National Data Service

TOWARDS
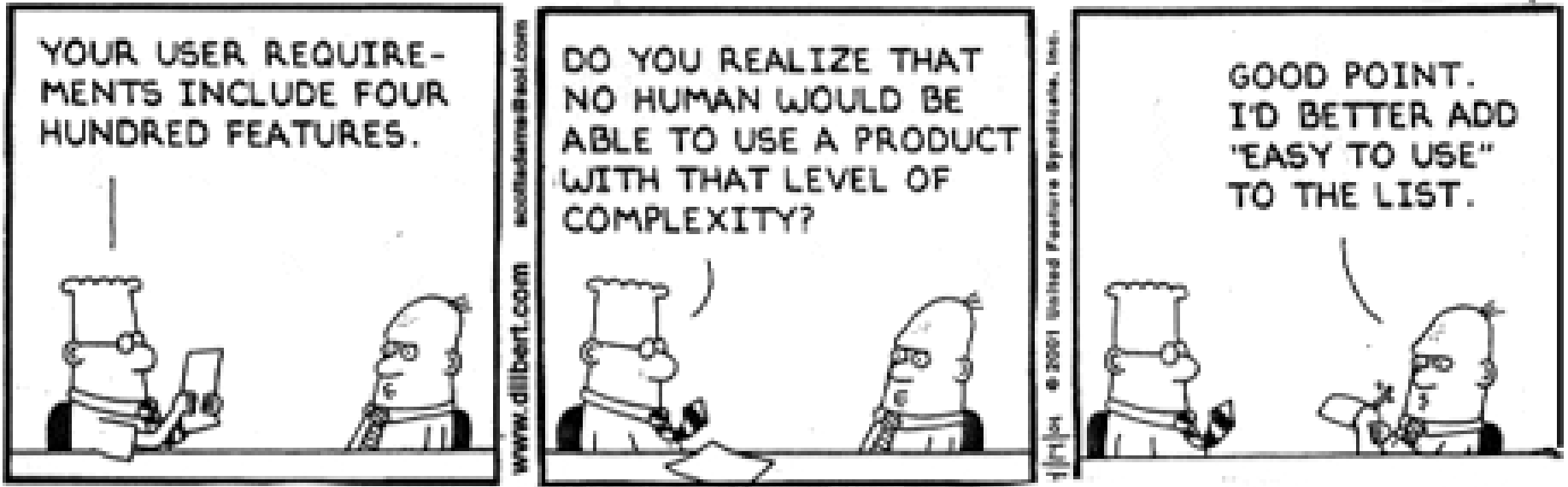# AN ARCHITECTURE
FOR NATIONAL DATA SERVICES

Ian Foster

Director, Computation Institute

Argonne National Laboratory
& The University of Chicago
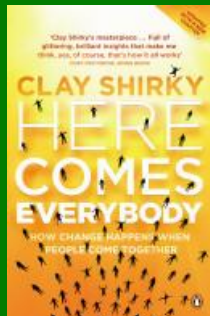
@ianfoster

ianfoster.org

Architecture?

