

# Globus research data management service: Capabilities, international experiences, and future plans

Ian Foster

[foster@uchicago.edu](mailto:foster@uchicago.edu)

University of Chicago/Argonne National Laboratory





Globus delivers ...

Fast and reliable big data transfer, sharing,  
publication, and discovery ...

... directly from your own storage systems ...

... via software-as-a-service, using existing  
identities.

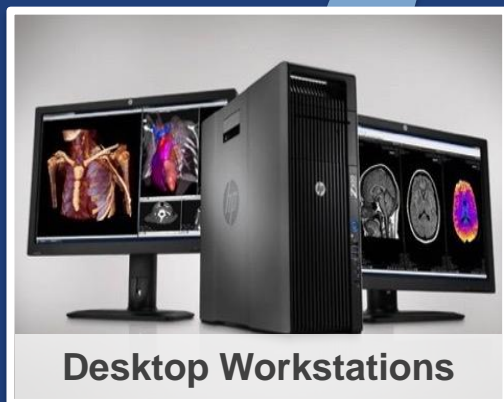
# Unifying access to data across tiers



Research Computing HPC



Personal Resources



Desktop Workstations



Mass Storage



Instruments



# Unifying access across storage interfaces

Current

IBM Spectrum Scale



ceph



Lustre®

HPSS



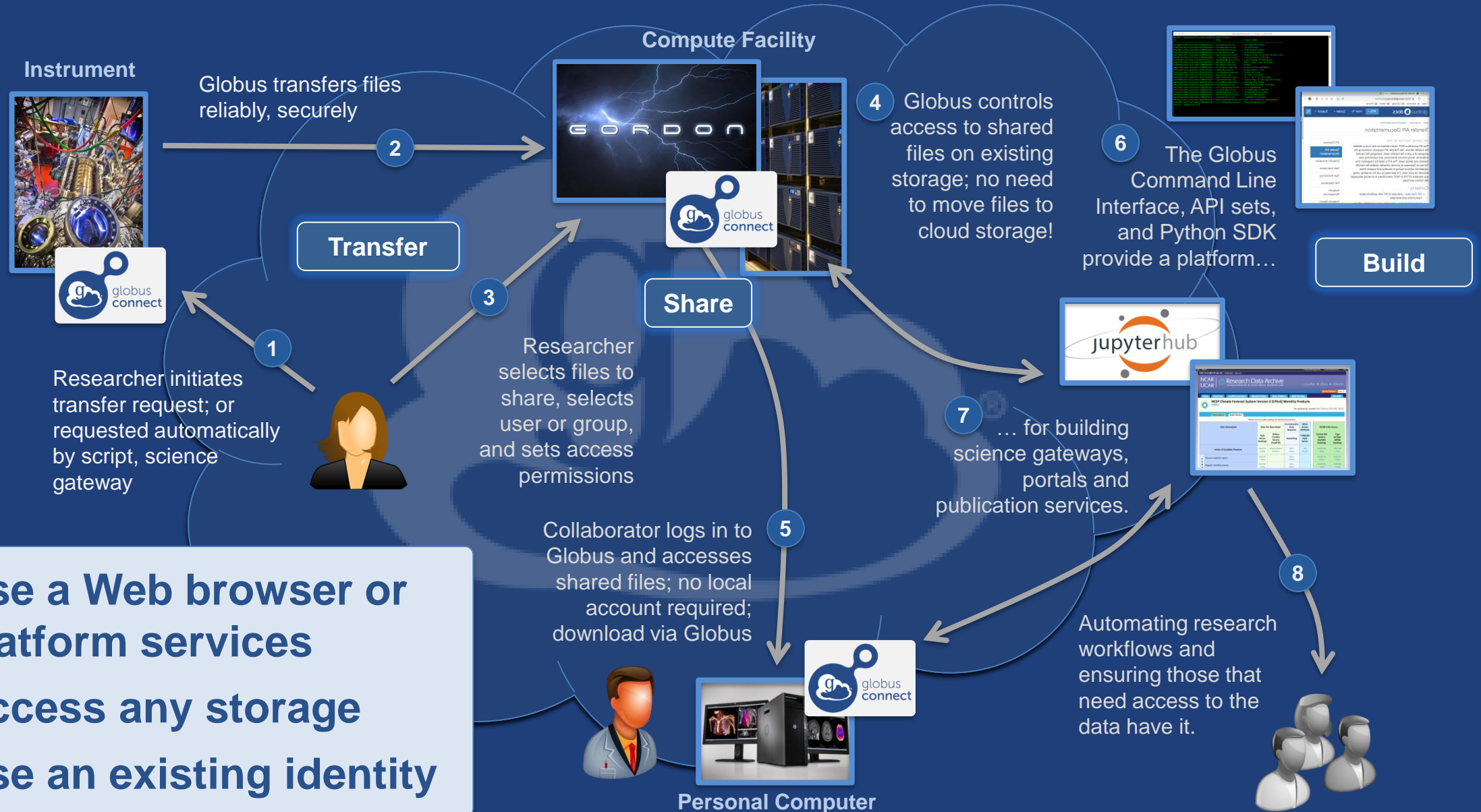
Planned



Microsoft Azure



# Globus SaaS / PaaS: Research data lifecycle



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

# Bridge to campus HPC

Move datasets to campus research  
computing center



Move results to laptop, department, lab...





# Bridge to national cyberinfrastructure

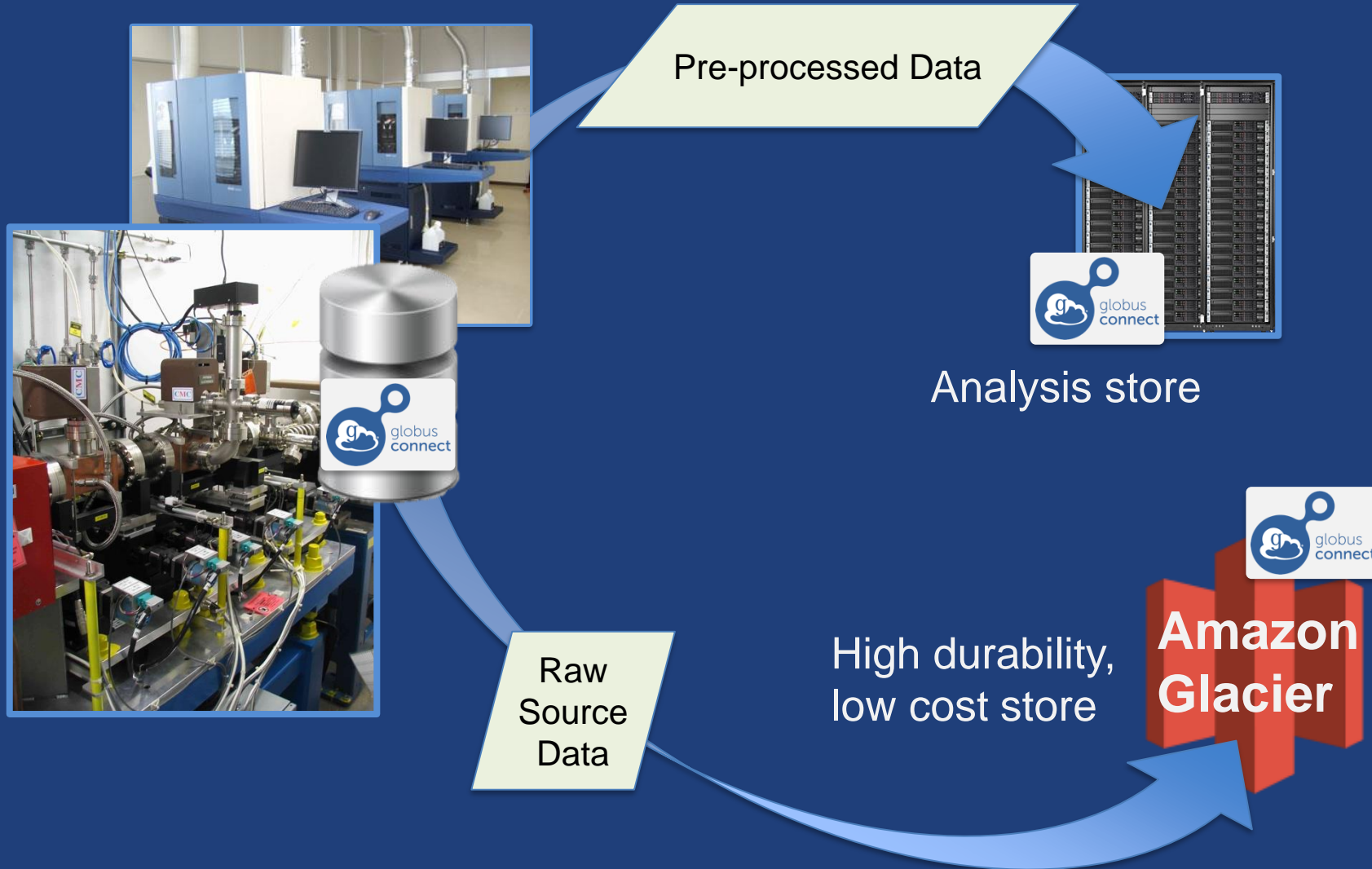
Move datasets to supercomputer,  
national facility



