

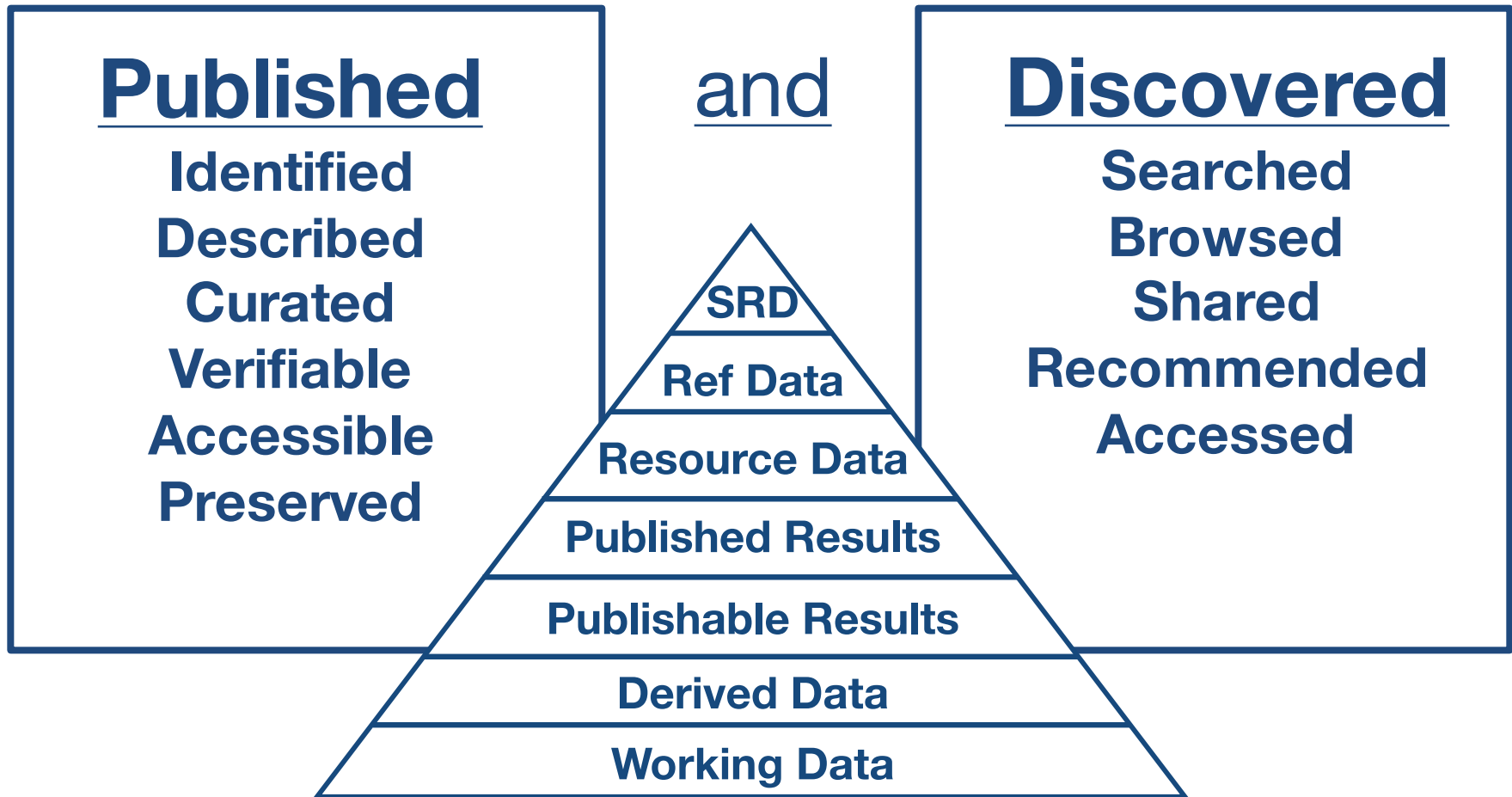
The Materials Data Facility

**Ben Blaiszik (blaiszik@uchicago.edu),
Kyle Chard (chard@uchicago.edu)
Ian Foster (foster@uchicago.edu)**

materialsdatafacility.org

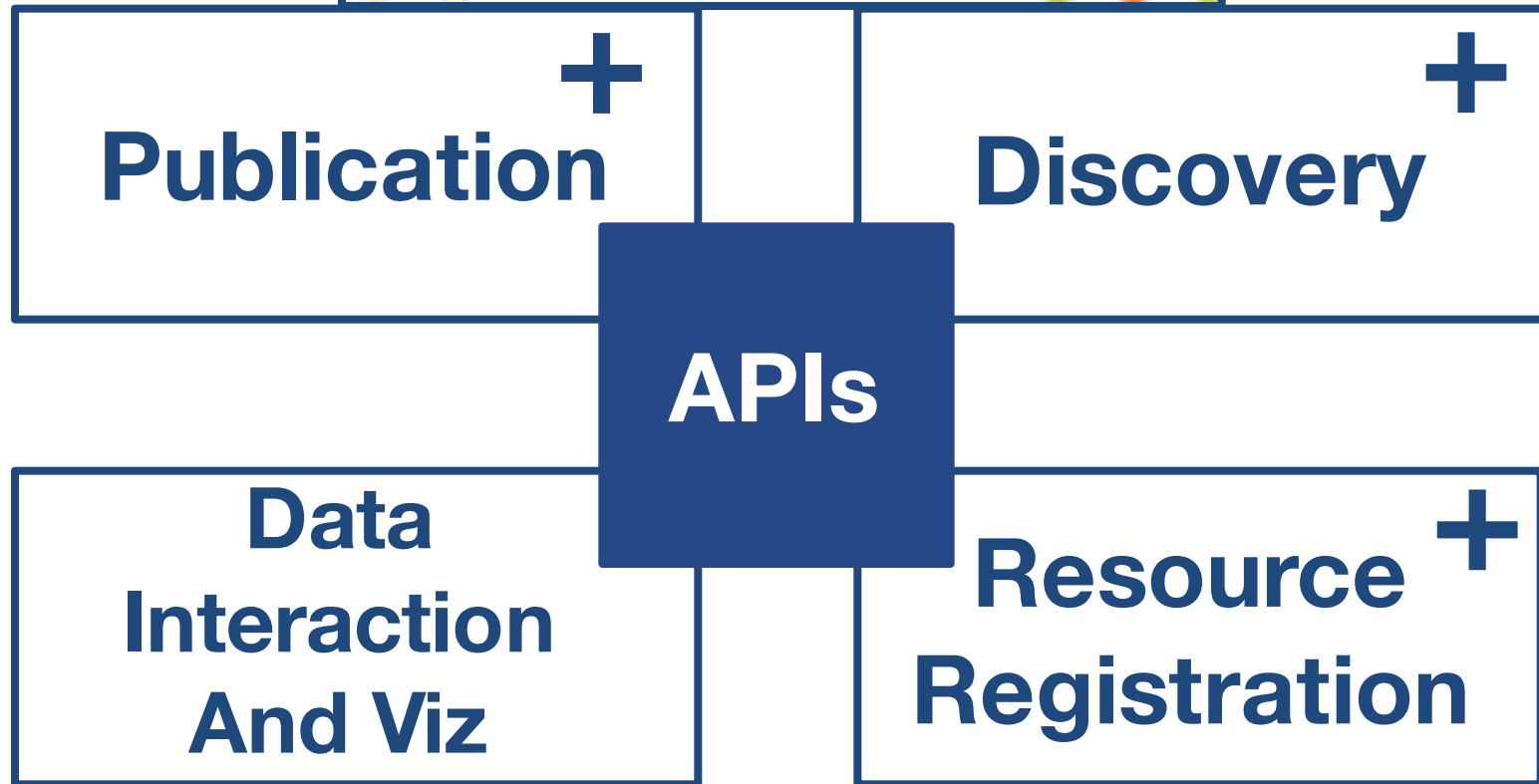
What is MDF?

We aim to make it simple for materials datasets and resources to be ...



**What infrastructure do
we need to effectively
support materials
researchers?**

Service Infrastructure



+ - Initial Foci

Publication


Deployed Nov. 2015

- Identify datasets with persistent identifiers (e.g. DOI)
- Describe datasets with appropriate metadata, and provenance
- Curate dataset metadata and data composition
- Verify dataset contents over time
- Preserve critical datasets in a state that increases transparency, replicability, and helps encourage reuse

Discovery

Coming late 2016-ish

- Search and query datasets in modern ways – e.g. via indexed metadata rather than remembering file paths
- Discover distributed materials resources (more later)



The screenshot shows a search interface for 'MDF' under the context 'TMS-2016-MDF'. The interface includes a search bar, a 'TOP HIT' section, and a list of folders and documents. The folders listed are 'mdf', 'MDF - Desktop', 'MDF - Google Drive', 'MDF - git', and 'mdf2iso'. The documents listed are '20151208-NCSA-PIRE-MDF', 'EZIDOrderForm-mdf', '20151006 - MDF - MGI Review - A...', 'BuildingMDF-bb', 'BuildingMDF', and 'BuildingMDF-2.docx'. The main content area displays a document titled 'The Materials Data Facility - Data Services to Advance Materials Science Research' by I. Foster, R. Ananthakrishnan, B. Blaiszik, K. Chard, J. Pruyne, J. Towns, and S. Tuecke. The document text describes the Materials Data Facility (MDF) and its goals, including enabling publication of large research datasets, granting the ability to publish data directly from local storage, building extensible domain-specific metadata, and developing publication workflows.

Future...

