

Delivering Research Data Management Services with Globus and a Science DMZ: Introduction and Service Overview

Steve Tuecke
Vas Vasiliadis
Raj Kettimuthu



SC14
New Orleans, LA | hpc
matters.





Presentations and other useful
information available at
globus.org/events/sc14/tutorial

goo.gl/s2Ew50



Thank you to our sponsors!



U.S. DEPARTMENT OF
ENERGY



THE UNIVERSITY OF
CHICAGO

Argonne
NATIONAL LABORATORY



powered by
amazon
web services



Agenda

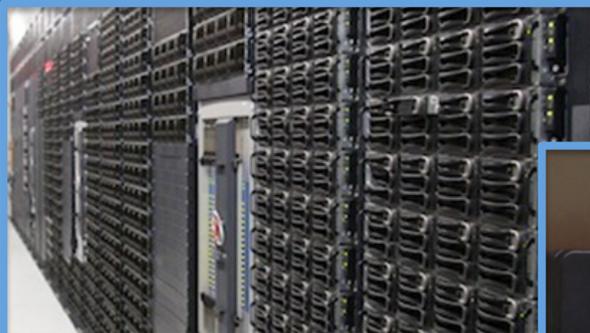
- **Research data management scenarios and challenges (15 min)**
- **Introduction to Globus (10 min)**
- **Demonstrations and Exercises**
 - Accessing Globus and Transferring Files (10 min)
 - File sharing and Group Management (20 min)
 - Data Publication and Discovery (20 min)
 - Globus Command Line Interface (10 min)
- **Globus: today and tomorrow (5 min)**



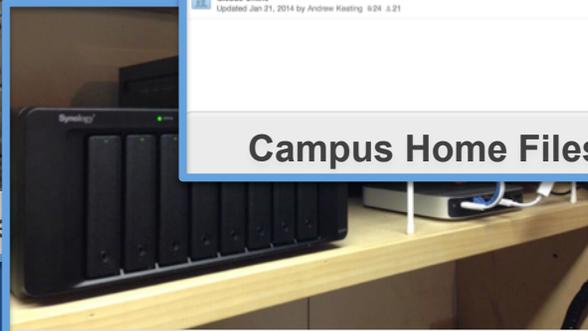
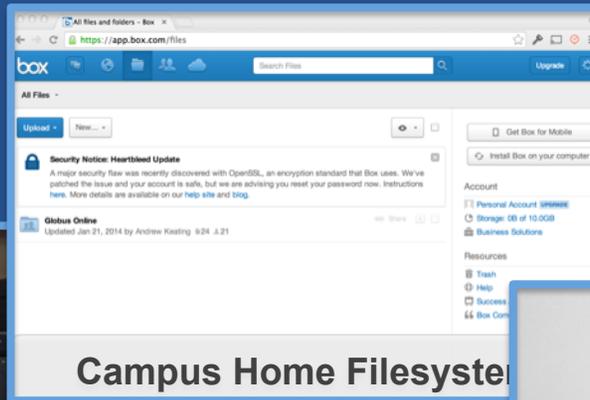
Research data management scenarios and challenge



“I need to easily, quickly, & reliably move or mirror portions of my data to other places.”