Principle 1: Reduce data friction

Make simple things easy

Make hard things possible

For example, **publish-then-filter**
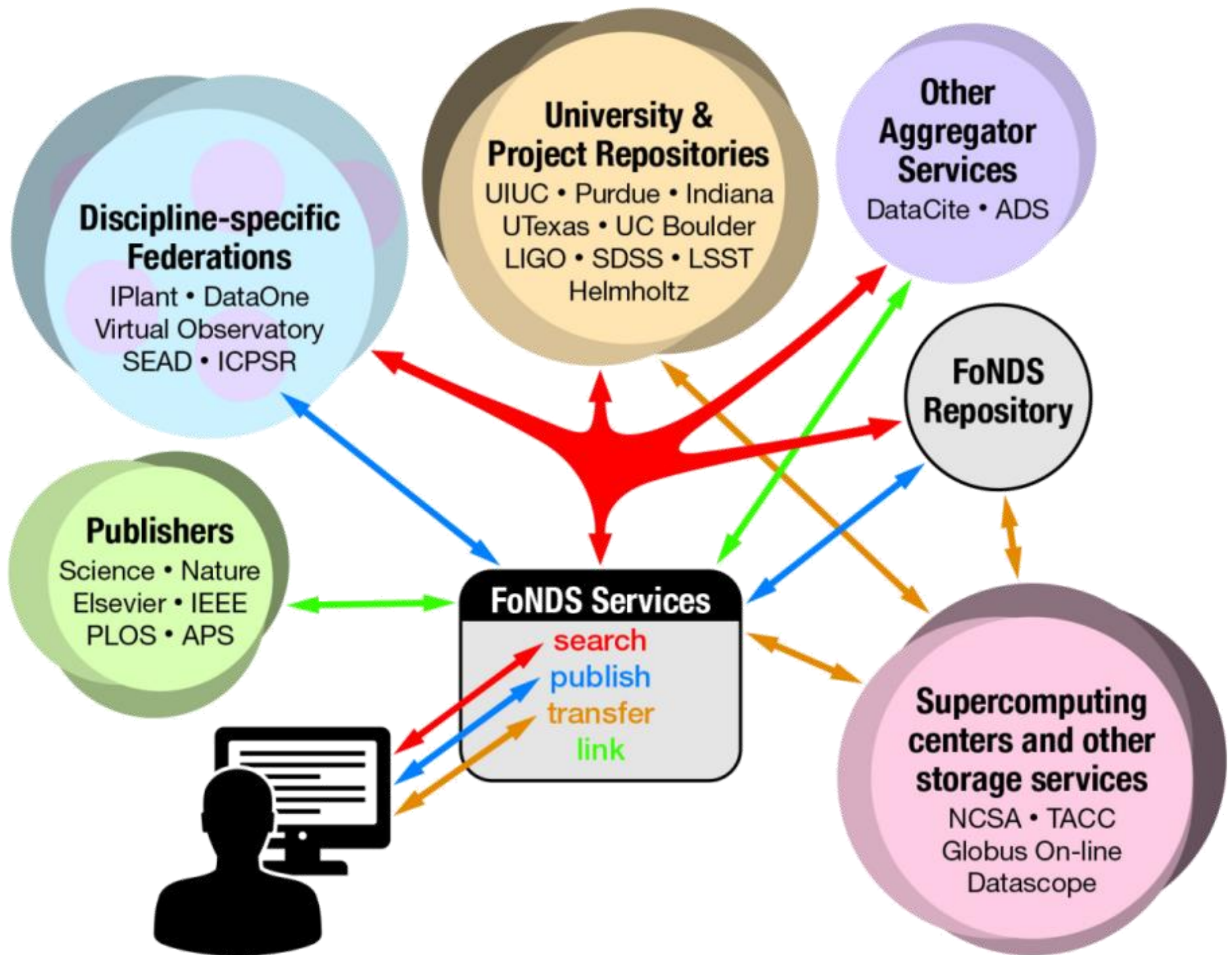
For example, cloud-hosted **software-as-a-service**

Principle 2:
Small pieces, loosely joined

- Storage systems
- Content management
- Analysis systems
- Registries
- Identity management
- Data movers
- … and many more …

- REST interfaces
- Open
- Simple
- Composable
- Extensible
- Versioned

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

Principle 3: Insist on stories

- For example:
  - "I need to store/backup/archive my data"
  - "I need to transfer/mirror my data"
  - "I need to share my data"
  - "I need to publish my data"
  - "I need to discover published data"
  - "I need to analyze my data"
- Good stories are detailed, urgent, popular

We have much to build on and/or integrate with

- Agave
- Brown Dog
- DataCite
- DATAone
- Dataverse
- Earth System Grid
- Globus
- Globus Connect

- InCommon
- iPlant
- ORCID
- SEAD
- XSEDE
- Zenodo
- Many more
- Many many more!