Move results to campus...

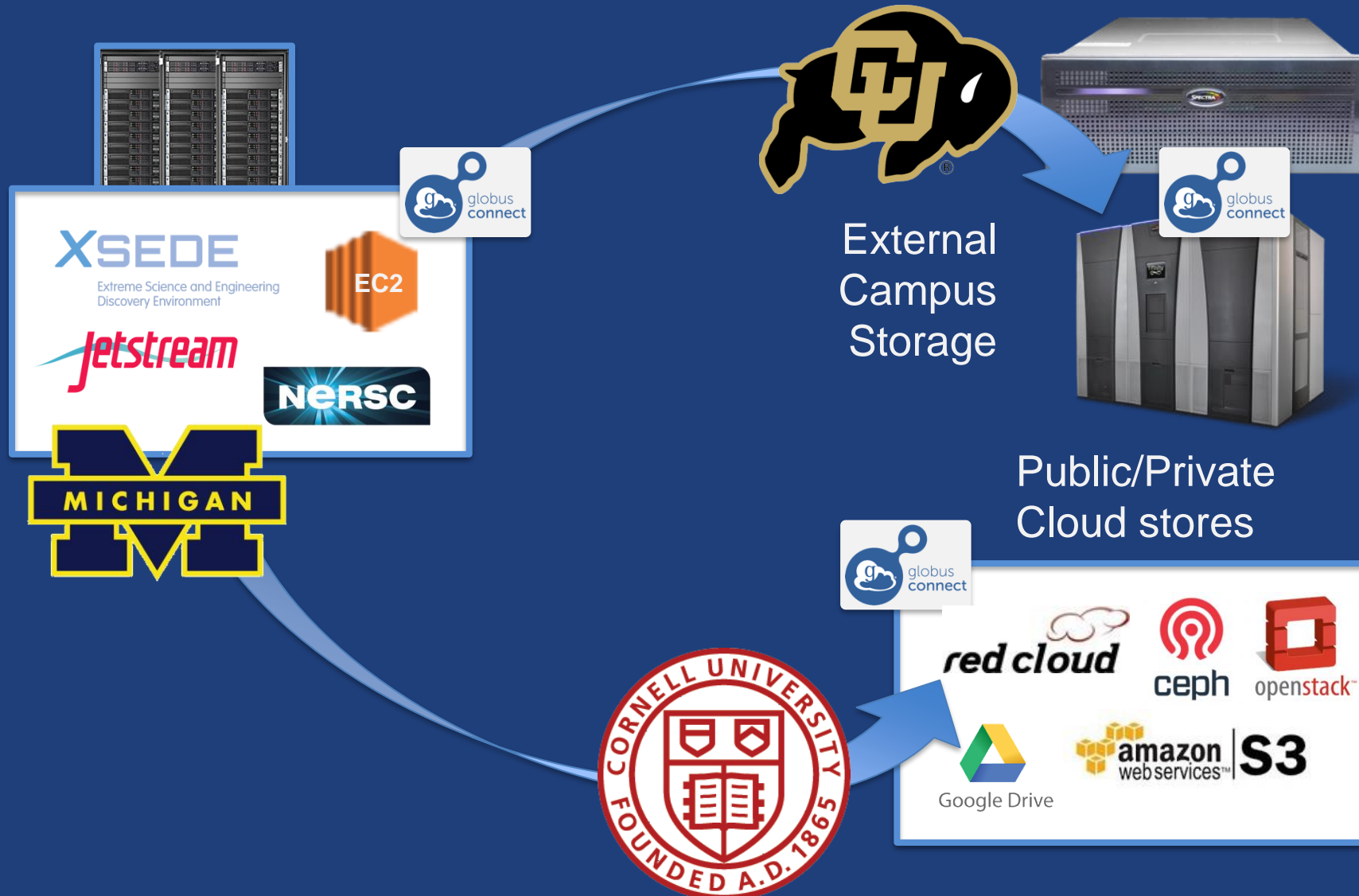


# Bridge to instruments



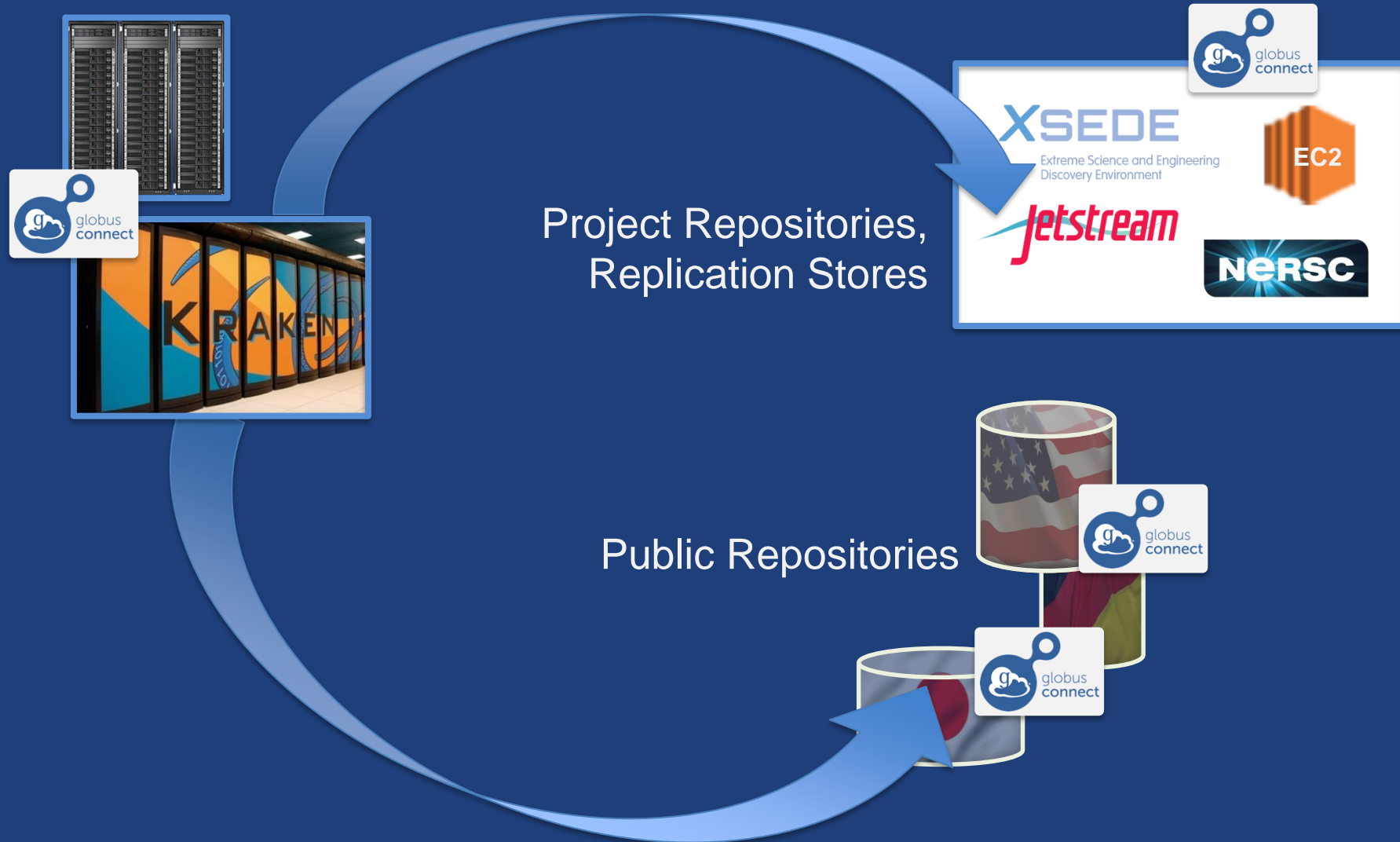


