



Globus for Repositories

Susan Borda - Data Workflows Specialist
University of Michigan - Library

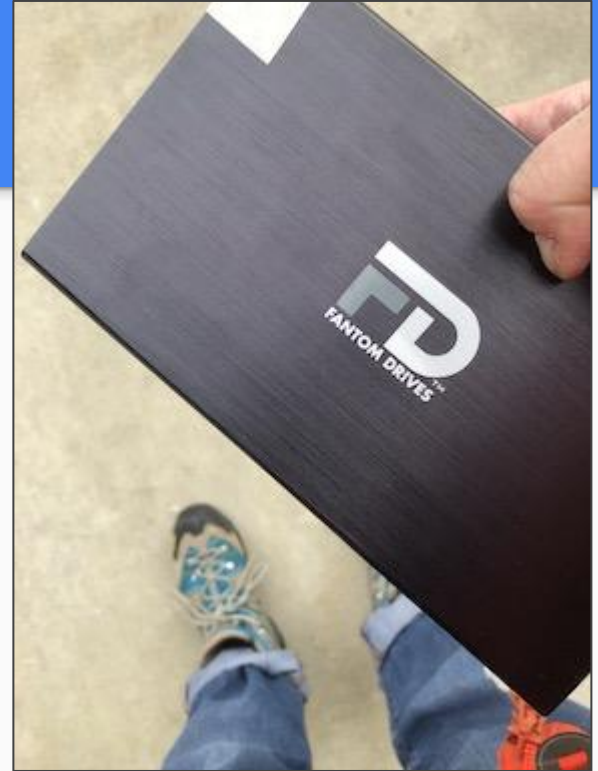


So you want to put your **LARGE** data set in the repository...

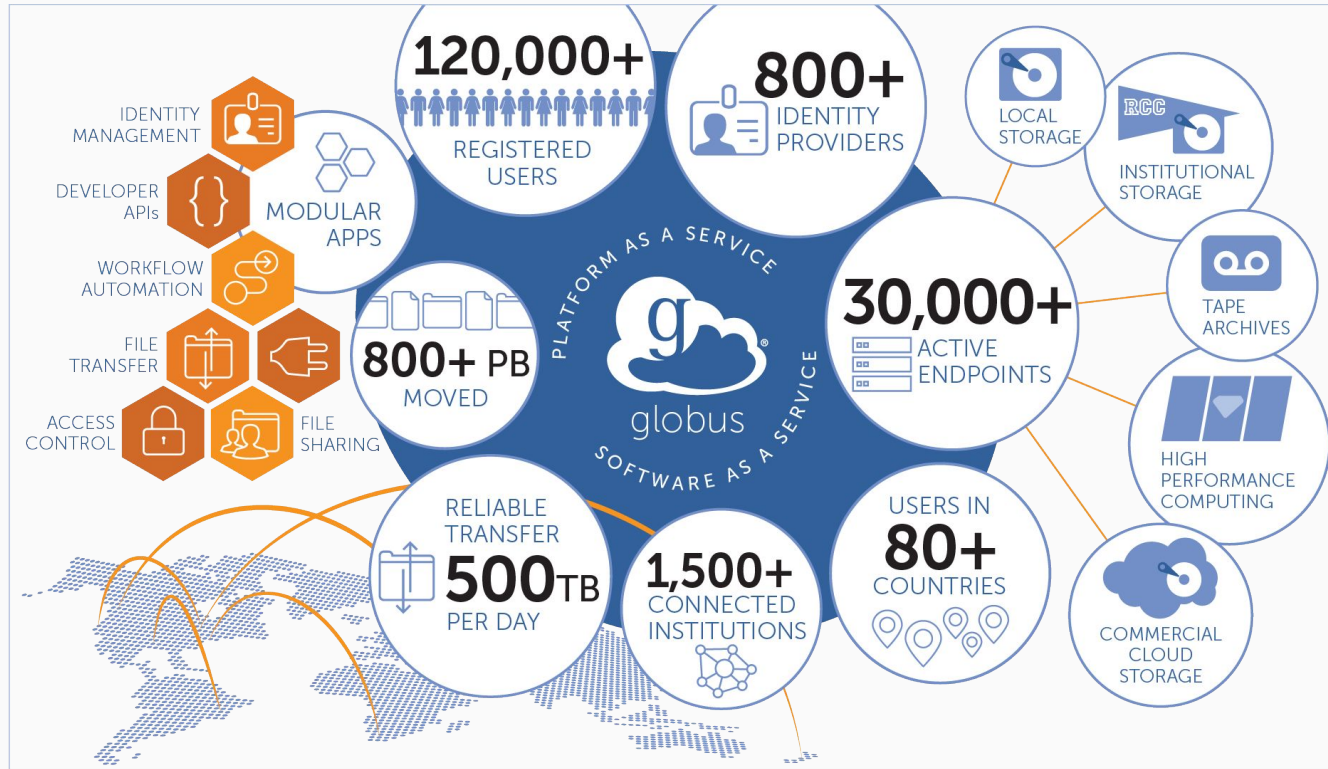
- Where is the data?
- How large is the data?
- ~~What kind of data is it?~~
- Where does it need to go?
- How to retrieve the data

