

Building the Model Research Data Portal

Ian Foster

foster@uchicago.edu













Developer Workshop: Building the Modern Research Data Portal

New high-speed networks make it possible, in principle, to transfer and share research data at tremendous speeds and scales—but have also proved challenging to use in practice. Two new technologies now allow us to translate this potential into reality: Science DMZ architectures provide frictionless end-to-end network paths; and Globus APIs allow programmers to create powerful research data portals that leverage these paths for data distribution, synchronization, and other useful purposes.

Come to Chicago in April to learn more!

Introduction, Concepts, and Components

IMPERIAL 2

Led by: TBD

We will introduce the Modern Research Data Portal and set the context for how Globus and the ScienceDMZ combine to deliver unique data management capabilities. This will include:

- Overview of use cases: Common patterns like data publication/distribution, orchestration of data flows, etc.
- Overview of the Globus platform: Architecture and brief overview of available services
- Introduction to the Globus Auth API: Authenticating and authorizing a client
- Introduction to the Globus Transfer API: Make your first call and move data with Globus
- Introduction to the Python SDK for using Globus Auth and Transfer



Thank you to our sponsors!



U.S. DEPARTMENT OF
ENERGY



THE UNIVERSITY OF
CHICAGO

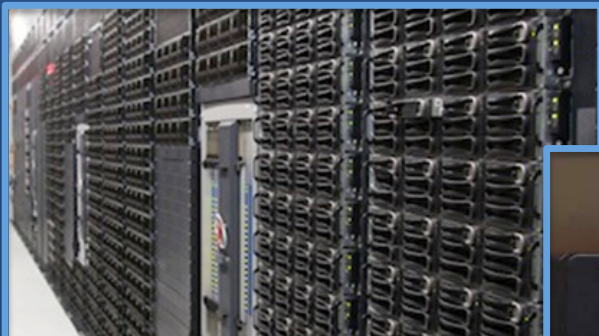
Argonne
NATIONAL LABORATORY



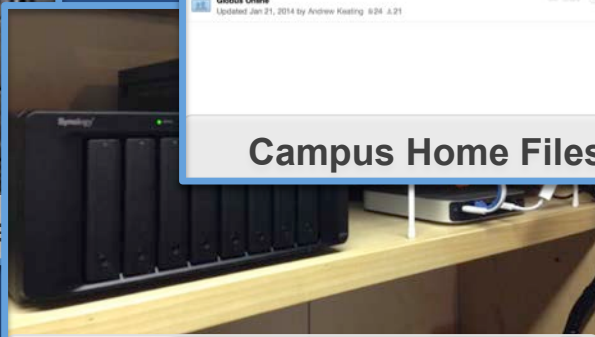
powered by
amazon
web services



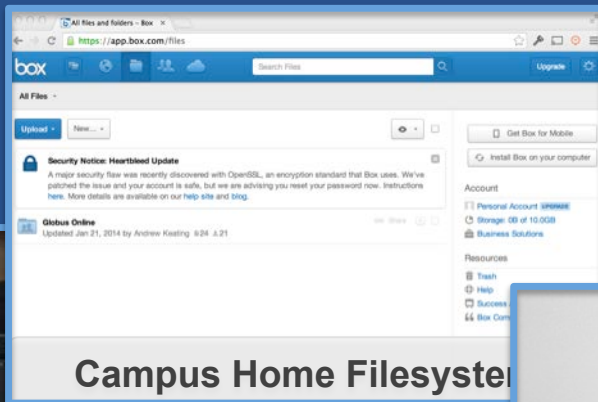
“I need to easily, quickly, & reliably move portions of my data to other locations.”



