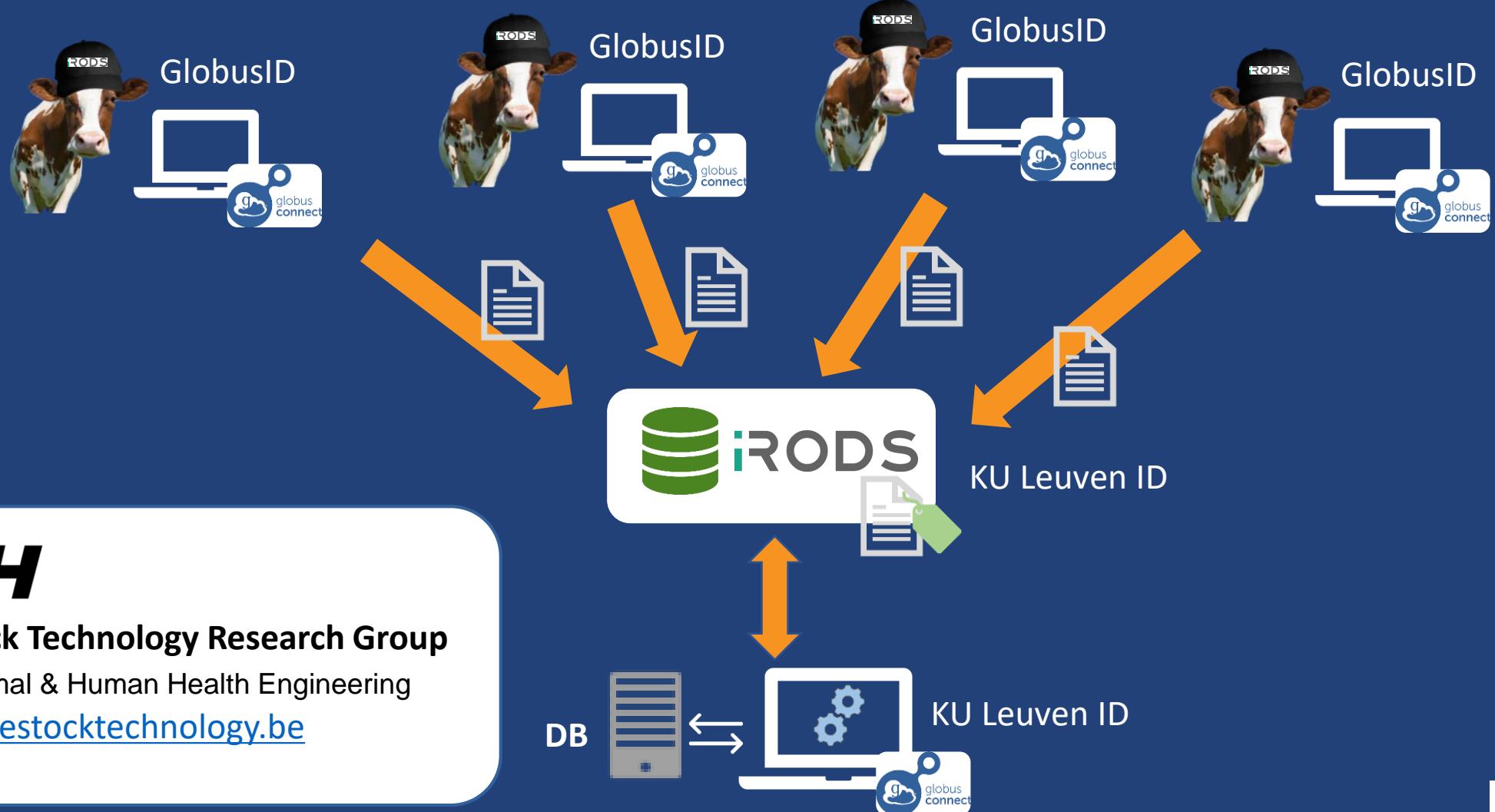


VSC



KU LEUVEN

Ingest (PoC): milk quality control at dairy farms



Configuration: iRODS

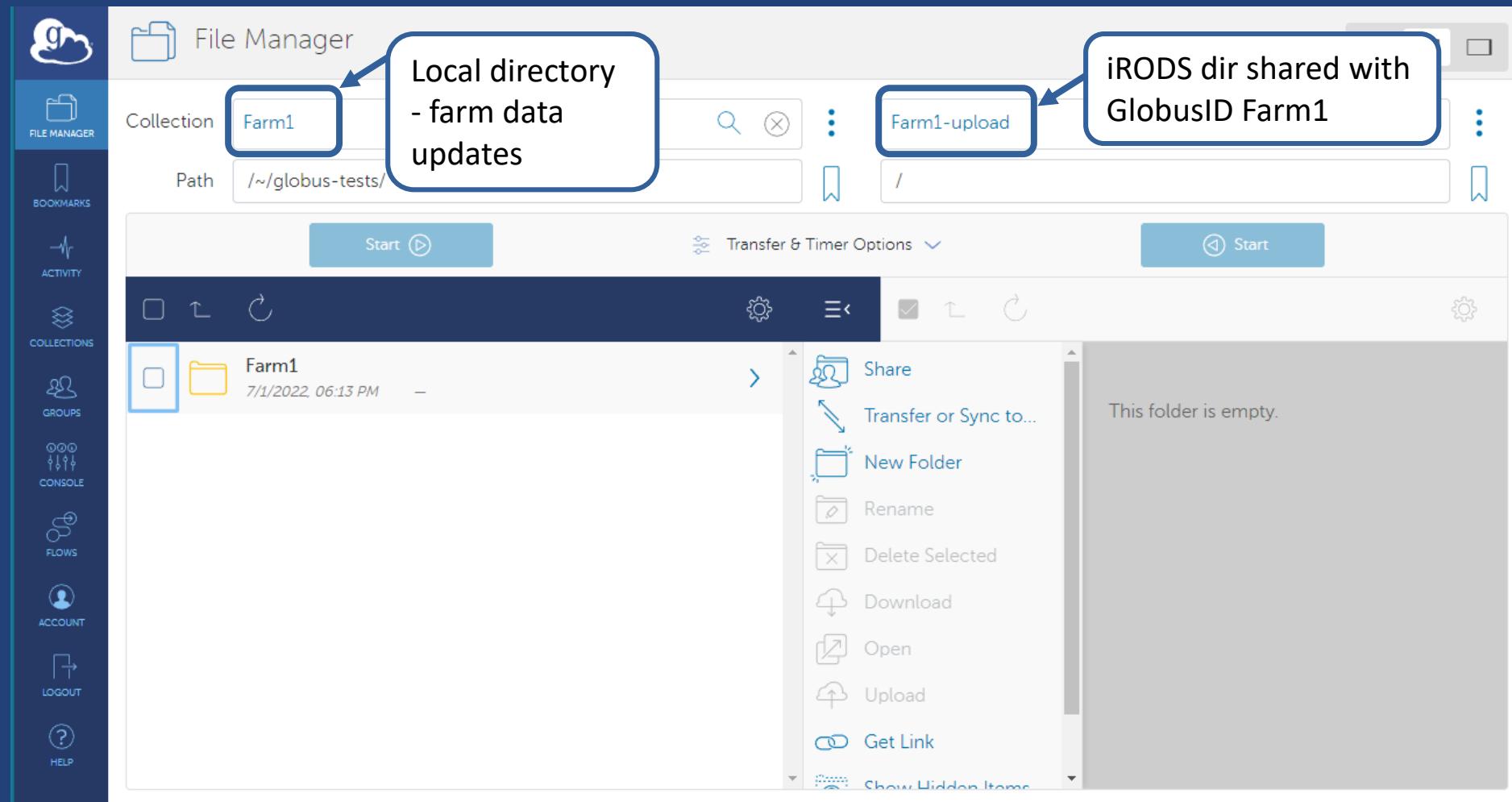
The screenshot shows the metalnx web interface. On the left, the navigation bar includes 'Users', 'Groups', 'Collections', 'Search', 'Templates', 'Shared Links', 'Favorites', 'Tickets', 'Public', and 'Trash'. The main area shows a file tree under 'Farm-uploads' with two entries: 'Farm1-upload' and 'Farm2-upload'. A callout box points to these entries with the text: 'An upload directory per Farm is created in iRODS'. To the right, a detailed view of 'Farm1-upload' is shown. It has tabs for 'Overview', 'Permissions', and 'Roles'. Under 'Shared With', there is a table with one row for 'Ingrid Barcena Roig (ibrtest@globusid.org)'. The 'Permissions' column for this user shows 'READ' and 'WRITE' checkboxes, both of which are checked. A second callout box points to this row with the text: 'The iRODS directory is shared with the GlobusID user with write permissions'.