Spotlight for all data you have access to regardless of location

Resource Registration

Coming Q1 2016
via NIST

- Find existing, widely distributed, materials resources
- Register new resources into the network

The screenshot shows the NIST Materials Resource Registry dashboard. At the top left is the NIST logo. The navigation bar includes links for Home, Services, Dashboard (which is underlined), Help, Contact, and API. Below the navigation bar is a large black banner with the text "Materials Resource Registry" in white. Underneath the banner, there is a section titled "Select a Resource type to add" in blue text. To the right of this section is a button labeled "My Resources" with a small icon of a folder. Below the text, there are six icons representing different resource types: Organization (a blue building icon), Data Collection (a green and black calendar icon), Dataset (a black grid icon), Service (an orange hexagon with "API" inside), Informational (a blue laptop icon), and Software (a green mouse icon). Each icon is positioned above its corresponding label.

NIST

Home Services » Dashboard Help Contact API

Materials Resource Registry

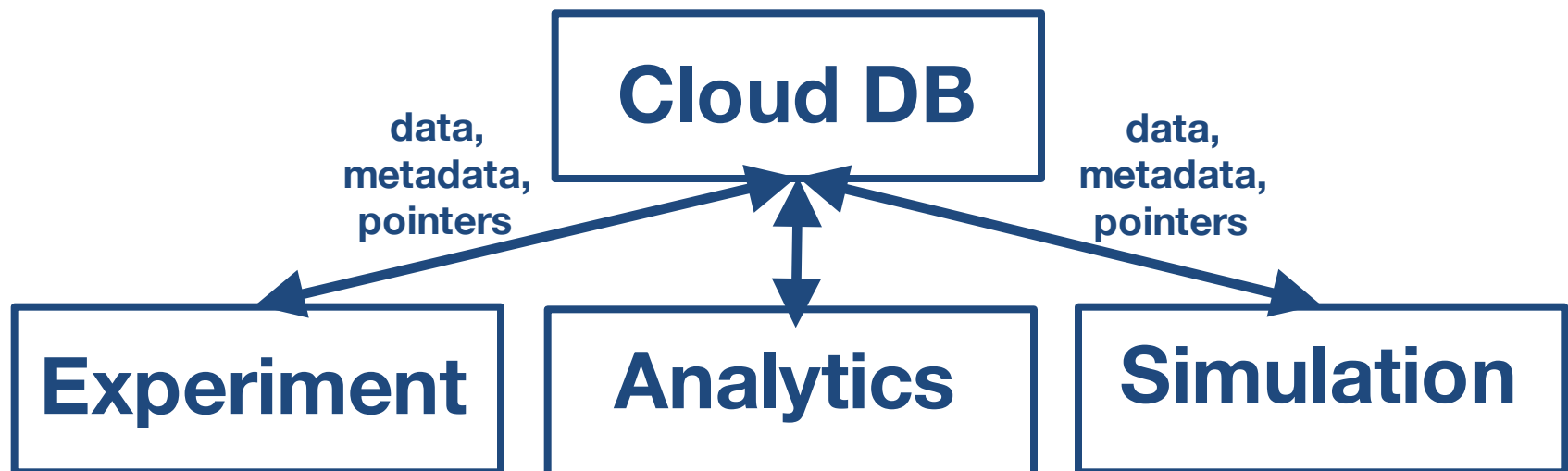
Select a Resource type to add

My Resources

Organization Data Collection Dataset Service Informational Software

Data Interaction And Viz

- Data-driven experiments using HPC resources and workflow technologies
- Real-time interaction with data regardless of data location (pending appropriate data access) and data size
- (future) Machine learning across datasets and storage locations
- (future) Automated discovery support



Understanding Incentives is Critical

Increasing Impact

- Increase paper citations¹
- Add dataset citation capabilities

Meeting Award Requirements

- Simplify DMP compliance

Smoothing Dislocations

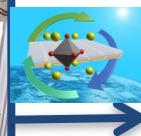
- Enable simple sharing among collaborators (near and far)
- Ease transitions between students
- Lessen need for *ad hoc* resource sharing (e.g. via group websites)

¹ Citation increase 30 (10.7717/peerj.175) - 60% (10.1371/journal.pone.0000308) [caveat bio research]

**So where are
we now?**

Publication

Materials Data Publication/Discovery is Often a Challenge

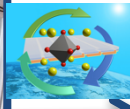


Materials Data Publication/Discovery is Often a Challenge

Want to Publish



Want to Discover / Use



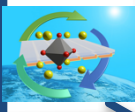
- Need networked storage, sometimes many TB ?
- Need to uniquely identify data for search/cite ?
- Need custom metadata descriptions ?
- Need a data curation workflow
- Need automation capabilities

Materials Data Publication/Discovery is Often a Challenge

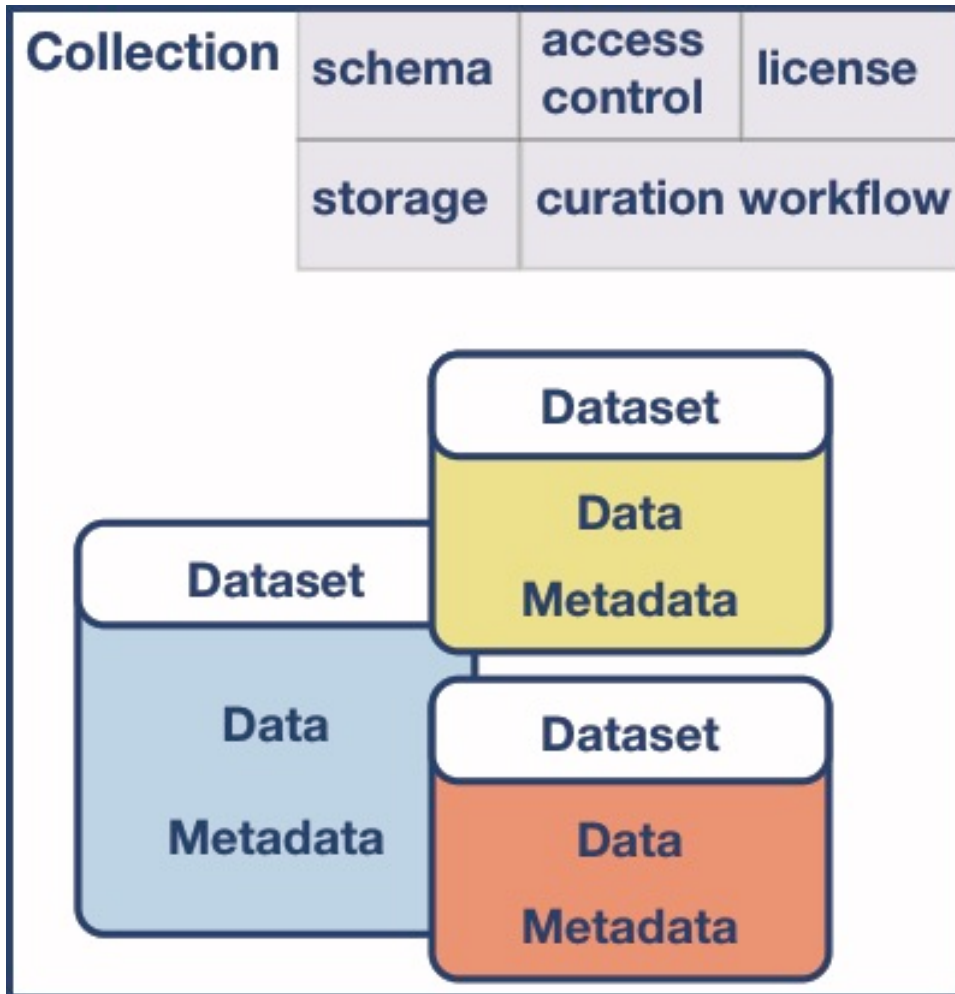
Want to Publish



Want to Discover / Use



Collection Model



- **Collections might be a research group or a research topic...**

- **Collections have specified**

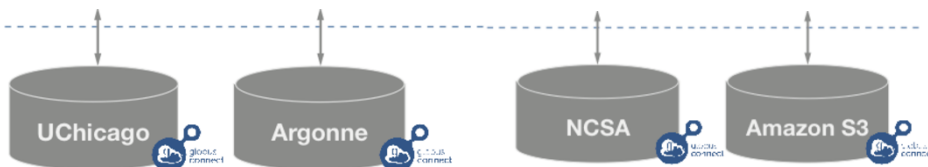
- Mapping to storage endpoint
 - Currently handled as automatically created shared endpoints
- Metadata schemas
- Access control policies
- Licenses
- Curation workflows

- **Collections contain**

- Datasets
 - Data
 - Metadata

- **Metadata Persistence**

- Metadata log file with dataset
- Metadata replicated in search index



Publish Large Datasets

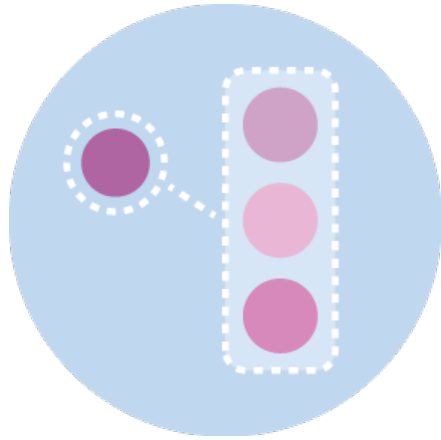


- Leverages Globus production capabilities for file transfer, user authentication, and groups

127,289,944,744 MB
TRANSFERRED

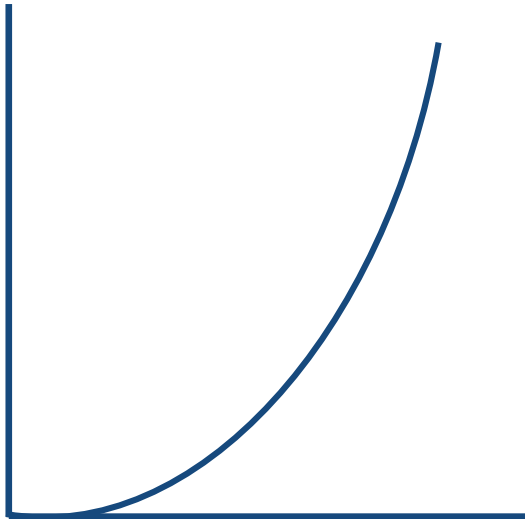
- **100 TB of reliable storage @ NCSA, and more storage at Argonne**
 - Globus endpoint at `nca#mdf`
 - Expandable to PBs as necessary
 - Automated tape backup for reliability (in progress)
- **Optionally use your own local or institutional storage**

Uniquely Identify Datasets



- **Associate a unique identifier with a dataset**
 - DOI, Handle
- **Improve dataset discovery and citability**
 - Aligning incentives and understanding the culture will be critical to driving adoption

Dataset Downloads



Time

Future...

