

Globus 2017: Beyond Big Data

GlobusWorld 2017 Keynote

Steve Tuecke





#globus2017

@globusonline



Thank you to our GlobusWorld sponsors



Support for student registrations provided by ESnet

Cloud transforms software, platform delivery

Software as a service: **SaaS**
(web & mobile apps)



NETFLIX



Platform as a service: **PaaS**



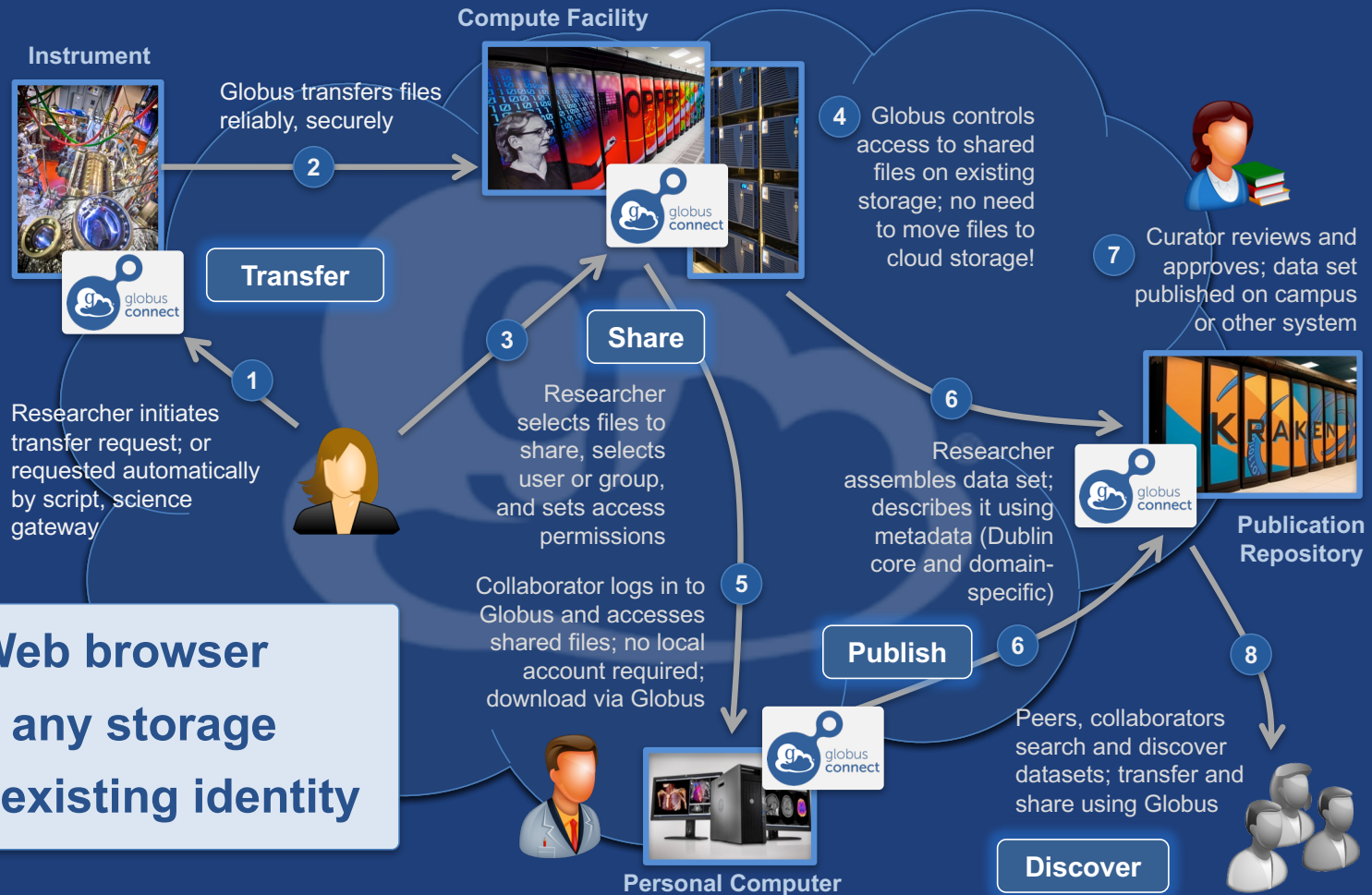
Infrastructure as a service: **IaaS**



PaaS enables more rapid, cheap, and scalable delivery of powerful (SaaS) apps



Globus SaaS: Research data lifecycle



- Use a Web browser
- Access any storage
- Use an existing identity



Globus by the numbers

48

most server endpoints on one campus

258 PB
transferred

43 billion
tasks processed

56,000
registered users

350

100TB+ users

10,000
active users

3 months

longest running managed transfer

10,000

active endpoints

300+

federated identities

1 PB

largest single transfer to date

5,119

active shared endpoints

99.5%

uptime



New features and enhancements



Storage connectors

- **Standard storage connectors (Posix)**

- Linux, Windows, MacOS
- Lustre, GPFS, OrangeFS, etc.

- **Premium storage connectors**

AWS S3

Ceph RadosGW (S3 API)

Spectra Logic BlackPearl

Google Drive

HPSS

HDFS (in progress)

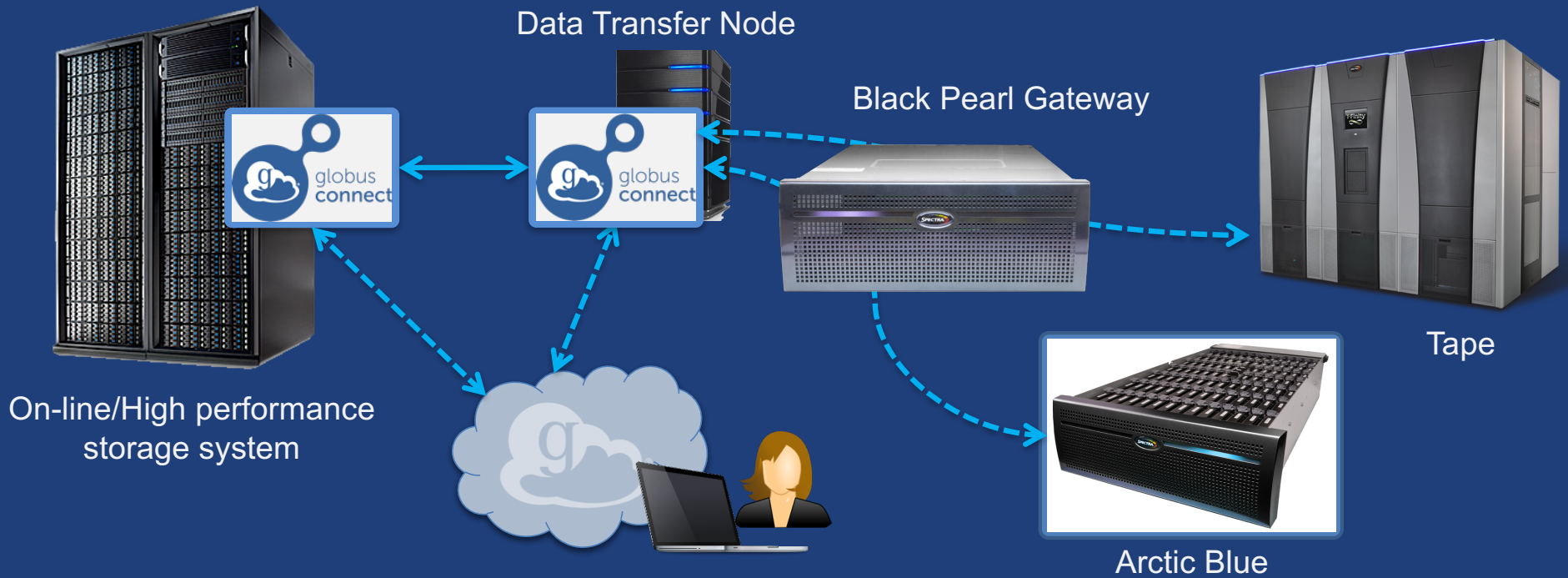
iRODS (in progress)

HGST Active Archive (in progress)

docs.globus.org/premium-storage-connectors



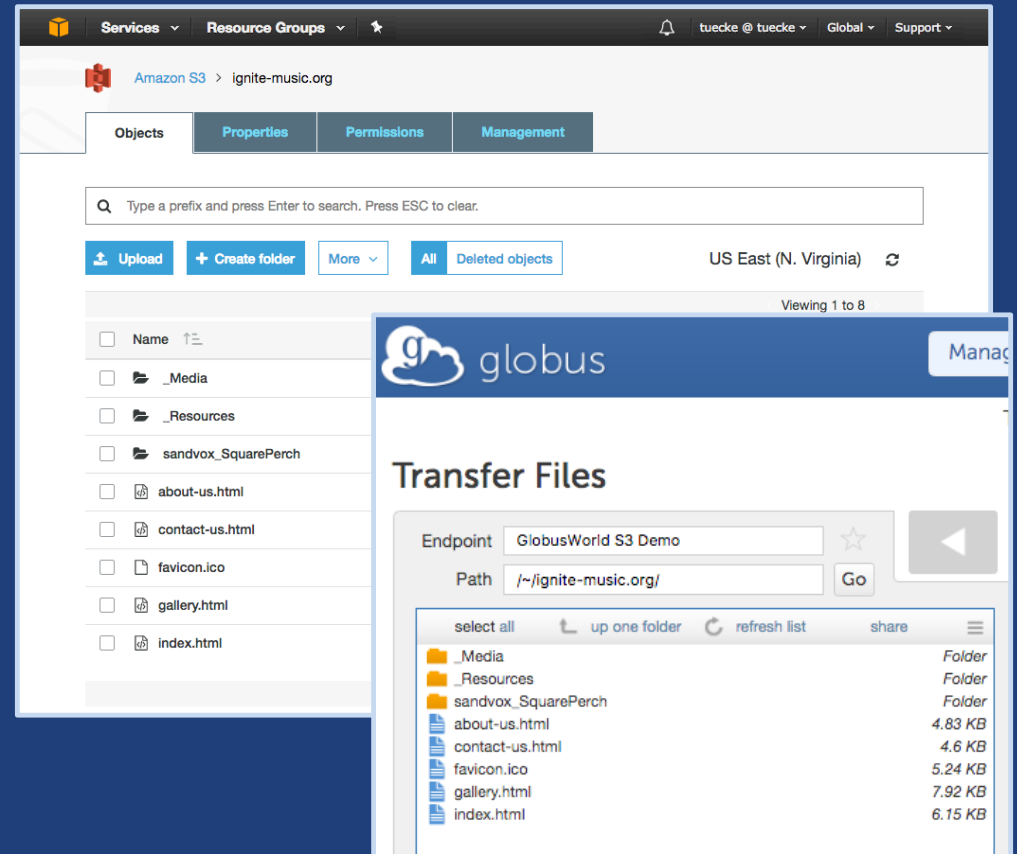
Spectra BlackPearl + Globus Solution



docs.globus.org/premium-storage-connectors/black-pearl/

AWS S3, Ceph RadosGW

- Now generally available
- Admin sets up credentials
- User self-registration coming



