

# DataFlow - A Data Bridge to Air-Gapped Instruments

Greg Shutt<sup>1</sup>, Marshall McDonnell<sup>2</sup> and Suhas Somnath<sup>3</sup>

<sup>1</sup>Software Services Development Group, National Center for Computational Sciences

<sup>2</sup>Applications Engineering Group, Computer Science and Mathematics Division

<sup>3</sup>Data Lifecycle and Scalable Workflows Group, National Center for Computational Sciences

ORNL is managed by UT-Battelle LLC for the US Department of Energy

DOE Data Days  
June 1, 2022



U.S. DEPARTMENT OF  
**ENERGY**



# Challenges in Everyday Research

- Observational instruments produce large datasets
- Instruments are often off-network
- Large datasets need strong compute resources for analysis



# Today's Solution

- USB drives for data movement
- Personal laptops for data storage and analysis
- Unnecessarily long experiment iterations
- Researchers need simple data movement and management



# DataFlow

## **Two primary components:**

1. Web application software
2. Hardware bridge





# DataFlow

- No upper limit on file sizes
- User authentication provided by Globus
- Instrument and storage can be in different facilities



# DataFlow

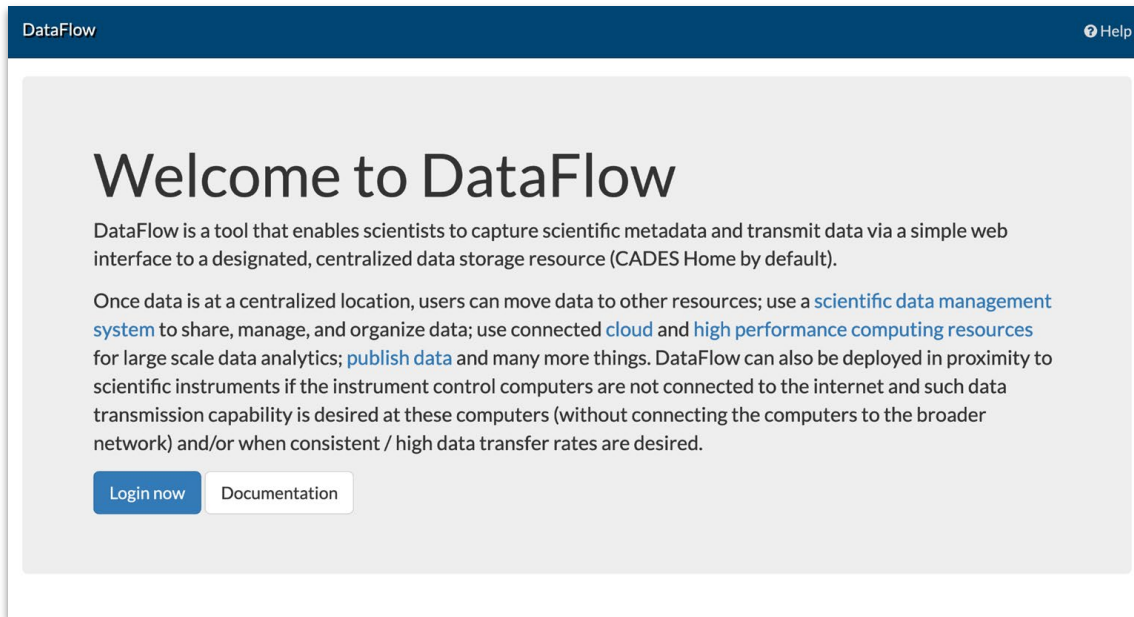
## Two methods of interaction

1. Web-based GUI
2. API with Python library



# DataFlow Demo

<https://dataflow.ornl.gov>



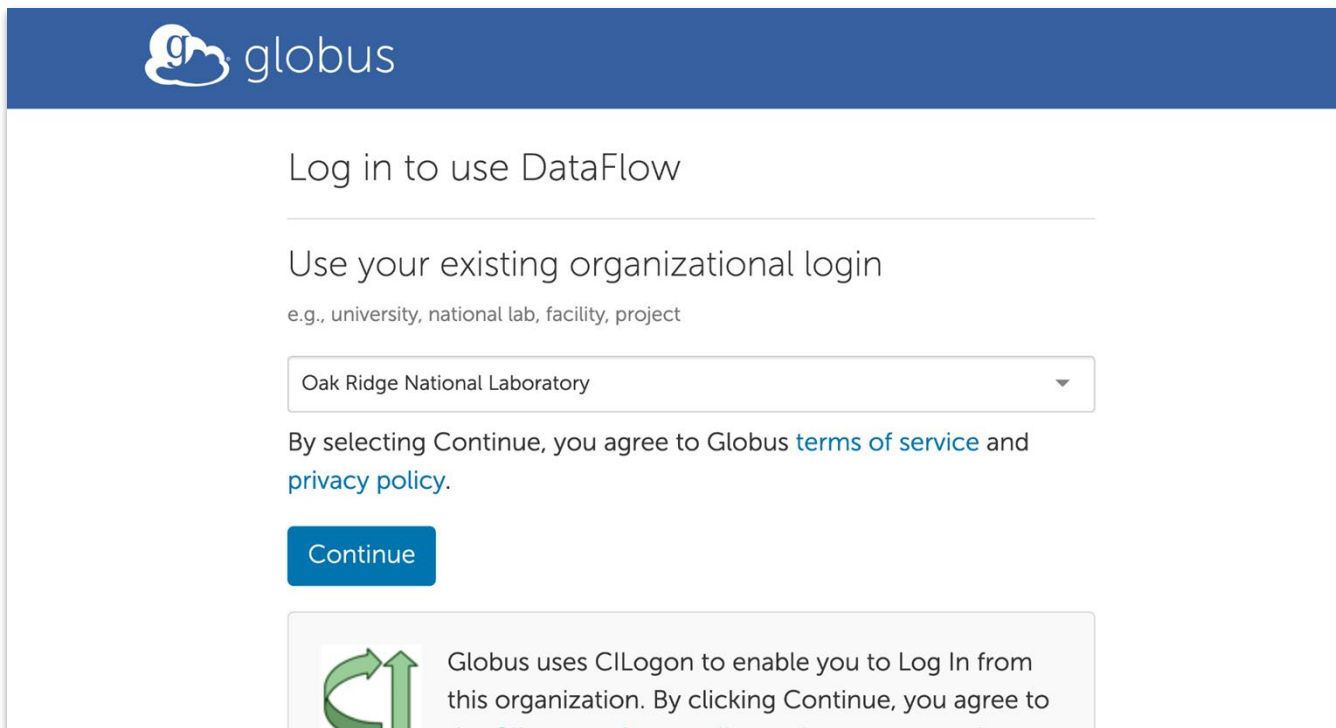
The screenshot shows the DataFlow web interface. At the top, there is a dark blue header with the text "DataFlow" on the left and a "Help" icon on the right. The main content area has a light gray background and features the heading "Welcome to DataFlow". Below the heading, there are two paragraphs of text. The first paragraph describes DataFlow as a tool for capturing scientific metadata and transmitting data. The second paragraph explains how users can move data to other resources, use a scientific data management system, and utilize cloud and high performance computing resources. At the bottom of the main content area, there are two buttons: "Login now" (a blue button) and "Documentation" (a white button with a blue border).

Initial DataFlow screen