Where is the data?

- External hard drive
- Personal server in their office
- Lab server
- On a centrally managed server
- Attached to an HPC environment



Still the fastest way to get 1TB across campus!

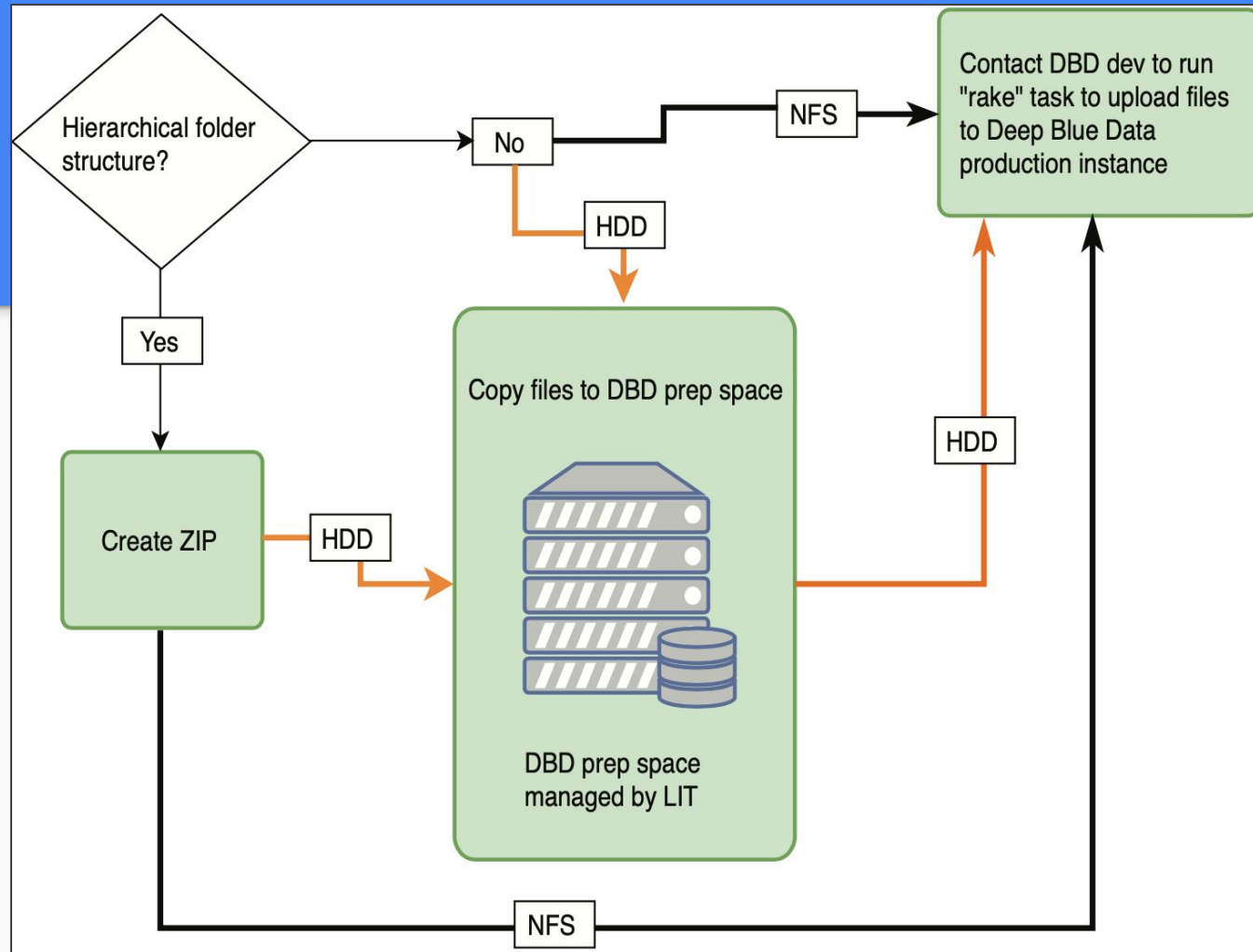


<https://www.globus.org/>

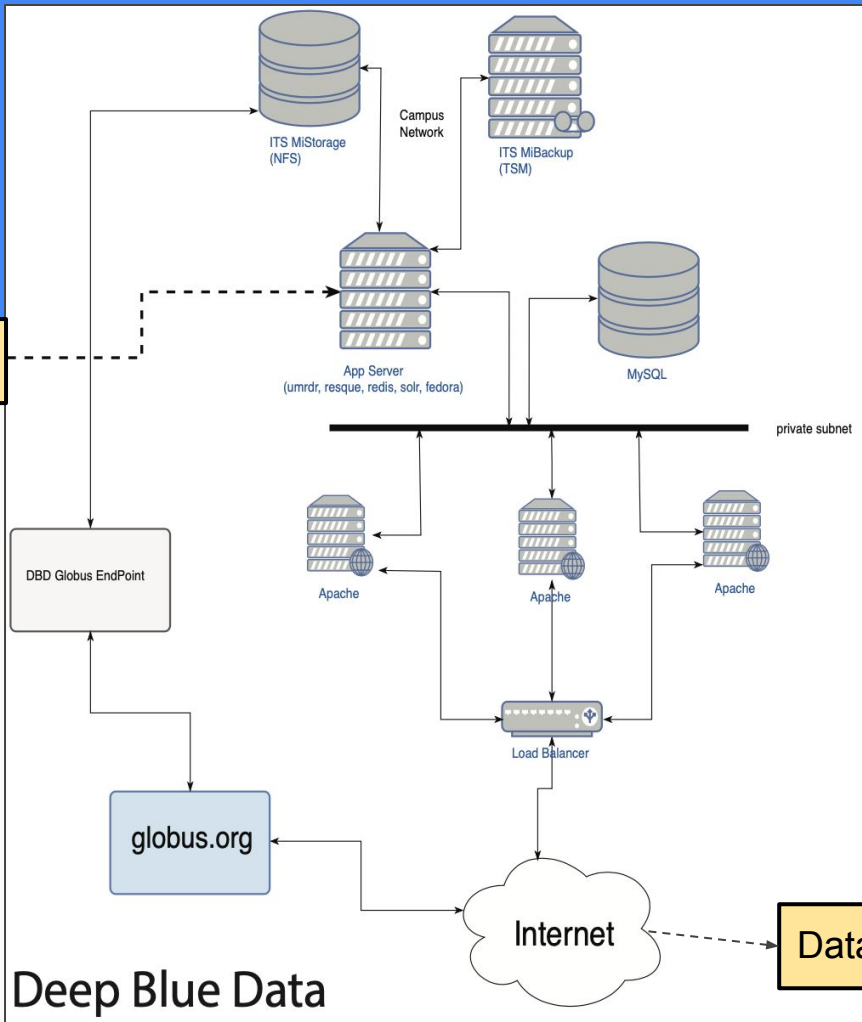
Helloooo Globus!!

There is NOT a direct connection between Globus and Deep Blue Data for data ingest/upload.

Folder structures need to be zipped or tarred first.



Data ingest



Globus is integrated for data-set download only!

Deep Blue Data







Data download

Globus prevents the “Roach Motel” scenario.

Data checks in but doesn’t check out.



