Computation Institute

Networking materials data

Ian Foster

foster@anl.gov

ianfoster.org

THE UNIVERSITY OF CHICAGO

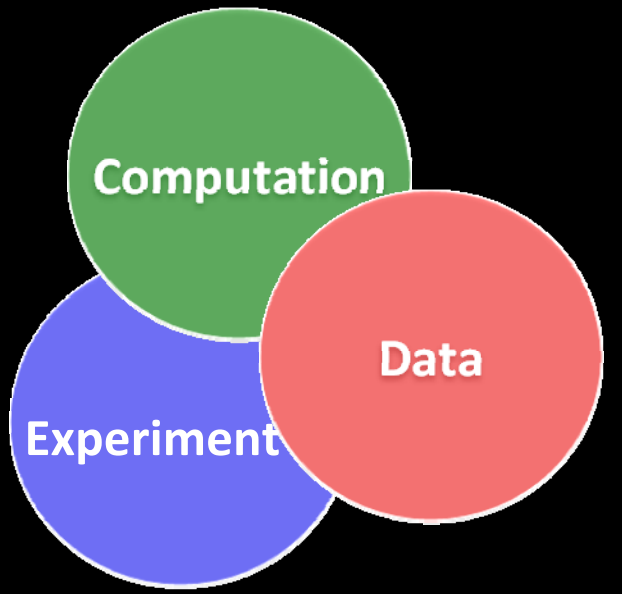
Argonne NATIONAL LABORATORY

computationinstitute.org

Materials Genome Initiative — Materials Innovation Infrastructure

A **data sharing system** to facilitate:

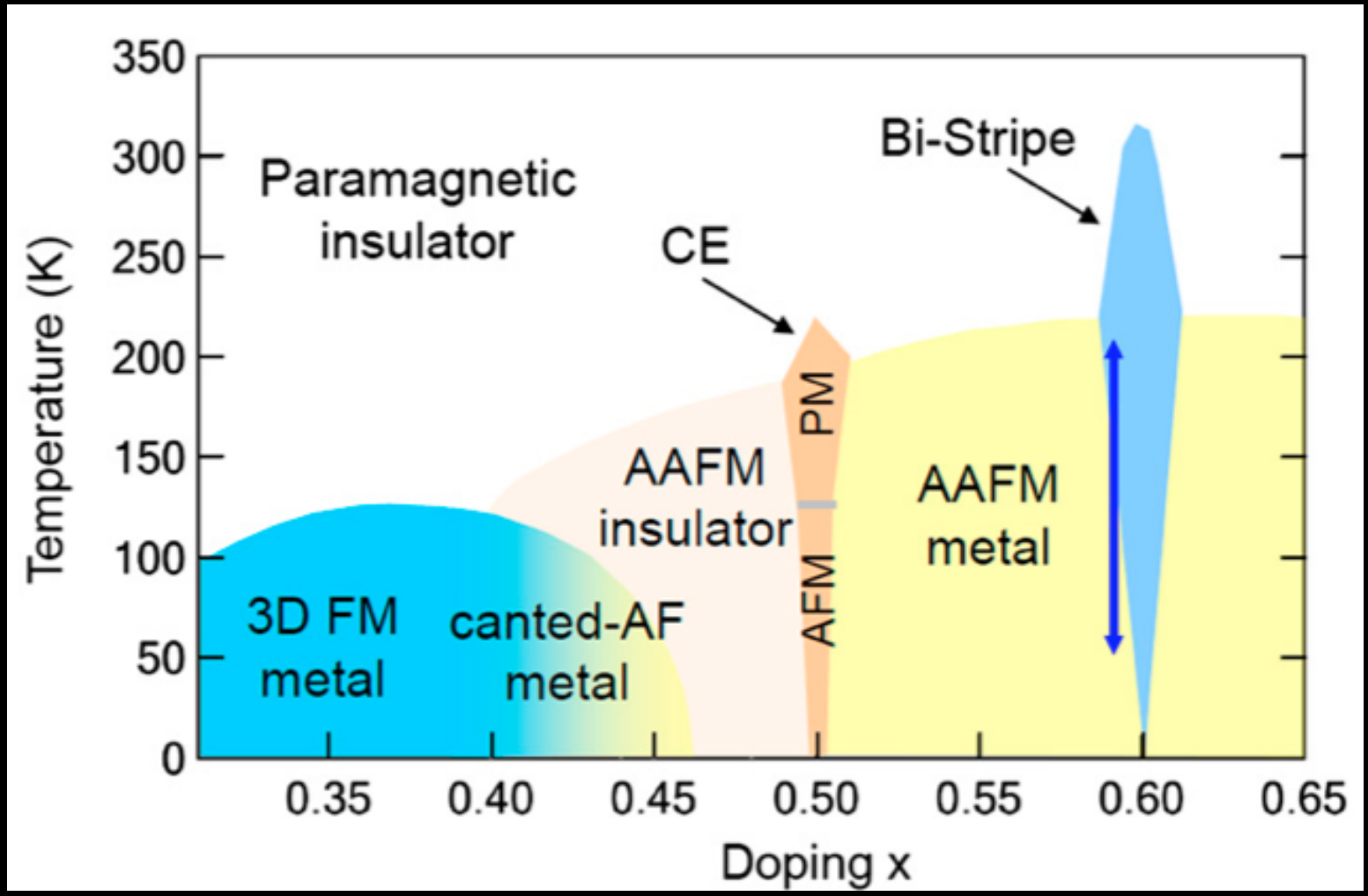
- Use of a broader set of data to render more accurate models
- Multi-disciplinary communication among scientists and engineers working on different stages of materials development
- Searches for advanced materials with specific, desired properties
- Curating and sharing of reliable computational code for modeling and simulation



Credit: Meredith Drosback, OSTP

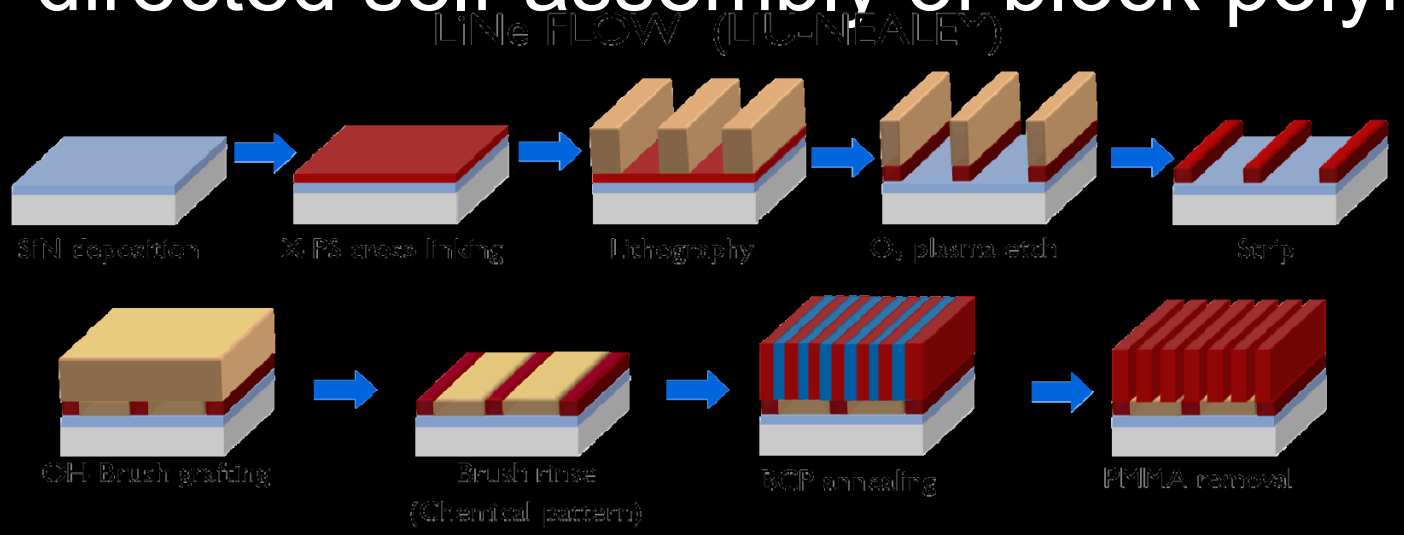
computationinstitute.org

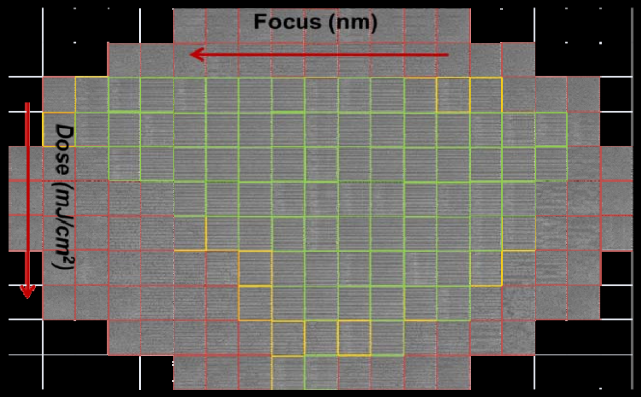
Linking simulation and experiment to study disordered structures



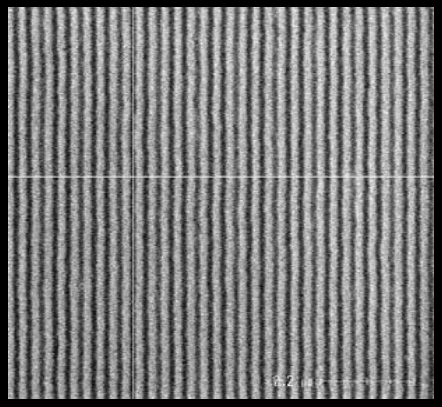
Diffuse scattering images from Ray Osborn et al., Argonne

Linking simulation, experiment to study directed self-assembly of block polymers





IMEC



300 nm wafers

Track processing

P. Nealey, J. De Pablo, et al.

computationinstitute.org

Materials Database Projects

- Materials Project (MIT/LBL)
- PRedictive Integrated Structural Materials Science (PRISMS) – Materials Commons
- Center for Hierarchical Materials Design (CHiMaD)
- Joint Center for Energy Storage Research (JCESR)
- NIST
- Etc.

Data: Rare treasure?

http://www.thejakartapost.com/news/2011/05/14/holy-water.html

Or chaotic deluge?

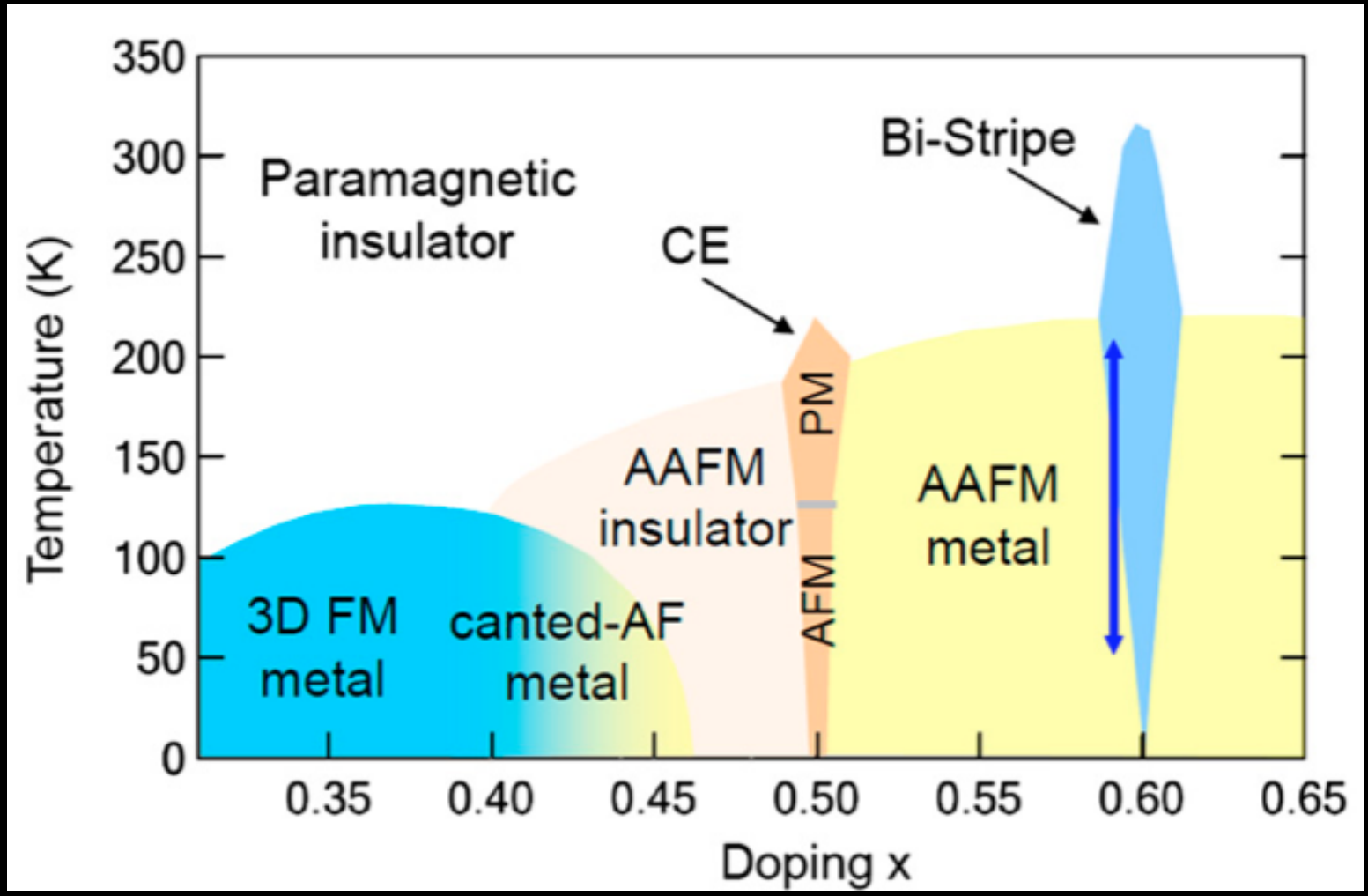
It's both …

We must manage the data deluge—both to enhance user productivity and to increase data capture