An upload directory per Farm is created in iRODS

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

User or Group	Path:	Permissions
Ingrid Barcena Roig (ibrtest@globusid.org)	/	READ WRITE

Configuration: farm (1/2)



Configuration: farm (2/2)

The screenshot shows the Globus File Manager interface with the following details:

- Collection:** Farm1
- Path:** /~/globus-tests/
- Transfer & Timer Options:** Set to 3 transfers.
- Transfer Settings:**
 - Note:** These settings will persist during this transfer.
 - sync - only transfer new or changed files (checkbox is highlighted with a blue box and callout "Select sync 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 (cannot be changed because it's defined by sharing config) (checkbox is highlighted with a blue box and callout "Transfer is forced to be encrypted (defined by the sharing config)")
 - Skip files on source with errors
 - Fail on quota errors
- Notification Settings:**
 - Disable success notification
 - Disable failure notification
 - Disable inactive notification
- Schedule Start:** 07/02/2022 06 00 AM
- Repeat:** hours (highlighted with a blue box and callout "A sync process is started every 6 hours")
 - every 6 hours
 - end never

Configuration: department

The screenshot shows the 'File Manager' interface with the following details:

- Collection:** VSC iRODS set.irods.icts.kuleuven.be
- Path:** /set/home/u0089722/ /set/home/project-farms/farm-uploads
- Transfer & Timer Options:** Deptserver → /project-farms/farm-uploads
- Label This Transfer:** Farm-upload-iRODStoDept
- Transfer Settings:**
 - Note:** These settings will persist during this session until changed.
 - sync - only transfer new or changed files (i) where the checksum is different
 - 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:** 07/05/2022 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 data sync to
keep nearly real time
processing of the data

Monitoring

The screenshot displays a monitoring interface with a sidebar and two main panels.

Left Sidebar:

- Activity
- Tasks
- Timers** (highlighted with a blue box)
- FILE MANAGER
- BOOKMARKS
- ACTIVITY
- COLLECTIONS
- GROUPS

Main Panel - List of scheduled 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
Farm1-upload-4h	-	7/1/2022, 07:50 PM	0 - no limit set	every 10 minutes	n/a

Main Panel - Timer Details:

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

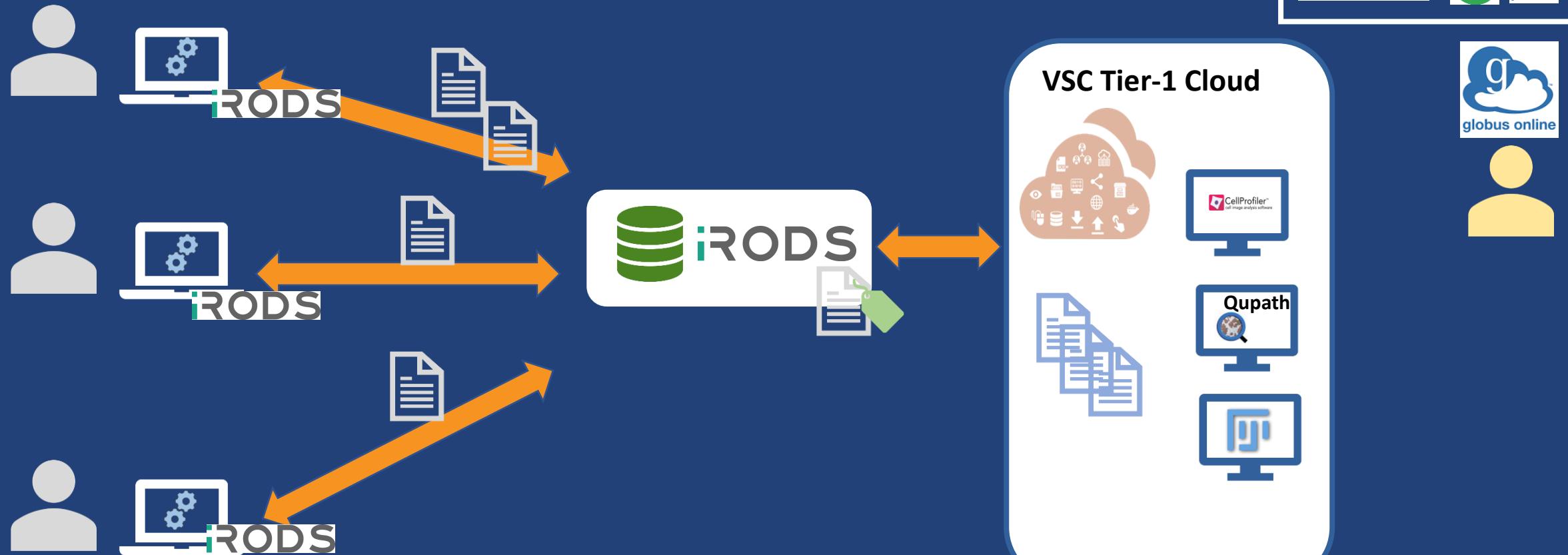
Timer Log (highlighted with a blue box)

All Requests Requests with Faults

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

7/1/2022, 07:50 PM - Timer Farm1-upload-4h, run 1

Share (soon PoC)





Questions?