Files (Count: 14; Size: 3.56 TB)

	Title	Original Upload	Last Modified	File Size	Access	Actions
	2010_gitm_input_files.tgz	2017-11-22	2017-11-22	102 MB	Open Access	Select an action ▾
	gitm_170809.tgz	2017-11-22	2017-11-22	26.9 MB	Open Access	Select an action ▾
	gitm_201001_all_output.tgz	2018-01-25	2018-12-13	312 GB	Open Access	Select an action ▾
	gitm_201002_all_output.tgz	2018-01-25	2018-12-14	273 GB	Open Access	Select an action ▾
	gitm_201003_all_output.tgz	2018-01-25	2018-12-14	315 GB	Open Access	Select an action ▾
	gitm_201004_all_output.tgz	2018-01-25	2018-12-15	302 GB	Open Access	Select an action ▾

Download All Files (To download individual files, select them in the “Files” panel above)

Total work file size of 3.56 TB is too large to download directly. Consider using Globus (see below).

Files are ready

[Download Data from Globus](#)

In Deep Blue Data

https://deepblue.lib.umich.edu/data/concern/data_sets/ft848r062?locale=en

In Globus

Collection:

Path:

File Name	Date	Time	Size	Actions
2010_gitm_input_files.tgz	4/2/2021	10:37 AM	106.70 MB	
gitm_170809.tgz	4/3/2021	3:19 PM	28.18 MB	
gitm_201001_all_output.tgz	4/3/2021	2:13 PM	335.01 GB	
gitm_201002_all_output.tgz	4/3/2021	4:14 PM	293.60 GB	
gitm_201003_all_output.tgz	4/3/2021	7:29 AM	338.35 GB	
gitm_201004_all_output.tgz	4/2/2021	4:59 PM	181.63 GB	
gitm_201005_all_output.tgz	4/3/2021	1:12 AM	335.38 GB	
gitm_201006_all_output.tgz	4/2/2021	8:47 AM	324.82 GB	

Search for a collection to begin

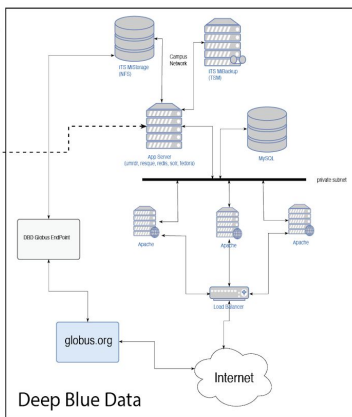
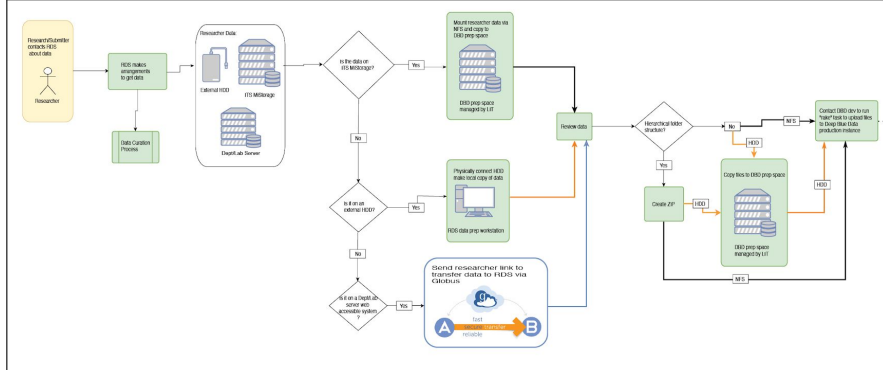
Get started by taking a short tour.

Data Sharing Workflow for Large Datasets (>1TB) with Globus

Susan Borda - Data Workflows Specialist University of Michigan - sborda@umich.edu

Research Data Services

Workflow for Large Datasets



Complete workflow

Globus File Download Process

1. Files (65.3 GB in 24 files)

File Name	Date Uploaded	File Size	Access	Actions
com_eia_us_present.nc	2017-08-25		Open Access	Details
com_eia_us_present.nc	2017-08-25		Open Access	Details
com_eia_us_present.nc	2017-08-25		Open Access	Details
com_eia_us_present.nc	2017-08-25		Open Access	Details
com_eia_us_present.nc	2017-08-25		Open Access	Details
com_eia_us_present.nc	2017-08-25		Open Access	Details
com_eia_us_present.nc	2017-08-25		Open Access	Details
com_eia_us_present.nc	2017-08-25		Open Access	Details
com_eia_us_present.nc	2017-08-25		Open Access	Details
com_eia_us_present.nc	2017-08-25		Open Access	Details