Research Computing HPC Cluster



Lab Server



Personal Laptop



Desktop Workstation



XSEDE Resource



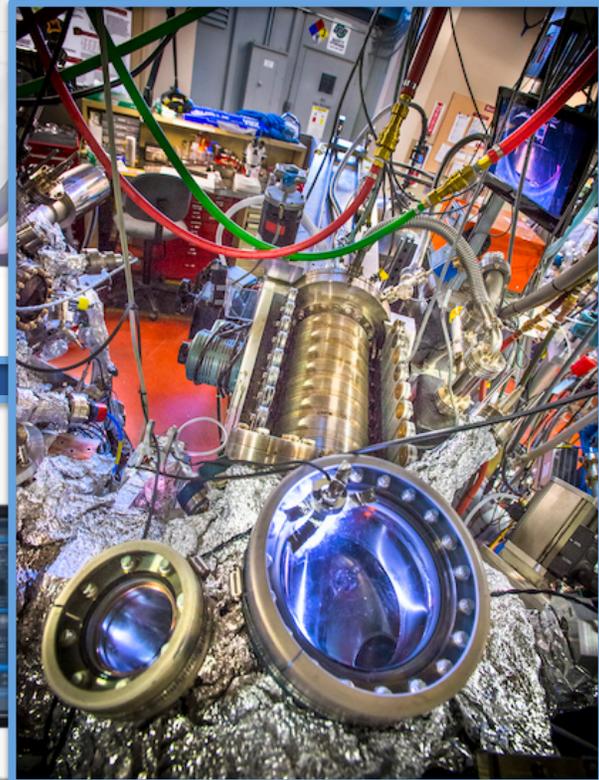


“I need to get data from a scientific instrument to my analysis server.”

MRI



Advanced Light Source



Next Gen Sequencer



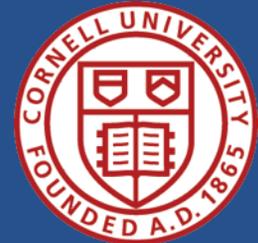
Light Sheet Microscope



“I need to easily and securely share my data with my colleagues at other institutions.”



iComputation
Institute





“I need a good place to store / backup / archive my (big) research data, at a reasonable price.”



Campus Store



Mass Store



Public Cloud Archive



“I need to publish my data so that others can find it and use it.”

Reference Dataset



Scholarly Publication



Active Research Collaboration



Globus introduction and demonstration



Globus is...

Research data management...

...delivered via SaaS



Globus delivers...

Big data transfer, sharing,
publication, and discovery...

...directly from your own
storage systems



It's about the user experience...