- Your work has been cited 153 times in the last year
- Researchers from 30 institutions have downloaded your datasets

Share Data with Flexible ACLs



- **Share data publicly, with a set of users, or keep data private**

Leverage Curation Workflows



- **Collection administrators can specify the level of curation workflow required for a given collection e.g.**
 - **No curation**
 - **Curation of metadata only**
 - **Curation of metadata and files**

Customize Metadata



- **Build a custom metadata schema for your specific research data**
- **Re-use existing metadata schemas**
- **Working in conjunction with NIST researchers to define these schemas**

Future...

- **Can we build a system that allows schema:**
 - **Inheritance**
 - E.g. a schema “polymers” might inherit and expand upon the “base material” of NIST
 - **Versioning**
 - E.g. Understand contextually how to map fields between versions
 - **Dependence**
 - E.g. Allows the ability to build consensus around schemas

Discover Research Datasets



- **Search on file metadata, custom metadata, and indexed file-level data**
- **Goal: Intuitive search (e.g. Google-style) with support for more complex range queries and faceting (e.g. Amazon-style)**

MaterialsDataFacility.org

The screenshot shows the Materials Data Facility website. At the top, there is a navigation bar with the logo on the left and links for 'ABOUT', 'GET STARTED', 'FEATURES', and 'HOW IT WORKS' on the right. Below the navigation bar is a large blue banner featuring a central diagram of a folder icon surrounded by various scientific symbols (molecules, crystals, etc.) connected by lines. Underneath this diagram is the heading 'WHAT IS MDF?' followed by a paragraph explaining that the Materials Data Facility (MDF) is a scalable repository for materials scientists to publish, preserve, and share research data. It also mentions that the repository provides a focal point for the materials community and is a pilot project funded by NIST. Below this text is a dark blue bar with logos for 'Funded and supported by NIST' (National Institute of Standards and Technology, U.S. Department of Commerce) and 'CHIMaD' (Center for Hierarchical Materials Design). The next section is 'GET STARTED', which contains two green buttons: 'Publish Your Data' and 'Search for Data'. Below these buttons is a link: 'Don't have a Globus account? Sign up here!'. The 'FEATURES' section follows, with three icons and corresponding text: 1. 'Publication of large datasets' with an icon of a document and a line graph, stating that MDF offers access to petabytes (PB) of reliable and high performance data storage via NCSA. 2. 'Customizable metadata descriptions' with an icon of a document and a network diagram, stating that collection owners can define and use their own materials-specific metadata schemas. 3. 'Flexible access control' with an icon of a padlock and a key, stating that published datasets may be private, shared with a particular group of users, or shared publicly.

MDF

Submission

Walkthrough

Example Use Case

Publishing Big, Remote Data

Group has taken 50 TB of data at APS need to send back to home inst. For analysis and archiving

Bundle multiple experimental runs with metadata and provenance

PI wants to verify dataset data/metadata before pub.

Want a citable DOI to share the raw and derived data with the community

Want their data to be discoverable by free text search and custom metadata



MDF Collection Home



[Publish](#) [Log In](#) [Sign Up](#)

Materials Data Facility Community home page

Browse

Issue Date	Author	Title	Subject
------------	--------	-------	---------

Discover

Author	Subject	Date issued
Cahill, D.G. 12	Amorphous solids, thermal conduct... 3	2010 - 2015 14
Plante, Raymond 6	thermal conductivity, superlattic... 3	2000 - 2009 8
Felarca, Mario 5	ozone, thermal conductivity, chem... 2	1990 - 1999 13
Lee, S-M 5	thermal conductivity, single crys... 2	1987 - 1989 1
Pohl, R.O. 3	another test 1	
Pruyne, Jim 3	example test 1	
Venkatasubramanian, R. 3	Film 1	
Selinder, T.I. 2	pyrex temperature 1	
Watson, S.K. 2	test demo 1	
Ash, B.J. 1	thermal conductivity, ac techniqu... 1	
next >	next >	

Sub-communities within this community

[Computational Materials Science](#)

MDF Collections

globus blaiszik

Submit: Choose Collection ?

Select the collection you wish to submit an item to from the list below, then click "Next".

Collection

- Advanced Photon Source
- Center for Nanoscale Materials**
- Chemical Sciences and Engineering
- Institute of Molecular Engineering

Go to [Globus Publishing Home](#)
[Dashboard](#)

Recall: Policies Set at the Collection Level

- Required metadata, schemas
- Data storage location
- Metadata curation policies

MDF Metadata Entry

- **Scientist or representative describes the data they are submitting**
- **For this collection Dublin Core and a custom metadata template are required**

The screenshot shows the 'Submit: Describe this Dataset' form in the Globus interface. The form includes a navigation bar with 'Publish', 'Manage Data', 'Groups', 'Support', and 'blaiszik'. Below the navigation bar are links for 'Browse & Discover', 'Data Publication Dashboard', and 'Communities & Collections'. The form has a progress bar with steps: License, Describe (selected), Describe, Globus Transfer, Verify, and Complete. The main heading is 'Submit: Describe this Dataset' with a help icon. The instructions state: 'Please fill in the requested information about this submission below. In most browsers, you can use the tab key to move the cursor to the next input box or button, to save you having to use the mouse each time.' The form fields are: 'Title *' (Al-Cu Coarsening 4D Tomography Dataset); 'Authors *' (a list of names: Fife, Gibbs, Gulsoy, Park, Thornton, Voorhees, and a field for 'Last name, e.g. Smith'); 'Publication Year *' (Month: (No Month), Day: (empty), Year: 2014); and 'Publisher *' (Northwestern University). On the right side of the authors list, there are 'Remove Entry' buttons for each name and an 'Add More' button. At the bottom, there are navigation buttons: '< Previous', 'Cancel/Save', and 'Next >'. The footer contains the copyright information: '© 2010-2015 Computation Institute, University of Chicago, Argonne National Laboratory legal'.

MDF Custom Metadata

- **Scientist or representative describes the data they are submitting**
- **For this collection Dublin Core and a custom metadata template are required**

The screenshot shows the 'Submit: Describe this Dataset' form in the Globus Data Publication Dashboard. The form is titled 'Submit: Describe this Dataset' with a help icon. Below the title, it says 'Please fill further information about this submission below.' The form contains several input fields for metadata: 'Material' (Al-Cu), 'Volume Fraction Al' (15), 'Volume Fraction Cu' (85), 'Technique' (x-ray tomography), 'Pixel size (µm)' (1.4), 'Beam energy (keV)' (20), and 'Instrumentation' (Swiss Light Source - Tomographic Microscopy and Coherent Radiology Experiments beamline). There is also a 'Keywords' section with a list of keywords: 'in situ', '4D coarsening', 'aluminum-copper alloys', 'dynamic morphological evolution', and 'solid-liquid interfaces'. Each keyword has a 'Remove Entry' button, and there is an 'Add More' button at the bottom of the list. At the bottom of the form, there are three buttons: '< Previous', 'Cancel/Save', and 'Next >'. The top navigation bar includes 'Publish', 'Manage Data', 'Groups', 'Support', and 'blaiszik'. The breadcrumb trail shows 'Browse & Discover | Data Publication Dashboard | Communities & Collections'. The form has a progress bar at the top with steps: License, Describe, Describe, Globus Transfer, Verify, Complete.

Dataset Assembly

- Shared endpoint is auto-created on collection-specified data store
- Scientist transfers dataset files to a unique publish endpoint
- Dataset may be assembled over any period of time
- When submission is finished, dataset will be rendered immutable via checksum

The screenshot displays the Globus web interface for managing data transfers. At the top, the Globus logo and navigation links (Manage Data, Groups, Support, blaiszik) are visible. The main section is titled "Transfer Files" and includes a sub-header "Get Globus Connect Personal Turn your computer into an endpoint." Below this, two file selection panes are shown side-by-side. The left pane, representing the source endpoint "blaiszik#macbookpro", shows a path "/~/Desktop/blaiszik-macbookpro/Voorhet" and contains two files: "20A_post_0004.h5" (3.19 GB) and "20A_post_0005.h5" (3.15 GB). The right pane, representing the destination endpoint "globuspublish#jcpublish-test", shows a path "/mdf_voorhees_72/results/" and contains the same two files. Large blue text overlays the panes, reading "(e.g. APS)" on the left and "(e.g. UC Berkeley)" on the right. Below the panes, there are controls for "more options", "Label This Transfer", and a text input field. The "Activity" section at the bottom shows a green checkmark indicating a successful transfer from "blaiszik#macbookpro to globuspublish#jcpublish-test" completed a minute ago. It includes tabs for "Overview" and "Event Log", and a summary table with the following data:

Task ID	c1191a64-ef5d-11e4-ab4a-22000b92c6ec	Files	2
Source	blaiszik#macbookpro	Directories	1
Destination	globuspublish#jcpublish-test	Bytes Transferred	6.34 GB

Dataset Curation

- Optionally specified in collection configuration
- Can be approved or rejected (i.e. sent back to the submitter)

The screenshot shows the 'Perform Task' interface on the Globus Data Publication Dashboard. The page title is 'Perform Task'. Below the title, a message states: 'The following item has been submitted to collection **Voorhees Group X-Ray Tomography**. Please review the item, check that it meets the criteria for entry into the collection. After reviewing the item, please approve or reject the item using the controls at the bottom of the page.'