➔ **Network materials data**

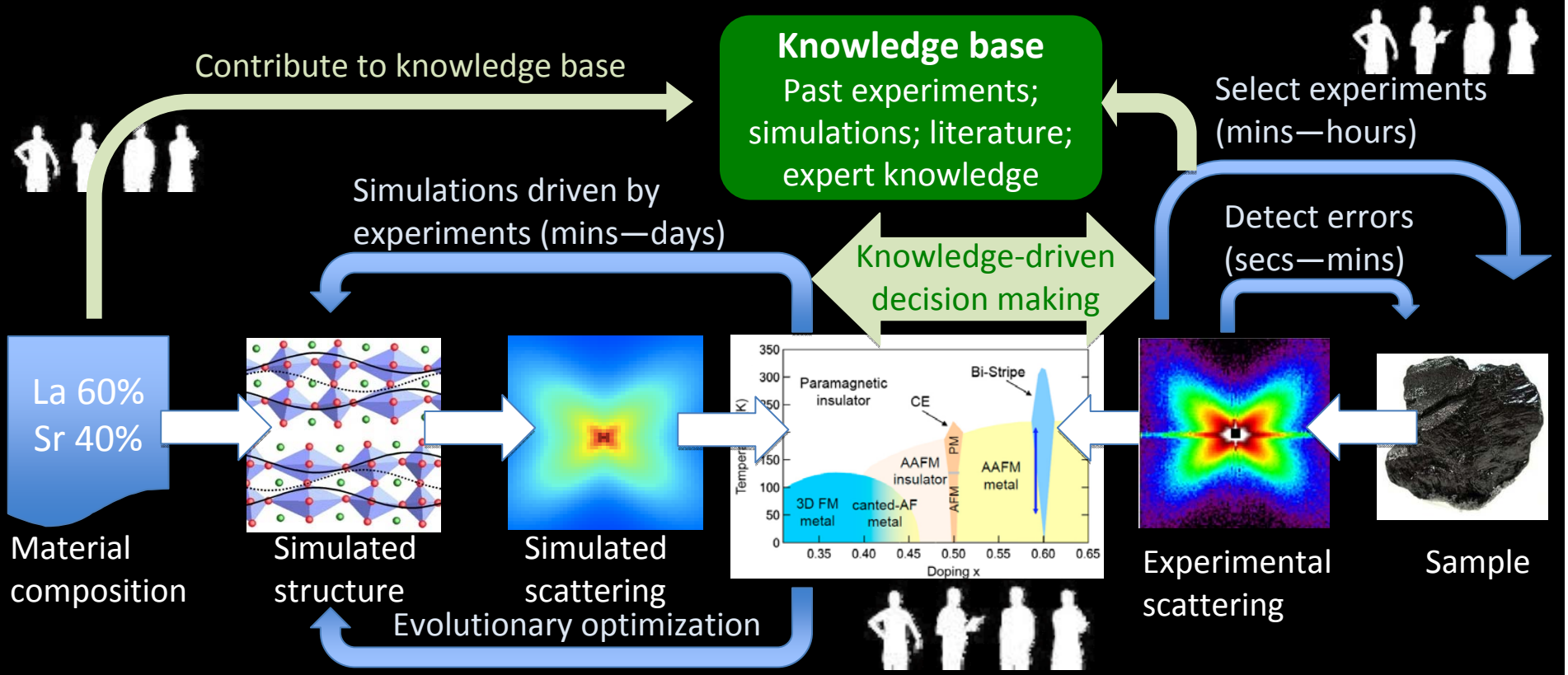
Wellington bucket fountain: https://www.youtube.com/watch?v=_p_FNNDu16w

Linking simulation and experiment to study disordered structures



Diffuse scattering images from Ray Osborn et al., Argonne

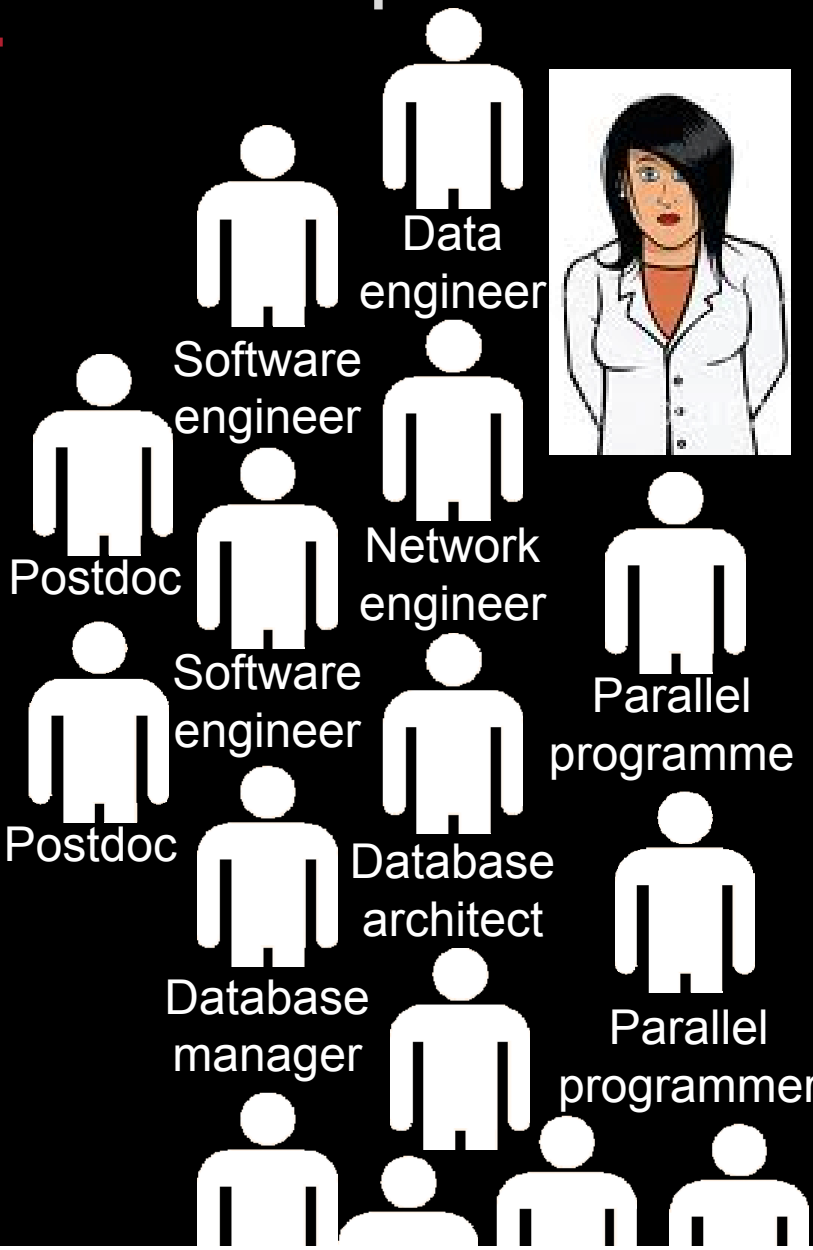
Linking simulation and experiment to study disordered structures

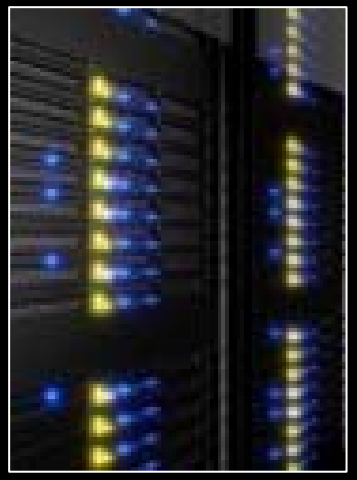


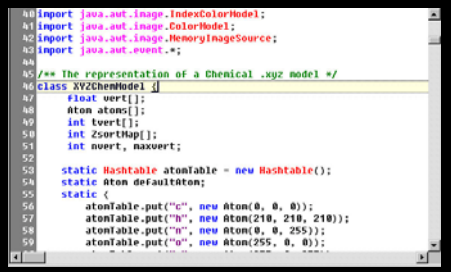
Diffuse scattering images from Ray Osborn et al., Argonne

computationinstitute.org

An expensive business …







A small business, 20 years ago

Personal assistant

IT department

Intern

Secretary

HR manager

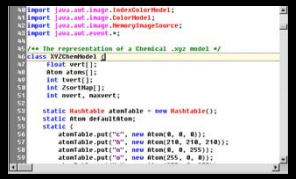
Accountant

Payroll

Marketing

Database manager

Shipping department



A small business, today

"Business cloud" { Reduce costs
Speed innovation
Reliable, scalable, simple }

amazon web services™
salesforce
oDesk Love the way you work.
KnowledgeTree
ADP
liveops
box
xero
Google Apps for Business

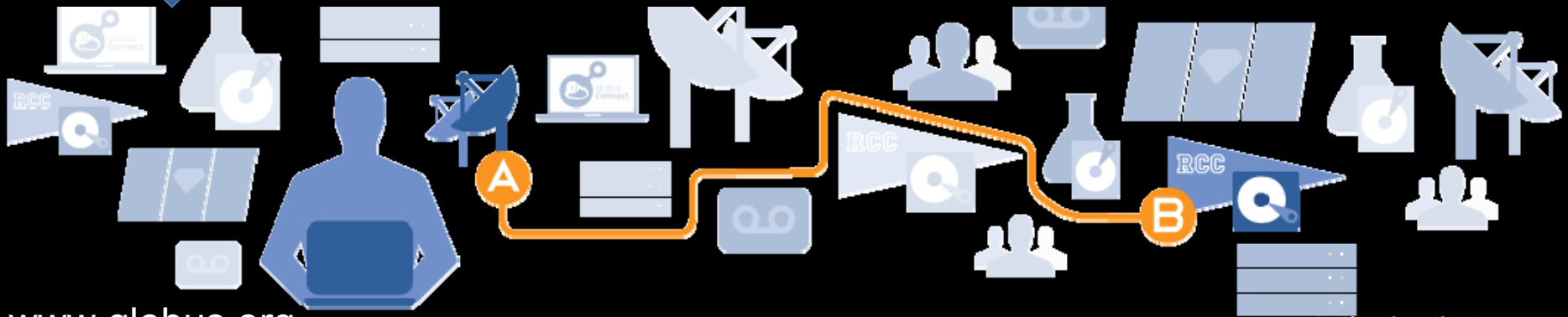
Can we do the same for research?

"Discovery cloud" { Reduce costs; Speed discovery; Reliable, scalable, simple }

?

Discovery cloud: Globus research data management services

File transfer
& sharing



www.globus.org

computationinstitute.org

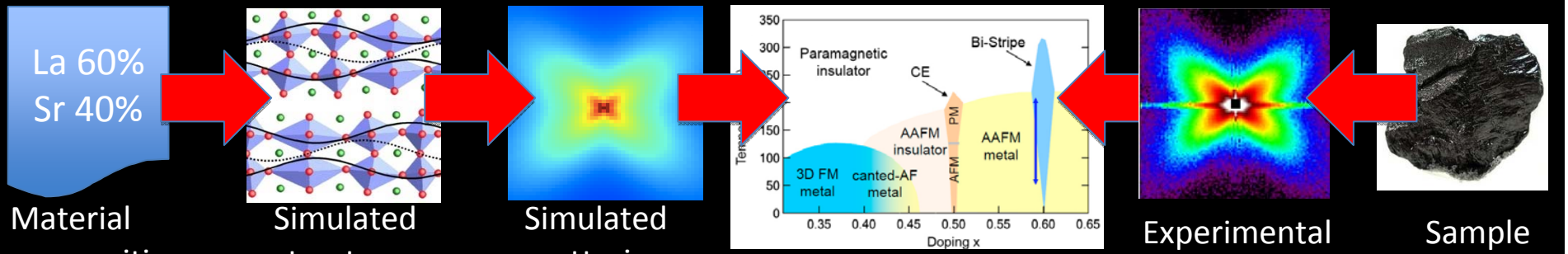
Linking simulation and experiment to study disordered structures

Globus transfer service

Cloud hosted: reliable, secure, fast

20K users, 3B files, 50 PB transferred

Available at **www.globus.org**



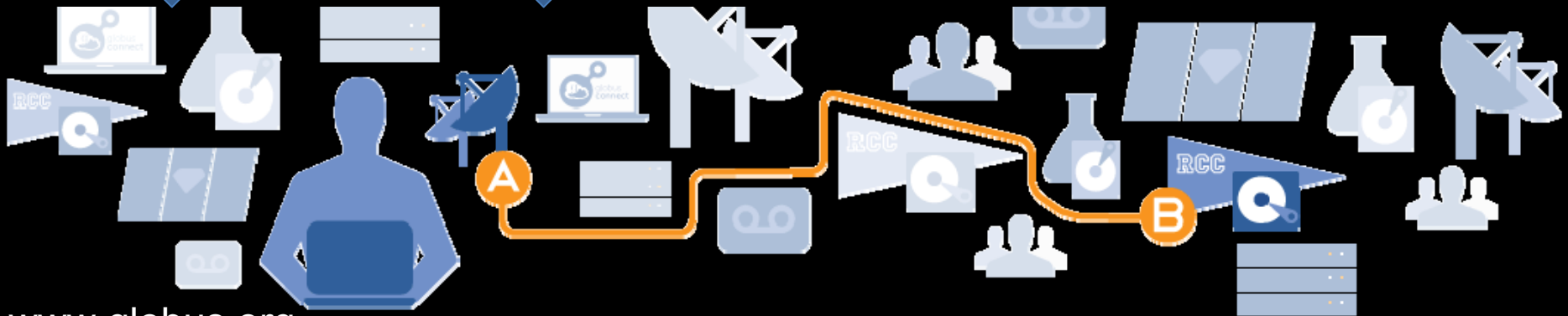
Diffuse scattering images from Ray Osborn et al., Argonne

computationinstitute.org

Discovery cloud: Globus research data management services

File transfer & sharing

Identity & group management

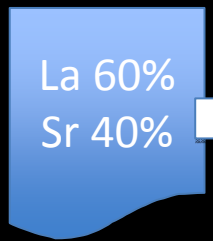


www.globus.org

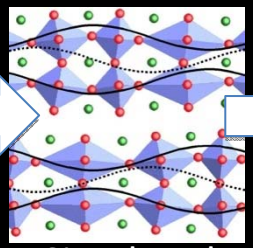
computationinstitute.org

Linking simulation and experiment to study disordered structures

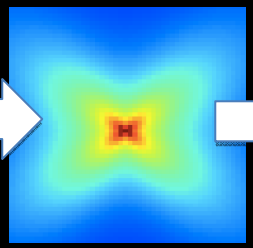
Globus sharing
Identities, groups, profiles
Cloud hosted