flickr ...for your photos

Gmail ...for your e-mail
by Google

NETFLIX ...for your entertainment

 **globus** ...for your research data



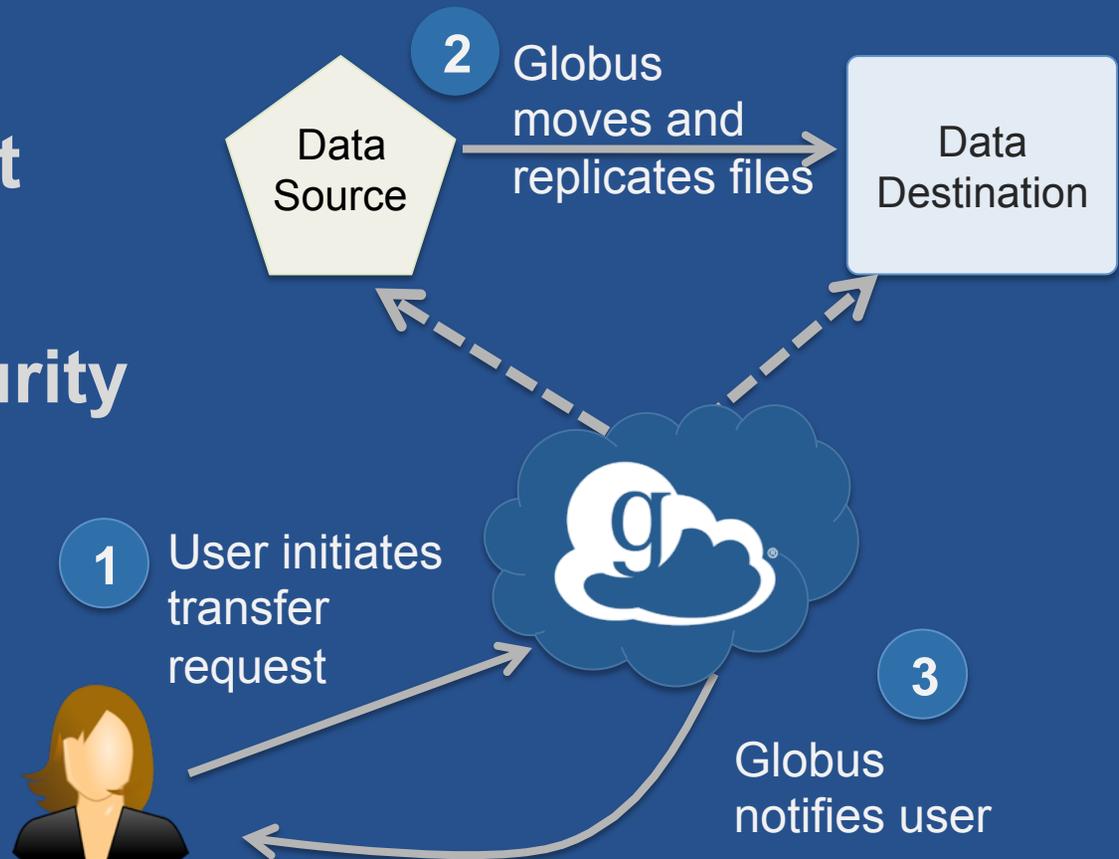
Globus is SaaS

- **Web, command line, and REST interfaces**
- **Reduced IT operational costs**
- **New features automatically available**
- **Consolidated support & troubleshooting**
- **Easy to add your laptop, server, cluster, supercomputer, etc. with Globus Connect**



Reliable, secure, high-performance *file transfer and replication*

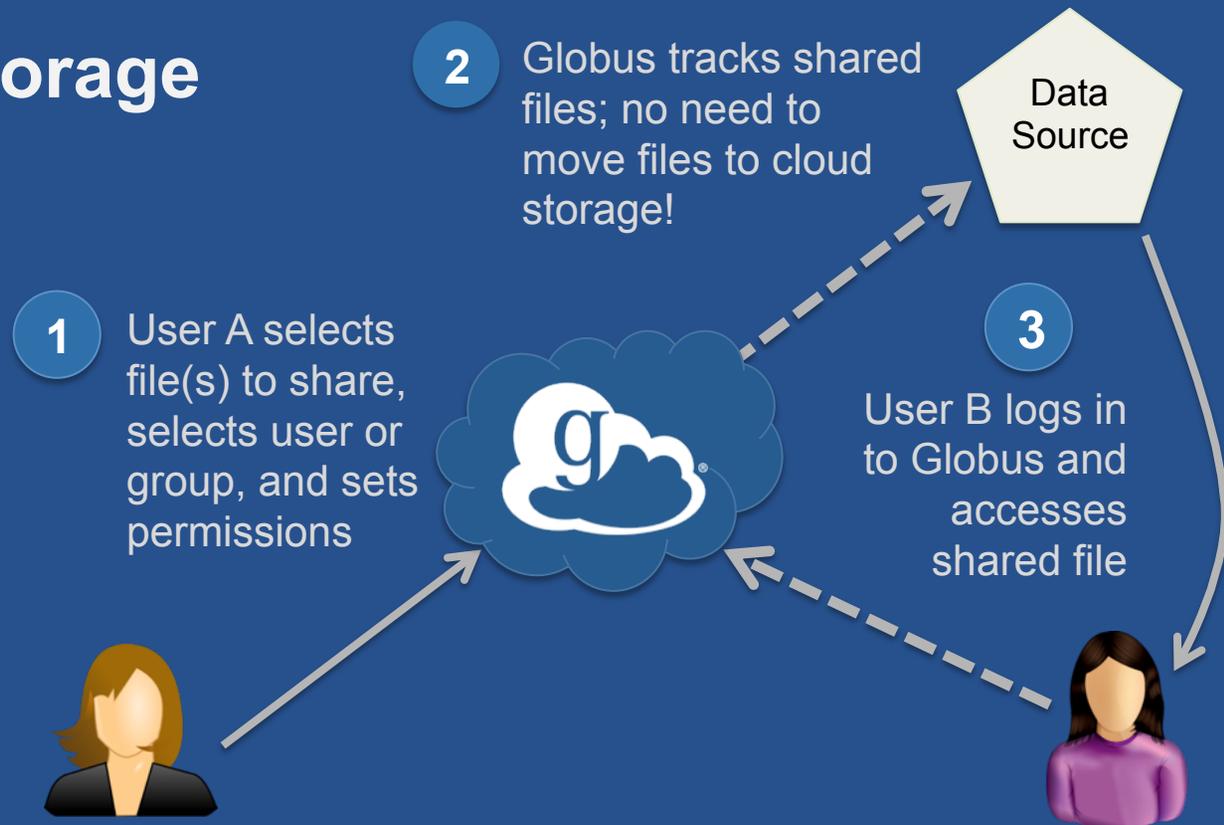
- “Fire-and-forget” transfers
- Automatic fault recovery
- Seamless security integration
- Powerful GUI and APIs