The item details are as follows:

Title:	Al-Cu Coarsening 4D Tomography Dataset
Authors:	Fife, J.L. Gibbs, J.W. Gulsoy, E.B. Park, C.-L. Thornton, K. Voorhees, P.W.
Keywords:	in situ 4D coarsening aluminum-copper alloys dynamic morphological evolution solid-liquid interfaces
Issue Date:	2014
Publisher:	Northwestern University

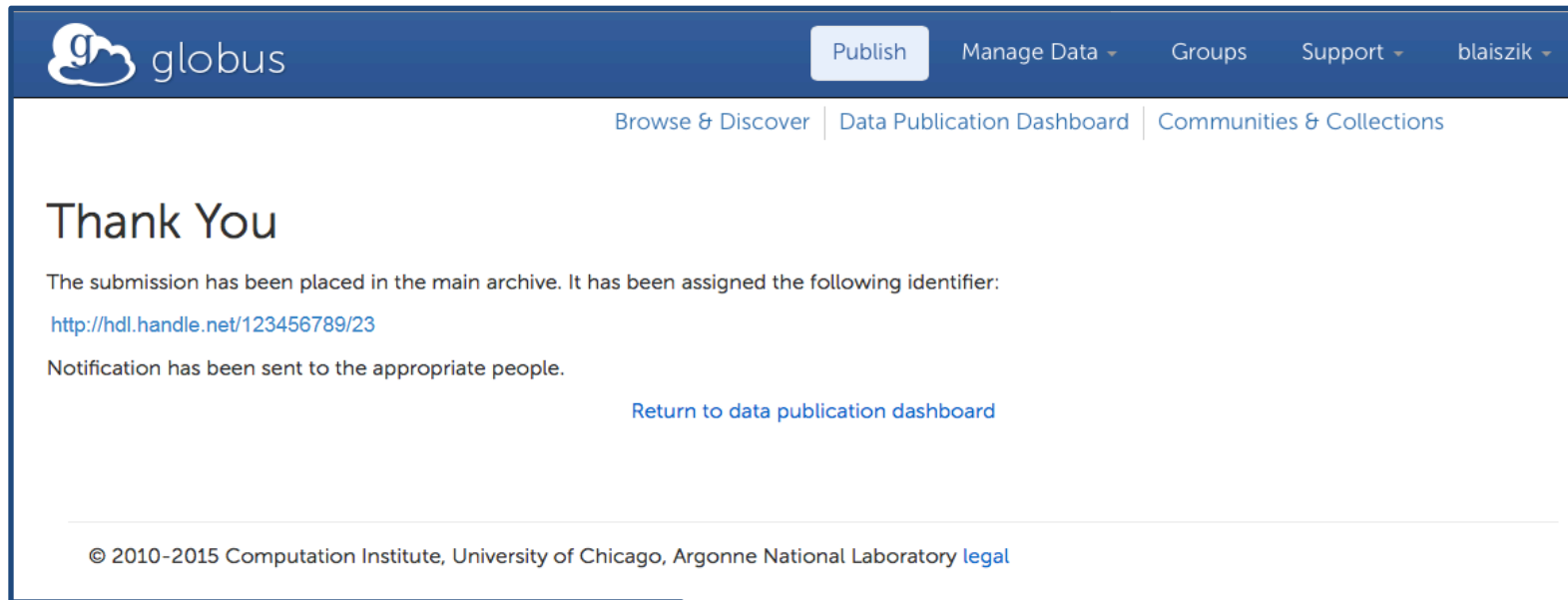
Files in This Item:

[globuspublish#jcpublish-test/mdf_voorhees_72/](#)

At the bottom, there are four action buttons with their respective descriptions:

Approve	If you have reviewed the item and it is suitable for inclusion in the collection, select "Approve".
Reject	If you have reviewed the item and found it is not suitable for inclusion in the collection, select "Reject". You will then be asked to enter a message indicating why the item is unsuitable, and whether the submitter should change something and re-submit.
Do Later	If you wish to leave this task for now, and return to the data publication dashboard, use this option.
Return Task to Pool	To return the task to the pool so that another user can perform the task, use this option.


Mint a Permanent Identifier




The screenshot shows the Globus Data Publication Dashboard. The top navigation bar is dark blue with the Globus logo and the text 'globus'. To the right of the logo are buttons for 'Publish', 'Manage Data', 'Groups', 'Support', and 'blaisik'. Below the navigation bar, there are links for 'Browse & Discover', 'Data Publication Dashboard', and 'Communities & Collections'. The main content area has a large heading 'Thank You'. Below this heading, it states: 'The submission has been placed in the main archive. It has been assigned the following identifier: <http://hdl.handle.net/123456789/23>'. Below this, it says 'Notification has been sent to the appropriate people.' and provides a link 'Return to data publication dashboard'. At the bottom of the page, there is a copyright notice: '© 2010-2015 Computation Institute, University of Chicago, Argonne National Laboratory [legal](#)'.

Can optionally be DOI or Handle

Dataset Record

Publish Manage Data ▾ Groups Support ▾ blaiszik ▾

[Browse & Discover](#) | [Data Publication Dashboard](#) | [Communities & Collections](#)



Please use this identifier to cite or link to this item: <http://bit.ly/1EGh9UL>

Title:	Al-Cu Coarsening 4D Tomography Dataset
Authors:	Fife, J.L. Gibbs, J.W. Gulsoy, E.B. Park, C.-L Thornton, K. Voorhees, P.W.
Keywords:	in situ 4D coarsening aluminum-copper alloys dynamic morphological evolution solid-liquid interfaces
Issue Date:	2014
Publisher:	Northwestern University
URI:	http://bit.ly/1EGh9UL
Appears in Collections:	Voorhees Group X-Ray Tomography

Admin Tools

[Configure...](#)


[Export Item](#)

[Export \(migrate\) Item](#)

[Export metadata](#)

Files in This Item:

[globuspublish#jcpublish-test/mdf_voorhees_72/](#)

[Show full item record](#) 

Items in Globus are protected by copyright, with all rights reserved, unless otherwise indicated.

Dataset Discovery

The screenshot shows the Globus dataset discovery interface. At the top, there is a navigation bar with the Globus logo, a search bar containing 'Voorhees', and several menu items: 'Publish', 'Manage Data', 'Groups', 'Support', and 'blaiszik'. Below the navigation bar, there are three tabs: 'Browse & Discover', 'Data Publication Dashboard', and 'Communities & Collections'. The main content area is divided into three sections: 'Search Results', 'Discover', and a pagination control.

Search Results

Collection results (1 result) [advanced search](#)

Results 1-2 of 2

Issue Date	Title	Author(s)
2014	Al-Cu Coarsening 4D Tomography Dataset	<i>Fife, J.L.; Gibbs, J.W.; Gulsoy, E.B.; Park, C.-L; Thornton, K.; Voorhees, P.W.</i>

previous 1 next

Discover

Author

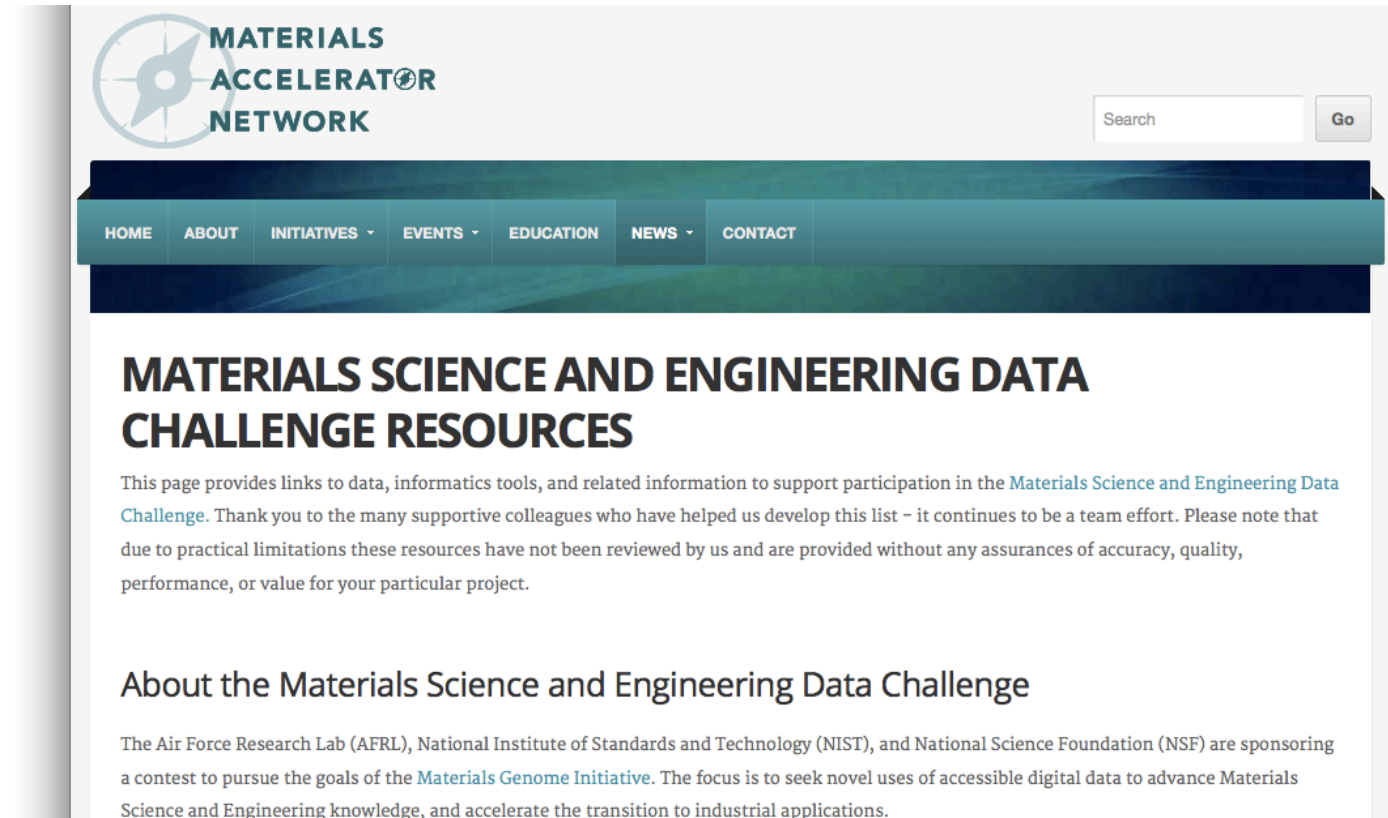
- Fife, J.L. 1
- Gibbs, J.W. 1
- Gulsoy, E.B. 1
- Park, C.-L 1
- Thornton, K. 1
- Voorhees, P.W. 1