# DataFlow Demo









The screenshot shows the Globus login interface. At the top is a blue header with the Globus logo (a white 'g' in a cloud) and the word 'globus'. Below the header, the text 'Log in to use DataFlow' is centered. A horizontal line separates this from the next section, 'Use your existing organizational login', followed by the example text 'e.g., university, national lab, facility, project'. A dropdown menu is open, showing 'Oak Ridge National Laboratory' with a downward arrow. Below the dropdown, a paragraph states: 'By selecting Continue, you agree to Globus [terms of service](#) and [privacy policy](#).' A blue 'Continue' button is positioned below this text. At the bottom of the page, there is a light gray box containing a green circular arrow icon and the text: 'Globus uses CILogon to enable you to Log In from this organization. By clicking Continue, you agree to...'

Login via Globus



# DataFlow Demo

The screenshot shows the DataFlow Datasets interface. At the top, there is a dark blue header with 'DataFlow' and 'Datasets' on the left, and 'Help' and 'Gregory Shutt Oak Ridge National Laboratory' on the right. Below the header, there is a 'Recent uploads' section with a search bar containing 'Search datasets...' and a '+ New dataset' button. The main content area shows a list of datasets with the following details:

Name
 My New Dataset Created May 20, 2022 at 9:26 am
 Test Created May 18, 2022 at 10:33 am
 Test Created May 18, 2022 at 9:52 am
 Dataflow Created Nov 2, 2021 at 1:35 pm
 Test2 Created Nov 2, 2021 at 12:59 pm
 Test Created Nov 2, 2021 at 11:28 am

Recently uploaded datasets



# DataFlow Demo

**Create new dataset**

**Dataset name**  
My New Dataset

**Instrument**  
Asylum Research Cypher West

**Metadata**

Name	Value
Temperature	110
Material	Titanium

[Add additional metadata...](#)

Cancel Next

Creating a new dataset



# DataFlow Demo

The screenshot displays the DataFlow Datasets interface. The top navigation bar includes 'DataFlow', 'Datasets', 'Help', and the user 'Gregory Shutt' from 'Oak Ridge National Laboratory'. The main content area is split into two panels. The left panel, titled 'Recent uploads', shows a list of datasets with a search bar and a table of results. The right panel, titled 'My New Dataset', displays the dataset's metadata and upload options.