Simple, secure *sharing* off existing storage systems

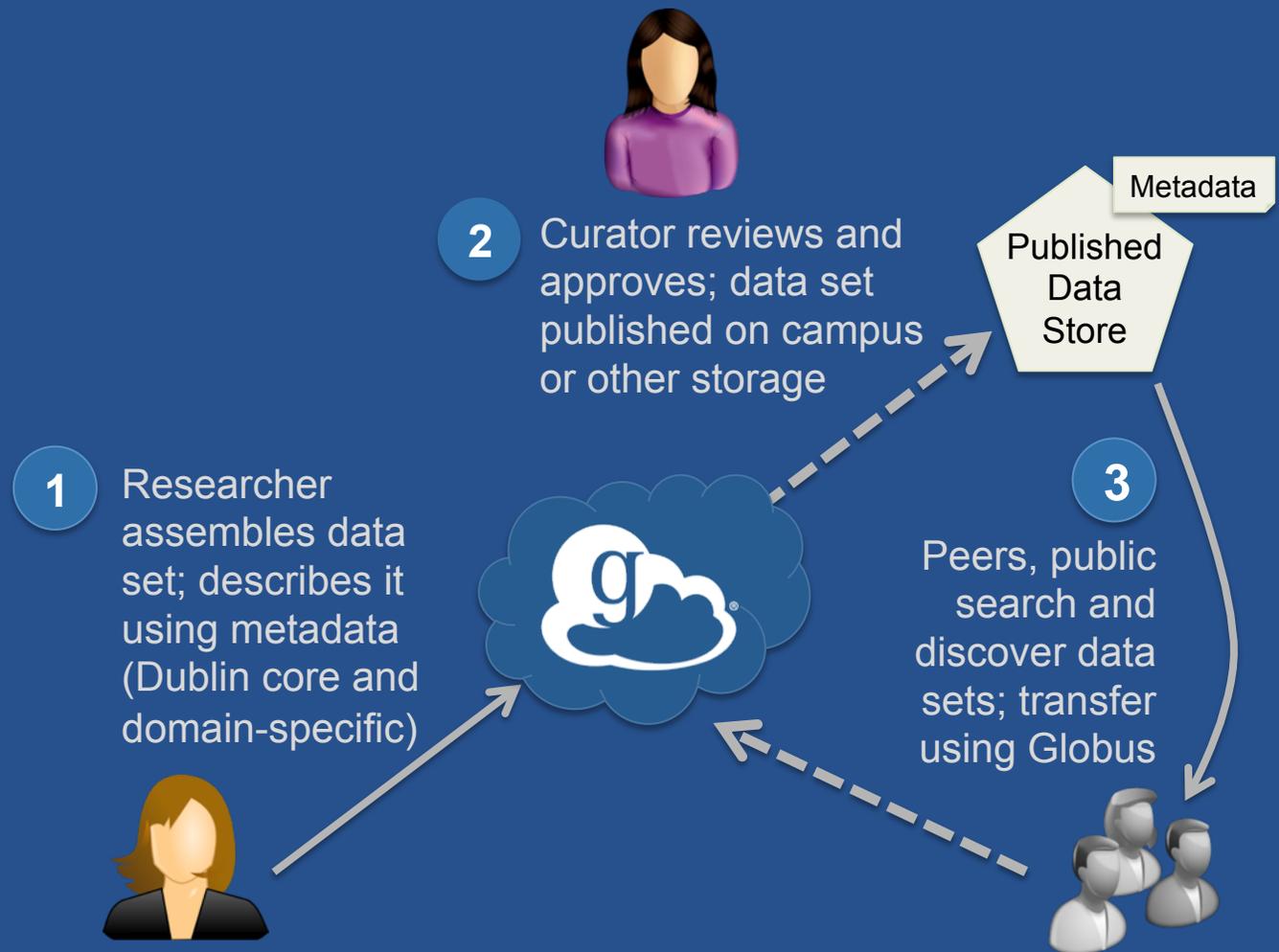
- Easily share large data with any user or group
- No cloud storage required