Research Computing HPC Cluster



Lab Server



Campus Home Filesystem



Personal Laptop



Desktop Workstation



XSEDE Resource



Public Cloud



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

MRI



Advanced Light Source



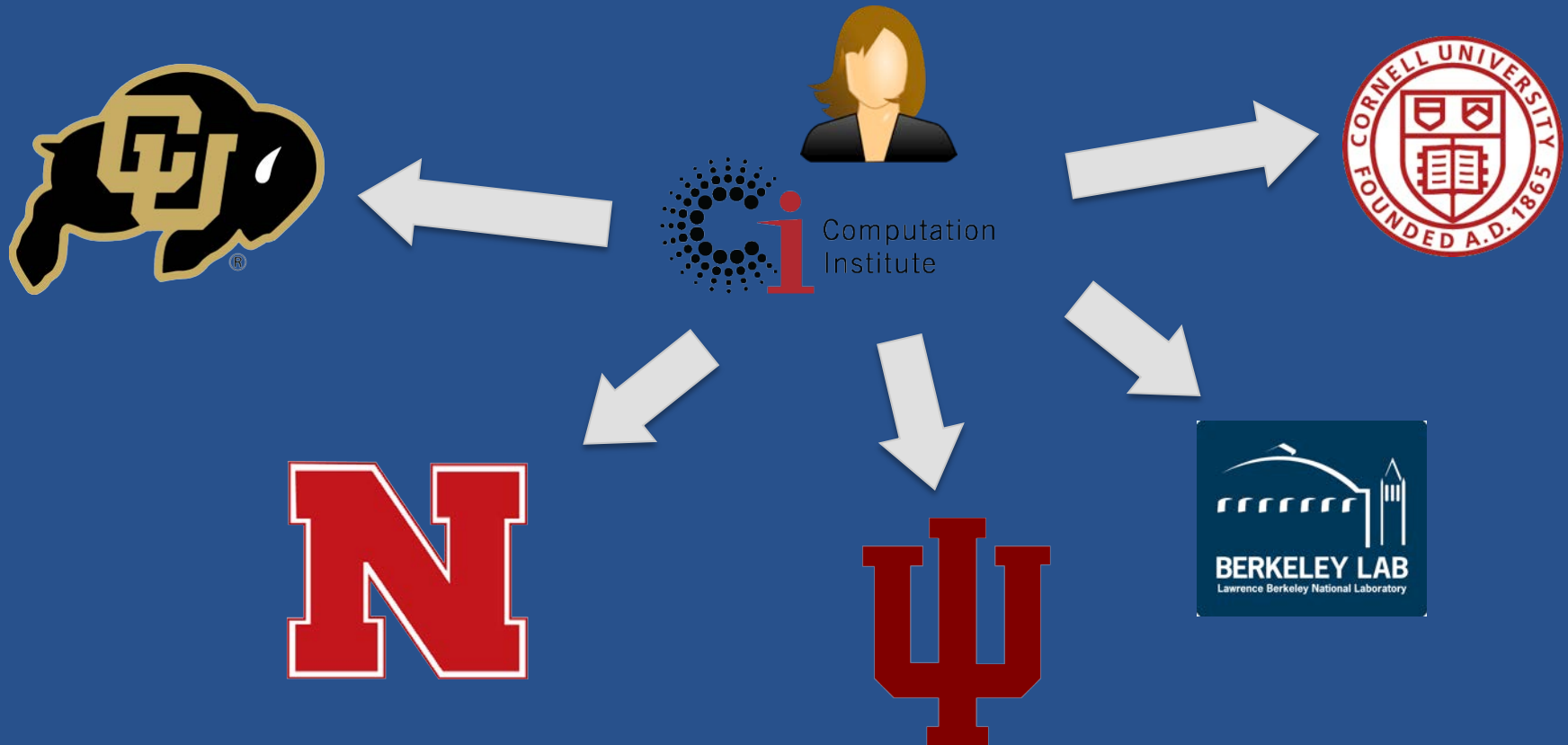
Next Gen Sequencer



Light Sheet Microscope



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





“I need to publish my data so others can find/use/validate/reproduce it.”