Globus demonstration

Globus cloud-hosted software-as-a-service for:

- Data transfer, sync, and sharing

- Identity and group management

- Data publication and discovery

Globus Connect software to integrate resources and institutions

What does it mean to **publish**?

Data is:

Identified

Described

Curated

Verifiable

Accessible

Preserved

What does it mean to **discover**?

I can:

Search

Browse

Access

the data

Data publication and discovery

National Data Service
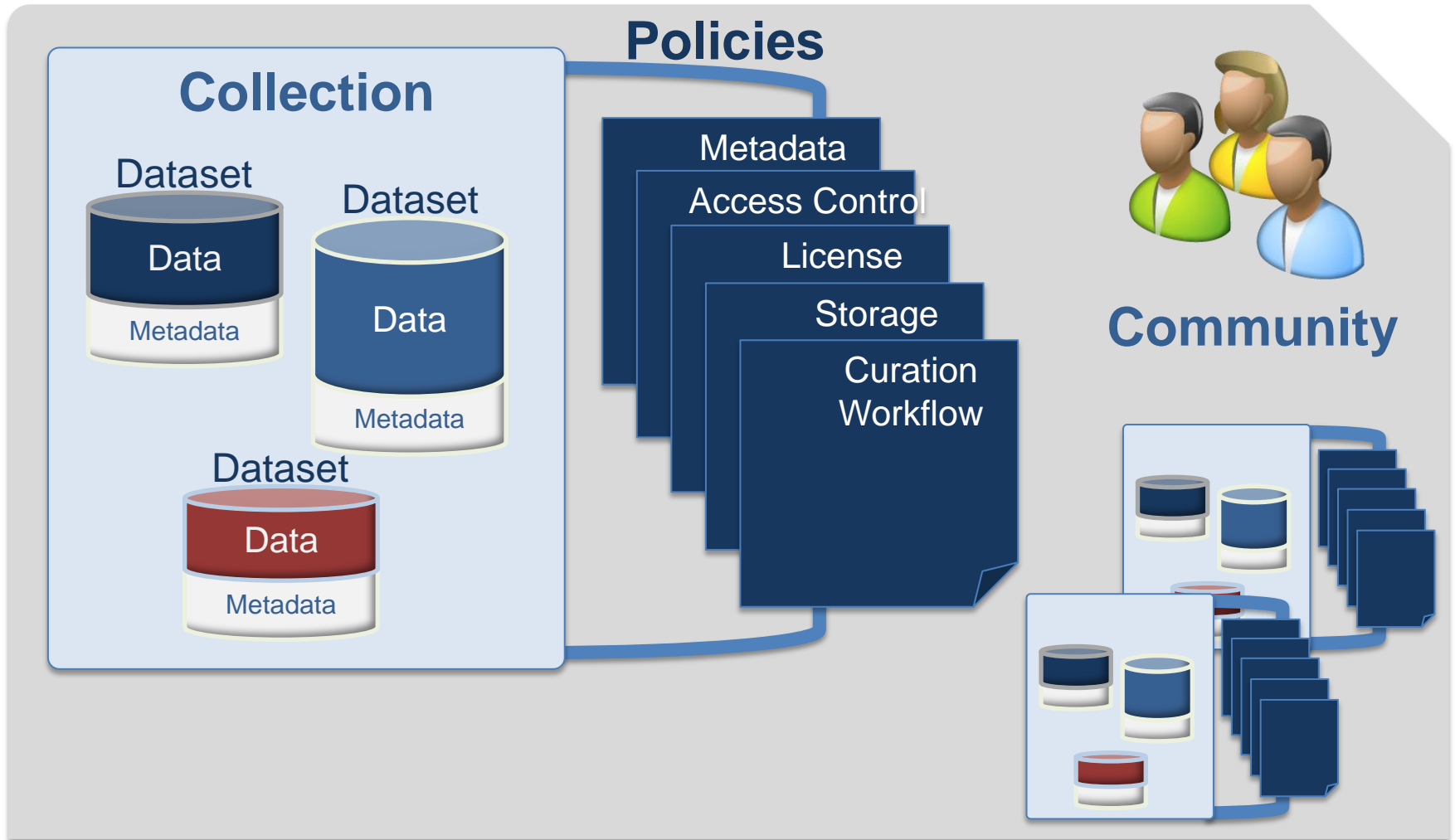
Takeaway messages
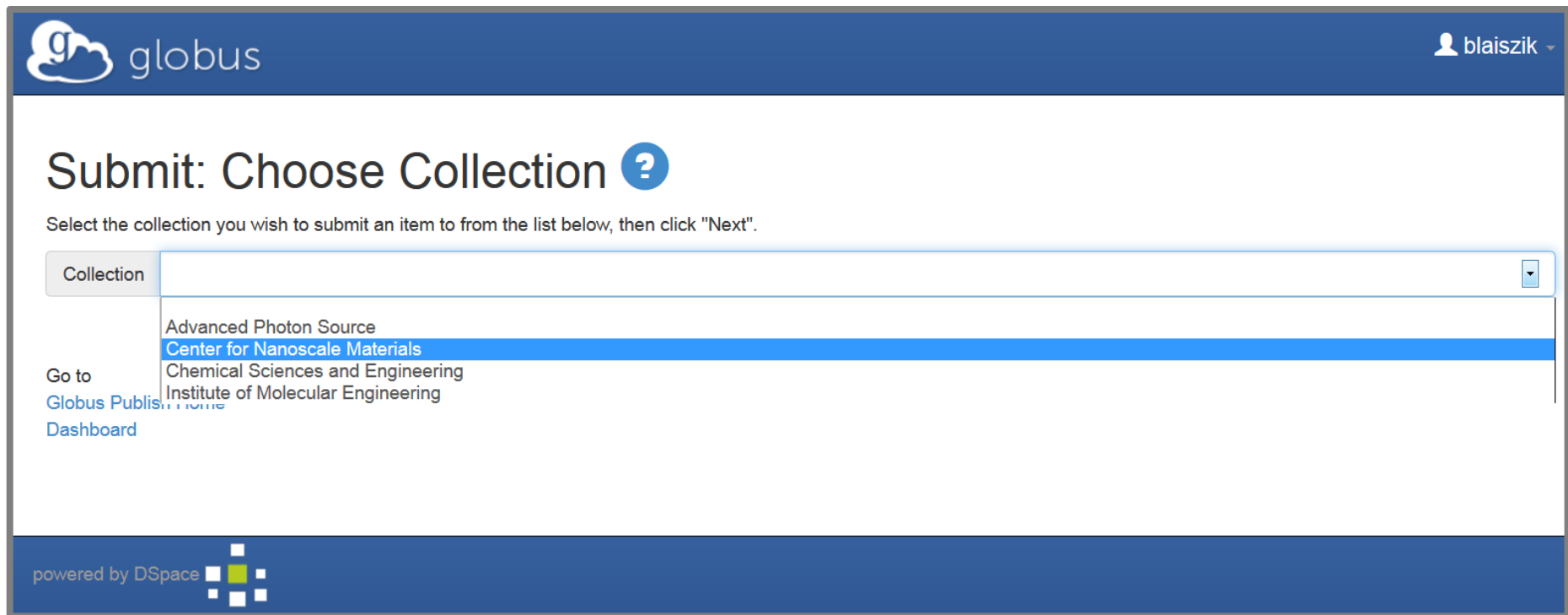
- Three principles:
  - Reduce data friction
  - Small pieces, loosely joined
  - Insist on stories
- We have strong components to start with
- We'd like from you:
  - Stories to inform and prioritize
  - Volunteers to deploy and explore