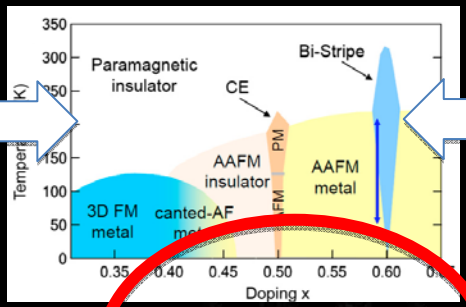
Material composition



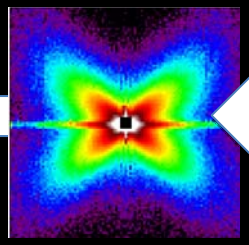
Simulated structure



Simulated scattering

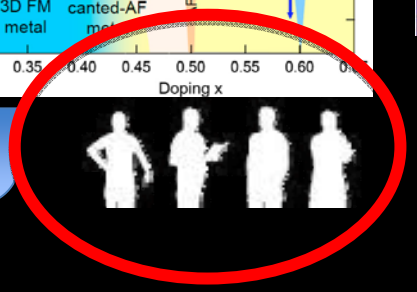


Evolutionary optimization



Experimental scattering

Sample



Diffuse scattering images from Ray Osborn et al., Argonne

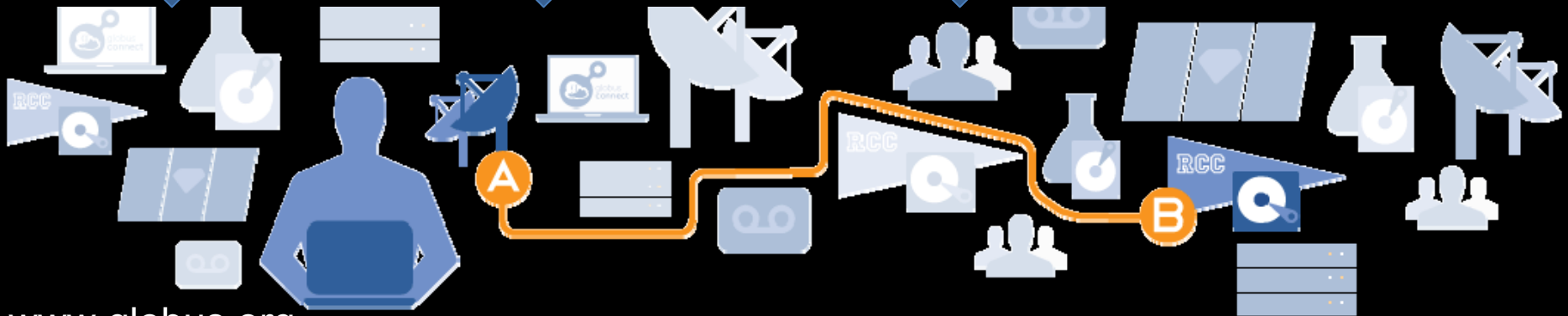
computationinstitute.org

Discovery cloud: Globus research data management services

File transfer & sharing

Identity & group management

Data publication & discovery



www.globus.org

computationinstitute.org

Linking simulation and experiment to study disordered structures

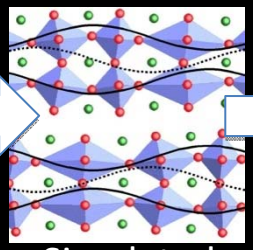
Contribute to knowledge base

Knowledge base
Past experiments; simulations; literature; expert knowledge

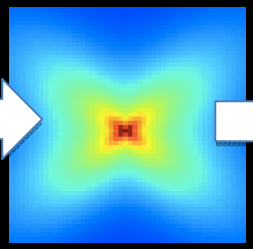
Knowledge-driven decision making

La 60%
Sr 40%

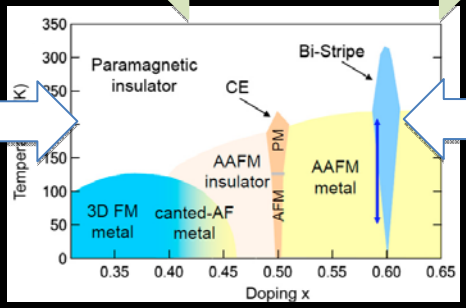
Material composition

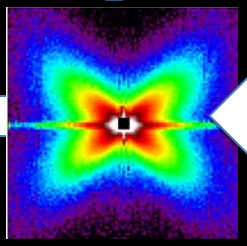


Simulated structure



Simulated scattering





Experimental scattering

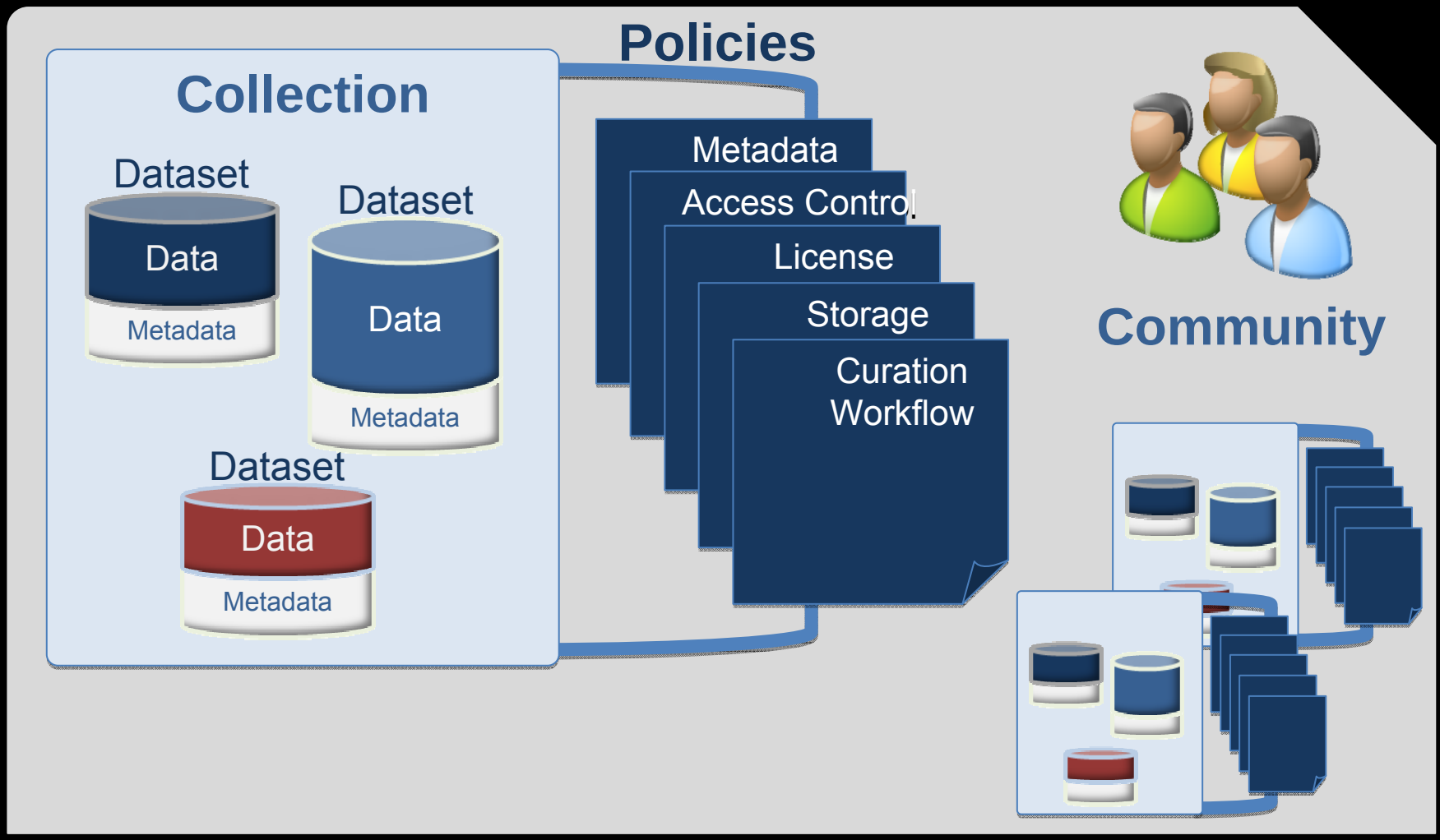
Sample

Globus data publication and discovery
Cloud hosted

Diffuse scattering images from Ray Osborn et al., Argonne

computationinstitute.org

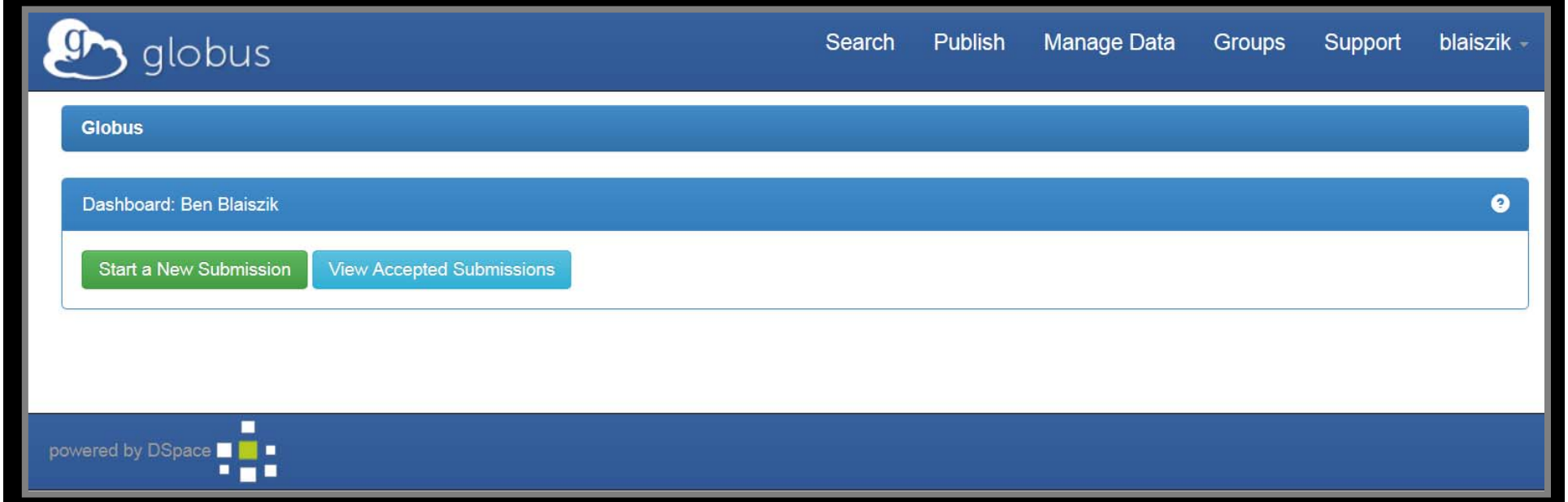
Data publication and discovery



We are looking for pilot users!

computationinstitute.org

Publish dashboard



Start a new submission

globus

blaiszik

Describe | Describe | Upload | Verify | License | Complete

Describe submission: 1) Dublin Core

Submit: Describe this Item

Please fill in the requested information about this submission below. In most browsers, you can use the tab key use the mouse each time.

Enter the names of the authors of this item below.

Authors

Last name, e.g. Smith	First name(s) + "Jr", e.g. Donald Jr	
Blaiszik	Ben	Remove Entry
Kramer	Sharlotte	Remove Entry
Grady	Martha	Remove Entry
Sottos	Nancy	Remove Entry
White	Scott	Remove Entry
		Add More

Enter ORCID of the authors of this item below.

ORCID

http://orcid.org/0000-0002-5326-4789	Remove Entry	http://orcid.org/0000-0002-7346-7621	Remove Entry
http://orcid.org/0000-0001-2034-2932		http://orcid.org/0000-0124-4356-1923	Add More

Enter the main title of the item.

Title *

Autonomic Restoration of Electrical Conductivity

Please give the date of data publication. You can leave out the day and/or month if they aren't applicable.

Date of Issue *

Month: May | Day: 22 | Year: 2012

Enter the standard citation for the publication associated with this item.

Citation

szik, Benjamin J., et al. "Autonomic restoration of electrical conductivity."Advanced Materials 24.3 (2012): 398-401.

Enter the standard citation DOI for the publication associated with this item.

Citation DOI

10.1002/adma.201102888

Cancel/Save | Next >

powered by DSpace

globus

blaiszik

Describe | Describe | Upload | Verify | License | Complete

Describe submission: 2) Science metadata

Submit: Describe this Item

Please fill further information about this submission below.

Enter appropriate subject keywords or phrases below.

Subject Keywords

self-healing — Remove Entry

microcapsules — Remove Entry

cicuit — Remove Entry

Add More

Enter the names of any sponsors and/or funding codes in the box below.

Sponsors

This material is based upon work supported as part of the Center for Electrical Energy Storage - Tailored Interfaces, an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Award Number (919 DOE ANL 9F-31921 NS).

Enter a description for this item in the box below.

Description

Thermomechanical failure of conductive pathways in highly integrated circuits results in loss of function that is often impossible to repair and remains a long-standing problem hindering advanced electronic packaging. Prior approaches to restoration of conductivity rely on external intervention in the form of heating or manual delivery of relatively low conductivity materials. Here, we demonstrate autonomic healing of an electrical circuit with nearly full recovery of conductance (ca. 99%) less than one millisecond after damage. The rapid restorative

Enter the name of experiment for this item below.

Experiment

self-healing-10wtper

Enter the names of materials used in this experiment below.

Material

Gallium — Remove Entry

Indium

Gold — Remove Entry

circuitboard

Add More

Enter the energy density used in this experiment.