Scholarly
Publication

Reference
Dataset



Research
Community
Collaboration



Research data management today

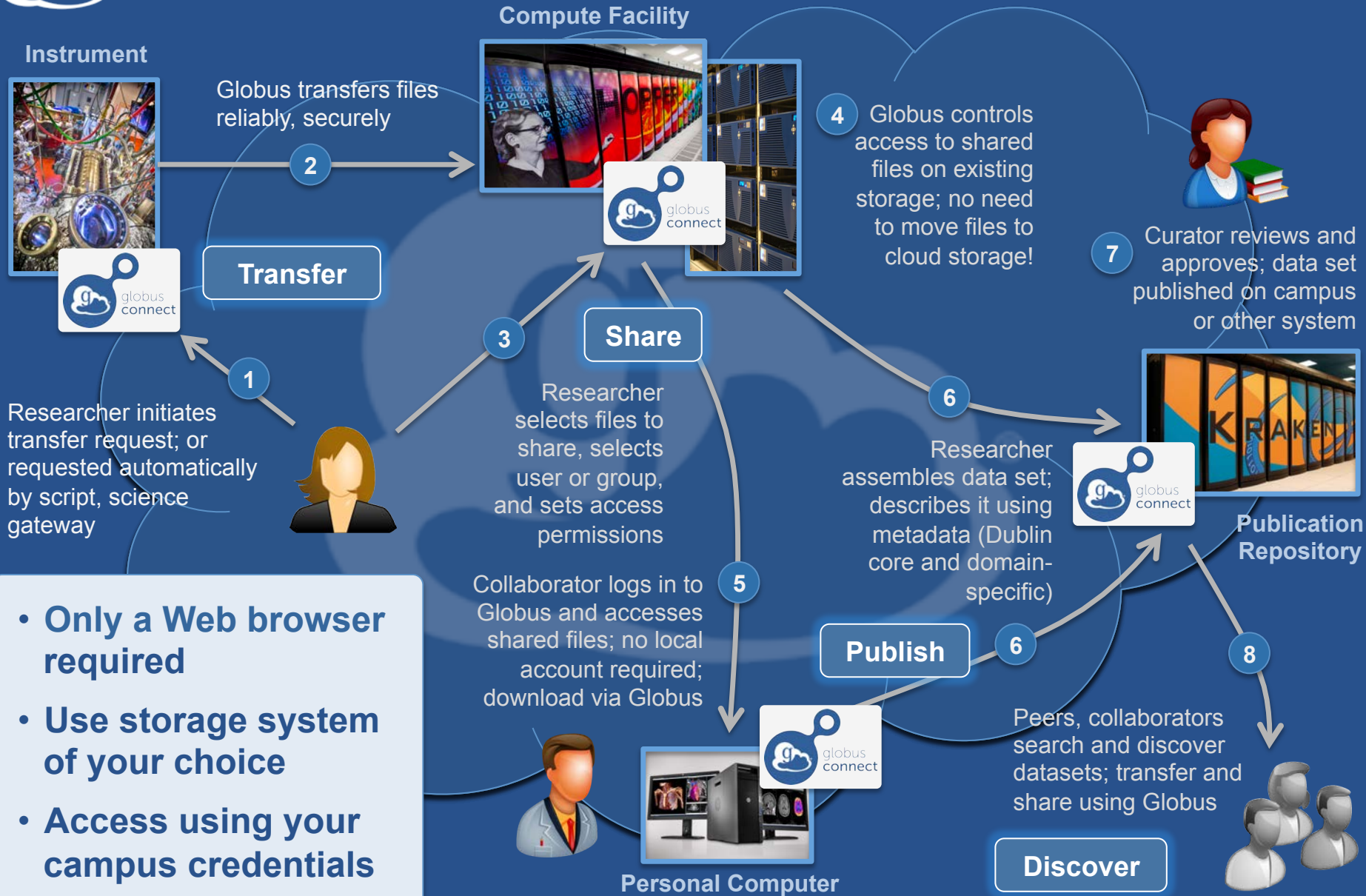


Index?





Globus and the research data lifecycle



- Only a Web browser required
- Use storage system of your choice
- Access using your campus credentials



Globus delivers...

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

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

...via software-as-a-service



Globus is SaaS

- **Easy to access via Web browser**
 - Command line, REST interfaces for flexible automation and integration
- **New features automatically available**
- **Reduced IT operational costs**
 - Small local footprint (Globus Connect)
 - Consolidated support and troubleshooting



Lowering collaboration overhead

- **Grant collaborators access to data on systems without requiring local accounts**
- **No need to replicate or move data to separate system/cloud just for sharing**
- **Researchers manage “virtual” ACLs...**
- **Respect local system access controls**



Group Management

- User-managed group creation and management
- Flexible control over roles, policies, workflows, and visibility
- Groups can be used in authorization decisions

Example: kBase

- Every kBase user added to `kbase_users`
- Subgroups also created
- Groups used for access control

The screenshot shows a user interface with a dark navigation bar at the top containing the KB logo, 'Dashboard', 'Groups', 'ian's account', and 'Log Out'. Below the navigation bar, there are three main sections: 'My Groups' with a link to 'kbase_users', 'Browse Groups' with the description 'Browse and join groups', 'My Account' with the description 'View and change your account settings, including contact information and security credentials', and 'KBase Site' with the description 'Return to the main KBase website'. Each section is accompanied by a circular icon representing the respective function.