**Recent uploads** (Showing datasets: All (6 results))

Name
My New Dataset Created May 20, 2022 at 9:26 am
Test Created May 18, 2022 at 10:33 am
Test Created May 18, 2022 at 9:52 am
Dataflow Created Nov 2, 2021 at 1:35 pm
Test2 Created Nov 2, 2021 at 12:59 pm
Test Created Nov 2, 2021 at 11:28 am

**My New Dataset**

**Dataset**

- Name:** My New Dataset
- Created by:** Gregory Shutt
- Instrument:** Asylum Research Cypher West

**Metadata**

- Temperature:** 110
- Material:** Titanium

**Files** **Filters** Found 0 files

My New Dataset

**Upload folder** **Upload files**

Newly-created dataset, ready for upload



# DataFlow Demo

The screenshot shows the DataFlow Datasets interface. The top navigation bar includes 'DataFlow', 'Datasets', 'Help', and the user 'Gregory Shutt' from 'Oak Ridge National Laboratory'. The main content area is divided into three sections:

- Recent uploads:** A list of datasets with search and close icons. The top item is 'My New Dataset' (Created May 20, 2022 at 9:26 am).
- Metadata:** A panel for the selected dataset showing:
  - Name: My New Dataset
  - Created by: Gregory Shutt
  - Instrument: Asylum Research Cypher West
  - Temperature: 110
  - Material: Titanium
- Files:** A section with 'Files Filters' and 'Found 1 file'. It includes 'Upload folder' and 'Upload files' buttons. Below is a table for 'My New Dataset':

Filename	Size	Date
dataset.tar.gz	13.3 MB	May 20, 2022

File successfully uploaded



# API Benefits

- Enable automated data upload for long-running experiments
- Monitor and automatically upload all new data



# DataFlow API Demo

```
4 from ordflow import API
5
6 # create an example dataset
7 response = API(api_key).dataset_create(
8     "My new dataset with nested metadata",
9     metadata={"Sample": "PZT",
10              "Microscope": {
11                  "Vendor": "Asylum Research",
12                  "Model": "MFP3D"
13              },
14              "Temperature": 373
15             }
16 )
17
```

Creating a dataset



# DataFlow API Demo

```
4 from ordflow import API
5
6 # upload a file to the dataset
7 response = API(api_key).file_upload("/Users/g5s/Downloads/olivia.jpeg", dataset_id)
8
```

Uploading a file





# DataFlow API Demo

The screenshot displays the DataFlow web interface. The top navigation bar includes 'DataFlow', 'Datasets', 'Help', and the user 'Gregory Shutt' from 'Oak Ridge National Laboratory'. The main content area is divided into a left sidebar and a right main panel.

**Recent uploads** (Search datasets...):

- My new dataset with nested metadata (Created May 20, 2022 at 11:27 am)
- My New Dataset (Created May 20, 2022 at 9:26 am)
- Test (Created May 18, 2022 at 10:33 am)
- Test (Created May 18, 2022 at 9:52 am)
- Dataflow (Created Nov 2, 2021 at 1:35 pm)
- Test2 (Created Nov 2, 2021 at 12:59 pm)
- Test (Created Nov 2, 2021 at 11:28 am)

**Metadata** (Instrument: Not set):

- Sample: PZT
- Microscope-Vendor: Asylum Research
- Microscope-Model: MFP3D
- Temperature: 373

**Files Filters** (Found 1 file):

My new dataset with nested metadata

Filename	Size	Date
olivia.jpeg	815.6 kB	May 20, 2022

Buttons: Upload folder, Upload files

Dataset and file created



# Future Work

- Additional adapters: S3, DropBox, OneDrive
- Adding hooks to integrate into workflow systems



# Contact

For more information, please contact:

Greg Shutt [shuttgl@ornl.gov](mailto:shuttgl@ornl.gov)

Suhas Somnath [somnaths@ornl.gov](mailto:somnaths@ornl.gov)

This research used resources of the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.