Energy Density (mAh/g)

2000

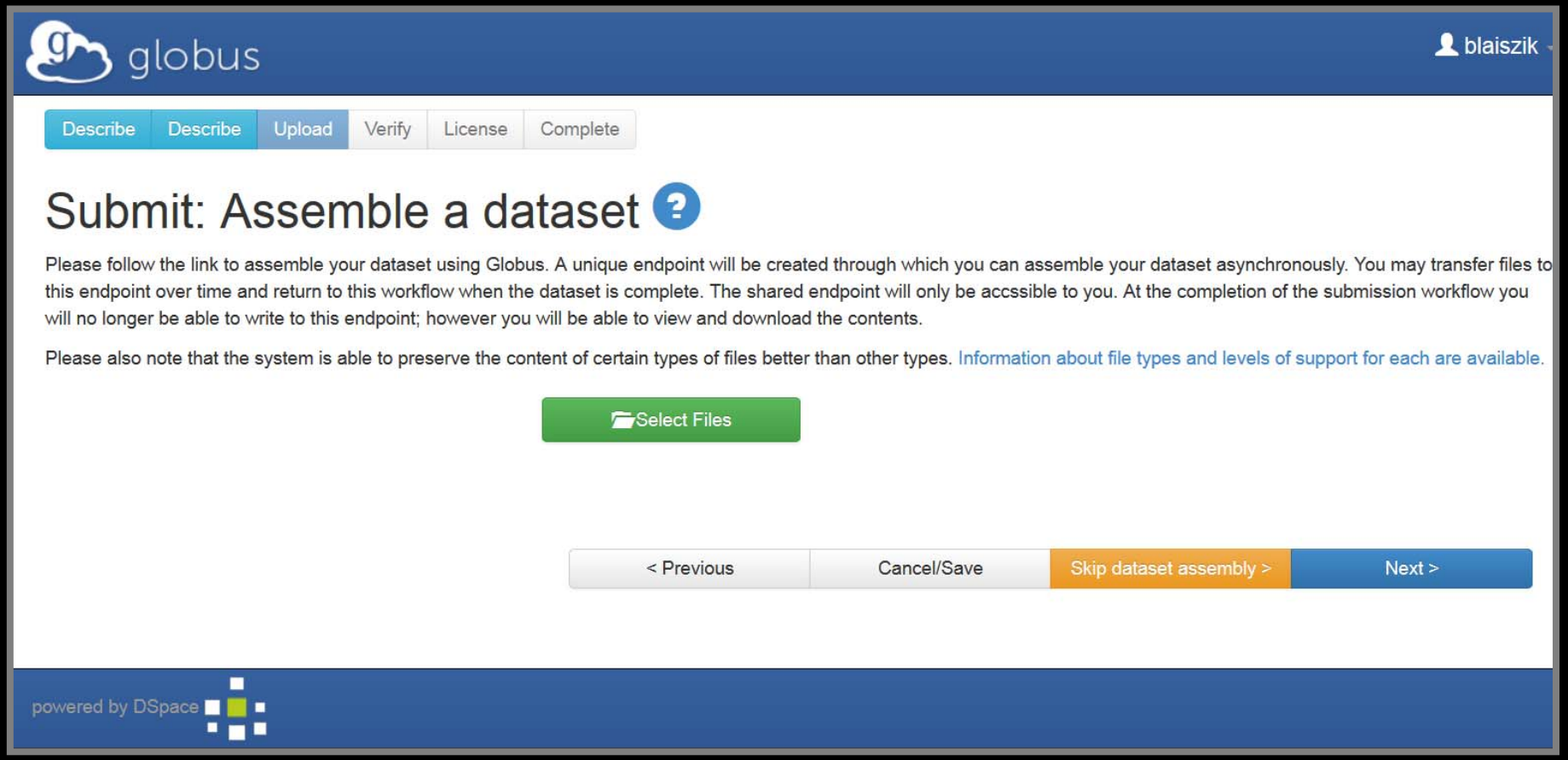
Enter the Argonne GUP that this experiment was conducted under.

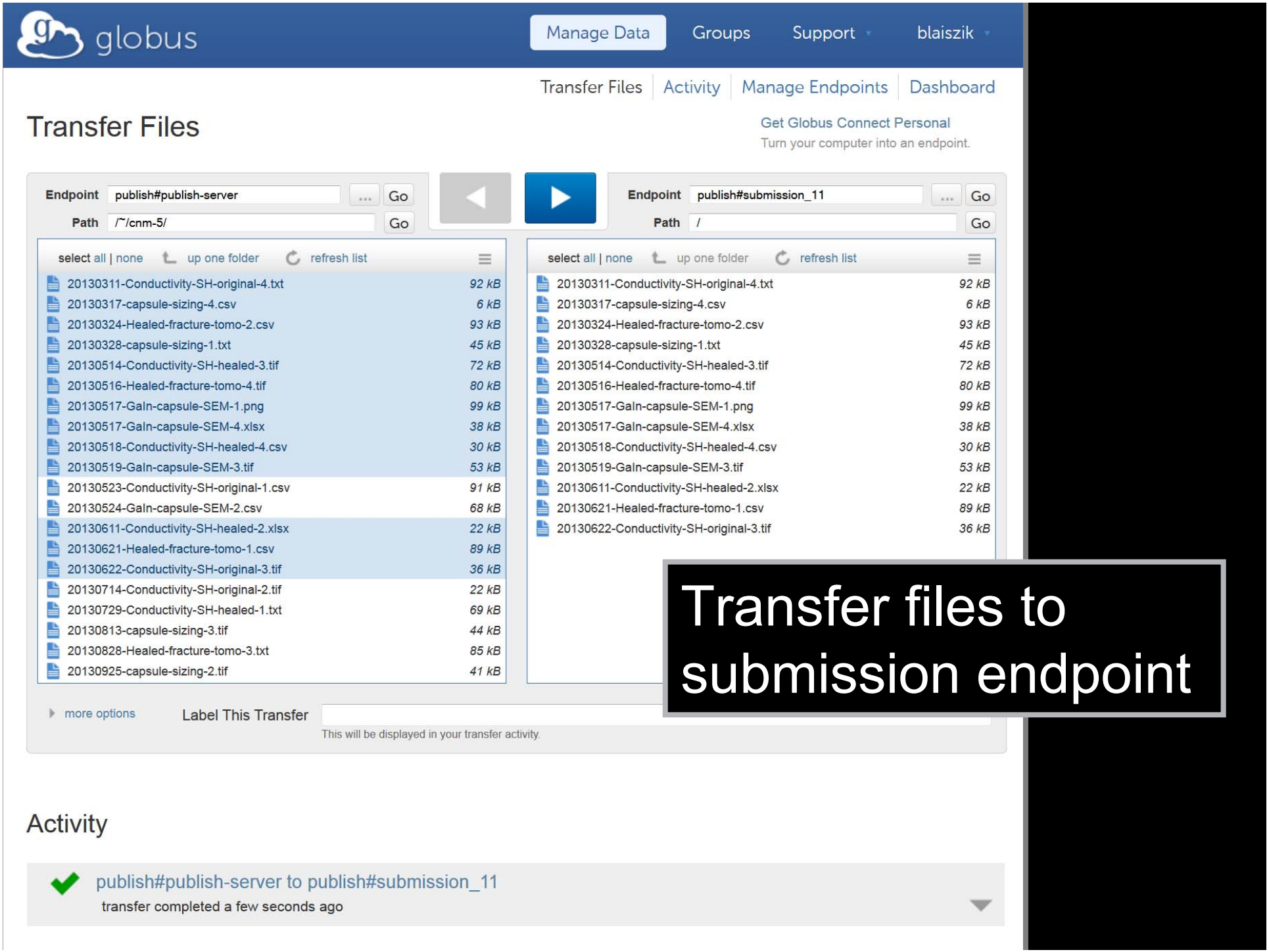
GUP

345-455-2543

< Previous | Cancel/Save | Next >

Assemble the dataset




globus
Manage Data
Groups
Support
blaiszik
Transfer Files
Activity
Manage Endpoints
Dashboard
Transfer Files
Get Globus Connect Personal
Turn your computer into an endpoint.
Endpoint publish#publish-server
Path /~/cnm-5/
Go
select all | none
up one folder
refresh list
20130311-Conductivity-SH-original-4.txt 92 kB
20130317-capsule-sizing-4.csv 6 kB
20130324-Healed-fracture-tomo-2.csv 93 kB
20130328-capsule-sizing-1.txt 45 kB
20130514-Conductivity-SH-healed-3.tif 72 kB
20130516-Healed-fracture-tomo-4.tif 80 kB
20130517-GaIn-capsule-SEM-1.png 99 kB
20130517-GaIn-capsule-SEM-4.xlsx 38 kB
20130518-Conductivity-SH-healed-4.csv 30 kB
20130519-GaIn-capsule-SEM-3.tif 53 kB
20130523-Conductivity-SH-original-1.csv 91 kB
20130524-GaIn-capsule-SEM-2.csv 68 kB
20130611-Conductivity-SH-healed-2.xlsx 22 kB
20130621-Healed-fracture-tomo-1.csv 89 kB
20130622-Conductivity-SH-original-3.tif 36 kB
20130714-Conductivity-SH-original-2.tif 22 kB
20130729-Conductivity-SH-healed-1.txt 69 kB
20130813-capsule-sizing-3.tif 44 kB
20130828-Healed-fracture-tomo-3.txt 85 kB
20130925-capsule-sizing-2.tif 41 kB
Endpoint publish#submission_11
Path /
Go
select all | none
up one folder
refresh list
20130311-Conductivity-SH-original-4.txt 92 kB
20130317-capsule-sizing-4.csv 6 kB
20130324-Healed-fracture-tomo-2.csv 93 kB
20130328-capsule-sizing-1.txt 45 kB
20130514-Conductivity-SH-healed-3.tif 72 kB
20130516-Healed-fracture-tomo-4.tif 80 kB
20130517-GaIn-capsule-SEM-1.png 99 kB
20130517-GaIn-capsule-SEM-4.xlsx 38 kB
20130518-Conductivity-SH-healed-4.csv 30 kB
20130519-GaIn-capsule-SEM-3.tif 53 kB
20130611-Conductivity-SH-healed-2.xlsx 22 kB
20130621-Healed-fracture-tomo-1.csv 89 kB
20130622-Conductivity-SH-original-3.tif 36 kB
more options
Label This Transfer
This will be displayed in your transfer activity.
Activity
publish#publish-server to publish#submission_11
transfer completed a few seconds ago
Transfer files to submission endpoint

globus

blaiszik

Describe | Describe | Upload | Verify | License | Complete

Submit: Dataset Assembled Successfully

Your dataset was successfully assembled.

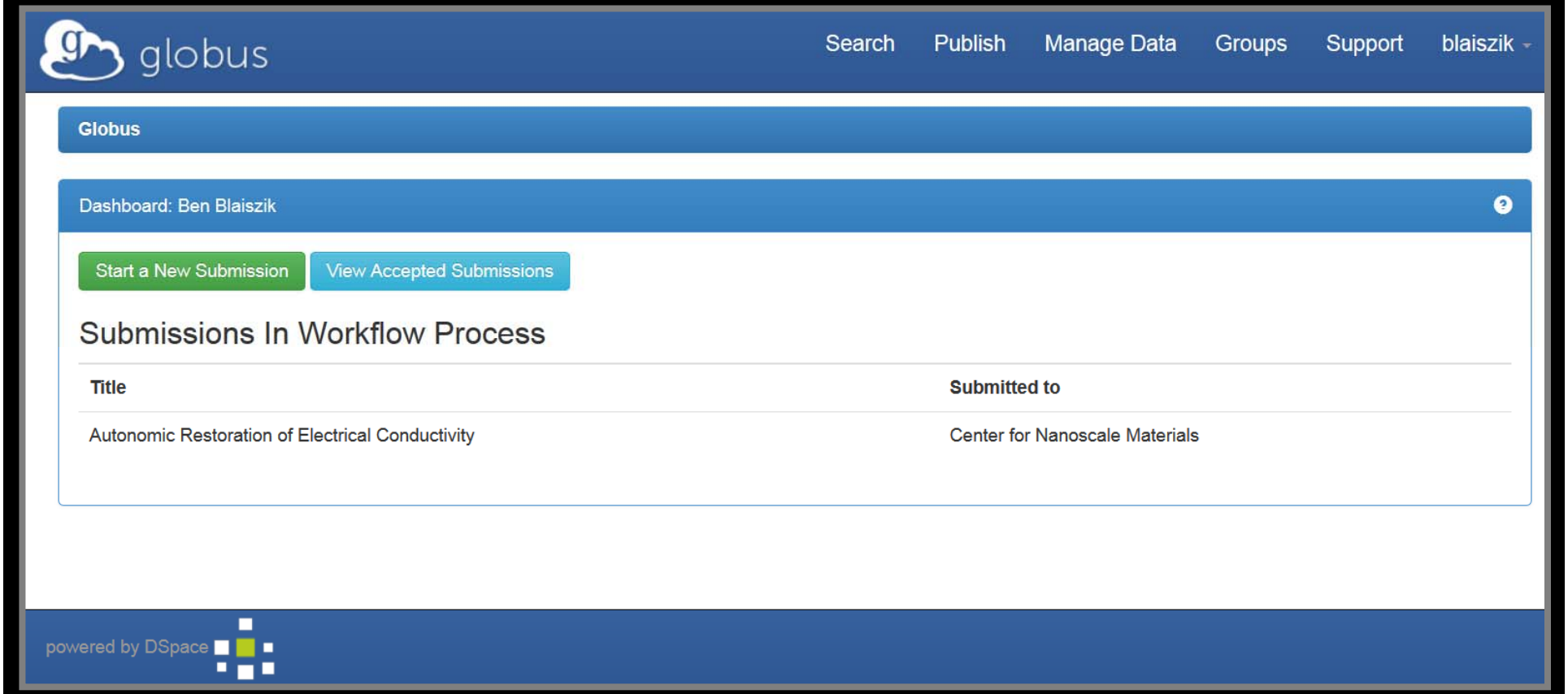
The table below shows the files which are included in this item.