2. Click below to start downloading Regional Climate Model simulations to Globus. If you would like to be sent an email when the files are available @ in your email address:

Email: sborda@umich.edu

3. sborda@umich.edu 12/22/17

Globus download is now available:
Work: Regional Climate Model simulations
At: https://transfer.deepblue.lib.umich.edu/datastore/camrsm_works/2560327-1331-4956-9f95-9837-3262552
By: Shawn Allison, Bryan, Alex
Discovered by: sborda@umich.edu
Globus Link: https://www.globus.org/apps/transfer?origin_ip=96bdc48-a0ff-11e7-a6dd-2200a02523ba&origin_auth=52f0e6a6a0932f2d0e8da54672651027-1331-4956-9f95-9837-3262552&f

4. Transfer Files

Endpoint	Size	Created
com_eia_us_present.nc	4.98 GB	2017-08-25
com_eia_us_present.nc	4.98 GB	2017-08-25
com_eia_us_present.nc	4.98 GB	2017-08-25
com_eia_us_present.nc	4.98 GB	2017-08-25
com_eia_us_present.nc	4.98 GB	2017-08-25
com_eia_us_present.nc	4.98 GB	2017-08-25
com_eia_us_present.nc	4.98 GB	2017-08-25
com_eia_us_present.nc	4.98 GB	2017-08-25
com_eia_us_present.nc	4.98 GB	2017-08-25
com_eia_us_present.nc	4.98 GB	2017-08-25
com_eia_us_present.nc	4.98 GB	2017-08-25
com_eia_us_present.nc	4.98 GB	2017-08-25
com_eia_us_present.nc	4.98 GB	2017-08-25
com_eia_us_present.nc	4.98 GB	2017-08-25
com_eia_us_present.nc	4.98 GB	2017-08-25
com_eia_us_present.nc	4.98 GB	2017-08-25
com_eia_us_present.nc	4.98 GB	2017-08-25
com_eia_us_present.nc	4.98 GB	2017-08-25
com_eia_us_present.nc	4.98 GB	2017-08-25
com_eia_us_present.nc	4.98 GB	2017-08-25
com_eia_us_present.nc	4.98 GB	2017-08-25
com_eia_us_present.nc	4.98 GB	2017-08-25
com_eia_us_present.nc	4.98 GB	2017-08-25
com_eia_us_present.nc	4.98 GB	2017-08-25
com_eia_us_present.nc	4.98 GB	2017-08-25

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



4a. Activity

4b. Overview

Task ID: 2560327-1331-4956-9f95-9837-3262552

Owner: shawn.allison@umich.edu

Source: 2560327-1331-4956-9f95-9837-3262552

Destination: sborda@umich.edu

Condition: ACTIVE

Requested: 2018-01-16 09:02 pm

Deadline: 2018-01-17 09:02 pm

Transfer Status: verify file integrity after transfer, transfer is not interrupted, overwriting all files on destination

Flux: 12.47 GB/s
Effective Speed: 57.78 MB/s

5. Globus File List

Name	Date Modified	Size	Type
com_eia_us_present.nc	2017-08-25 09:00 PM	4,973,859	Document
com_eia_us_present.nc	2017-08-25 09:00 PM	4,973,859	Document
com_eia_us_present.nc	2017-08-25 09:00 PM	3,995,029	Document
com_eia_us_present.nc	2017-08-25 09:00 PM	3,995,029	Document
com_eia_us_present.nc	2017-08-25 09:00 PM	473,168	Document
com_eia_us_present.nc	2017-08-25 09:00 PM	4,973,859	Document
com_eia_us_present.nc	2017-08-25 09:00 PM	4,973,859	Document
com_eia_us_present.nc	2017-08-25 09:00 PM	4,973,859	Document
com_eia_us_present.nc	2017-08-25 09:00 PM	4,973,859	Document
com_eia_us_present.nc	2017-08-25 09:00 PM	3,995,029	Document
com_eia_us_present.nc	2017-08-25 09:00 PM	3,774,029	Document



Globus Code on Github:
<https://github.com/mlibrary/deepblue/tree/master/app/jobs>

Questions?
sborda@umich.edu