The screenshot displays the Globus interface for Amazon S3. The top navigation bar includes 'Services', 'Resource Groups', and user information. The main content area shows the 'Amazon S3' bucket for 'ignite-music.org'. Below the search bar, there are buttons for 'Upload', 'Create folder', and 'More'. A table lists the bucket's contents, including folders like '_Media', '_Resources', and 'sandvox_SquarePerch', and files like 'about-us.html', 'contact-us.html', 'favicon.ico', 'gallery.html', and 'index.html'. An inset window titled 'Transfer Files' shows the 'Endpoint' set to 'GlobusWorld S3 Demo' and the 'Path' as '/~/ignite-music.org/'. Below this, a list of files and folders is shown with their respective sizes.

Name	Type	Size
_Media	Folder	
_Resources	Folder	
sandvox_SquarePerch	Folder	
about-us.html	File	4.83 KB
contact-us.html	File	4.6 KB
favicon.ico	File	5.24 KB
gallery.html	File	7.92 KB
index.html	File	6.15 KB

docs.globus.org/premium-storage-connectors/aws-s3/

Google Drive

Connect Your Google Drive to Globus

Make the data on your Google Drive accessible as an endpoint on Globus

What do you want to make accessible as a Globus endpoint?

This access will be limited to you unless you decide to share your connected Google Drive with others by adding Globus sharing permissions to your endpoint.

Google Account: stevetuecke@gmail.com (with Globus Identity tuecke@globus.org)

Google Drive Folder: /My Drive

Globus endpoint information:
Endpoint Display Name: Google Drive stevetuecke@gmail.com

Transfer Files

Endpoint: Google Drive stevetuecke@gmail.com | Path: /My Drive/Demo Share/

Endpoint: UChicago RCC Midway | Path: /-/datasets/

Transfer Settings:

- sync - only transfer new or changed files
- delete files on destination that do not exist
- preserve source file modification times
- verify file integrity after transfer
- encrypt transfer

Google Drive

My Drive > Demo Share

Name	Owner	Last modified
Baltimore-Washington International Airport Climate Data	me	11:41 AM
experiment1	me	11:38 AM
experiment2	me	11:39 AM
LaGuardia Airport Climate Data	me	11:41 AM
Miami International Airport Climate Data	me	11:41 AM
myproject	me	Apr 7, 2017
1M.dat	me	Apr 7, 2017
9606sort.fn	me	11:40 AM
index.html	me	Apr 7, 2017
myproject.tar	me	11:42 AM

Manage Endpoints

Google Drive stevetuecke@gmail.com

Share With: user, all users

Identity/E-mail: ranantha@uchicago.edu

Permissions: read, write

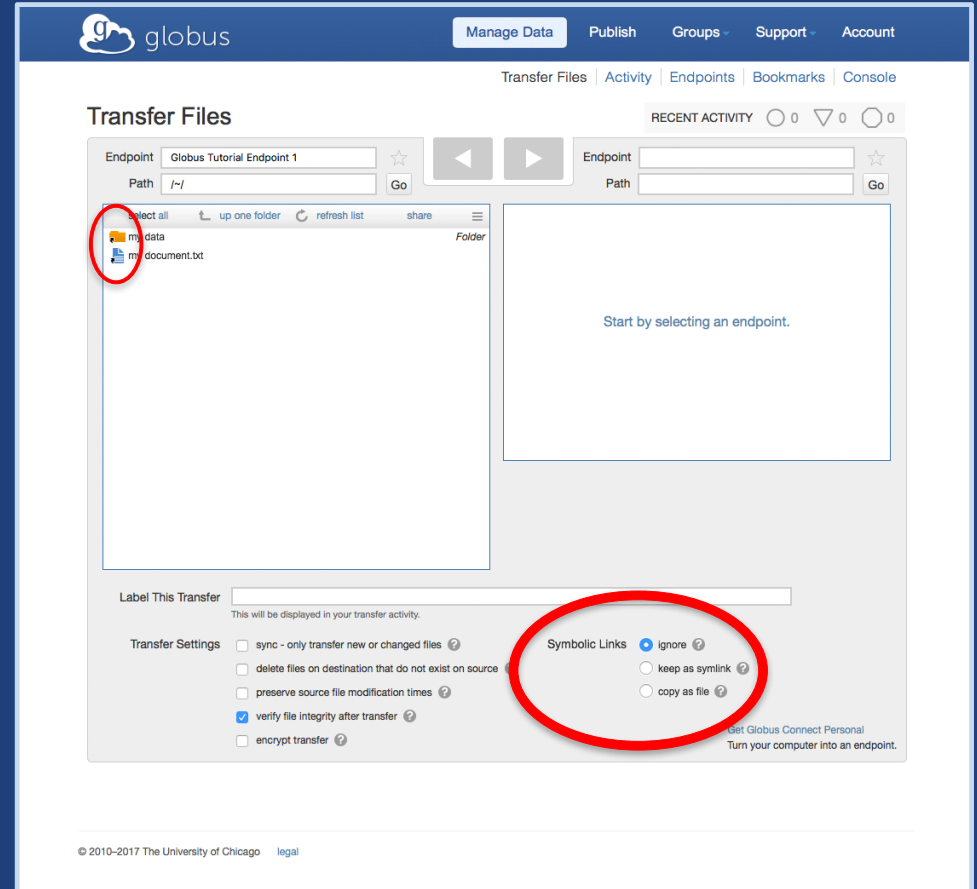
docs.globus.org/premium-storage-connectors/google-drive/



Demonstration:
Google Drive
AWS S3

Symbolic links

- On transfer and sync
- Options:
 - Ignore
 - Keep as symlinks
 - Copy as files
- Coming soon



The screenshot displays the Globus Transfer Files interface. At the top, there is a navigation bar with the Globus logo and links for Manage Data, Publish, Groups, Support, and Account. Below this, there are tabs for Transfer Files, Activity, Endpoints, Bookmarks, and Console. The main area is titled "Transfer Files" and shows a file browser for "Globus Tutorial Endpoint 1". The path is "/~/" and the file list includes a folder named "data" and a file named "document.txt". A red circle highlights the "data" folder. To the right, there is a large empty box with the text "Start by selecting an endpoint." Below the file browser, there is a "Label This Transfer" field and a "Transfer Settings" section. The "Symbolic Links" section is highlighted with a red circle and contains three radio button options: "ignore" (selected), "keep as symlink", and "copy as file". Other transfer settings include "sync - only transfer new or changed files", "delete files on destination that do not exist on source", "preserve source file modification times", "verify file integrity after transfer" (checked), and "encrypt transfer".

© 2010-2017 The University of Chicago [legal](#)



Globus Auth

- ORCID
- OpenID Connect compliance
- Email aliases at InCommon domains

Account

Identities

Primary Identity
tuecke@uchicago.edu

Linked Identities
0000-0003-2038-2512@orcid.org
stevetuecke@uchicago.edu
tuecke@globus.org
tuecke@globusid.org
tuecke@uchicago.edu@rda.ucar.edu
tuecke@xsede.org

The image shows two overlapping screenshots. The background screenshot is an ORCID authorization page. It features the ORCID logo and the text "Globus Auth has asked for the following access to your ORCID Record". Below this is a permission list with a checked box for "Allow this permission". A "Sign in with ORCID" button is visible. The foreground screenshot is a Globus login modal. It has a "globus" logo and "Globus Account Log In" text. The main heading is "Log in to use Globus Web App". It prompts the user to "Use your existing organizational login" and provides a dropdown menu currently set to "ORCID". Below the dropdown is a "Continue" button. At the bottom of the modal are "Sign in with Google" and "Sign in with ORCID ID" buttons. A "Deny" button is also visible at the bottom of the modal.