# Bridge to collaborators



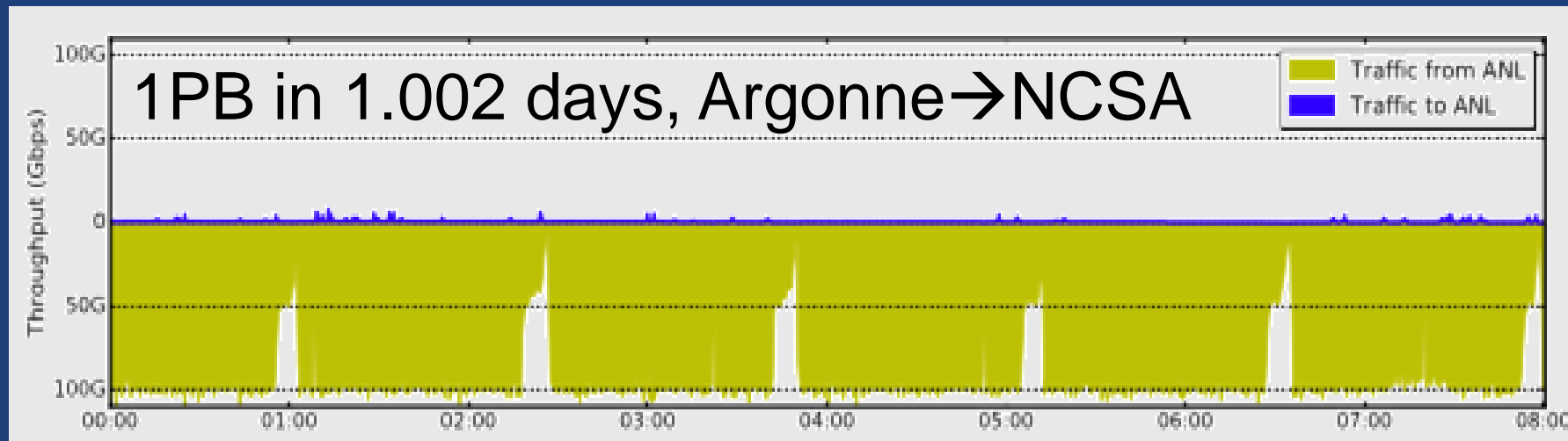
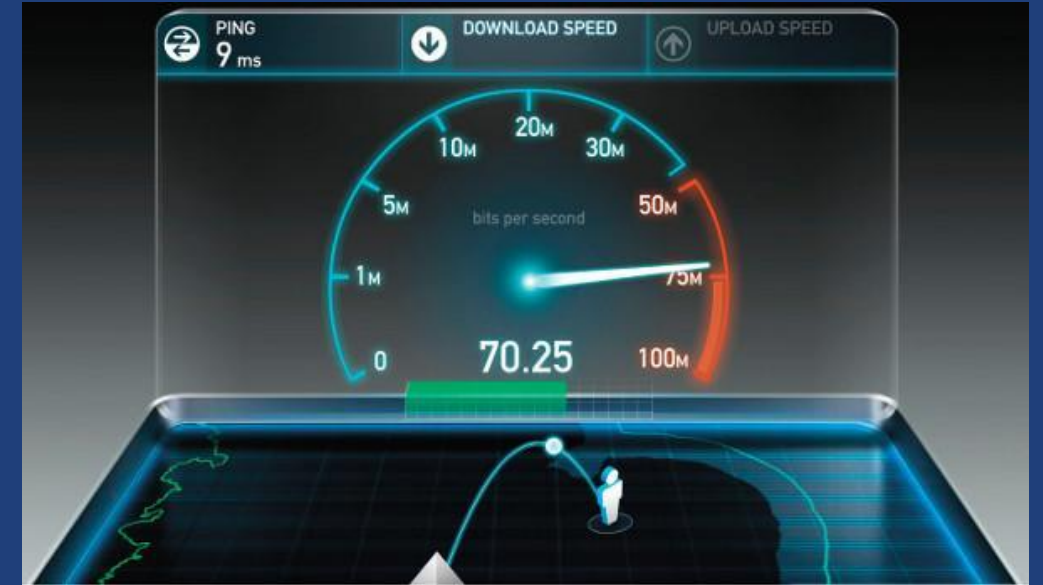


# Bridge to community/public



# Performance