Subject

- 4D coarsening 1
- aluminum-copper alloys 1
- dynamic morphological evolution 1
- in situ 1
- solid-liquid interfaces 1

Registering Materials Resources

Materials Resource Registry



The screenshot shows the Materials Accelerator Network website. The header includes the logo (a compass rose) and the text "MATERIALS ACCELERATOR NETWORK". A search bar with a "Go" button is located in the top right. The navigation menu contains links for HOME, ABOUT, INITIATIVES, EVENTS, EDUCATION, NEWS, and CONTACT. The main content area features the title "MATERIALS SCIENCE AND ENGINEERING DATA CHALLENGE RESOURCES" and a paragraph of introductory text. Below this is a section titled "About the Materials Science and Engineering Data Challenge" with a paragraph of descriptive text.

MATERIALS ACCELERATOR NETWORK

Search

HOME ABOUT INITIATIVES ▾ EVENTS ▾ EDUCATION NEWS ▾ CONTACT

MATERIALS SCIENCE AND ENGINEERING DATA CHALLENGE RESOURCES

This page provides links to data, informatics tools, and related information to support participation in the [Materials Science and Engineering Data Challenge](#). Thank you to the many supportive colleagues who have helped us develop this list - it continues to be a team effort. Please note that due to practical limitations these resources have not been reviewed by us and are provided without any assurances of accuracy, quality, performance, or value for your particular project.

About the Materials Science and Engineering Data Challenge

The Air Force Research Lab (AFRL), National Institute of Standards and Technology (NIST), and National Science Foundation (NSF) are sponsoring a contest to pursue the goals of the [Materials Genome Initiative](#). The focus is to seek novel uses of accessible digital data to advance Materials Science and Engineering knowledge, and accelerate the transition to industrial applications.

<http://acceleratornetwork.org/mse-challenge/>

Materials Science Data Challenge

Materials Resource Registry

Data Resources

Computed Data

- [AFLOW database](#)
- [Computational Materials Data \(CMD\) Network](#)
- [Harvard Clean Energy Project](#)
- [Materials Project](#)
- [National Institute of Standards and Technology \(NIST\) Interatomic Potentials Repository Project](#)
- [Open Knowledgebase of Interatomic Models \(KIM\) or OpenKIM](#)
- [Open Quantum Materials Database \(OQMD\)](#)

Experimental (and possibly computed) Data

- [3D Materials Atlas](#)
- [American Mineralogist Crystal Structure](#)

Data Mining Tools

- [Best Data Mining Tools by Quora](#)
- [Citrine](#). See also their blog posts on machine learning for the materials scientist [part 1](#) and [part 2](#).
- [Dream3D](#)
- [Fiji \(ImageJ\)](#)
- [Granta \(Material Intelligence\)](#)
- [Massive Online Analysis \(MOA\)](#)
- [Materials Knowledge Systems in Python \(PyMKS\)](#)
- [Matlab](#)
- [Matlab Toolbox for Dimensionality Reduction by Laurens van der Maaten](#)
- [nanoHUB](#)
- [Nutonian Eureka](#)
- [Open Science Commons Materials \(OSCM\)](#)

Places to Publish, Share (and Find) Data

Journals with Data Focus

- [Data in Brief \(DiB\) \(Elsevier\)](#) . See also Harvard Dataverse [DiB section](#).
- [Harvard Dataverse](#)
- [Integrating Materials and Manufacturing Innovation \(IMMI\) \(see Data Descriptor article type\)](#)
- [Materials Discovery \(Elsevier\)](#)
- [Open Data journals at Elsevier](#). Part of a number of projects at Elsevier supporting the Materials Genome Initiative. See also Elsevier's page on their resources for the MS&E Data Challenge.
- [Scientific Data \(Nature Publishing Group\)](#)

Data Repositories and Data Sharing Tools

- [Citrine](#) (see their [blog](#) for details on their support of datasets for the Challenge)

Materials Accelerator Network

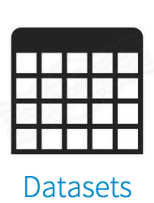
Materials Resource Registry



Materials Resource Registry

Search Results for 'compound'

[Edit Search Criteria](#) [New Search](#)



Detailed Results View

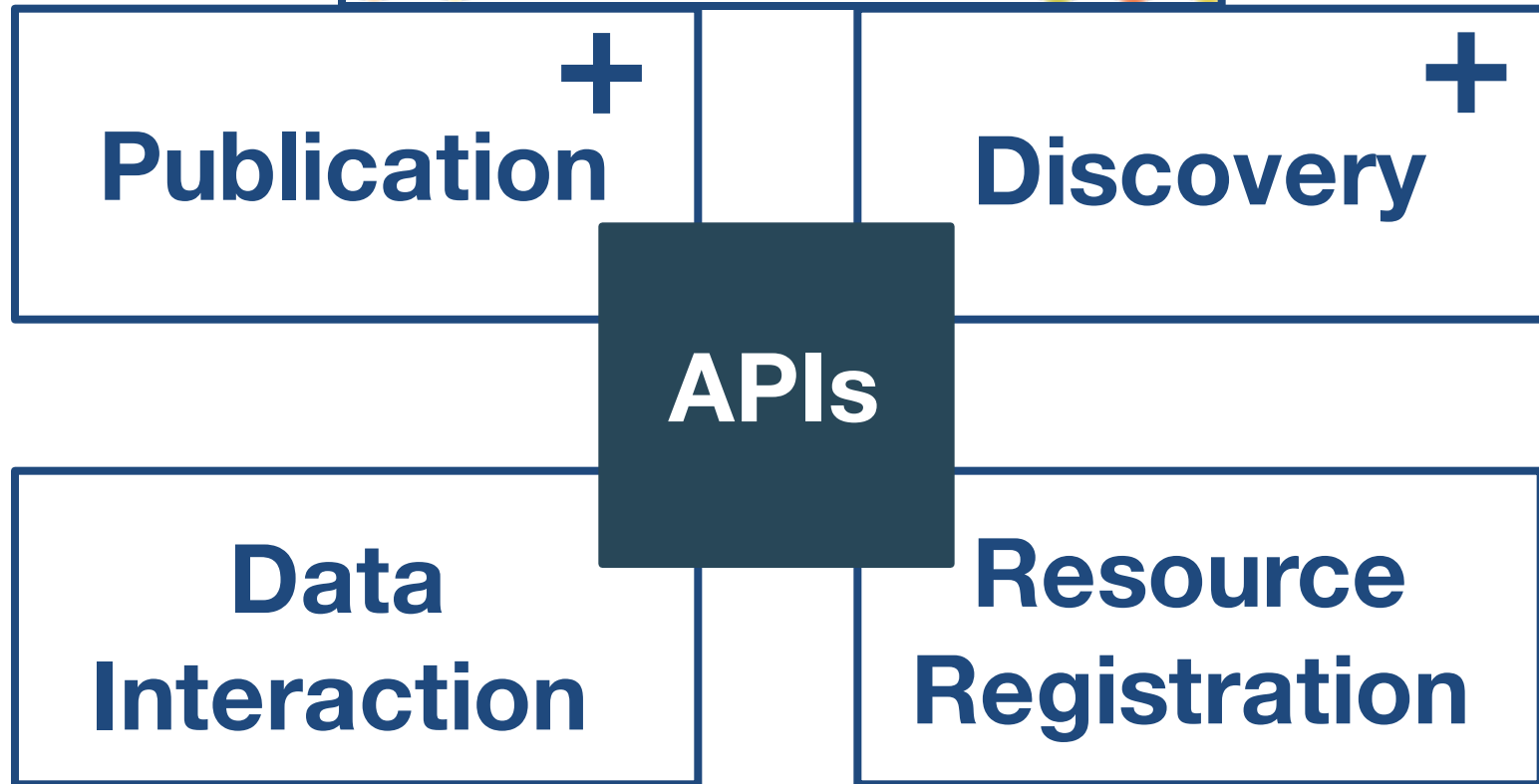
Resource Type:

- All Resources
- Organization
- Data Collection
- Repository
- Project Archive
- Database

AFLOW		Resource Details	Go To
Publisher	AFLOW Consortium		
Resource Type	Repository		
Material Science	Material Types: Metal, Semiconductor, Organic Morphology/Structures: Crystalline, Bulk		