New Command Line Interface

- Transfer and Auth
- Replaces old SSH-based command line shell
- Uses Python SDK
- Open source

```
$ 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
  --jmespath, --jq TEXT  A JMESPath expression to apply to json output.
                        Takes precedence over any specified '--format' and
                        forces the format to be json processed by this
                        expression
  --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.
  delete        Submit a Delete Task
  endpoint      Manage Globus Endpoint definitions
  get-identities Lookup Globus Auth Identities
  list-commands List all CLI Commands
  login         Login to Globus to get credentials for the Globus CLI
  logout        Logout of the Globus CLI
  ls            List Endpoint directory contents
  mkdir         Make a directory on an Endpoint
  rename        Rename a file or directory on an Endpoint
  task          Manage asynchronous Tasks
  transfer      Submit a Transfer Task
  version       Show the version and exit
  whoami        Show the currently logged-in identity.
```

github.com/globus/globus-cli

docs.globus.org/cli

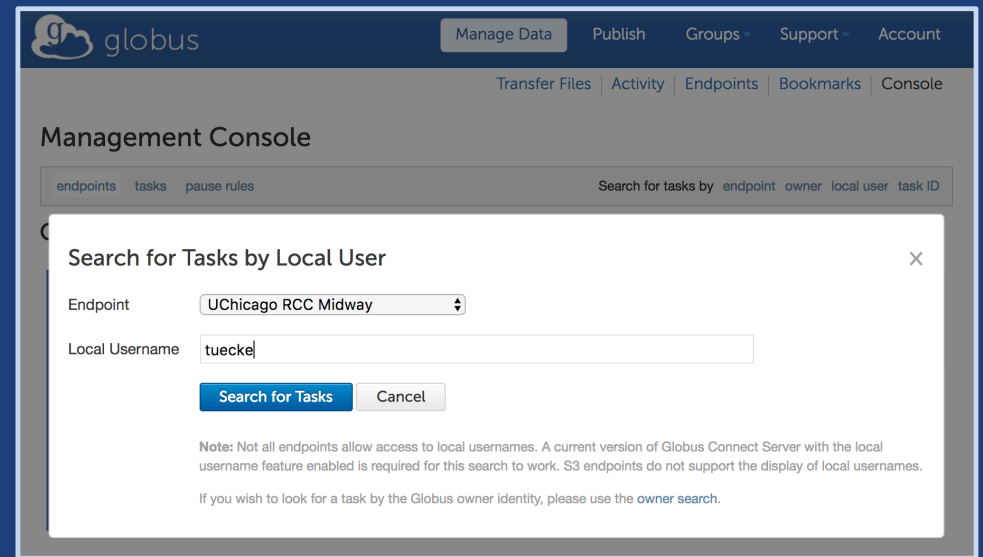


Demonstration:

Command Line Interface

Management Console

- **Updated user interface**
 - Shared endpoints
 - MC on shared endpoint
 - Host endpoint admin can see all shares & permissions
 - Search by local account



- **Self-help management console access via roles**
 - Activity manager
 - Activity monitor



Demonstration **Management Console**

**How do we empower the
research community to create an
integrated ecosystem of
services and applications?**

A busy year of training

- **GlobusWorld Tour:**
5 stops, over 200 people
- **Lots of open source
developer materials**
- **Contact us to host event**



Yale



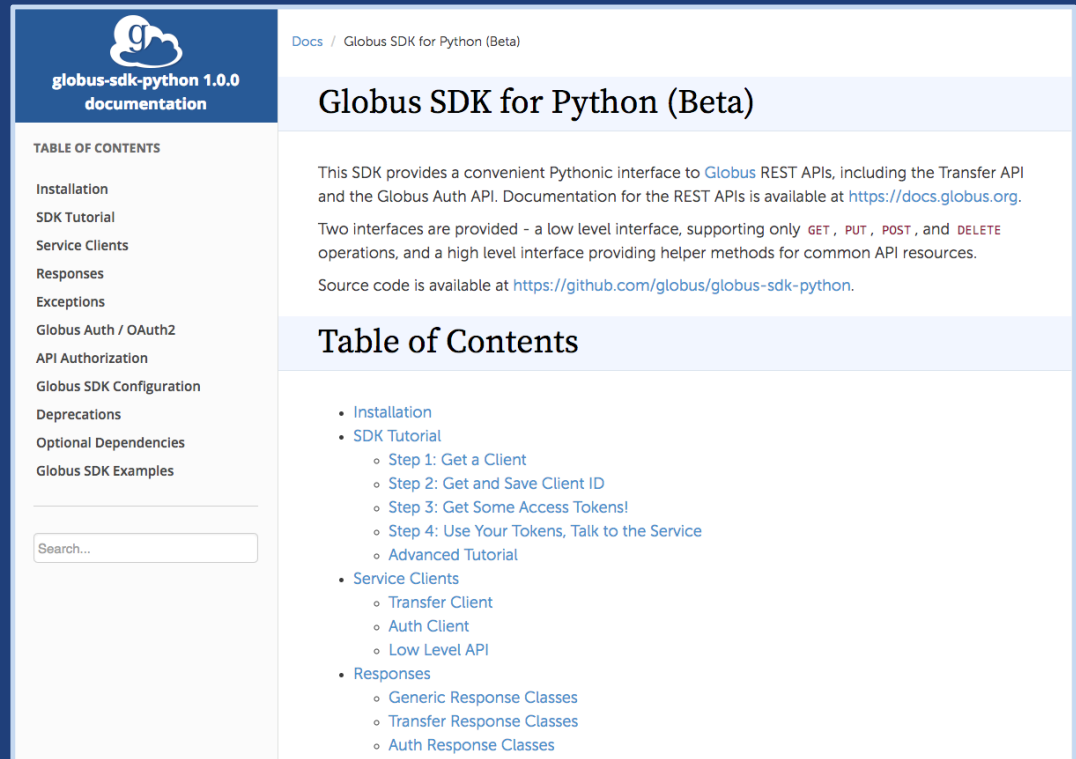
PennState
May 16, 2017



NEW YORK UNIVERSITY
May 31 – June 1, 2017

Globus Python SDK

- Python SDK for Globus PaaS
- Transfer and Auth APIs
- Open source



The screenshot shows the documentation page for the Globus SDK for Python (Beta). The page has a dark blue header with the Globus logo and the text "globus-sdk-python 1.0.0 documentation". Below the header is a "TABLE OF CONTENTS" section with a search bar. The main content area is titled "Globus SDK for Python (Beta)" and contains a paragraph describing the SDK's purpose and interfaces, followed by a "Table of Contents" section listing various topics.

Docs / Globus SDK for Python (Beta)

Globus SDK for Python (Beta)

This SDK provides a convenient Pythonic interface to [Globus REST APIs](#), including the Transfer API and the Globus Auth API. Documentation for the REST APIs is available at <https://docs.globus.org>.

Two interfaces are provided - a low level interface, supporting only `GET`, `PUT`, `POST`, and `DELETE` operations, and a high level interface providing helper methods for common API resources.

Source code is available at <https://github.com/globus/globus-sdk-python>.

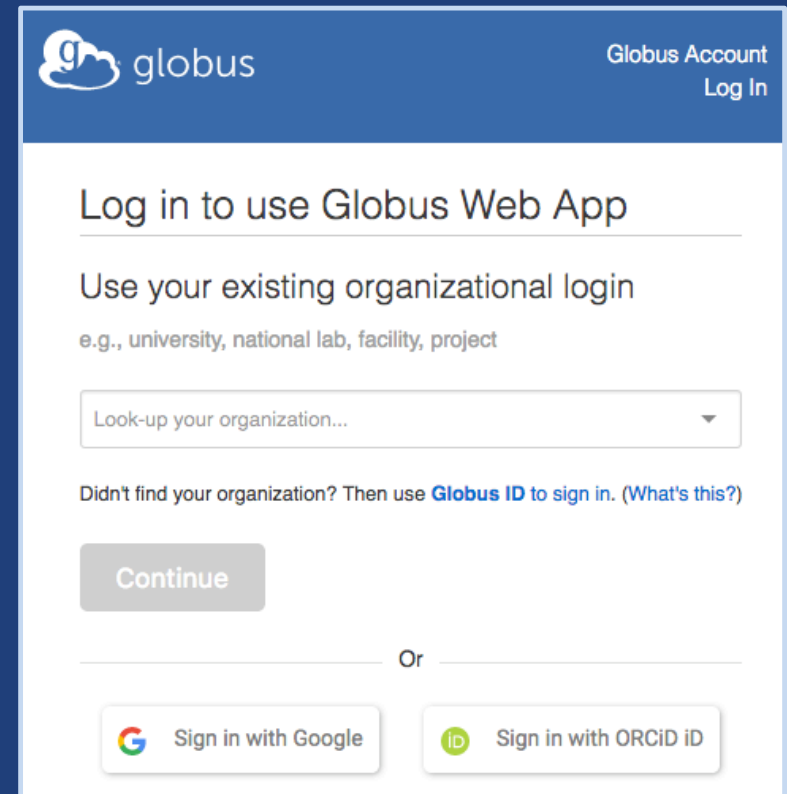
Table of Contents

- Installation
- SDK Tutorial
 - Step 1: Get a Client
 - Step 2: Get and Save Client ID
 - Step 3: Get Some Access Tokens!
 - Step 4: Use Your Tokens, Talk to the Service
 - Advanced Tutorial
- Service Clients
 - Transfer Client
 - Auth Client
 - Low Level API
- Responses
 - Generic Response Classes
 - Transfer Response Classes
 - Auth Response Classes

github.com/globus/globus-sdk-python

Globus Auth: Identity and Access Management PaaS

- **Federated identity login to apps**
 - Web, mobile, desktop, command line
- **Protect all REST API communications**



The screenshot shows the Globus Auth login page. At the top left is the Globus logo, and at the top right are the links "Globus Account" and "Log In". The main heading is "Log in to use Globus Web App". Below this, it says "Use your existing organizational login" with examples "e.g., university, national lab, facility, project". There is a dropdown menu labeled "Look-up your organization...". Below the dropdown is a link: "Didn't find your organization? Then use [Globus ID](#) to sign in. (What's this?)". A "Continue" button is positioned below the link. At the bottom, there are two buttons: "Sign in with Google" and "Sign in with ORCID iD", separated by the word "Or".



Globus PaaS: Federated Identity Login

XSEDE
Extreme Science and Engineering
Discovery Environment

Account

Jetstream Web App would like to:

Access all Jetstream resources

By clicking "Allow", you allow Jetstream to access your information and services. You can revoke access at any time.

FaceBase Data Browser Resources About Projects Help

Sign Up Log In

Selected by: Organism Human, Mouse, Zebrafish

Showing 1-25 of 653 results, sort by: Select an attribute View: [List View] [Grid View]

Download: CSV

Gene expression microarray - mouse E10.5 maxillary arch

Release Date: Steve Potter
Investigator: Steve Potter
Summary: Gene expression microarray using tissue obtained by laser capture microdissection (LCM) from the maxillary arch

Search within attributes...

CHOOSE ATTRIBUTES

- Organism
- Experiment Type
- Data Type
- Mouse Gene
- Mouse Age Stage
- Mouse Anatomic Source
- Mouse Genotype

XSEDE
Extreme Science and Engineering
Discovery Environment

Globus Account Log In

Log in to use XSEDE

Use your existing organizational login

e.g., university, national lab, facility, project

XSEDE
Extreme Science and Engineering
Discovery Environment

Globus Account Log In

Log in to use XSEDE Moodle Prod

Use your existing organizational login

e.g., university, national lab, facility, project

XSEDE

Didn't find your organization? Then use **Globus ID** to sign in. (What's this?)