Primary bitstream	File		Size	Description		File Format	
	20130311-Conductivity-SH-original-4.txt	Remove	94208 bytes	*None*	Change	Text (known)	Change
	20130317-capsule-sizing-4.csv	Remove	6144 bytes	*None*	Change	Unknown (unsupported)	Change
	20130324-Healed-fracture-tomo-2.csv	Remove	95232 bytes	*None*	Change	Unknown (unsupported)	Change
	20130328-capsule-sizing-1.txt	Remove	46080 bytes	*None*	Change	Text (known)	Change
	20130514-Conductivity-SH-healed-3.tif	Remove	73728 bytes	*None*	Change	TIFF (known)	Change
	20130516-Healed-fracture-tomo-4.tif	Remove	81920 bytes	*None*	Change	TIFF (known)	Change
	20130517-GaIn-capsule-SEM-1.png	Remove	101376 bytes	*None*	Change	image/png (known)	Change
	20130517-GaIn-capsule-SEM-4.xlsx	Remove	38912 bytes	*None*	Change	Microsoft Excel XML (known)	Change
	20130518-Conductivity-SH-healed-4.csv	Remove	30720 bytes	*None*	Change	Unknown (unsupported)	Change
	20130519-GaIn-capsule-SEM-3.tif	Remove	54272 bytes	*None*	Change	TIFF (known)	Change
	20130611-Conductivity-SH-healed-2.xlsx	Remove	22528 bytes	*None*	Change	Microsoft Excel XML (known)	Change
	20130621-Healed-fracture-tomo-1.csv	Remove	91136 bytes	*None*	Change	Unknown (unsupported)	Change
	20130622-Conductivity-SH-original-3.tif	Remove	36864 bytes	*None*	Change	TIFF (known)	Change