Curated *publication* of data, with relevant metadata for *discovery*

- Identify
- Describe
- Curate
- Verify
- Access
- Preserve





Demonstration: Accessing Globus File Transfer



Exercise 1: Sign up & transfer files

1. Go to: www.globus.org/signup
2. Create your Globus account
3. Validate e-mail address
4. Optional: Login with your campus/
InCommon identity
5. Install Globus Connect Personal
6. Move file(s) from `esnet#lbl-diskpt1` to your laptop



Demonstration: File Sharing Group Management



Exercise 2: File sharing and Group Management

- 1. Join the “SC14 Tutorial Users”**
 - Go to “Groups” and search for “SC14”
- 2. Create a shared endpoint on your laptop**
- 3. Grant your neighbor permissions on your shared endpoint**
- 4. Access your neighbor’s shared endpoint**
- 5. Optional: Create group, and grant share access**



Demonstration: Data Publication Discovery (sneak preview)



Exercise 3: Publish a data set

1. Create a new data set
2. Enter some metadata
3. Assemble data set from the endpoint `sc14#tutorial_data` (or another endpoint)
4. Complete the workflow and submit
5. Curators (a.k.a. presenters) will “review” your submission and publish
6. Browse the published data set



Demonstration: Globus Command Line Interface (CLI)