XSEDE
Extreme Science and Engineering
Discovery Environment

Globus Account Log In

Log in to use XSEDE JIRA

Use your existing organizational login

e.g., university, national lab, facility, project

Look-up your organization...

Didn't find your organization? Then use **Globus ID** to sign in. (What's this?)

Or

KBase
PREDICTIVE BIOLOGY
DOE Systems Biology Knowledgebase

Home About News Developer Zone KBase Labs Contact Us

The new Systems Biology Knowledgebase (KBase) is a collaborative effort designed to accelerate our understanding of microbes, microbial communities, and plants. It will be a community-driven, extensible and scalable open-source software framework and application system. KBase will offer free and open access to data, models and simulations, enabling scientists and researchers to build new knowledge and share their findings.

Latest News

- KBase at International Plant and Animal Genome XXI
Posted by salazar Jan 09, 2013
- KBase Team at Argonne for November Build
Posted by riharris Nov 30, 2012
- November Build at Argonne
Posted by salazar Nov 23, 2012

What can GPCR Data Explorer

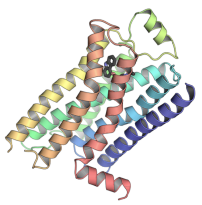
- ✓ C
- ✓ O
- ✓ U
- ✓ D
- ✓ M
- ✓ D
- ✓ E

Welcome to the GPCR Data Explorer. The GPCR Data Explorer contains data generated during the process of GPCR structure determination. These data will represent the entire process of GPCR structural biology including all associated constructs and related experimental data generated from three leading academic sites: the University of Southern California in Los Angeles, iHuman Institute at ShanghaiTech University, and Shanghai Institute of Materia Medica, a member of the Chinese Academy of Sciences.

Click one of the links below to start exploring GPCR data.

Search Targets: [USC](#) | [iHuman](#) | [SIMM](#) | [All](#)

Search Constructs: [USC](#) | [iHuman](#) | [SIMM](#) | [All](#)





Globus PaaS Usage (past 19 days)

4665

Non-Globus
IdP Logins

3829
Unique Identities

152 IdPs

2062

Non-Globus Webapp
Logins

1. Jetstream
2. Blue Waters
3. Compute Canada
4. GPCR
5. CIRM
6. XSEDE Moodle
7. KBase

Top 10 IdPs

XSEDE

Google



THE UNIVERSITY OF
CHICAGO



PURDUE
UNIVERSITY

JOHNS HOPKINS
UNIVERSITY

NCSA

ILLINOIS

UCLA

Globus PaaS In Action: Research Data Portal

The image shows two overlapping browser windows. The background window is the NCAR Research Data Archive website, displaying the 'NCEP Climate Forecast System Version 2 (CFSv2) Monthly Product' page. The foreground window is the 'Globus Download' portal from the Jet Propulsion Laboratory, showing instructions for downloading data via web browser or script.

NCAR Research Data Archive Website:

- Header: NCAR | Research Data Archive | Computational & Information Systems Lab
- Navigation: Home, Find Data, Ancillary Services, About/Contact, Data Citation, Web Services
- Page Title: NCEP Climate Forecast System Version 2 (CFSv2) Monthly Product ds094.2
- Buttons: Description, Data Access

Globus Download Portal:

- Header: Jet Propulsion Laboratory | California Institute of Technology
- Text: Welcome, Lukasz. | Register a New Project | My Profile | Log out
- Section: Globus Download
- Text: Globus uses the high performance, reliable GridFTP protocol to transfer many files at once from their remote archives to your desktop.
- Options:
 - Option 1: Web Download (Start Web Download)
 - Option 2: Script Download (Start Script Download)
- Text: After starting the file transfer request, please use the Globus site to monitor the download.
- List of files to be downloaded (e.g., /cmip5_css02_data/cmp5/output1/CMCC/CMCC-CM/historical/day/atmos/day/r11p1/clr/1/ct_day_CMCC-CM_historical_r11p1_19980101-19981231.nc)

Table from NCAR Website:

Data Description		Data File Downloads		Customizable Data Requests	Other Access Methods
		Web Server Holdings	Globus Transfer Service (GridFTP)	Subsetting	THREDDS Data Server
Union of Available Products		Web File Listing	Request Globus Invitation	Get a Subset	TDS Access
p r o d	Diurnal monthly means	Web File Listing		Get a Subset	
	Regular monthly means	Web File Listing		Get a Subset	GLADE File Listing HPSS File Listing



Globus PaaS In Action: Automation

DMagic ANL APS

DMagic latest

Search docs

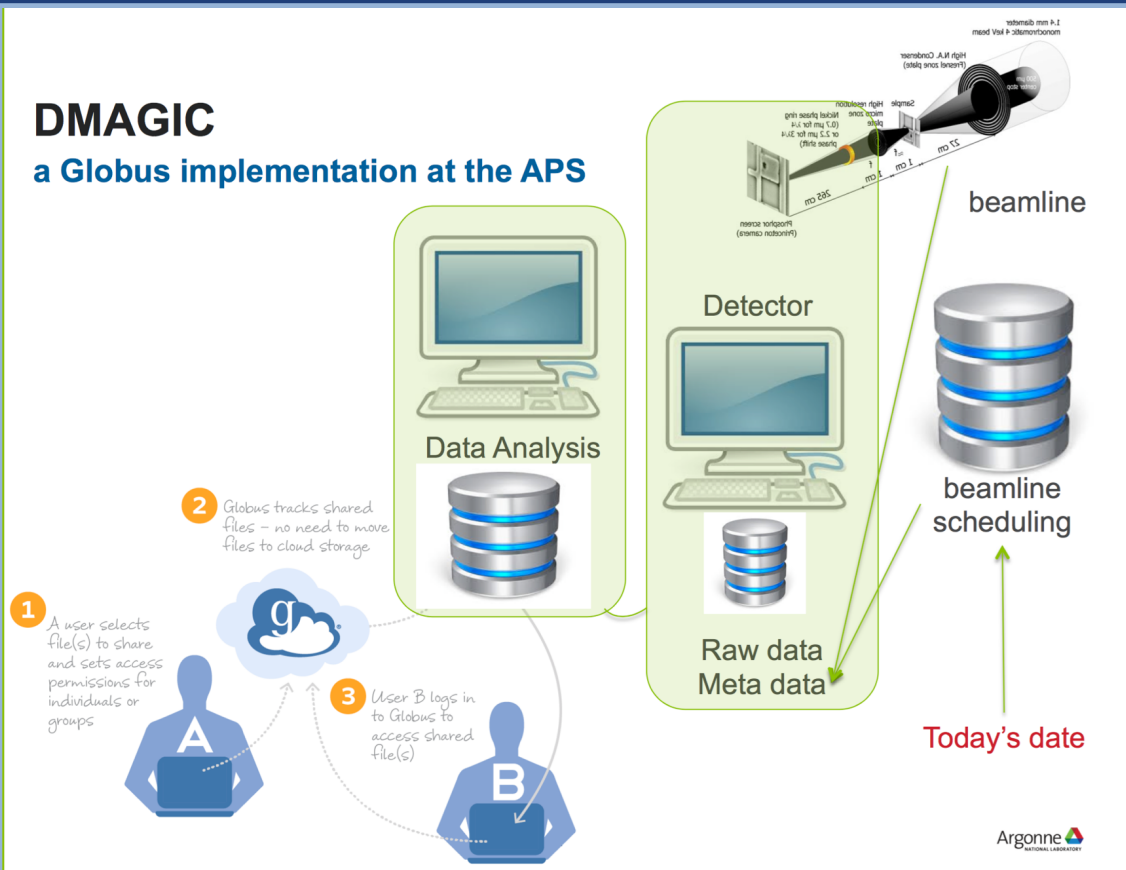
- About DMagic
- Install directions
- Development
- API reference
- Examples
- Frequently asked questions

Docs » DMagic [Edit on GitHub](#)

DMagic is an open-sourced Python toolbox to perform data management and data sharing for users of the Imaging Group of the Advanced Photon Source.

dmagic.readthedocs.io

DMAGIC a Globus implementation at the APS



Courtesy of Francesco De Carlo, Argonne National Laboratory (2016)



Globus PaaS In Action: Automation

Spot Suite Portal LBNL ALS

spot.nersc.gov

← → ↻ ⓘ Not Secure spot.nersc.gov ☆ ⓘ ⓘ




SPOT About ▾ Status & Tools ▾ Data Browser Simulation ▾ Username Password Sign in

Welcome to the SPOT Suite Portal for Light-Source Data and Simulations.

Log in with your NERSC account at the top right to gain access to data and NERSC compute resources.

This site is a work in progress. Please report any problems with the page to spot-help@lists.lbl.gov

[Learn more. >](#)



© LBL LDRD 2013 Please report any problems with the page to spot-help@lists.lbl.gov. [Web Policies](#).



Globus Auth: Native App Support

```
(globus) ubuntu@ip-172-31-58-24:~/globus/native-app-examples$ python example_copy_paste_refresh_token.py
Native App Authorization URL:
https://auth.globus.org/v2/oauth2/authorize?code_challenge=3tsuESy-QgRetkC_N_YutELskg0yvCbBGFYEnkThIFc&state=_default&redirect_uri=https%3A%2F%2Fauth.globus.org%2Fv2%2Fweb%2Fauth-code&response_type=code&client_id=079bdf4e-9666-4816-ac01-7eab9dc82b93&scope=openid+email+profile+urn%3Aglobus%3Aauth%3Ascope%3Atransfer.api.globus.org%3Aall&code_challenge_method=S256&access_type=offline
Enter the auth code:
```

```
(globus) ubuntu@ip-172-31-58-24:~/globus/native-app-examples$ python example_copy_paste_refresh_token.py
Native App Authorization URL:
https://auth.globus.org/v2/oauth2/authorize?code_challenge=00D4uKtJL8ZFFDjC&state=_default&redirect_uri=https%3A%2F%2Fauth.globus.org%2Fv2%2Fweb%2Fauth-code&response_type=code&client_id=079bdf4e-9666-4816-ac01-7eab9dc82b93&scope=openid+email+profile+urn%3Aglobus%3Aauth%3Ascope%3Atransfer.api.globus.org%3Aall&code_challenge_method=S256&access_type=offline
Enter the auth code: svnZ5oJ65AM6hbz50IzHj1izvDhvrJ

.globus/
shares/
.bash_logout
.bashrc
.profile

Doing a second directory listing with a new access token:
.globus/
shares/
.bash_logout
.bashrc
.profile
(globus) ubuntu@ip-172-31-58-24:~/globus/native-app-examples$ _
```

github.com/globus/native-app-examples

App registration

globus  developers

[Register your app with Globus](#)

[Developer documentation](#)

© 2010-2017 The University of Chicago [Legal](#)

developers.globus.org

App Registration

App Name

Scopes

Redirects
one per line
must be HTTPS

Required Identity Provider Require a specific identity provider

Privacy Policy

Terms & Conditions

Native App Will be used by a native application

For information on registering and using Apps with Globus Auth see the [documentation](#).