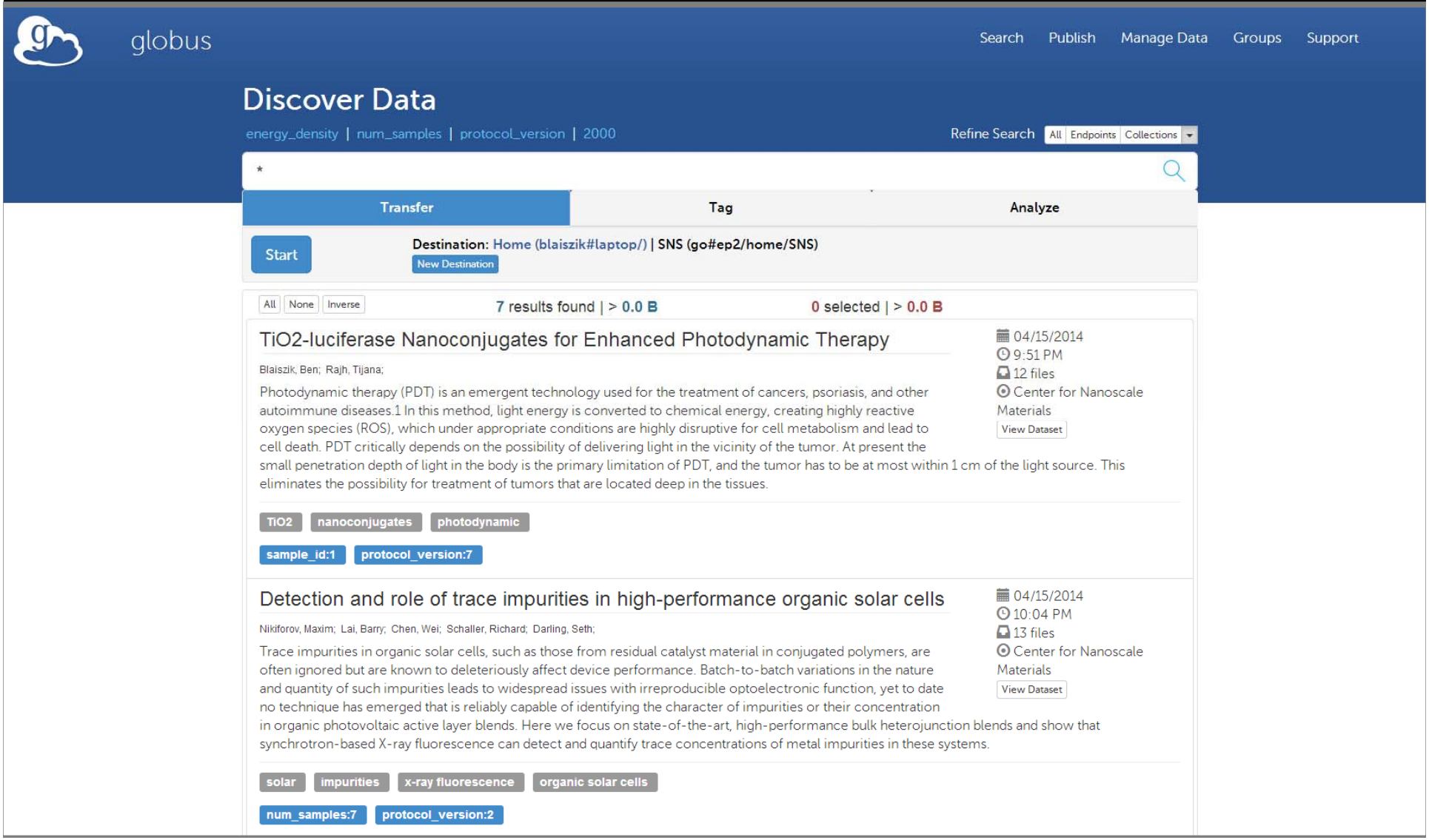
Add Another File

Check dataset is assembled correctly

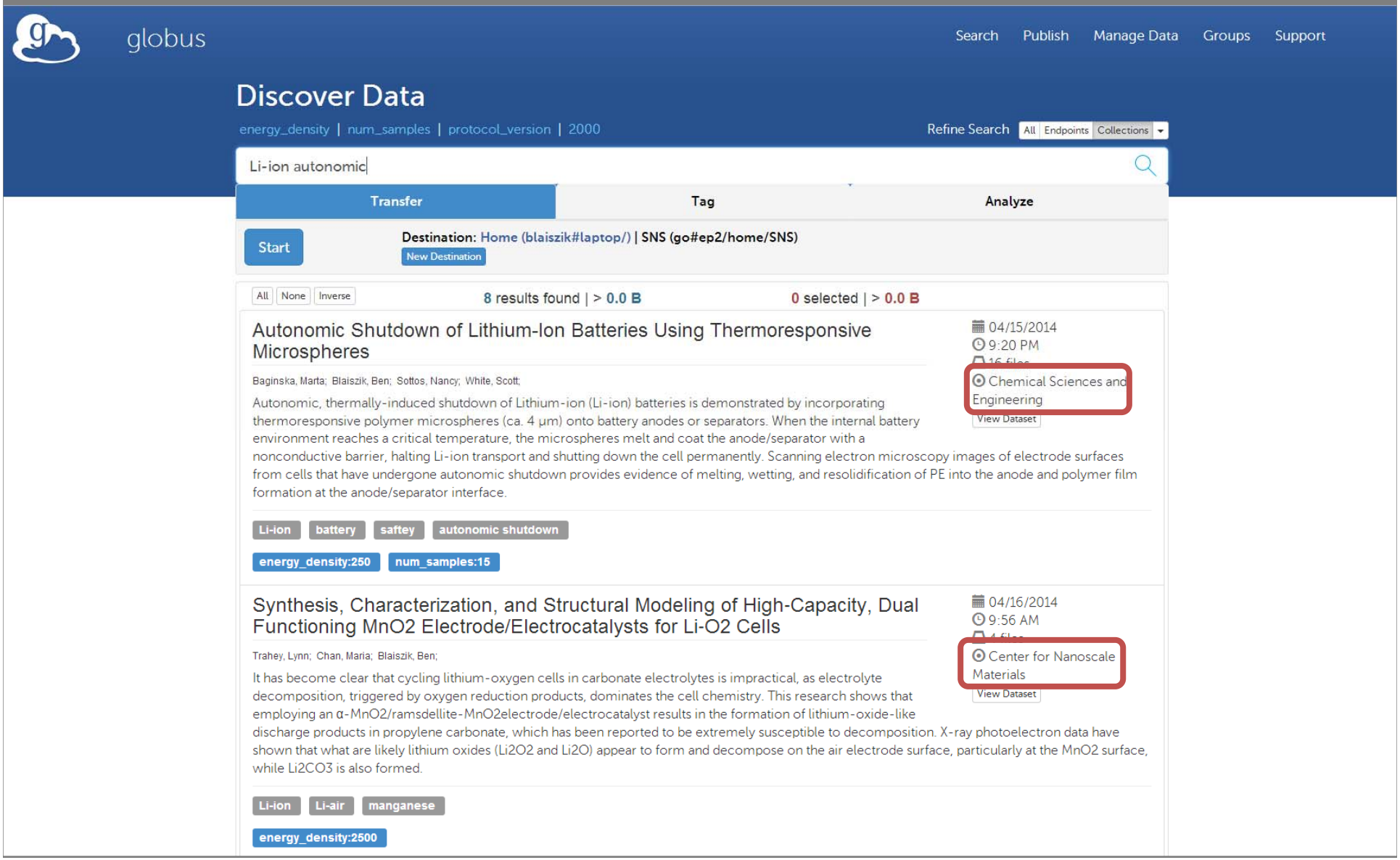
Submission now in curation workflow



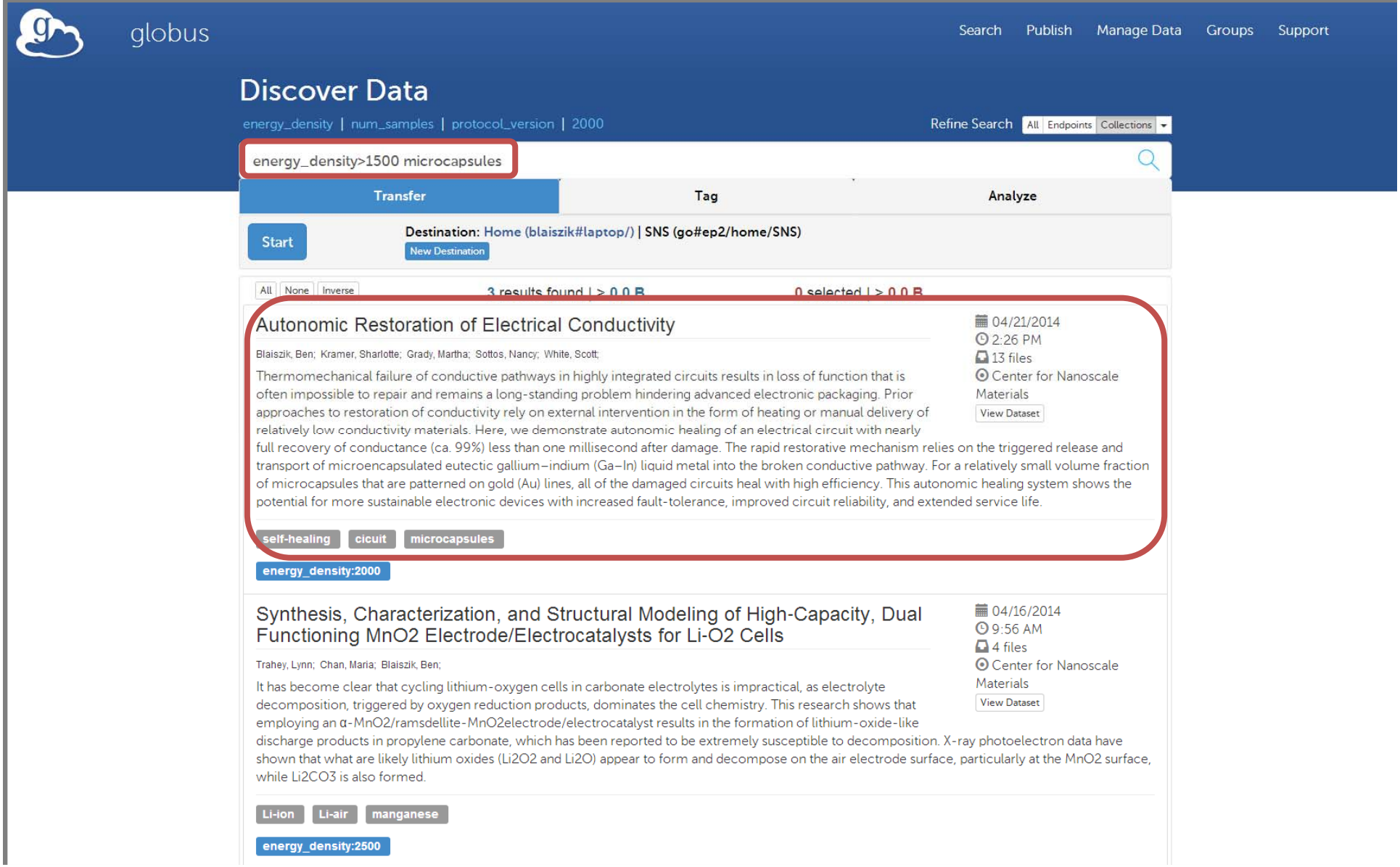
Search published datasets



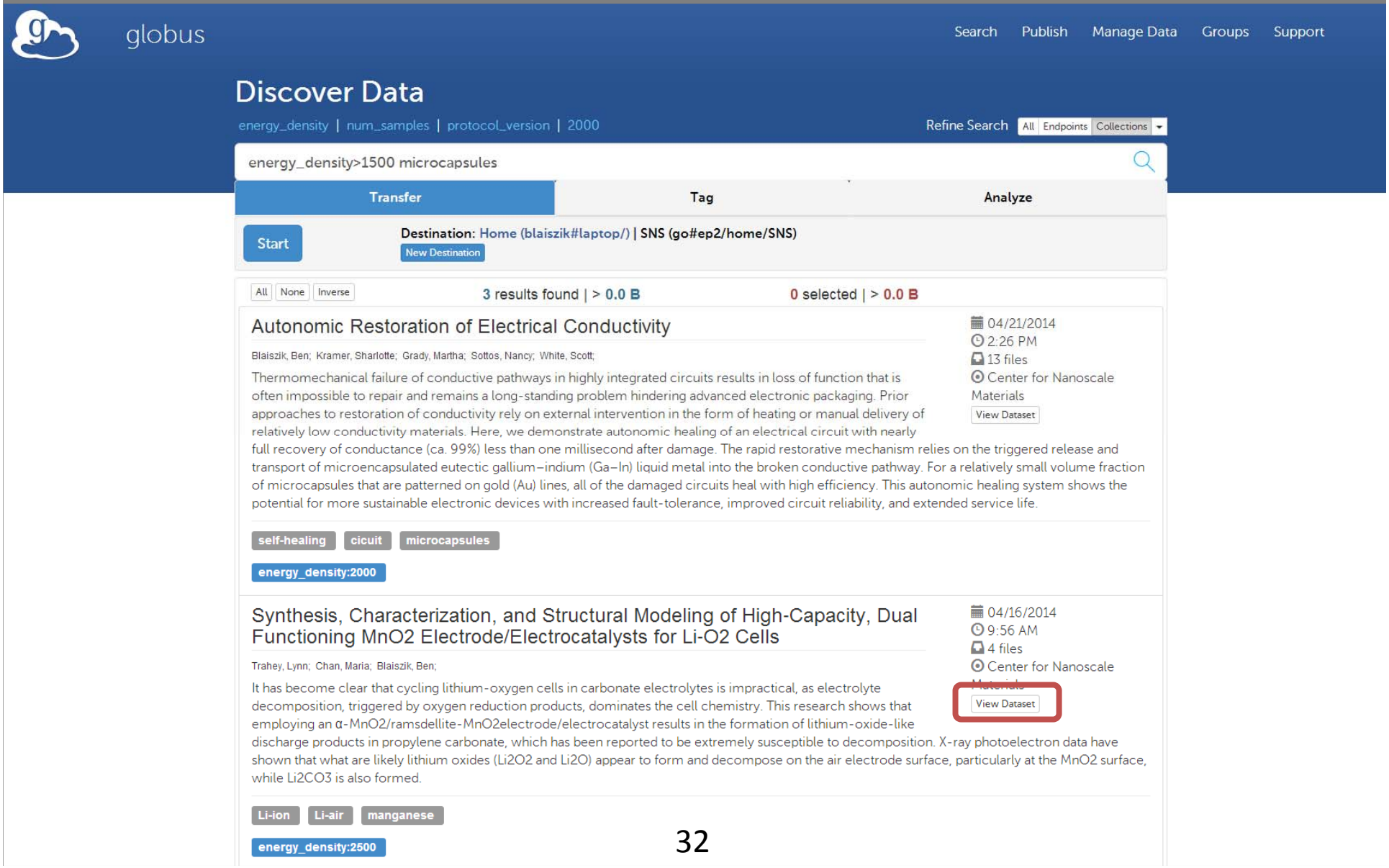
Search across collections



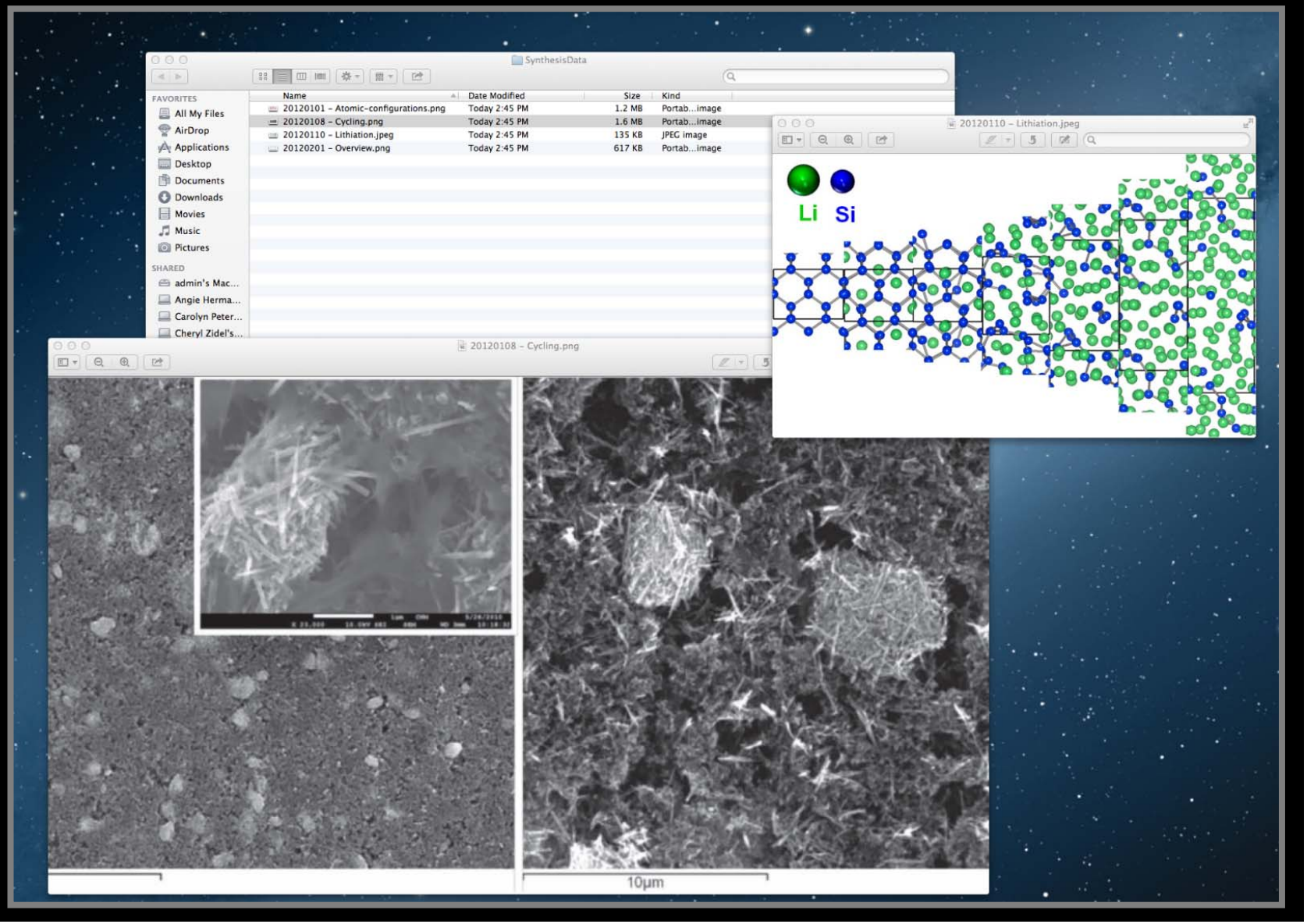
Discover a published dataset

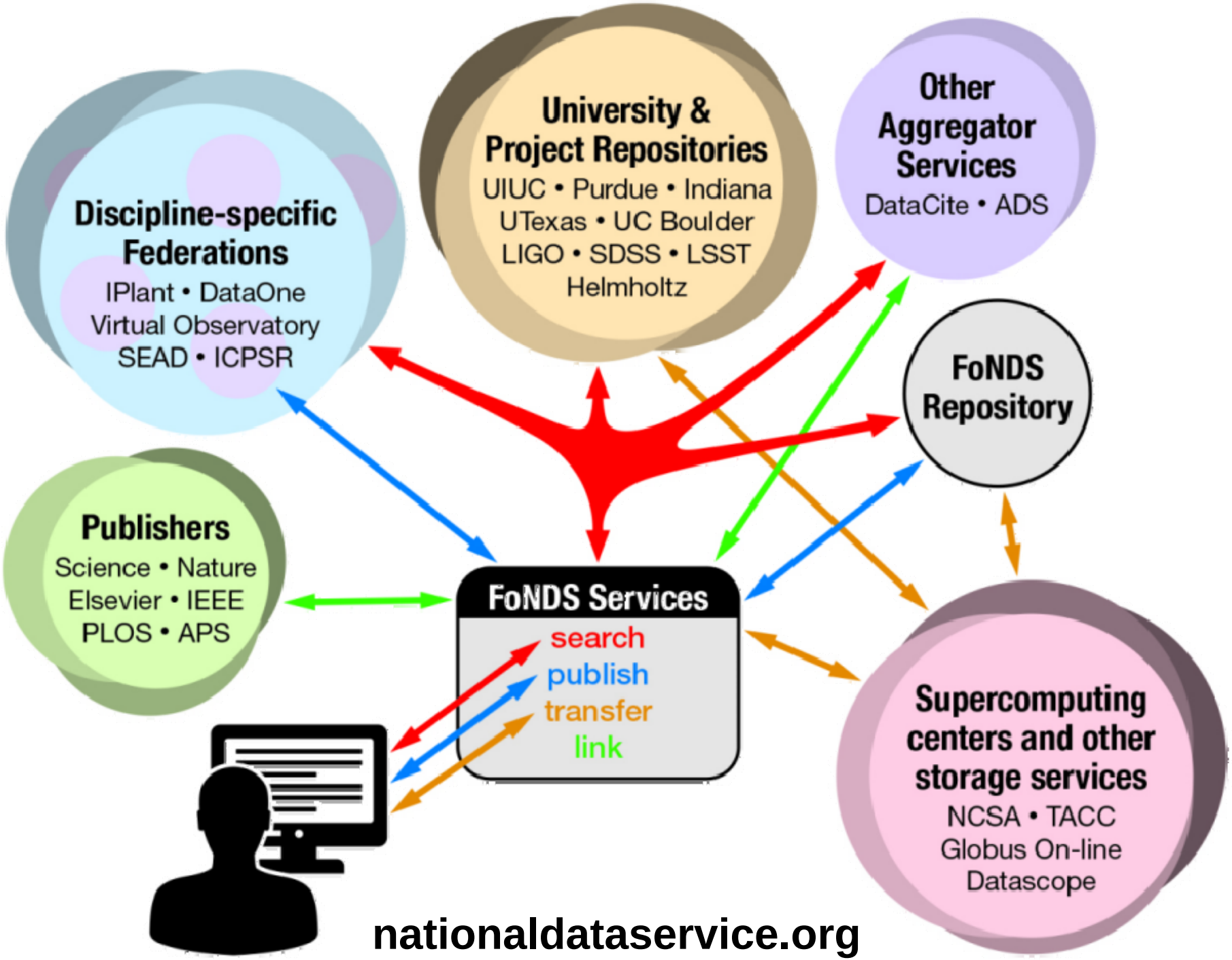


Select a published dataset



View downloaded dataset




Discipline-specific Federations
IPlant • DataOne
Virtual Observatory
SEAD • ICPSR
University & Project Repositories
UIUC • Purdue • Indiana
UTexas • UC Boulder
LIGO • SDSS • LSST
Helmholtz
Other Aggregator Services
DataCite • ADS
FoNDS Repository
Publishers
Science • Nature
Elsevier • IEEE
PLOS • APS
FoNDS Services
search
publish
transfer
link
Supercomputing centers and other storage services
NCSA • TACC
Globus On-line
Datascope
nationaldataservice.org

MATERIALS DATA FACILITY LAUNCHED IN SUPPORT OF MATERIALS GENOME INITIATIVE

06.19.14 - *Permalink*

The National Data Service Consortium is launching a materials data facility for the advancement of materials science research through open data access and sharing.

On the third anniversary of President Barack Obama establishing the Materials Genome Initiative (MGI)—a multi-agency effort to transform materials science research in the United States through a national infrastructure—a consortium of research universities, national laboratories, and academic publishers announced the Materials Data Facility today.

This new facility is being established as a pilot program under the National Data Service (NDS) and will provide a repository where scientists can preserve and share materials research data, produced by both simulations and experiments.

NDS is a new emerging vision for a national data infrastructure that enables the discovery, reuse, and publication of data for scientists and researchers across all disciplines. Sharing in this vision, the Materials Data Facility will push the MGI's goals of doubling the pace of development of advanced materials research.

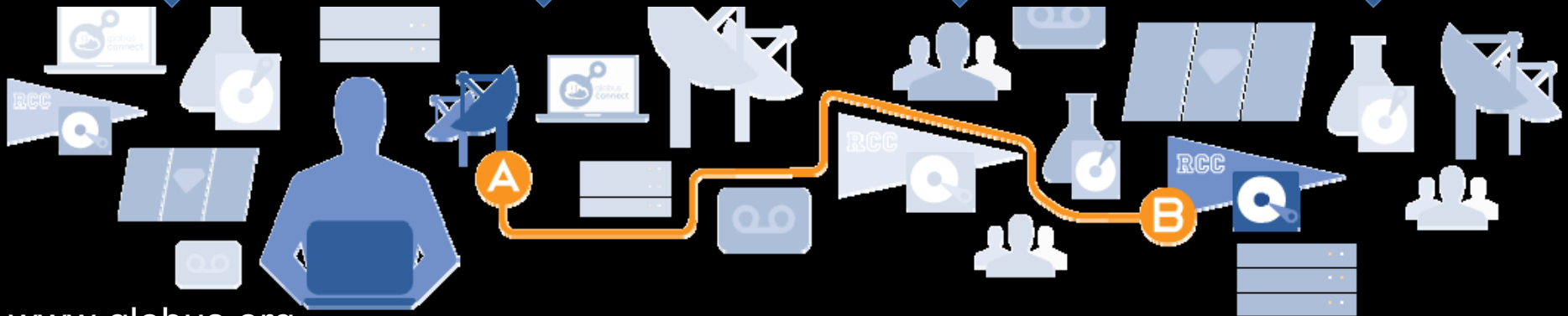
Discovery cloud: Globus research data management services