The screenshot shows a page titled 'kbase_users'. At the top, there are three buttons: 'Home', 'Members', and 'SubGroups'. Below these buttons, there is a list of subgroups: 'kbase_staff', 'kbase_microbes', and 'ENIGMA'.



Globus today...

4

major services

137 PB

transferred

20 billion

files processed

31,000

registered users

13

national labs
use Globus

10,000

active endpoints

~400

active daily users

99.95%

uptime

35+

institutional
subscribers

1 PB

largest single
transfer to date

3 months

longest
continuously
managed transfer

130


federated
campus identities




New Features: Globus Auth and HTTP



New features: Globus Auth

 **CI**

 globus

Request for Approval

Do you approve of granting the following clients access?

Globus Web Client needs to use your `chard@uchicago.edu` identity in order to:

- Manage identities in Globus Auth ⓘ
- Manage your Globus Groups ⓘ
- Transfer files using Globus Transfer ⓘ

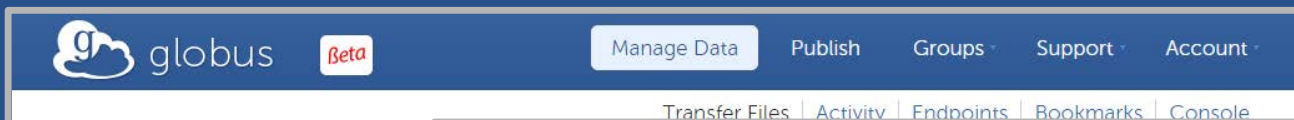
Globus Transfer needs to use your `chard@uchicago.edu` identity in order to:

- Manage your Globus Groups ⓘ
- View your identities on Globus Auth ⓘ

By clicking "Approve", you allow clients to use your information in the ways described above.



New features: HTTP



Transfer Files

Endpoint:

Path:

select none | up one folder | refresh

- hello.txt
- ianfoster.jpg

Label This Transfer:

This will be displayed in your transfer history

Transfer Settings

- sync - only transfer files that have changed
- delete files on destination
- preserve source file permissions
- verify file integrity
- encrypt transfer