Demo MRDP on EC2

Client ID
06e6b39f-c18e-47b3-84fd-848ad9dca85

Redirect URLs
https://globus.computationinstitute.org:5000
https://globus.computationinstitute.org:5000/authcallback
https://globus.computationinstitute.org
https://globus.computationinstitute.org/authcallback
https://34.197.76.8:5000/authcallback
https://34.197.76.8:5000
https://52.72.26.74:5000
https://52.72.26.74:5000/authcallback

Scopes
no scopes requested for this client

Required Identity Provider
specific identity provider not required

Privacy Policy URL
none provided

Terms and Conditions URL
none provided

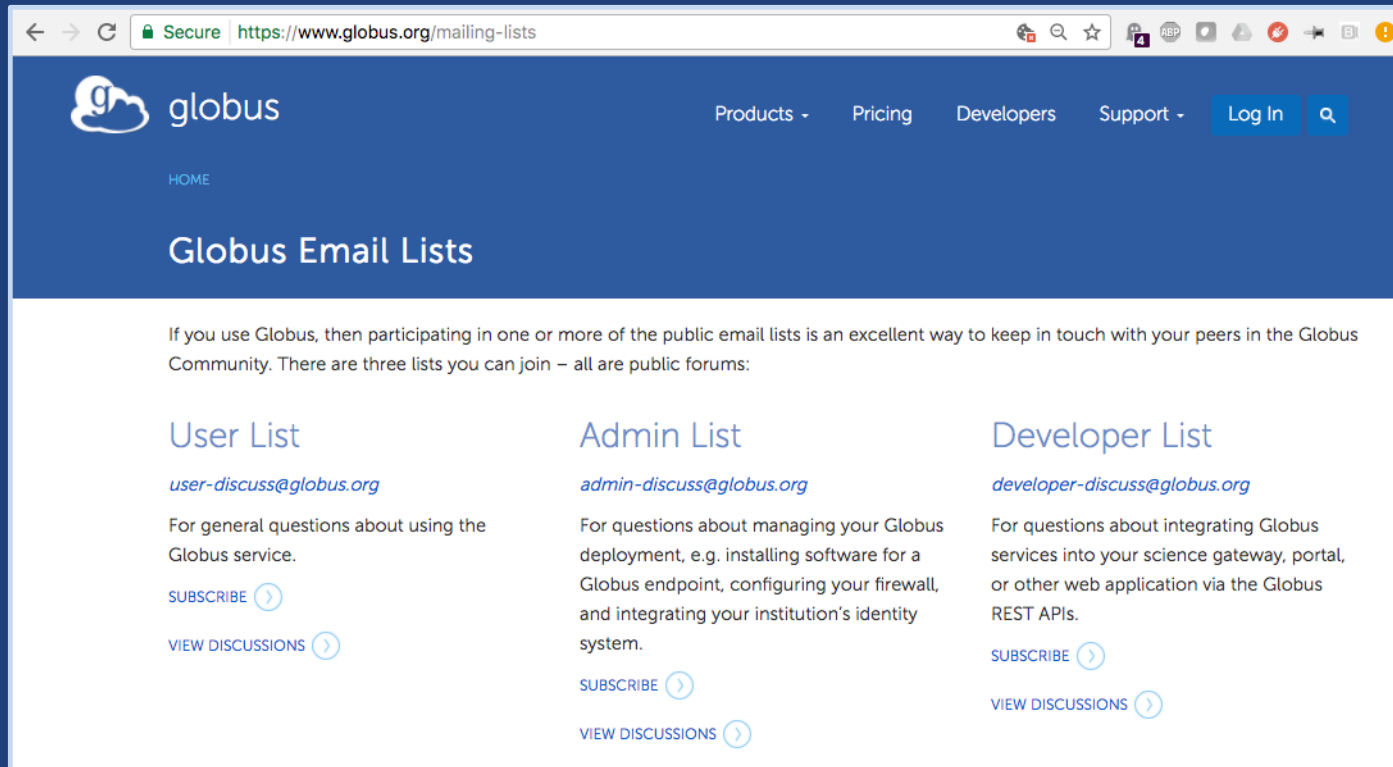
Client Identity Username
06e6b39f-c18e-47b3-84fd-848ad9dca85@clients.auth.globus.org

Client Secrets

Sample MRDP on EC2 accessed via Globus web site
created 6 months ago



Mailing lists



The screenshot shows a web browser window with the URL <https://www.globus.org/mailling-lists>. The page features a dark blue header with the Globus logo and navigation links for Products, Pricing, Developers, Support, and Log In. Below the header, the main content area is titled "Globus Email Lists" and includes a brief introduction: "If you use Globus, then participating in one or more of the public email lists is an excellent way to keep in touch with your peers in the Globus Community. There are three lists you can join – all are public forums:". Three columns describe the lists: "User List" (for general questions, email: user-discuss@globus.org), "Admin List" (for deployment and configuration questions, email: admin-discuss@globus.org), and "Developer List" (for integration questions, email: developer-discuss@globus.org). Each column includes "SUBSCRIBE" and "VIEW DISCUSSIONS" buttons with right-pointing arrows.

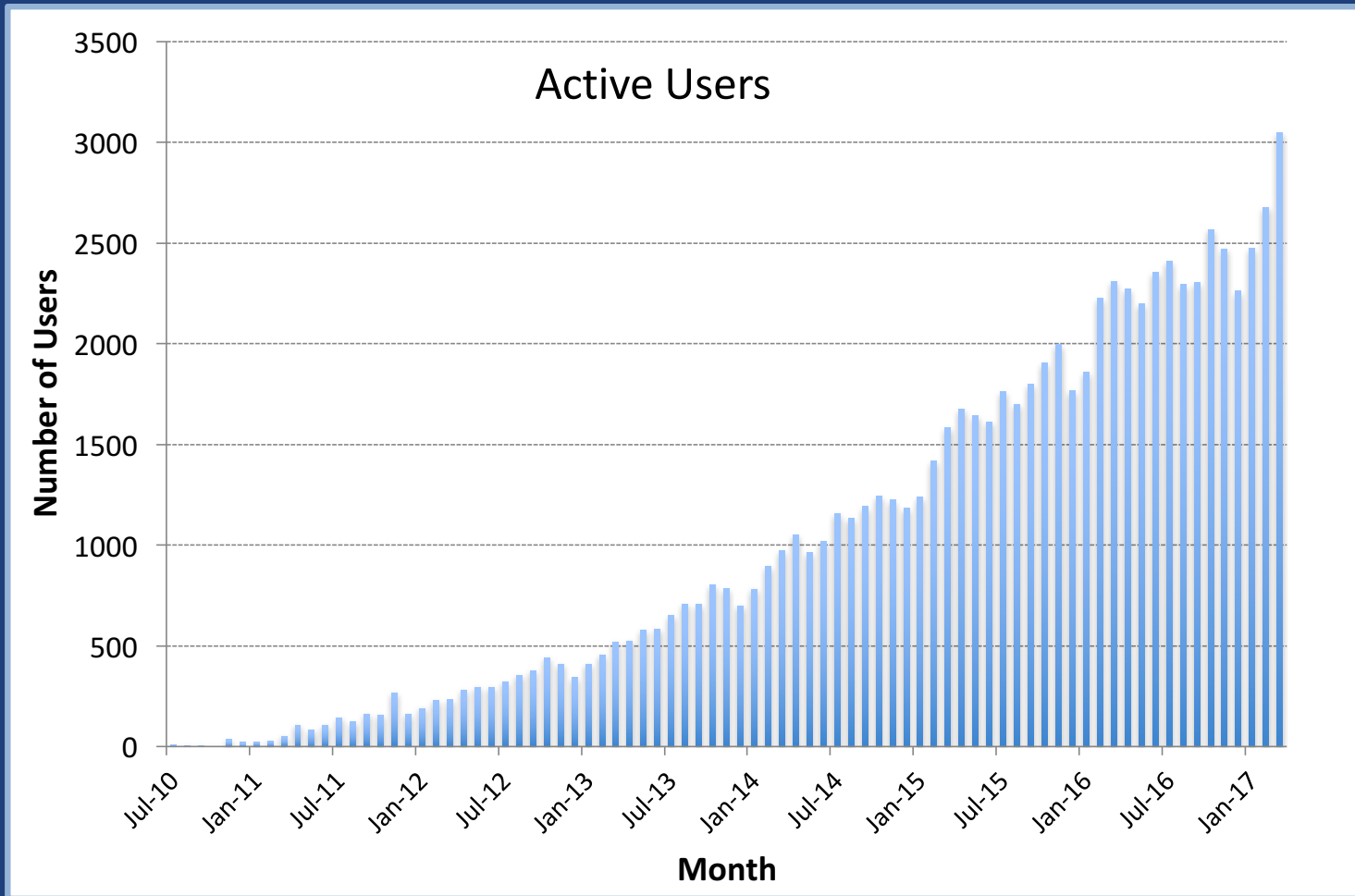
[globus.org/mailling-lists](https://www.globus.org/mailling-lists)