is more than just bandwidth

- ✓ Specialized protocols
- ✓ Auto-configuration
- ✓ Parallel DTNs
- ✓ File system optimizations
- ✓ Tape system optimizations





**Protected data further  
complicate the picture...**

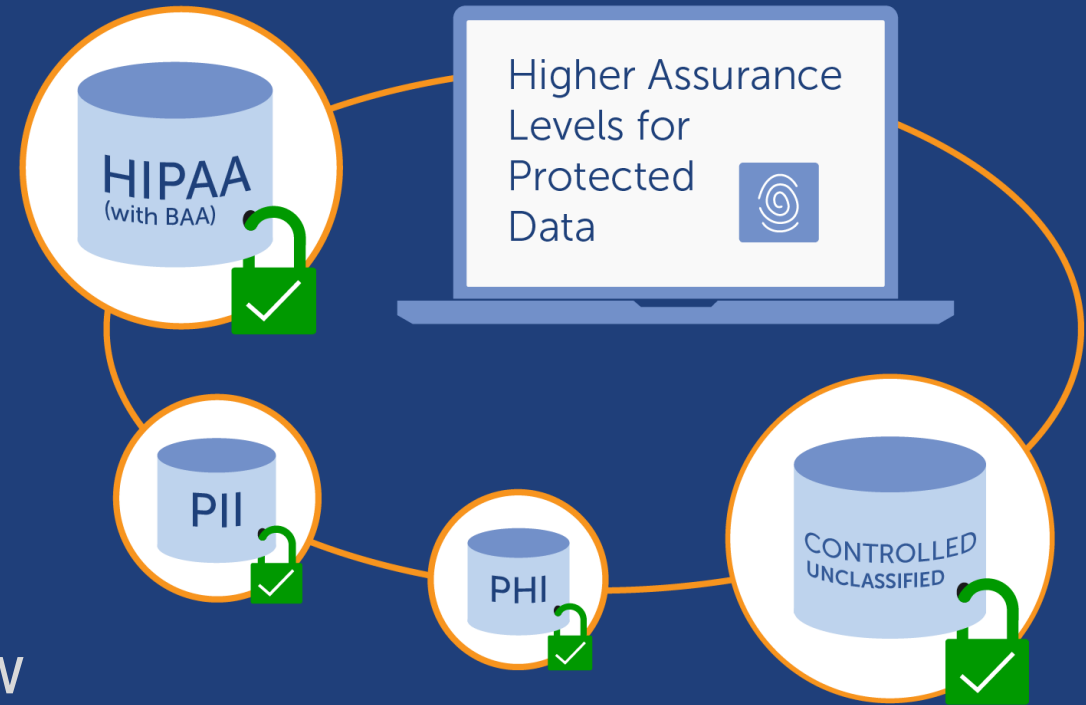
**! Share  $\leftrightarrow$  Protect !**