Browse Results

Service Infrastructure



+ - Initial Foci

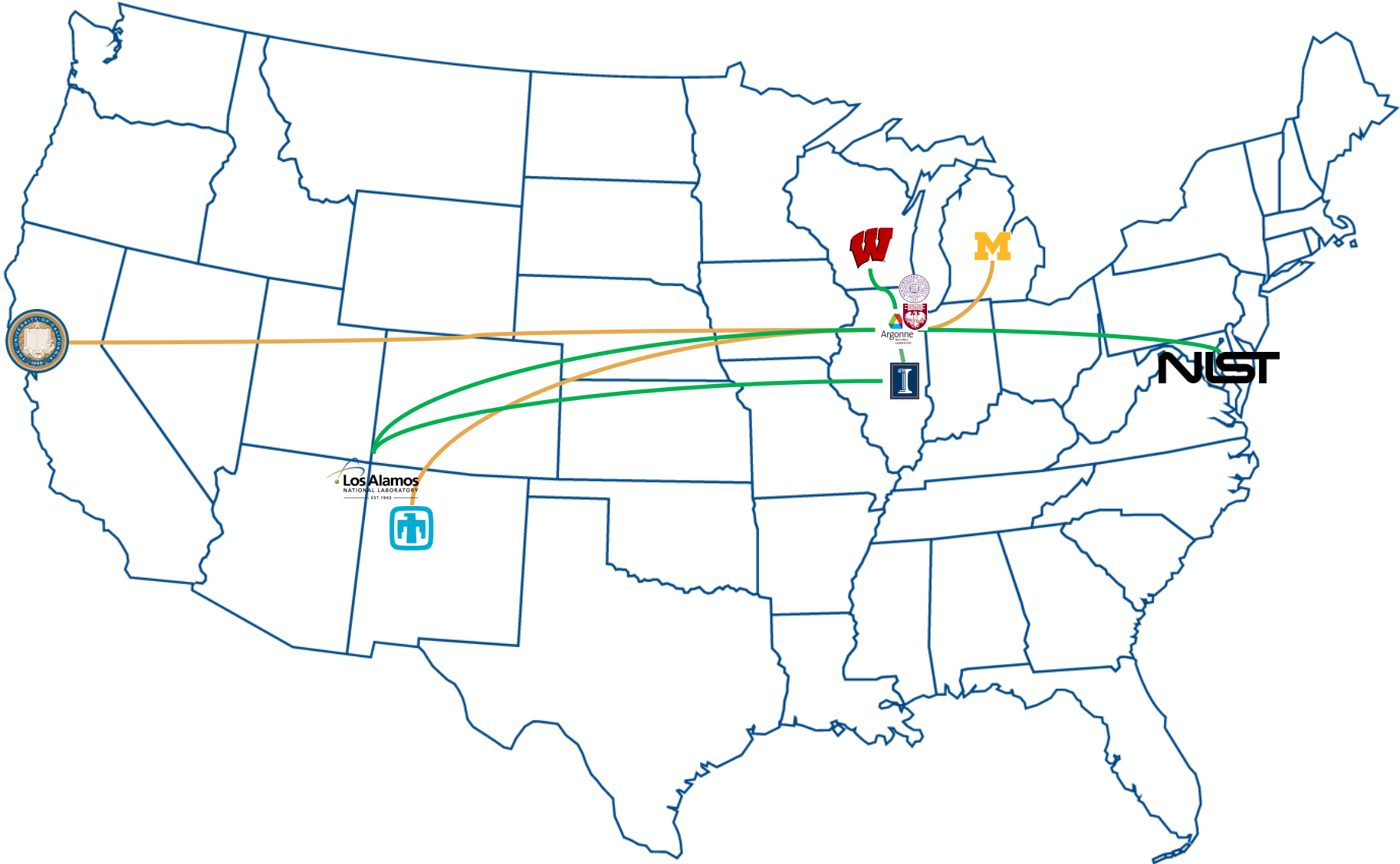
What's Available?

- **Web interface to support data publication via Globus platform (identify management, user groups, optimized big data transfer)**
- **100 TB of storage at NCSA (scalable to >1 PB) more at Argonne (?)**
- **Help with developing metadata schemas to describe your research datasets**

Yet another publication system?

- Software as a Service
- Self service management
 - Identifiers, policies, submission and curation workflows, storage, metadata, access control
- Remote storage
- Supports arbitrarily large datasets
- Powerful search

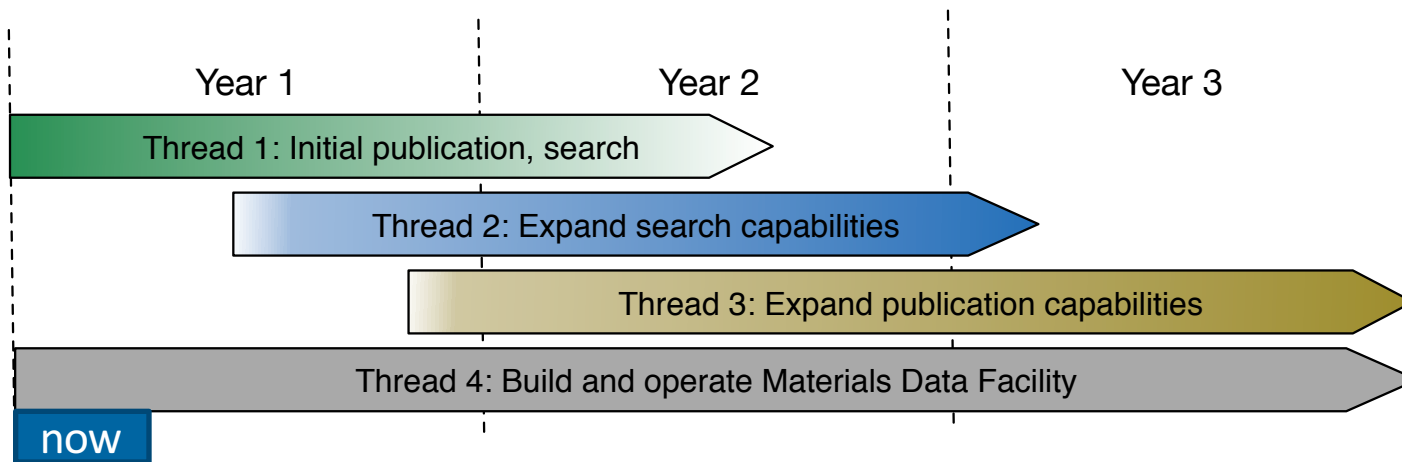
Current Interactions



What are we looking for?

- **Early adopters, willing to get their hands dirty with the service and give honest feedback**
- **Key datasets of all sizes, shapes, raw or derived, that might help us understand the process better**
- **Currently working with researchers from UIUC, NWU, UC Berkeley, UW-Madison, UMichigan, Argonne**

Next Steps



- **Identify datasets to pilot publication pipelines and build schema repository**
- **Engage with researches working with materials data to understand use cases and learn friction points**
- **Please talk to us if you have data you want to share, publish, discover, ...**
- **Globus tutorials (identity, transfer, sharing):**
<https://github.com/globusonline/globus-tutorials>

Thanks to Our Sponsors!

The logo for the National Institute of Standards and Technology (NIST), consisting of the letters "NIST" in a bold, black, sans-serif font.

U.S. DEPARTMENT OF
ENERGY



THE UNIVERSITY OF
CHICAGO

Globus delivers...

**big data transfer, sharing,
publication, and discovery...**

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

...via software-as-a-service

Globus is SaaS

- **Access made easy via Web browser**
 - **Command line, REST interfaces, python clients for flexible automation and integration**
- **New features automatically available without user updates**
- **Reduced IT operational costs**
 - **Small local footprint (Globus Connect)**
 - **Consolidated support and troubleshooting**

User Experience

flickr ...for your photos



...for your office docs



...for your entertainment



globus

...for your research data

Globus Platform-as-a-Service (PaaS)

Identity management

- create and manage a unique identity linked to external identities for authentication

User groups

- Manage user group creation and administration flows
- Share data with user groups

Data publication

Data transfer

- High-performance data transfer from a web browser
- Optimize transfer settings and verify transfer integrity
- Add your laptop to the Globus cloud with Globus Connect Personal

Data sharing

- Share directly from your storage device (laptop or cluster)
- File and directory-level ACLs