Sustainability Update

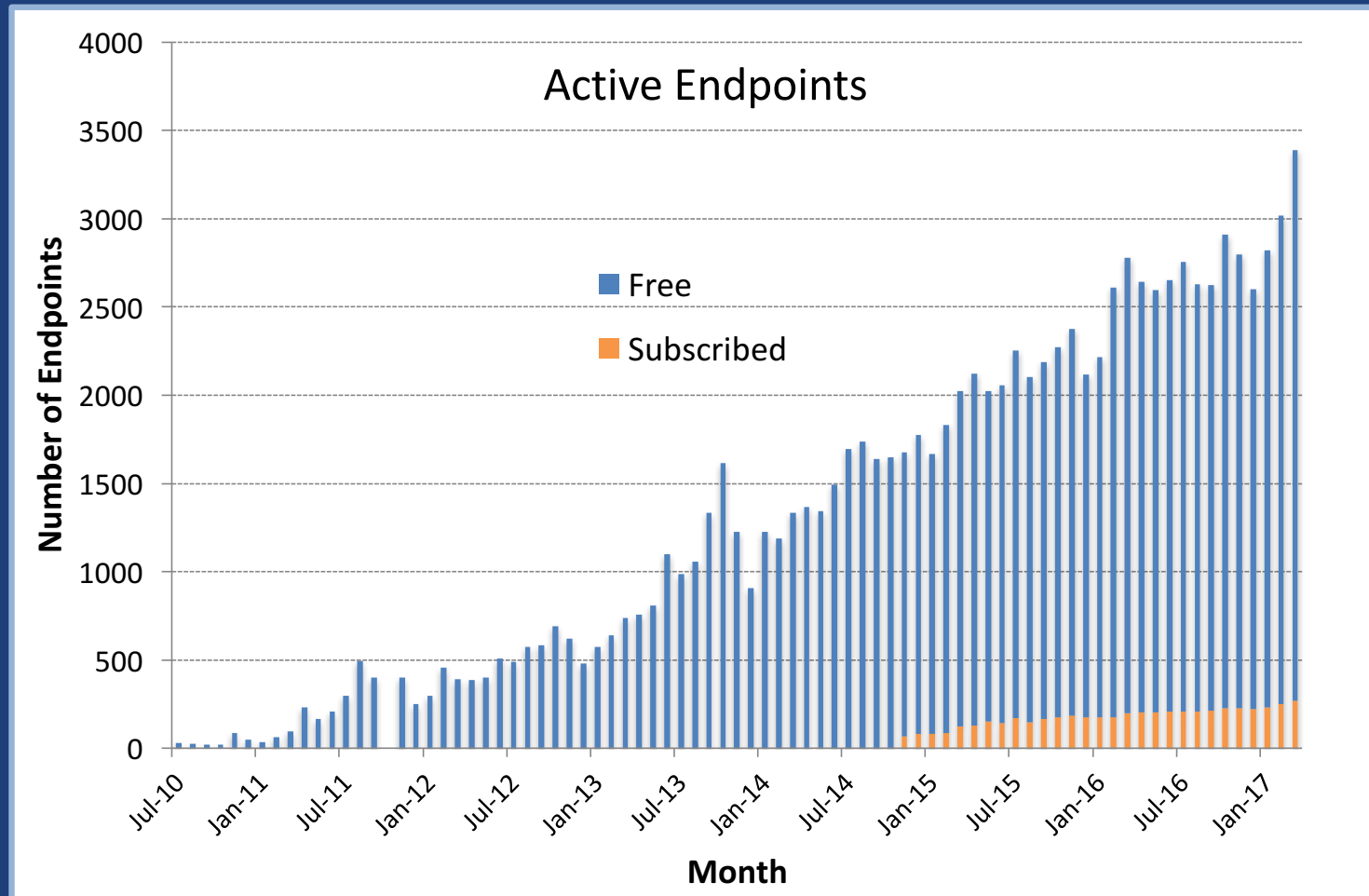


User/usage growth trend continues





Subscription growth slow and steady



THANK YOU, subscribers!



NEW YORK UNIVERSITY



THE UNIVERSITY OF CHICAGO



MICHIGAN STATE UNIVERSITY



Dartmouth

SIMONS FOUNDATION





Greatly simplified subscription model helps

- **Moved from usage based to unlimited use model**
 - Encourage campus wide adoption
 - Remove uncertainty from budgeting process
- **Pricing based on Carnegie Classification**
 - Proxy for scope/scale of research supported by Globus
 - Transparency and equity across broad range of institutions
- **Multi-year ramp-up, as necessary**



Our vision for the future





Mission

Increase the efficiency and effectiveness of researchers engaged in data-driven science, scholarship, and policy

Globus Auth: Identity and Access Management PaaS


- **Planned enhancements:**
 - Access to attributes from IdPs
 - Support more InCommon and eduGAIN identity providers by relaxing ePPN requirements
 - Level of Assurance policy support

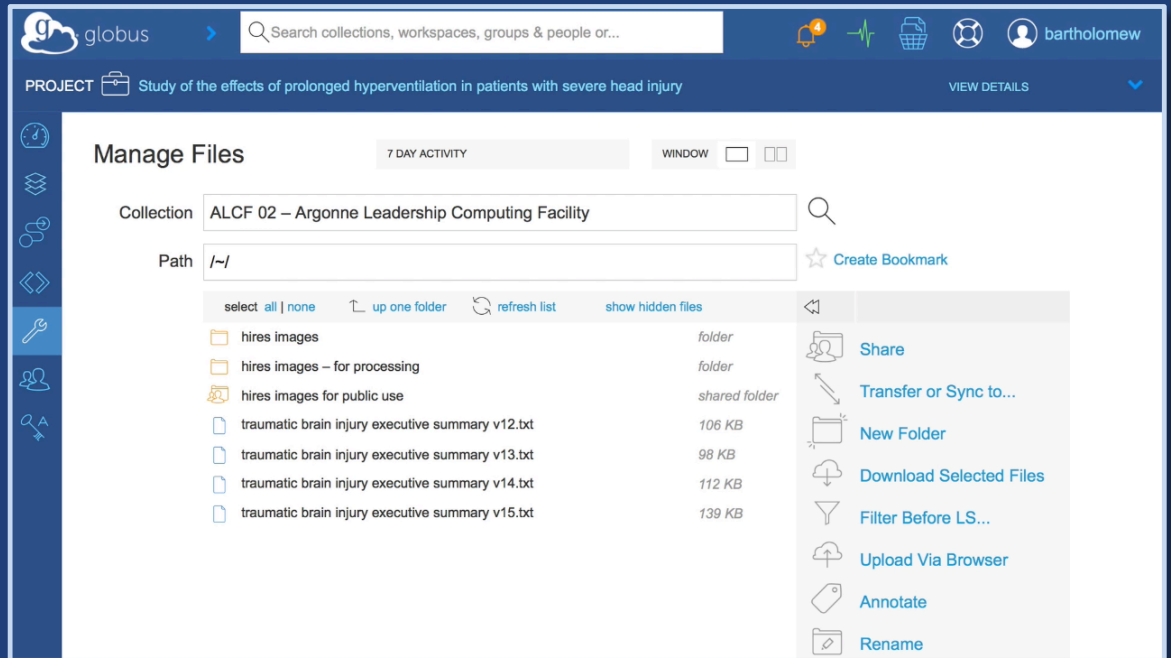


The screenshot shows the Globus InCommon login interface. At the top left is the Globus logo, and at the top right are the links "Globus Account" and "Log In". The main heading is "InCommon" in a large, bold, blue font. Below it, the text reads "Log into use Globus Web App". A large blue plus sign is positioned above the text "Use your existing organizational login" and "e.g., university, national lab, facility, project". Below this is a search box containing the text "Library, your organization". The "eduGAIN" logo is prominently displayed in the center. Below the search box, a note says "Didn't find your organization? Then use Globus ID to sign in. (What's this?)". A "Continue" button is located below the search box. At the bottom, there are two buttons: "Sign in with Google" and "Sign in with ORCID iD".

= 1,500+ IdPs

Discover, access, and manage all of your data

- **New interface** 
- **Complete file management for all data**
- **File upload, download, view, edit**
- **Storage connectors**



The screenshot displays the Globus 'Manage Files' interface. At the top, there's a search bar and a user profile for 'bartholomew'. Below that, the project name 'Study of the effects of prolonged hyperventilation in patients with severe head injury' is shown. The main area is titled 'Manage Files' and shows the current collection as 'ALCF 02 - Argonne Leadership Computing Facility' and the path as '/-/'. A table lists the following items:

Item Name	Type	Size
hires images	folder	
hires images - for processing	folder	
hires images for public use	shared folder	
traumatic brain injury executive summary v12.txt	file	106 KB
traumatic brain injury executive summary v13.txt	file	98 KB
traumatic brain injury executive summary v14.txt	file	112 KB
traumatic brain injury executive summary v15.txt	file	139 KB

On the right side, a menu of actions is available: Share, Transfer or Sync to..., New Folder, Download Selected Files, Filter Before LS..., Upload Via Browser, Annotate, and Rename.



Collections: The evolution of endpoints

globus > Search collections, workspaces, groups & people or...