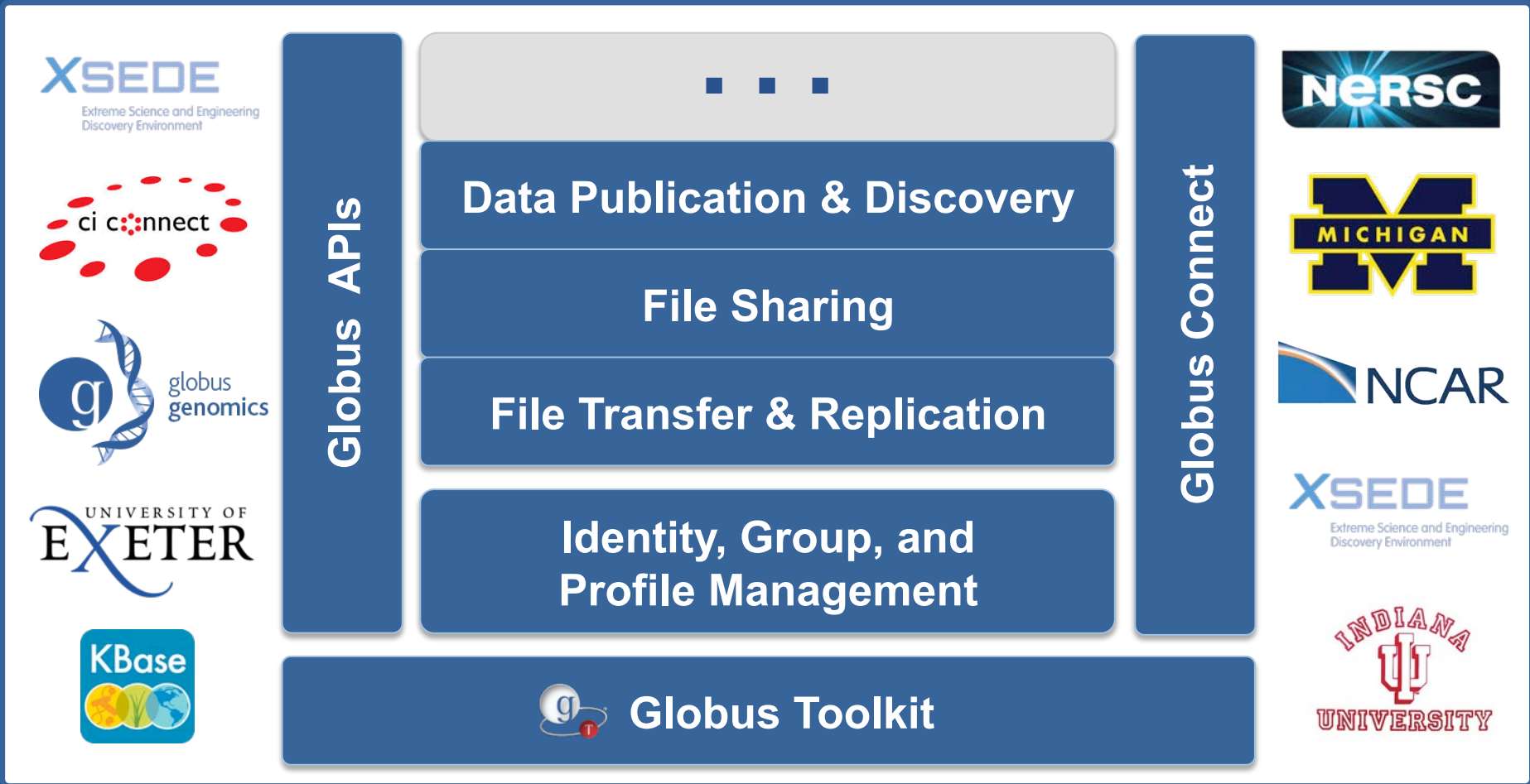




Leveraging the Globus Platform



Globus Platform-as-a-Service





APIs and Clients

- **Globus Nexus (identity, profile, groups)**
 - API: <http://globusonline.github.io/nexus-docs/api.html>
 - Clients:
 - <https://github.com/globusonline/python-nexus-client>
 - <https://github.com/globusonline/java-nexus-client>
- **Globus transfer**
 - API: <https://docs.globus.org/api/transfer/>
 - Clients:
 - <https://github.com/globusonline/transfer-api-client-python>
 - <https://github.com/globusonline/transfer-api-client-java>
- **Globus publication (coming soon)**