File transfer & sharing

Identity & group management

Data publication & discovery

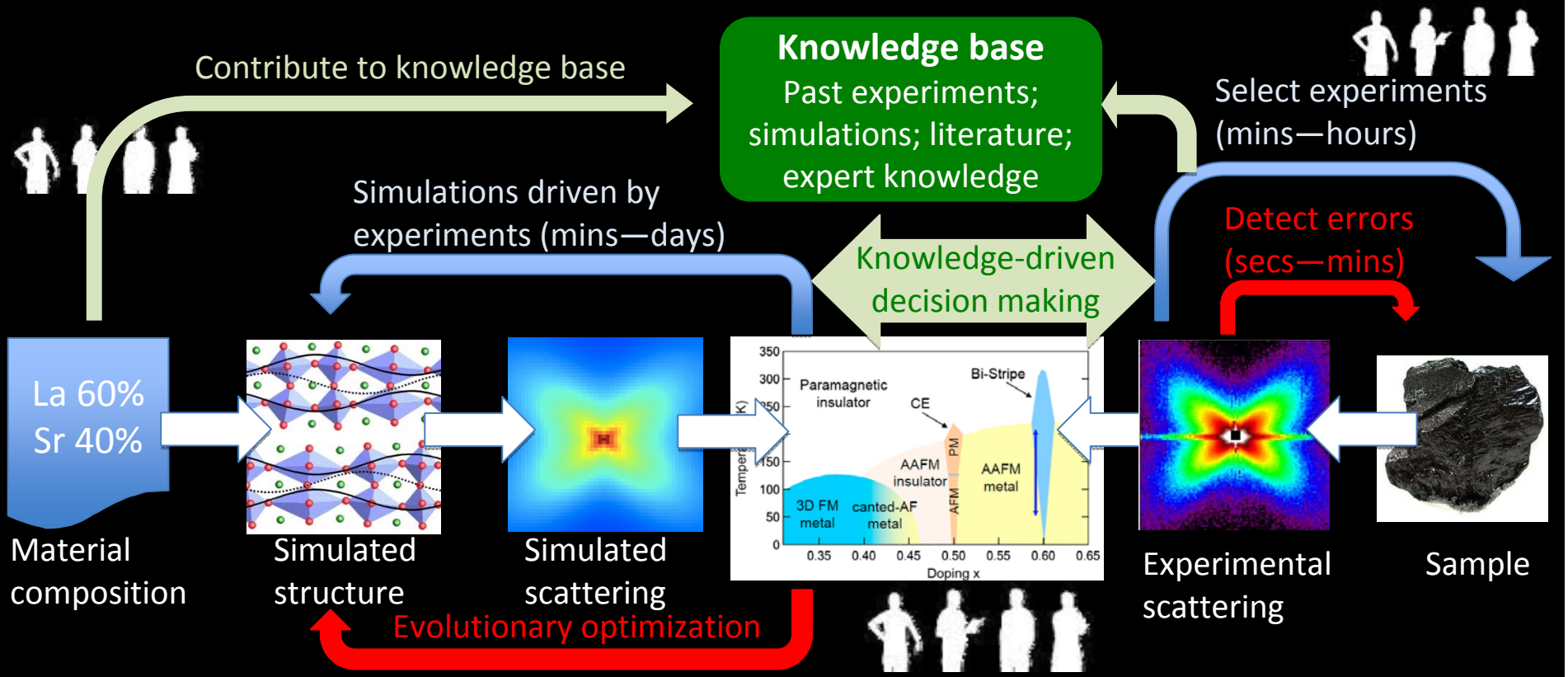
Simulation & data analysis



www.globus.org

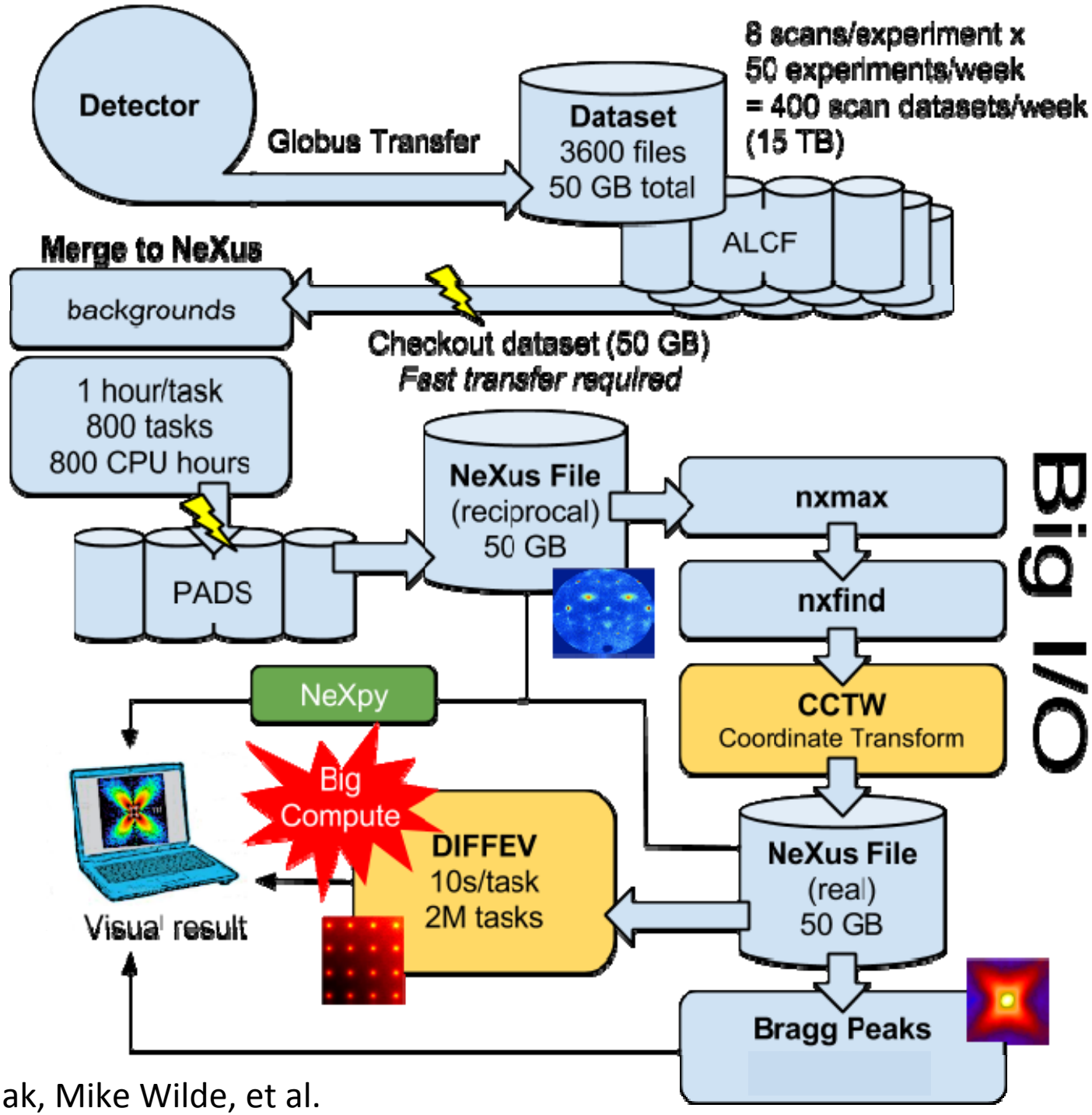
computationinstitute.org

Linking simulation and experiment to study disordered structures



Diffuse scattering images from Ray Osborn et al., Argonne

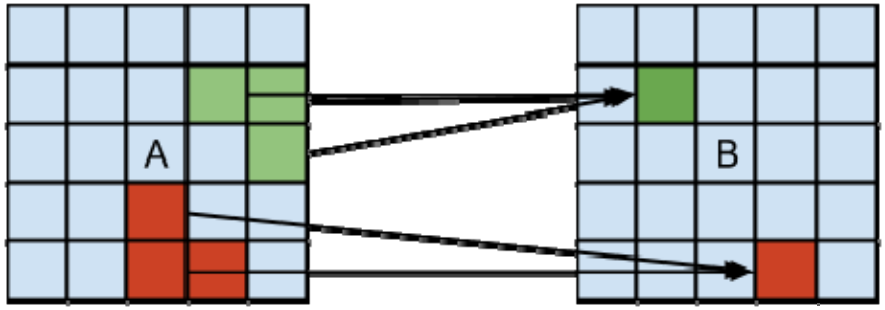
computationinstitute.org



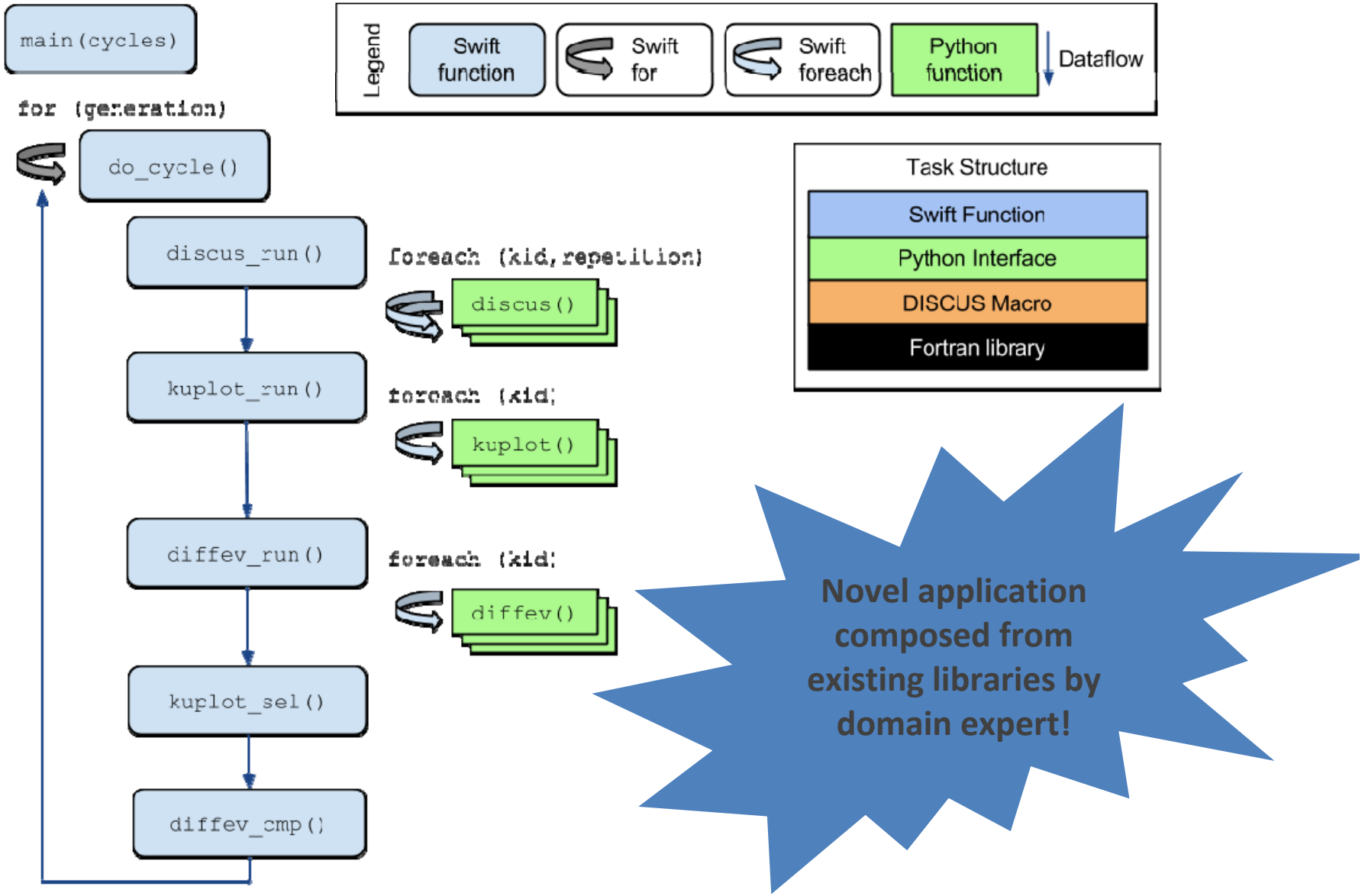
Justin Wozniak, Mike Wilde, et al.

CCTW: Swift/T application (C++)

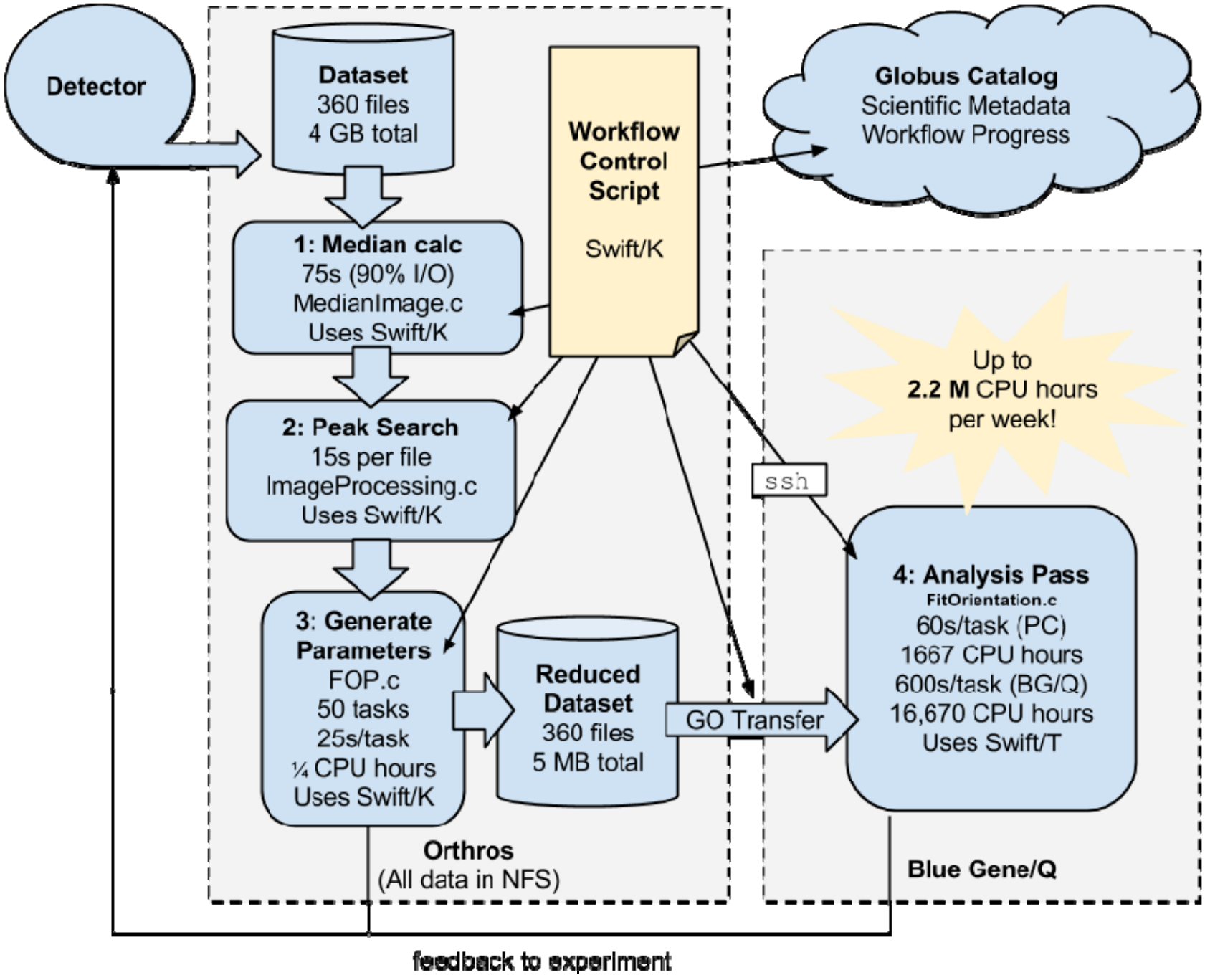
```
bag<blob> M[];
foreach i in [1:n] {
    blob b1= cctw_input("pznpt.nxs");
    blob b2[];
    int outputId[];
    (outputId, b2) = cctw_transform(i, b1);
    foreach b, j in b2 {
        int slot = outputId[j];
        M[slot] += b;
    }
    foreach g in M {
        blob b = cctw_merge(g);
        cctw_write(b);
    }
```



DIFFEV: Genetic algorithm via dataflow



Justin Wozniak, Mike Wilde, et al.