National Data Service

The following are backup slides in case network failure prevents the live demonstration

Publish dashboard

Start a new submission

Describe submission: 1) Dublin Core

Describe submission: 2) Science metadata

globus

Describe | Describe | Upload | Verify | License | Complete

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 | cicuit | 🗑 Remove Entry |
| microcapsules | 🗑 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 | Gold | 🗑 Remove Entry |
| Indium | | 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

Transfer files to submission endpoint

globus

Describe | Describe | Upload | Verify | License | Complete

Submit: Dataset Assembled Successful

Your dataset was successfully assembled.
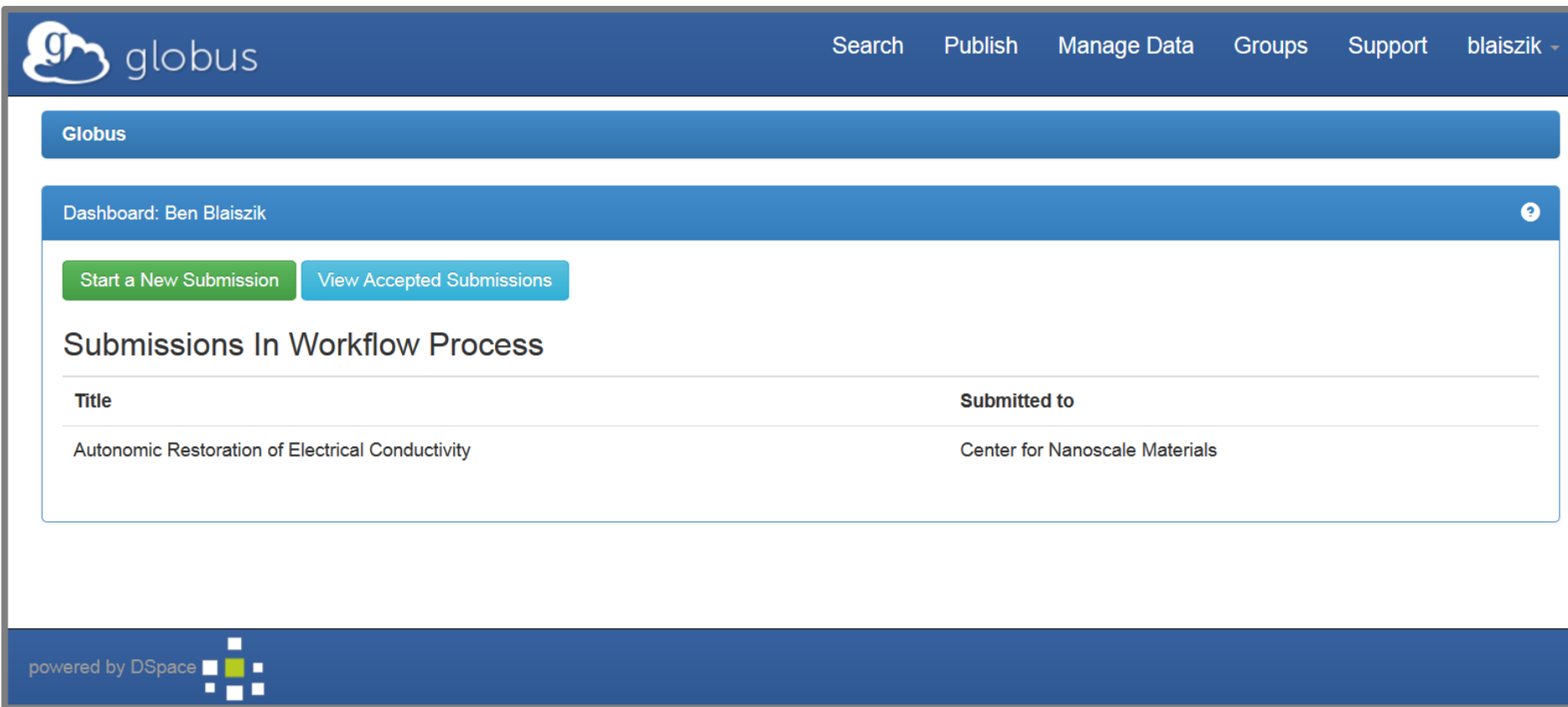
The table below shows the files which are included in this item.

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

Submission now in curation workflow

Search published datasets

Search across collections

Discover a published dataset

Select a published dataset

View downloaded dataset