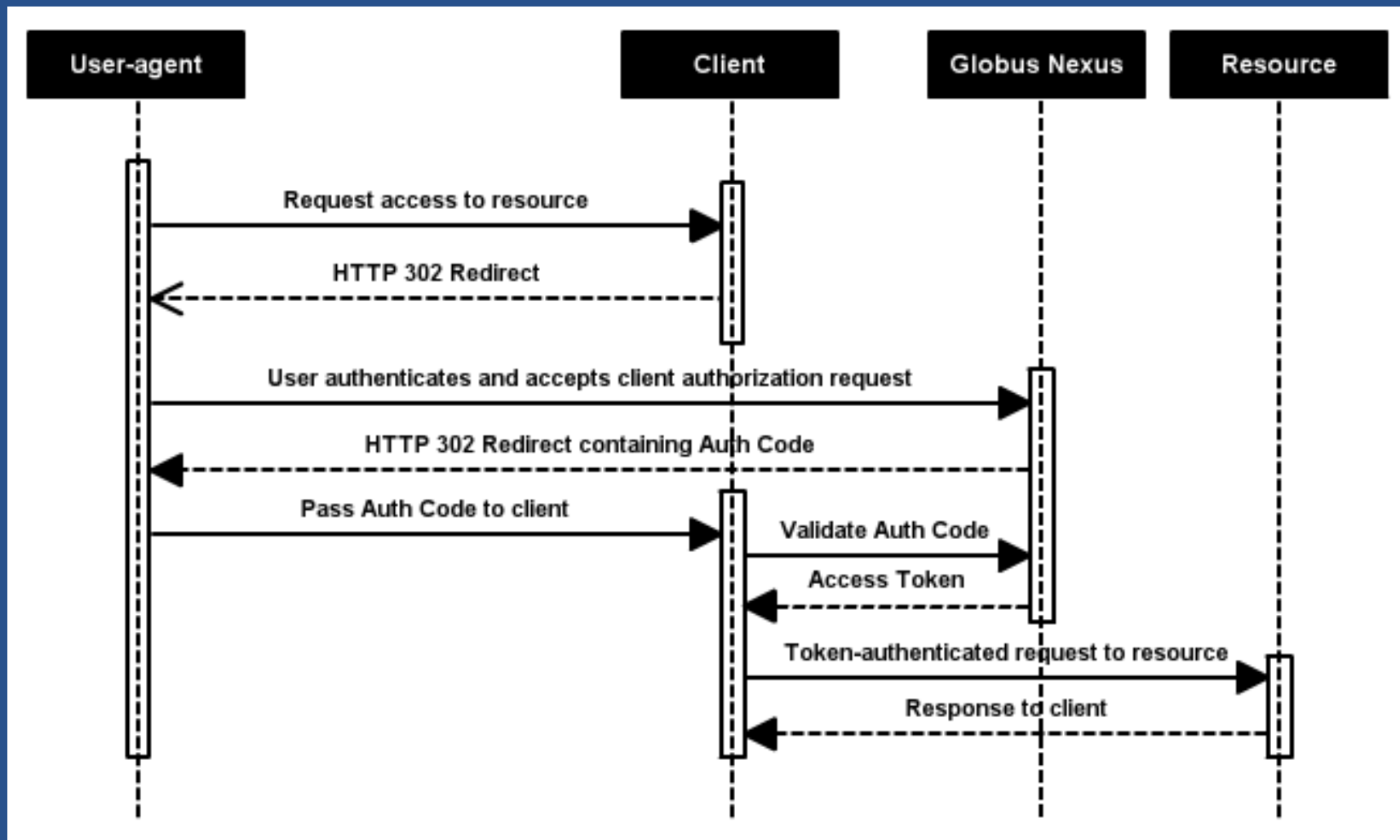


Developing a 3rd party application

- **Register client (URL and client username/password)**
- **Implement an OAuth client (or reuse an existing one)**
- **Store the resulting access token and make calls to Globus APIs on behalf of the authenticated user**



Globus OAuth workflow





Developing a 3rd party application

The image shows a stack of three browser windows illustrating the development of a third-party application for the Globus Data Publication Dashboard. The top window is the main dashboard at <https://publish.globus.org/jspui/PublishDashboard>. The middle window shows an OAuth sign-in process with a URL like https://www.globus.org/OAuth#response_type=code&client_id=globuspublish&redirect_uri=https%3A%2F%2Fpublish.globus.org%2Fjspui%2Fgoauth-login. The bottom window shows the main Globus website at <https://publish.globus.org>.

The dashboard interface includes a search bar, navigation links for "Manage Data", "Publish", "Groups", "Support", and "chard", and a "Data Publication Dashboard" section with buttons for "Submit a New Dataset" and "View My Published Datasets". A copyright notice at the bottom reads: © 2010-2015 Computation Institute, University of Chicago, Argonne National Laboratory legal.



Globus PaaS at NCAR

- **Research Data Archive at NCAR**
- **Integrate Globus for data downloads**
- **Shared endpoint with subfolder per request**
- **Single sign on via streamlined account provisioning**