Detector
Dataset
360 files
4 GB total
Workflow Control Script
Swift/K
Globus Catalog
Scientific Metadata
Workflow Progress
1: Median calc
75s (90% I/O)
MedianImage.c
Uses Swift/K
2: Peak Search
15s per file
ImageProcessing.c
Uses Swift/K
3: Generate Parameters
FOP.c
50 tasks
25s/task
¼ CPU hours
Uses Swift/K
Reduced Dataset
360 files
5 MB total
GO Transfer
ssh
Up to
2.2 M CPU hours
per week!
4: Analysis Pass
FitOrientation.c
60s/task (PC)
1667 CPU hours
600s/task (BG/Q)
16,670 CPU hours
Uses Swift/T
Orthros
(All data in NFS)
Blue Gene/Q
feedback to experiment

Simulation and data analysis: Point and click parallelism

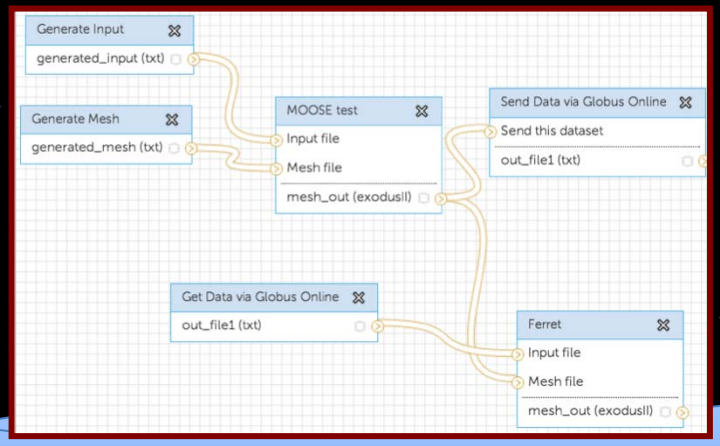
Capture domain knowledge: data and code

Data space
Local and remote datasets

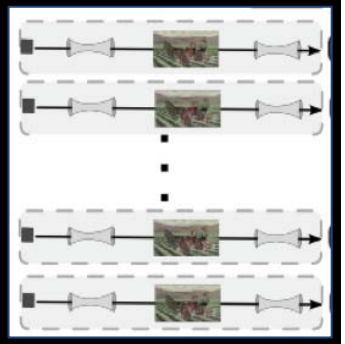
Tool shed
Simulation models & analysis

Workflows
Link data, tools in reusable form

Reusable workflows encode commonly used modeling and analysis pipelines



Large simulation campaigns



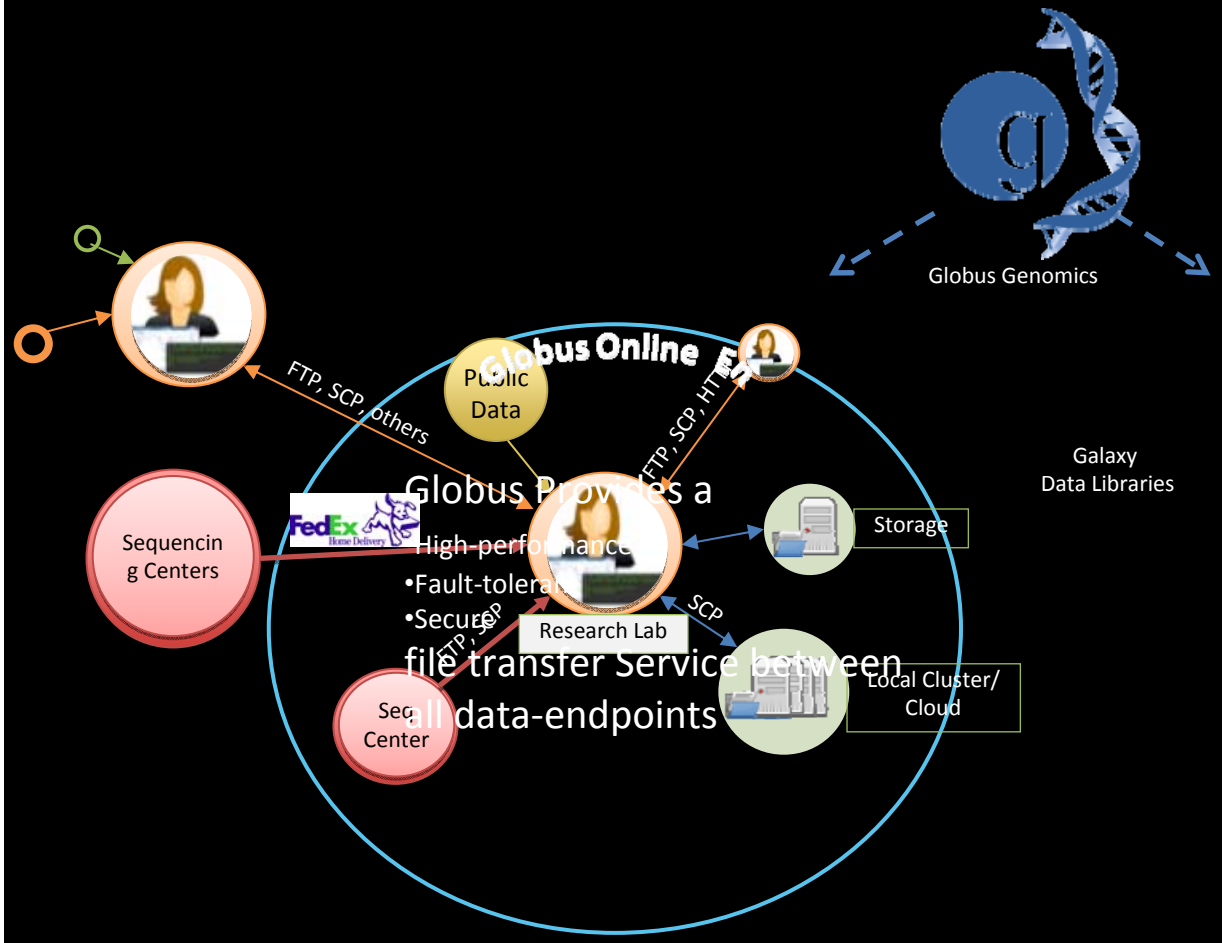
Hosted on Amazon cloud for reliable, on-demand access and scalability

Builds on widely used Galaxy, Globus, and Swift systems

galaxyproject.org □ globus.org □ swift-lang.org

computationinstitute.org

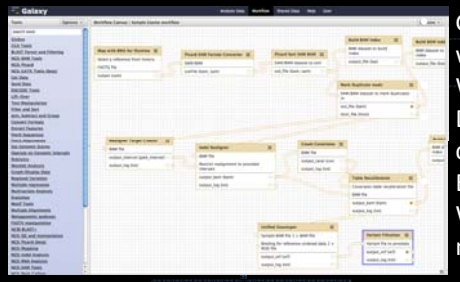
Experiences in genomics



Data Management

Galaxy Based Workflow Management System

Galaxy Data Libraries



Globus Integrated within Galaxy
Web-based UI
Drag-Drop workflow creation
Easily modify Workflows with new tools

Analytical tools are automatically run on scalable compute resources when possible

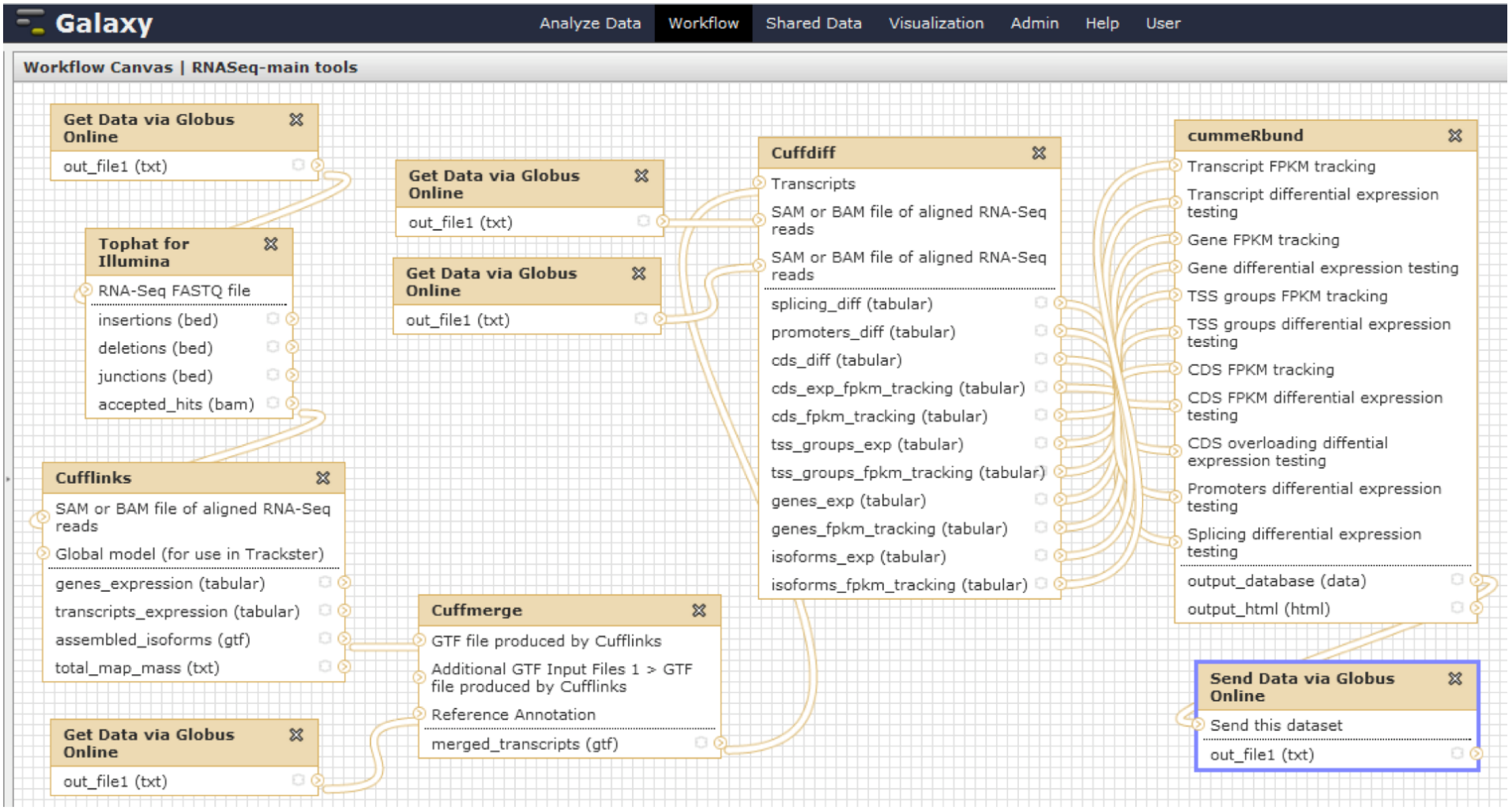
Amazon Web Services

Data Analysis

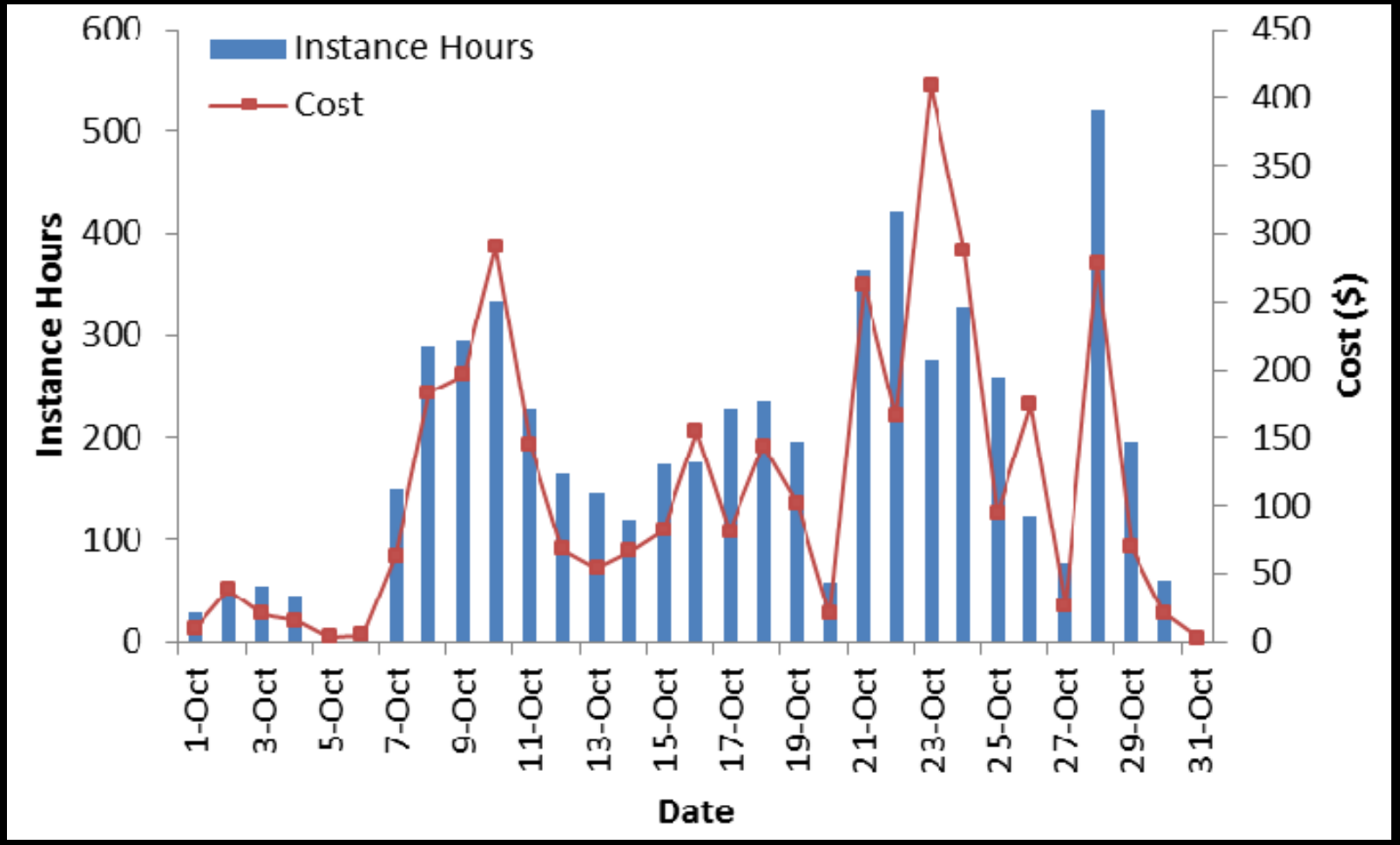
Globus Genomics: Ravi Madduri, Paul Davé, et al.

computationinstitute.org

RNA-Seq pipeline



Globus Genomics on AWS



800K Core hours in last 6 months

Discovery Cloud: Three common themes

1) Accelerate discovery via automation

2) Slash costs of trying new methods
   - No local software installation
   - No need to read manual
   - On-demand, elastic scalability
   - Low operational costs, proactive support

3) Make data preservation trivial

Take away messages

- Data has a dual nature: rare treasure and chaotic deluge

- Science must embrace this duality
  - **Treasure**: Store, curate, index, preserve
  - Deluge: Slash management costs, to both accelerate use & facilitate data preservation
- Cloud services can help in both areas

Thanks to great colleagues and collaborators

- Rachana Ananthakrishnan, Ben Blaiszik, Kyle Chard, Raj Kettimuthu, Ravi Madduri, Tanu Malik, Steve Tuecke, Justin Wozniak, and other CS colleagues
- Ray Osborn, Francesco de Carlo, Chris Jacobsen, Nicola Ferrier, and other Argonne scientists
- Juan de Pablo, Peter Voorhees, and other NIST CHiMaD participants

Thank you to our sponsors!

U.S. DEPARTMENT OF ENERGY

THE UNIVERSITY OF CHICAGO

Argonne NATIONAL LABORATORY

NIST

amazon web services™

computationinstitute.org