PROJECT Study of the effects of prolonged hyperventilation in patients with severe head injury VIEW DETAILS

Manage Files

7 DAY ACTIVITY WINDOW

Collection ALCF 02 – Argonne Leadership Computing Facility

Path Create Bookmark

select all | none up one folder refresh list show hidden files

	hires images	folder	
	hires images – for processing	folder	
	hires images for public use	shared folder	
	traumatic brain injury executive summary v12.txt	106 KB	
	traumatic brain injury executive summary v13.txt	98 KB	
	traumatic brain injury executive summary v14.txt	112 KB	
	traumatic brain injury executive summary v15.txt	139 KB	

- Share
- Transfer or Sync to...
- New Folder
- Download Selected Files
- Filter Before LS...
- Upload Via Browser
- Annotate
- Rename



Collection properties

- **Set of blobs (files), hierarchically named (folders)**
- **Rooted at a unique DNS name**
- **URL referenceable files, folders**
- **Accessible and manageable via:**
 - HTTPS: client/server file access
 - GridFTP: async bulk transfer
 - REST API: advanced operations
- **OAuth2 authentication and authorization via Globus Auth**
- **Collection-specific access policies**
- **Data is stored on a storage system, which determines storage policies such as durability and availability**
- **File change events** ⚡



Globus Connect Server v5

- **Collection model**
- **HTTPS access to storage**
- **Security improvements**
 - OAuth2 in GridFTP (no more X.509 user certificates or Myproxy!)
 - OpenID Connect identity provider
 - Credential expiration LoA policies
 - User credential management (e.g., for Google Drive, S3, Kerberos)
- **Kerberos protected file systems**
- **Directory listing with path expressions**



Globus Connect Server v5 install & config


- **Setup with any identity (GlobusID not required)**
- **Automatable installation and configuration**
- **Configuration API, CLI, GUI**
- **Scale-out deployment without shared file system**
- **Multiple storage systems simultaneously**
- **Single port GridFTP (no ephemeral ports)**
- **Distributed as Docker containers**

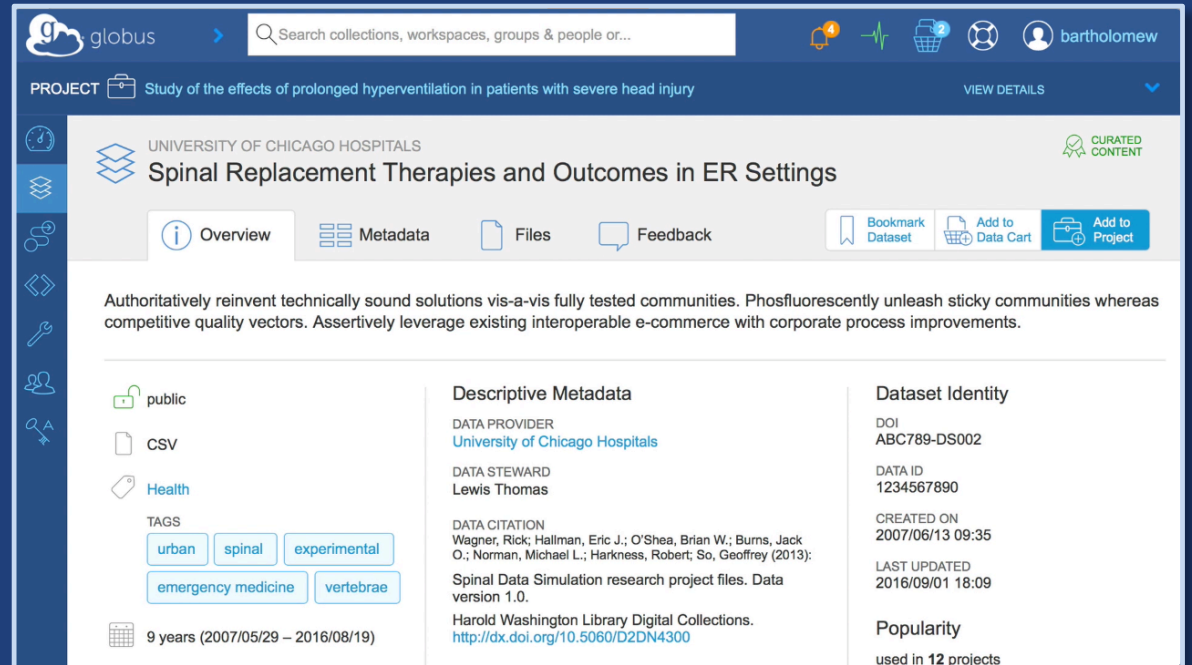


Share Data

- **Remove friction of sharing**
 - Guest collections where possible, e.g., Google Drive
 - Hybrid collections: Mapped access to home & project folders, else guest access
- **Enhanced sharing permissions**
 - permission expiration
 - permissions on files (not just folders)
 - sharing via URL possession
- **Storage connectors: share from anywhere**

Metadata

- Shine a light on dark data
- Unlock a broad  range of possibilities for new discovery and reproducibility



The screenshot shows the Globus interface for a project titled "Spinal Replacement Therapies and Outcomes in ER Settings" from the University of Chicago Hospitals. The page includes a search bar, navigation tabs for Overview, Metadata, Files, and Feedback, and a "CURATED CONTENT" badge. The main content area features a descriptive paragraph, a list of tags (public, CSV, Health), a "TAGS" section with buttons for "urban", "spinal", "experimental", "emergency medicine", and "vertebrae", and a date range of "9 years (2007/05/29 – 2016/08/19)". The "Descriptive Metadata" section lists the data provider as "University of Chicago Hospitals", the data steward as "Lewis Thomas", and provides a data citation and a DOI link. The "Dataset Identity" section shows the DOI "ABC789-DS002", data ID "1234567890", creation date "2007/06/13 09:35", and last updated date "2016/09/01 18:09". The "Popularity" section indicates the dataset is "used in 12 projects".

globus Search collections, workspaces, groups & people or...

PROJECT Study of the effects of prolonged hyperventilation in patients with severe head injury VIEW DETAILS

UNIVERSITY OF CHICAGO HOSPITALS Spinal Replacement Therapies and Outcomes in ER Settings CURATED CONTENT

Overview Metadata Files Feedback Bookmark Dataset Add to Data Cart Add to Project

Authoritatively reinvent technically sound solutions vis-a-vis fully tested communities. Phosphorescently unleash sticky communities whereas competitive quality vectors. Assertively leverage existing interoperable e-commerce with corporate process improvements.

public CSV Health

TAGS urban spinal experimental emergency medicine vertebrae

9 years (2007/05/29 – 2016/08/19)

Descriptive Metadata

DATA PROVIDER
University of Chicago Hospitals

DATA STEWARD
Lewis Thomas

DATA CITATION
Wagner, Rick; Hallman, Eric J.; O'Shea, Brian W.; Burns, Jack O.; Norman, Michael L.; Harkness, Robert; So, Geoffrey (2013):
Spinal Data Simulation research project files. Data version 1.0.
Harold Washington Library Digital Collections.
<http://dx.doi.org/10.5060/D2DN4300>

Dataset Identity

DOI
ABC789-DS002

DATA ID
1234567890

CREATED ON
2007/06/13 09:35

LAST UPDATED
2016/09/01 18:09

Popularity

used in 12 projects

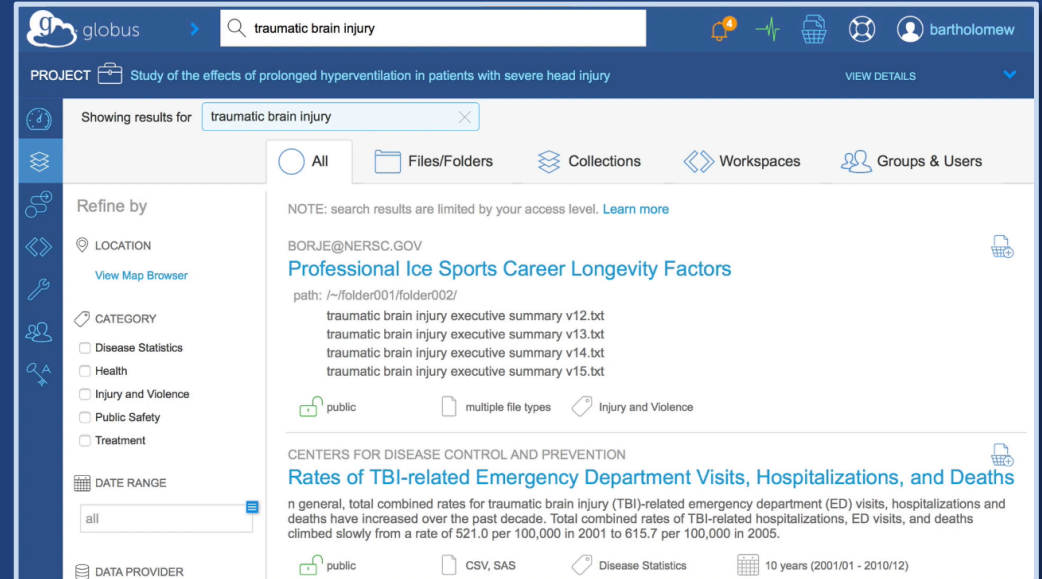


Give us your unwashed metadata...

- **Schema agnostic for broad applicability**
 - Leverage schemas where possible
 - But allow complete flexibility and evolution
- **Capture early and often**
 - Using interactive forms and automated tools
- **Don't get hung up on database tech**
 - Store metadata in files alongside data
 - ETL into useful formats and tools for use
- **Access control to metadata**
 - Independent of access control to data

Search

- **All data, irrespective of location**
 - Collections on Globus enabled storage
 - Non-Globus repositories
- **Access control aware**
 - Return results based on metadata that user is allowed to see
- **Flexible metadata indexing, schema agnostic**
- **Search and explore**
 - Plaintext search, facets, related, ratings, reviews, comments



The screenshot displays the Globus search interface. At the top, the search bar contains the query "traumatic brain injury". Below the search bar, the interface shows a list of search results. The first result is from "BORJE@NERSC.GOV" and is titled "Professional Ice Sports Career Longevity Factors". It lists several text files: "traumatic brain injury executive summary v12.txt", "traumatic brain injury executive summary v13.txt", "traumatic brain injury executive summary v14.txt", and "traumatic brain injury executive summary v15.txt". The second result is from the "CENTERS FOR DISEASE CONTROL AND PREVENTION" and is titled "Rates of TBI-related Emergency Department Visits, Hospitalizations, and Deaths". It includes a summary paragraph: "In general, total combined rates for traumatic brain injury (TBI)-related emergency department (ED) visits, hospitalizations and deaths have increased over the past decade. Total combined rates of TBI-related hospitalizations, ED visits, and deaths climbed slowly from a rate of 521.0 per 100,000 in 2001 to 615.7 per 100,000 in 2005." The interface also features a left sidebar with "Refine by" options such as "LOCATION", "CATEGORY", "DATE RANGE", and "DATA PROVIDER".