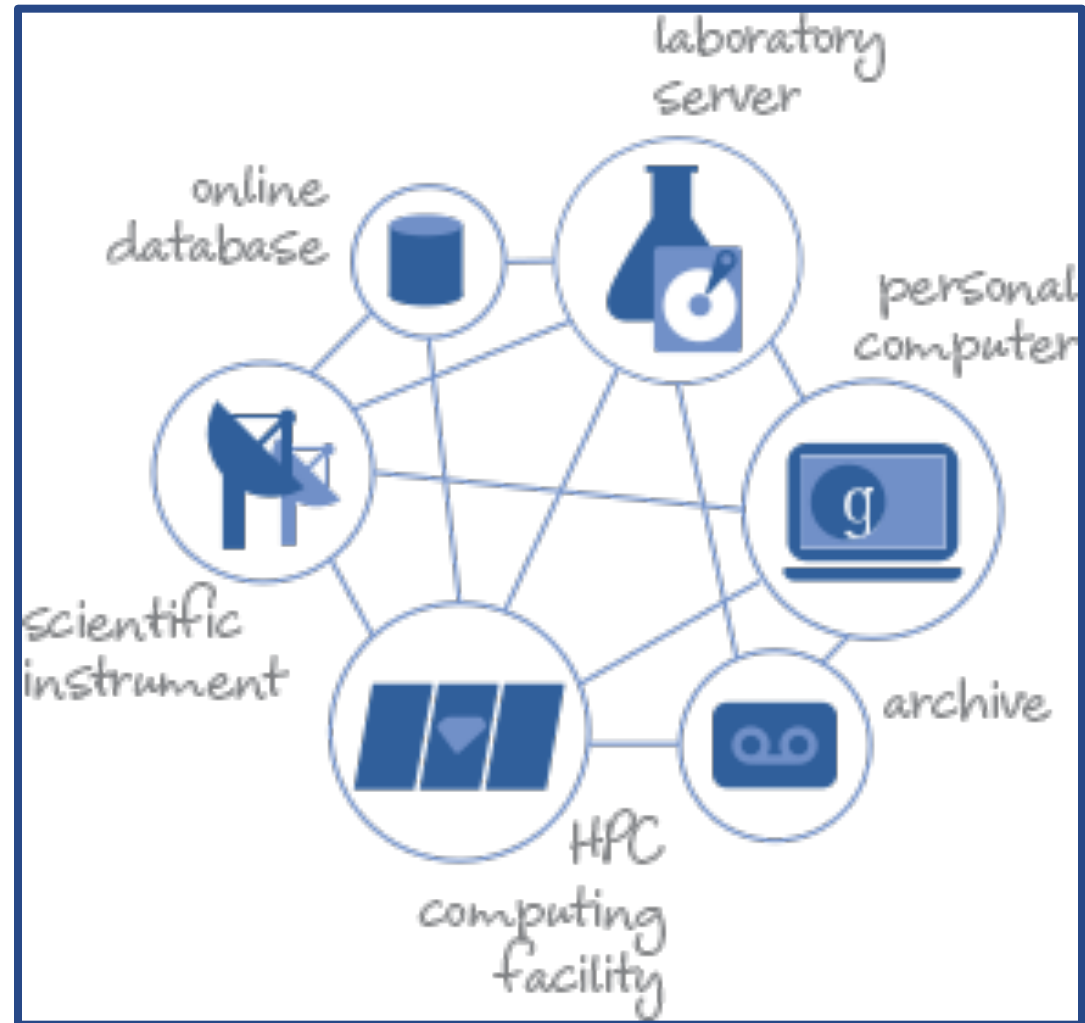
Background

Endpoint

- E.g. laptop or server running a Globus client (e.g. Dropbox client)
- Enables advanced file transfer and sharing
- Currently GridFTP, future GridFTP + HTTP

Some Key Features

- REST API for automation and interoperability
- Web UI for convenience
- Optimizes and verifies transfers
- Handles auto-restarts
- Battle tested with big data



Background

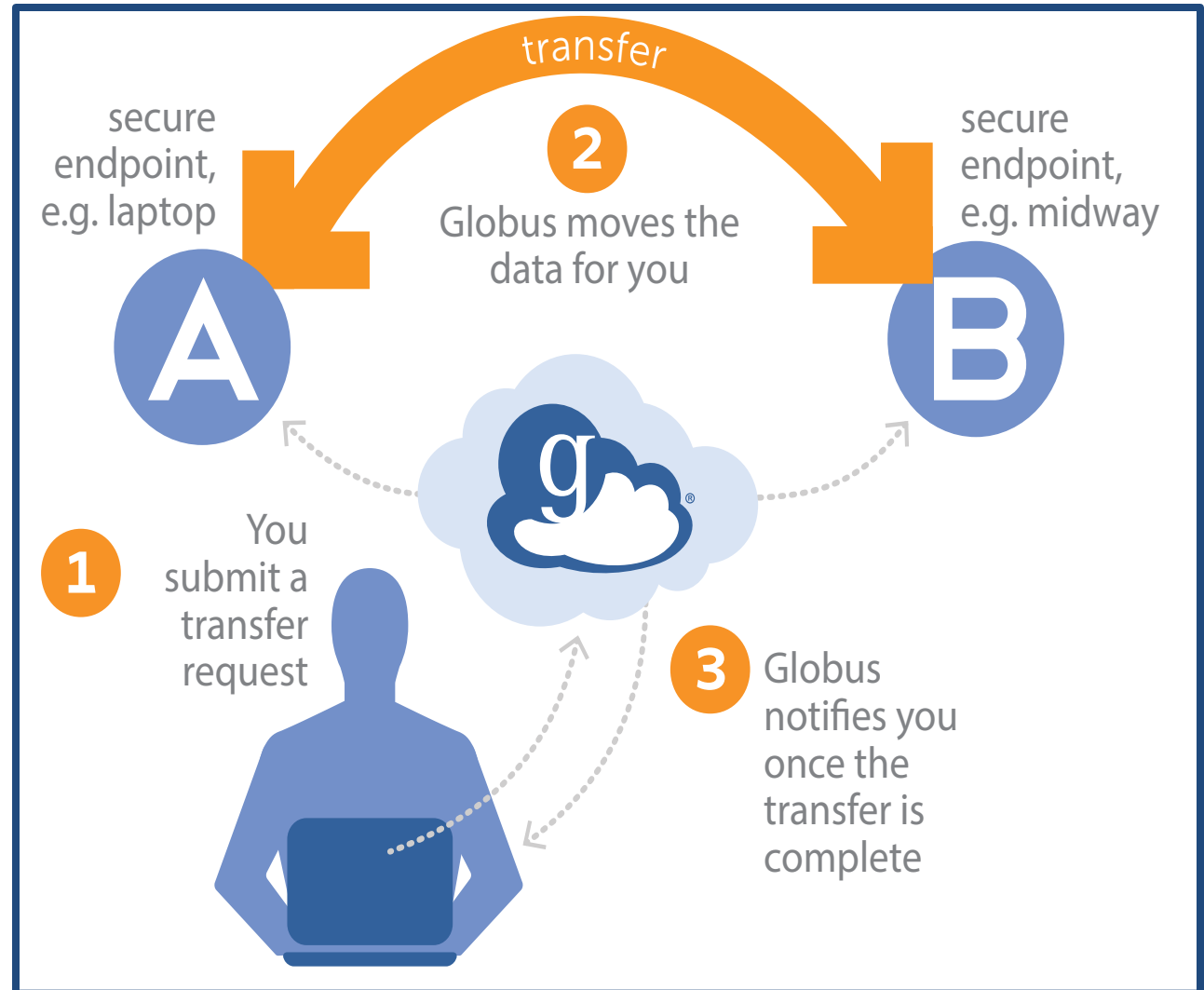
Endpoint

- E.g. laptop or server running a Globus client (e.g. Dropbox client)
- Enables advanced file transfer and sharing
- Currently GridFTP, future GridFTP + HTTP

Some Key

Features

- REST API for automation and interoperability
- Web UI for convenience
- Optimizes and verifies transfers
- Handles auto-restarts
- Battle tested with big data



Background

Endpoint

- E.g. laptop or server running a Globus client (e.g. Dropbox client)
- Enables advanced file transfer and sharing
- Currently GridFTP, future GridFTP + HTTP

Some Key Features

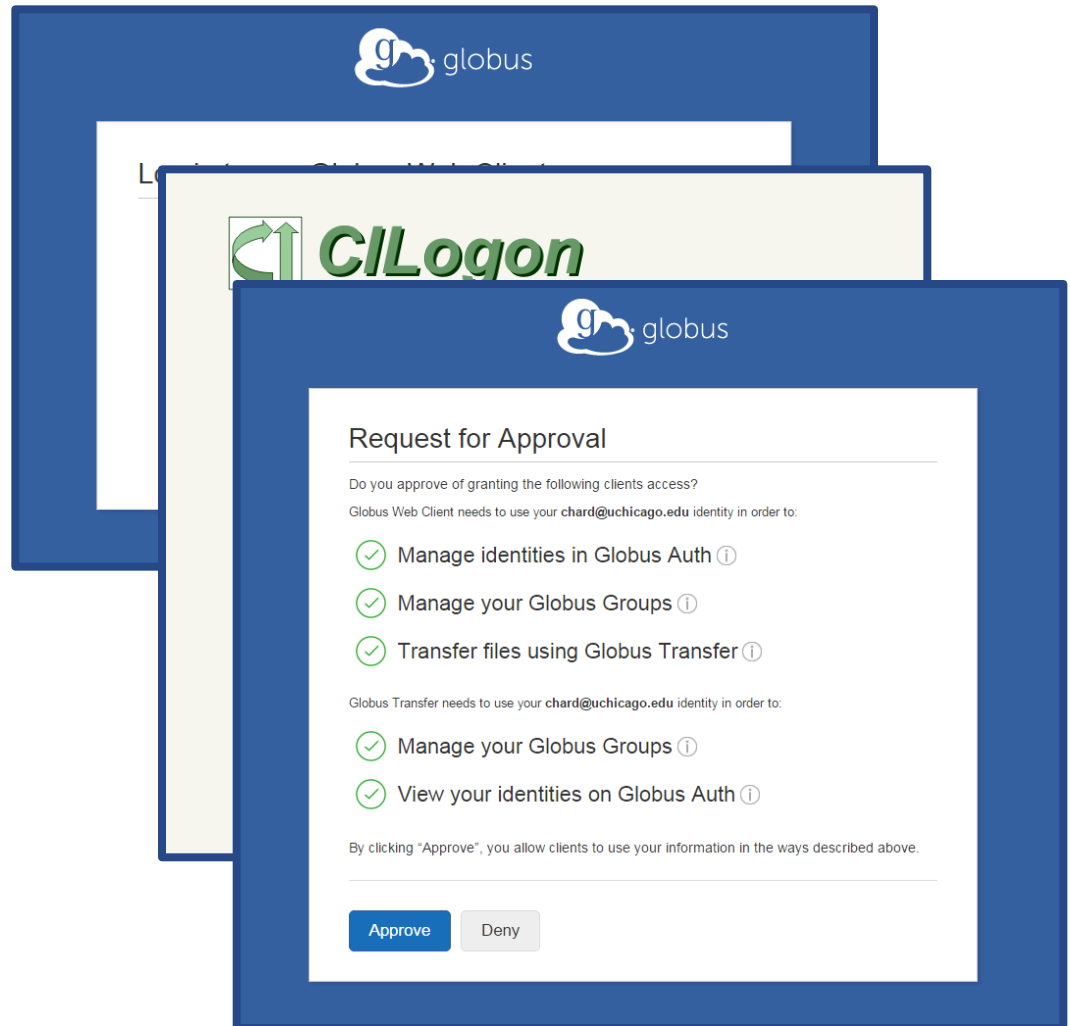
- REST API for automation and interoperability
- Web UI for convenience
- Optimizes and verifies transfers
- Handles auto-restarts
- Battle tested with big data