Find Data | **Ancillary Services** | **About/Contact**

All Datasets | Recently Added/Updated | Browse the RDA

- **GCMD Topic:**
 - Agriculture • Atmosphere • Biosphere • Climate
 - Oceans • Paleoclimate • Solid Earth • Spectral/En
- **Atmospheric Reanalysis Data:**
 - All Reanalysis Datasets • BPRC Arctic System Reanal
 - ECMWF ERA15 Reanalysis (ERA15) • ECMWF ERA40
 - ECMWF Interim Reanalysis (ERA-I) • JMA Japanese 2
 - JMA Japanese 55-year Reanalysis (JRA55) • NCEP Cl
 - NCEP North American Regional Reanalysis (NARR) •
 - NCEP/NCAR Reanalysis Project (NNRP) • NOAA-CIRE
- **Station Observations:**
 - Land Surface Air Temperature: Hourly, Monthly

Find Platform Observations datasets

CISL Research Data Archive

Managed by NCAR's Data Support Section
Data for Atmospheric and Geosciences Research

RDA





Branded sites

BLUE WATERS
REGISTERED PERSONAL COMPUTING

NCSA |

Reliable, high-performance, secure file transfer by Globus Online.

Blue Waters has partnered with the Globus Online file transfer service.

You may access this service by entering your Blue Waters username and password.

Sign In

Use Your NCSA Blue Waters login

NOTE: Please be aware that the faculty website for Blue Waters has been moved to a Globus Online account. If you had a Blue Waters account (which you can create one).

© 2011 Open Science Grid and University of Chicago

osg connect

Support | Resources | OSG Connect | Sign In / Register

Efficiently connect your science to cycles and data

osg connect

OSG Connect offers users simple access to distributed high throughput computing resources, and reliable, high-performance file transfer services.

Sign In

Use your iCommons / CILogon login

You will now be redirected to iCommons / CILogon's authentication page.

© 2011 Open Science Grid and University of Chicago

Research Computing Center

THE UNIVERSITY OF CHICAGO

Log In | Sign Up

Sign In

Use Your iCommons / CILogon login

You will now be redirected to iCommons / CILogon's authentication page.

© 2011 The University of Chicago
8000 S. Ellis Ave. Suite 100 Chicago, IL 60637
Phone: 773-799-2667
Chicago, IL

KBase
Knowledge Base
DOE Systems Biology Knowledgebase

Log In | Sign Up

Sign In

Use Your GlobusOnline login

Use Your iCommons / CILogon login

You will now be redirected to iCommons / CILogon's authentication page.

ENERGY
Office of Science

KBase is sponsored by the U.S. Department of Energy's Office of Biological and Environmental Research.

INDIANA UNIVERSITY Cyberinfrastructure Gateway

Sign In

Use Your iCommons / CILogon login

You will now be redirected to iCommons / CILogon's authentication page.

Copyright © The Trustees of Indiana University. All Rights Reserved.

UNIVERSITY OF EXETER

Sign In

Use Your Exeter login

You will now be redirected to Exeter's authentication page.

UNIVERSITY OF THE WILMINGTON

globus online

Manage Data | Catalog | News & Events | About | Support | Log In | Sign Up

Sign In

Use Your EG login

You will now be redirected to EG's authentication page.

Have you read the Privacy Policy and to be signed in to the beta Beta2.0 version? Yes No

MANAGE DATA
Data transfer
Data transfer activity
manage endpoints
Dashboard

OUR USERS
user guides & tools
User guides
papers
what others are saying

GROUPS
my groups
administrative groups

NEWS & EVENTS
Blog
events calendar

OUR COMMUNITY
support

HOW TO
quick start guides
tutorials
FAQs
videos & presentations

RESOURCES
globus content
documentation
API
release notes

IN PARTNERSHIP WITH
e2i

Subscribe to Globus News & Updates
Email Address
First Name

NERSC

Sign In

Use Your NERSC login

Use Your iCommons / CILogon login

You will now be redirected to iCommons / CILogon's authentication page.

ENERGY
Office of Science

BIRN ACCESS Group Management

Manage Data | Groups | Log In | Sign Up

ACCESS is a user and group management system used by the BIRN community to allow editing access to the BIRN sites.

Sign in with your Globus Online ID.

Are you working with a BIRN Working Group or user community and need editing access to the BIRN web? Sign up now.

BIRN Web 1
BIRN Web 2
BIRN Support

Username
Password

Forgot Your Password?
Reset an account? Sign Up Now

BIRN is supported by NSF grants ICS-0802776, ICS-0802768, ICS-0802770 and by the Information Technology Research Center grant ICS-0802761.



Thank you!