Demonstration **Search**

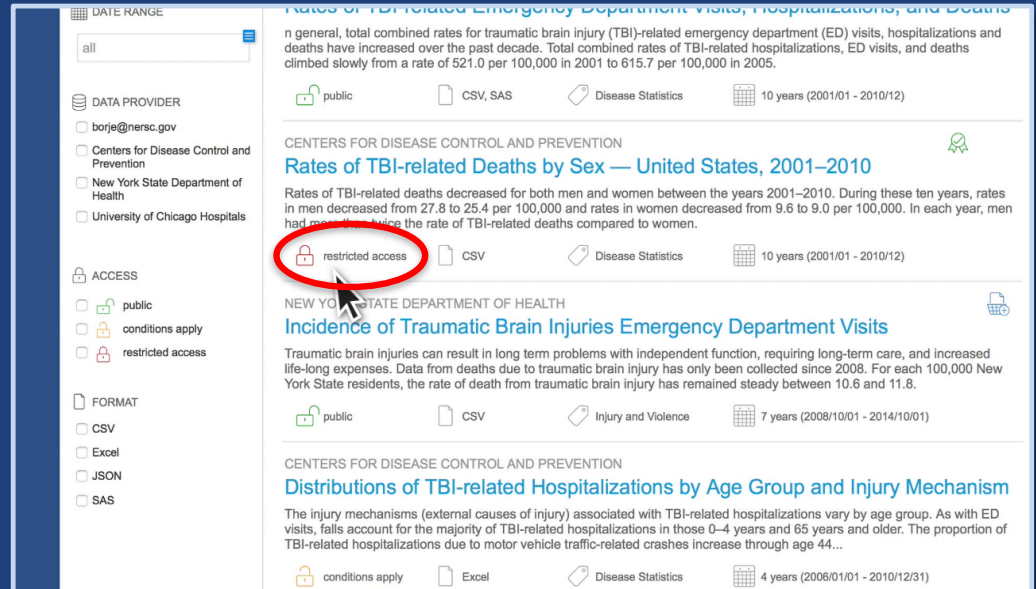


Search PaaS

- **Create custom indexes**
 - Deep indexing of science domain specific metadata
 - Custom facets
- **Publish metadata to multiple indexes**
 - Globus managed, general-purpose indexes
 - Community or project indexes
 - Individual lab or researcher indexes
- **Rich query interface**
 - Partial matches, fuzzy matches, range queries, filters, facets, ranking, boosting
 - Across one or more indexes
 - Access controlled
- **Embed Globus Search from your own apps and services**

Data stewardship

- Data steward manages access to restricted data
- Stewardship service manages workflows for:
 - Requesting access to data
 - Reviewing access proposals
 - Approve or deny requests
 - Export request and review
 - Reporting and audit



DATE RANGE
all

DATA PROVIDER
 borje@nersc.gov
 Centers for Disease Control and Prevention
 New York State Department of Health
 University of Chicago Hospitals

ACCESS
 public
 conditions apply
 restricted access

FORMAT
 CSV
 Excel
 JSON
 SAS

Rates of TBI-related Emergency Department Visits, Hospitalizations, and Deaths
In general, total combined rates for traumatic brain injury (TBI)-related emergency department (ED) visits, hospitalizations and deaths have increased over the past decade. Total combined rates of TBI-related hospitalizations, ED visits, and deaths climbed slowly from a rate of 521.0 per 100,000 in 2001 to 615.7 per 100,000 in 2005.
public CSV, SAS Disease Statistics 10 years (2001/01 - 2010/12)

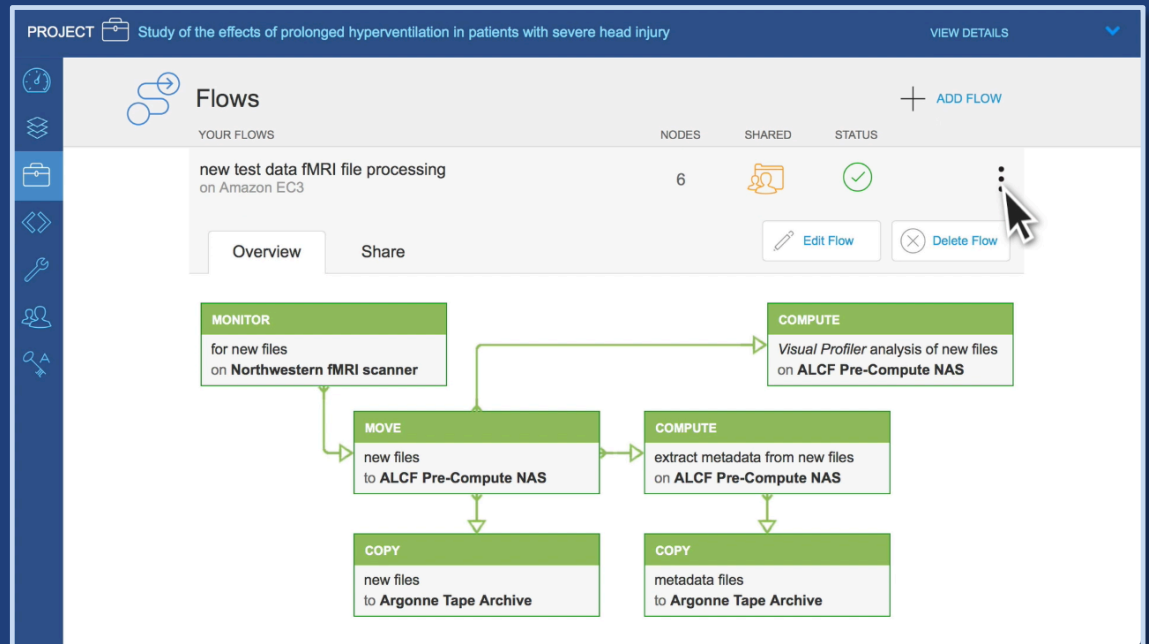
CENTERS FOR DISEASE CONTROL AND PREVENTION
Rates of TBI-related Deaths by Sex — United States, 2001–2010
Rates of TBI-related deaths decreased for both men and women between the years 2001–2010. During these ten years, rates in men decreased from 27.8 to 25.4 per 100,000 and rates in women decreased from 9.6 to 9.0 per 100,000. In each year, men had more than twice the rate of TBI-related deaths compared to women.
restricted access CSV Disease Statistics 10 years (2001/01 - 2010/12)

NEW YORK STATE DEPARTMENT OF HEALTH
Incidence of Traumatic Brain Injuries Emergency Department Visits
Traumatic brain injuries can result in long term problems with independent function, requiring long-term care, and increased life-long expenses. Data from deaths due to traumatic brain injury has only been collected since 2008. For each 100,000 New York State residents, the rate of death from traumatic brain injury has remained steady between 10.6 and 11.8.
public CSV Injury and Violence 7 years (2008/10/01 - 2014/10/01)

CENTERS FOR DISEASE CONTROL AND PREVENTION
Distributions of TBI-related Hospitalizations by Age Group and Injury Mechanism
The injury mechanisms (external causes of injury) associated with TBI-related hospitalizations vary by age group. As with ED visits, falls account for the majority of TBI-related hospitalizations in those 0–4 years and 65 years and older. The proportion of TBI-related hospitalizations due to motor vehicle traffic-related crashes increase through age 44...
conditions apply Excel Disease Statistics 4 years (2006/01/01 - 2010/12/31)

Flows

- Create event driven actions that can be chained together to automate tasks
- Data replication and migration, metadata extraction, analysis pipelines, *et al* ⚡





Data Publication: SaaS to PaaS

- **Monolithic publication flow**
→ **Composable services**
 - Collections
 - Flows
 - Persistent Identifiers
 - Metadata editor
 - Curation
- **Common usage scenarios**
 - Turn-key data publication
 - Integrate into institutional repositories
 - Custom flows throughout the research lifecycle

The screenshot shows the Globus Data Publication web interface. At the top, there is a navigation bar with the Globus logo, a search bar, and links for 'Manage Data', 'Publish', 'Groups', 'Support', and 'Account'. Below the navigation bar, there are links for 'Browse & Discover', 'Data Publication Dashboard', and 'Communities & Collections'. A search bar is prominently displayed. The main content area is titled 'Globus Data Publication' and includes a brief description of the service and a link to learn more. Below this, there are three columns of data: 'Communities', 'Discover', and 'Subject'. The 'Communities' column lists various data repositories like BD2K Data Repository and Materials Data Facility. The 'Discover' column lists authors such as GGCM, McInerney, D. J., and Moyer, E. J. The 'Subject' column lists topics like Agricultural Impacts, Climate Change, and CO2 effects. Each item in these columns has a small circular icon with a number inside, likely representing the number of publications or datasets associated with that item.

Communities	Discover	Subject
BD2K Data Repository Big Data to Knowledge Centers	GGCM 57	Agricultural Impacts 57
Globus	McInerney, D. J. 35	Climate Change 57
Materials Data Facility Community for the Materials Data Facility Collaboration	Moyer, E. J. 35	Climate Impacts 57
National Data Service	Sun, S. 35	CO2 effects 57
RDCEP Center for Robust Decision Making on Climate and Energy Policy	Moyer, E.J. 18	Farm system models 57
	Schwarzwald, K. 18	Food Security 57
	Zhorin, V. 18	Model Intercomparison 57
	Voorhees, Peter W. 9	climate 18
	Fife, Julie L. 5	cmip5 18



Near-term priorities

- **Globus Connect Server v5**
- **Premium storage connectors**
- **Collections**
- **New web interface**
- **Search PaaS**
- **Publication PaaS**
- **Compliance (HIPAA, FedRAMP)**

 Thank you to our sponsors



U.S. DEPARTMENT OF
ENERGY

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



THE UNIVERSITY OF
CHICAGO



Argonne
NATIONAL LABORATORY



powered by
amazon
web services





Thank you to our GlobusWorld sponsors



Support for student registrations provided by ESnet



Program Preview

- **Today**

- Lightning talks
- Globus Labs
- Cocktail reception

globusworld.org/program

- **Tomorrow**

- Introductory and advanced admin tutorials
- Developer tutorials
- Reception and poster session with NDS



#globus2017

@globusonline