The screenshot displays the Globus web interface. At the top, there's a navigation bar with 'Manage Data', 'Groups', 'Support', and 'blaiszik'. Below this, a secondary navigation bar includes 'Transfer Files', 'Activity', 'Manage Endpoints', and 'Dashboard'. The main content area is titled 'Transfer Files' and features a sub-header 'Get Globus Connect Personal Turn your computer into an endpoint.' The interface is split into two panels, each showing a file transfer progress. The left panel shows the source endpoint 'blaiszik#macbookpro' with the path '/~/Desktop/blaiszik-macbookpro/Voorhees'. The right panel shows the destination endpoint 'globuspublish#jcpublish-test' with the path '/mdf_voorhees_72/results/'. Both panels list two files: '20A_post_0004.h5' (3.19 GB) and '20A_post_0005.h5' (3.15 GB). Below the panels, there's a 'Label This Transfer' field and a note: 'This will be displayed in your transfer activity.' The bottom section is titled 'Activity' and shows a green checkmark indicating a successful transfer from 'blaiszik#macbookpro to globuspublish#jcpublish-test' completed a minute ago. Below this, there are tabs for 'Overview' and 'Event Log'. The 'Overview' tab displays transfer details: Task ID 'c1191a64-ef5d-11e4-ab4a-22000b92c6ec', Source 'blaiszik#macbookpro', Destination 'globuspublish#jcpublish-test', Files '2', Directories '1', and Bytes Transferred '6.34 GB'.

Identity Management

Key Features

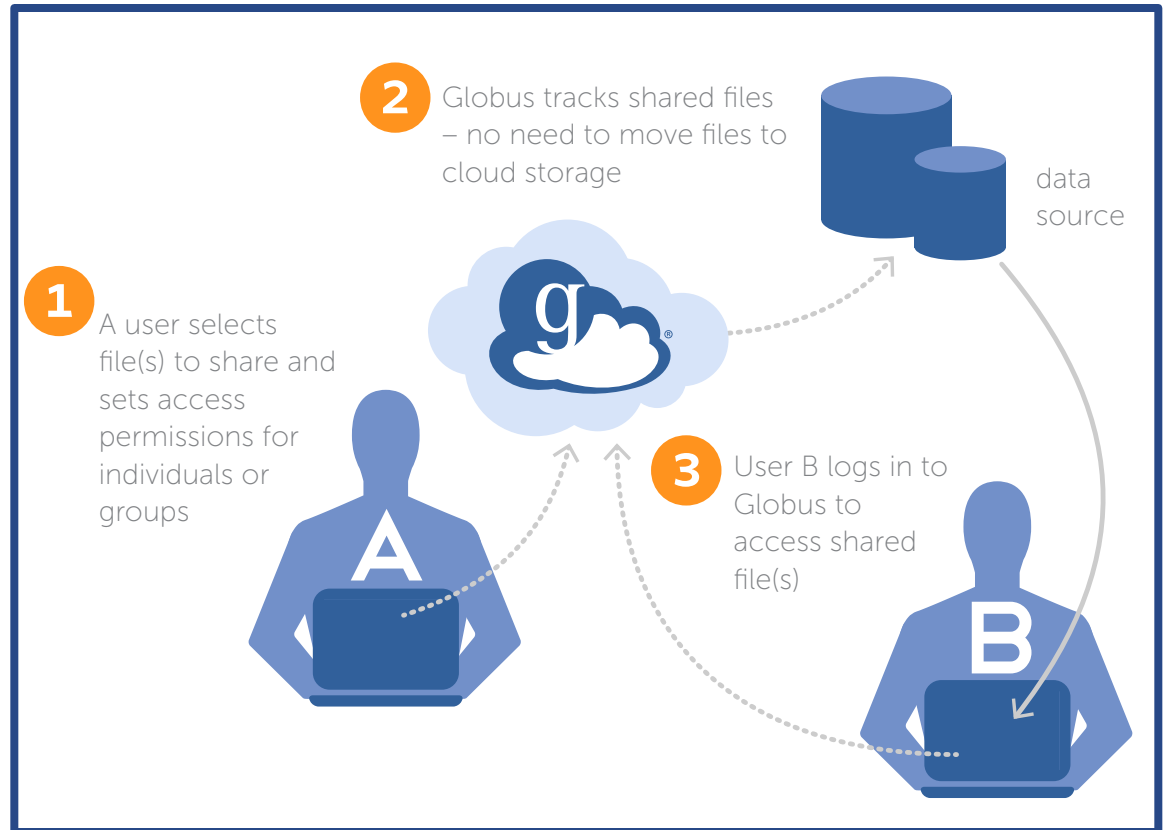
- Leverage institutional credentials
- Link multiple identities
- Standard OAuth2 flow



User Groups and Sharing

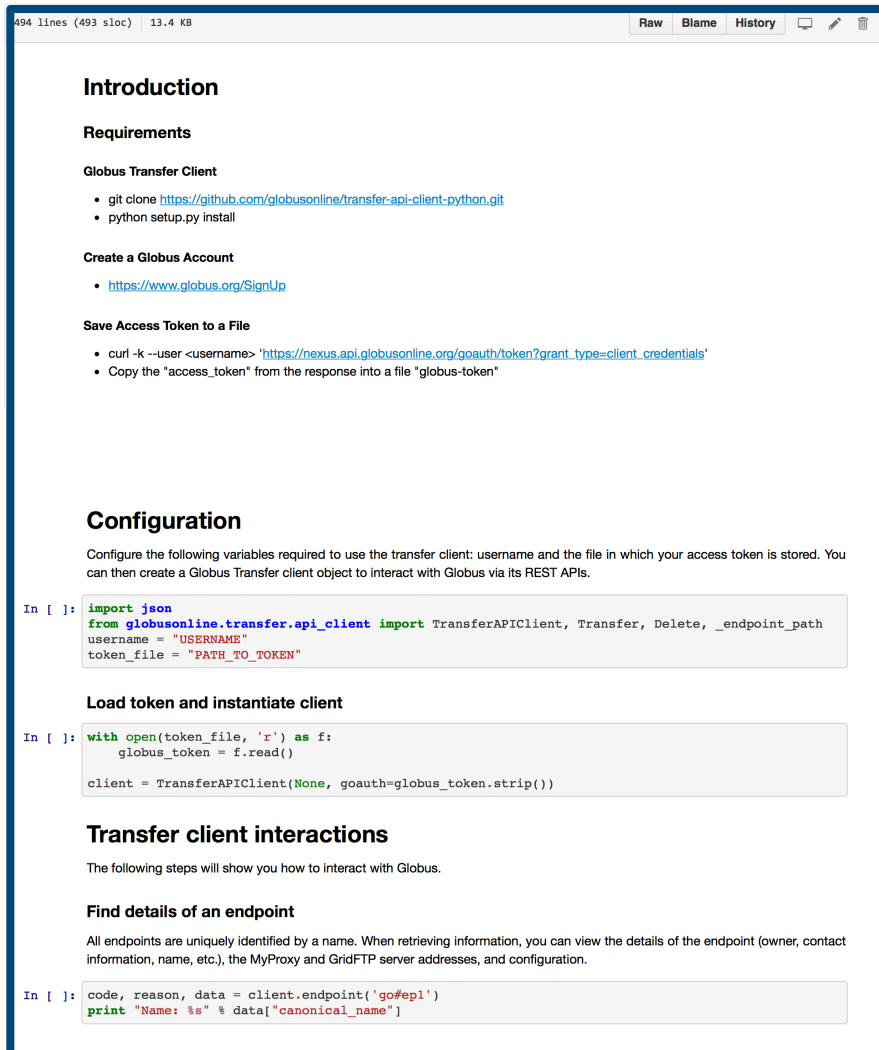
Key Features

- **Share without users requiring accounts on your systems.**
- **Share data in place, no need to move to the cloud**
- **Maintain security and access controls as defined by your resource provider.**
- **Groups: Manage permissions on a group basis rather than on an individual basis**



Globus PaaS Jupyter Notebooks

<https://github.com/globusonline/globus-tutorials>



494 lines (493 sloc) | 13.4 KB

Raw Blame History

Introduction

Requirements

Globus Transfer Client

- git clone <https://github.com/globusonline/transfer-api-client-python>
- python setup.py install

Create a Globus Account

- <https://www.globus.org/SignUp>

Save Access Token to a File

- curl -k --user <username> 'https://nexus.api.globusonline.org/goauth/token?grant_type=client_credentials'
- Copy the "access_token" from the response into a file "globus-token"

Configuration

Configure the following variables required to use the transfer client: username and the file in which your access token is stored. You can then create a Globus Transfer client object to interact with Globus via its REST APIs.

```
In [ ]: import json
from globusonline.transfer.api_client import TransferAPIClient, Transfer, Delete, _endpoint_path
username = "USERNAME"
token_file = "PATH_TO_TOKEN"
```

Load token and instantiate client

```
In [ ]: with open(token_file, 'r') as f:
        globus_token = f.read()

client = TransferAPIClient(None, goauth=globus_token.strip())
```

Transfer client interactions

The following steps will show you how to interact with Globus.

Find details of an endpoint

All endpoints are uniquely identified by a name. When retrieving information, you can view the details of the endpoint (owner, contact information, name, etc.), the MyProxy and GridFTP server addresses, and configuration.

```
In [ ]: code, reason, data = client.endpoint('go#ep1')
print "Name: %s" % data["canonical_name"]
```

Identity
management

User
groups

Data
transfer

Data
sharing

Create an Account

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. Clone repo
 - `git clone https://github.com/globusonline/globus-tutorials`
7. Move files from `kyle#ncsa-tutorial` endpoint to your laptop