Exercise 4: Globus CLI

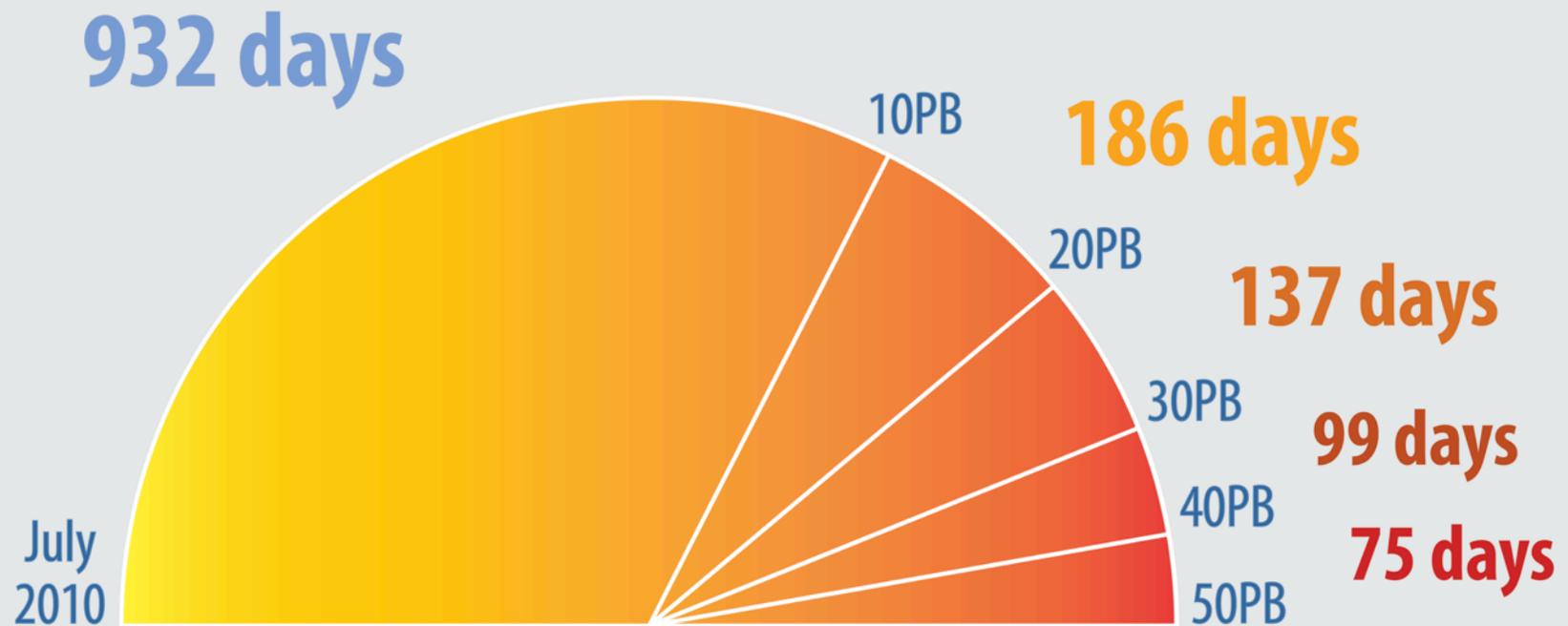
1. **Optional: Generate SSH key**
2. **See globus.org/account/ManageIdentities**
3. **Add your SSH key to your Globus identity**
4. **SSH to `cli.globusonline.org`**
5. **Check on status of earlier transfer(s)**
6. **Optional: Transfer a file using the `scp` command**



Globus: today and tomorrow



Moving the Needle



~70PB moved to date



We are a non-profit, delivering a production-grade service to the non-profit research community

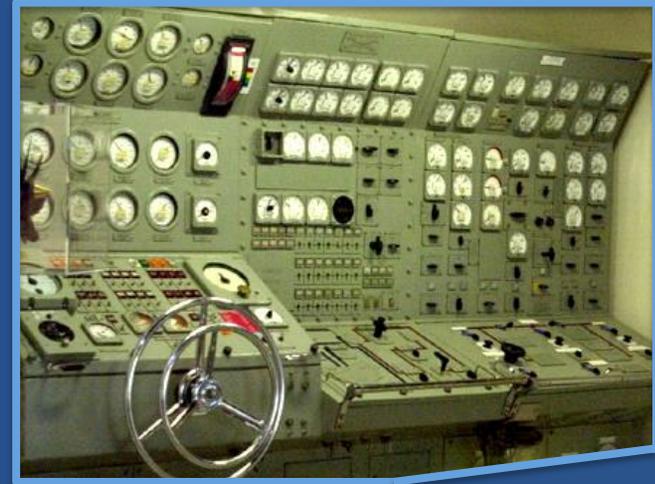
Our challenge:
Sustainability



Globus Subscriptions

- **Provider**

- Managed endpoints
- Host shared endpoints
- Management console
- Data publication*
- Amazon S3 endpoints
- Usage reports
- Priority support
- Mass Storage System optimization
- Transfer integration support