# New: Globus High Assurance

- Enables transfer and sharing of PHI data

- Features

- Additional authentication assurance (IdP locking)
- Policies for authentication frequency
- Session/Device isolation
- Encryption of user data in transit
- File names/paths can have PHI
- Comprehensive audit logging
- Sec controls: NIST 800-53, 800-171 Low
- Business Associate Agreement (BAA)







# Globus Automate: automating the research lifecycle

- **Simple state machine model**
  - Conditions, loops, fault tolerance, etc.
  - Propagates state through the flow
- **Standardized API for integrating custom event and action services**
  - Actions: synchronous or asynchronous
  - Custom Web forms prompt for user input
- **Actions secured with Globus Auth**

**Auth**



**Search**



**Manage**



**Execute**





# Building the connectome

## APS



1 Imaging



Lab Server 1

2 Acquisition



Lab Server 2

3 Pre-processing

## Argonne Leadership Computing Facility

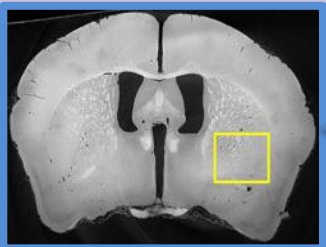


4 Preview/Center



5 User validation

## UChicago



8 Visualization

## Argonne JLSE



7 Publication



6 Reconstruction

Neuroanatomy reconstruction pipeline

9 Science!



# Globus by the numbers ...

**1960**

most shared  
endpoints  
at a single  
institution

**825 PB**

moved

**100 billion**

files processed

**1866**

active server  
endpoints

**120**

subscribers

**117,000**

registered users

**80**

countries with endpoints

**23,818**

active personal  
endpoints

**860**

identity providers

**2.9 PB**

largest single  
transfer to date

**6764**

active shared  
endpoints

**99.9%**

availability



# Thank you to our sponsors!

globus.org



THE UNIVERSITY OF  
**CHICAGO**



U.S. DEPARTMENT OF  
**ENERGY**



**NIST**

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





# THANK YOU, our 120+ subscribers!

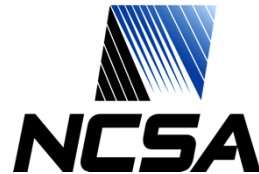


JOHNS HOPKINS  
UNIVERSITY



Yale

CORNELL  
UNIVERSITY



NEW YORK UNIVERSITY



THE UNIVERSITY OF  
CHICAGO



MICHIGAN STATE  
UNIVERSITY



NeSI  
New Zealand eScience  
Infrastructure



VirginiaTech  
*Invent the Future*



syngenta

NLS

Dartmouth

SIMONS FOUNDATION