- **Branded Web Site**

- **Alternate Identity Provider (InCommon is standard)**

globus.org/provider-plans



Some early Globus supporters

XSEDE

Extreme Science and Engineering
Discovery Environment



**Carnegie
Mellon
University**

**MICHIGAN STATE
UNIVERSITY**



Te Whare Wānanga o Tāmaki Makaurau



Information Sciences Institute



EMORY



**CORNELL
UNIVERSITY**



Ole Miss



THE UNIVERSITY OF
CHICAGO

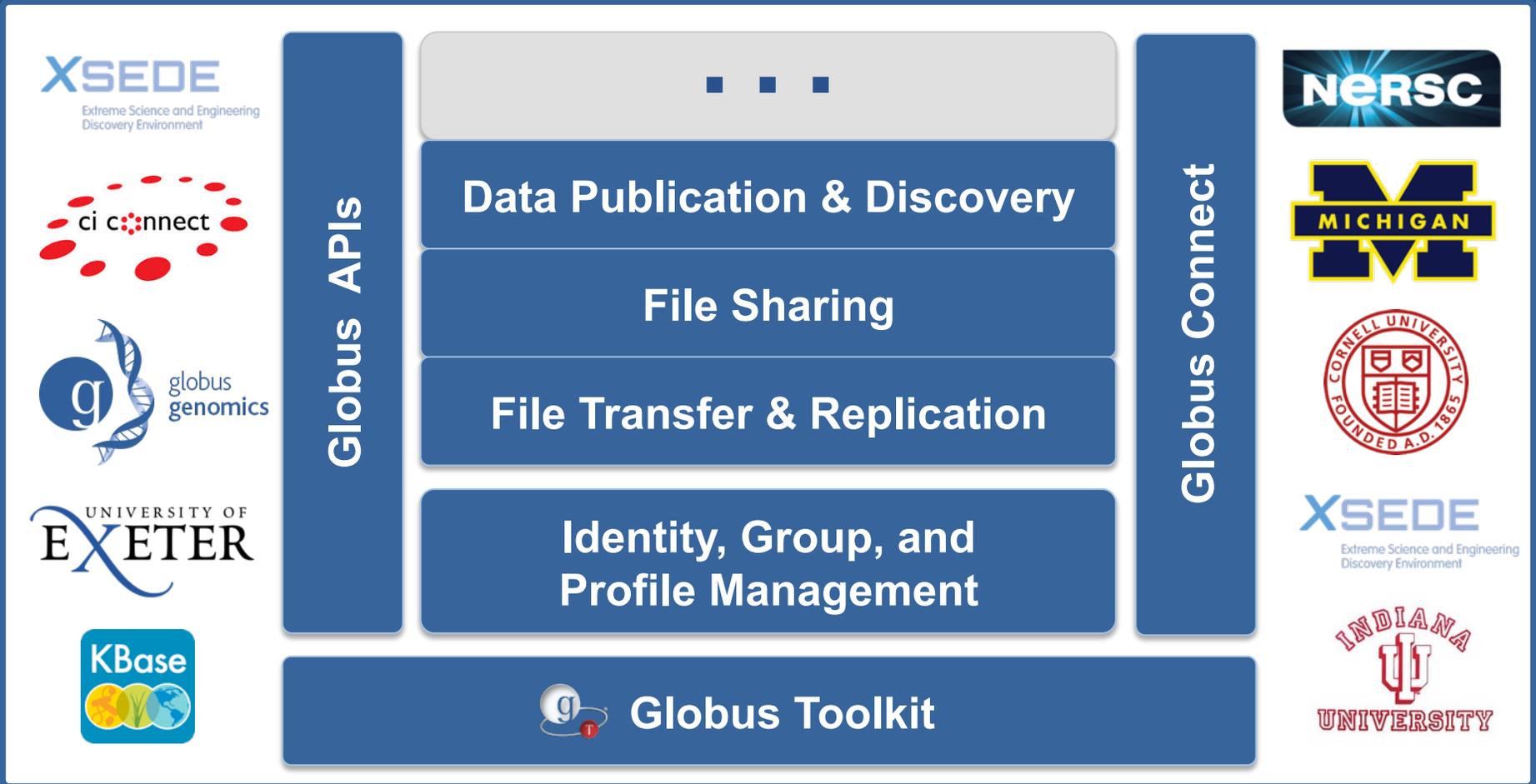


NEW YORK UNIVERSITY





Globus Platform-as-a-Service





End: Introduction and Service Overview