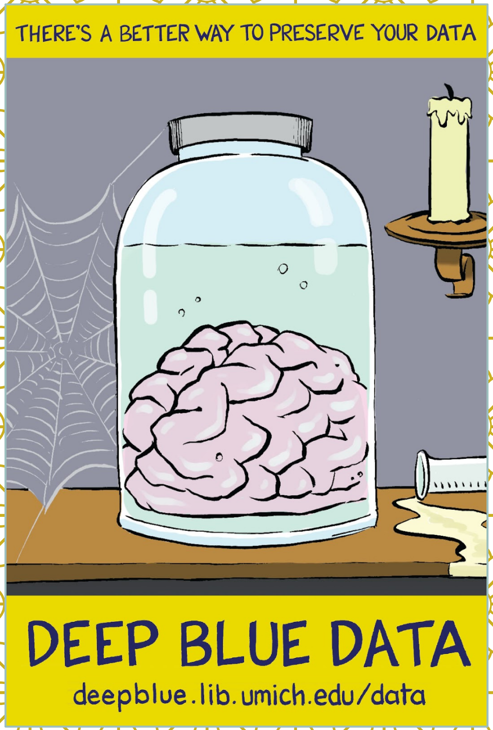

THERE'S A BETTER WAY TO PRESERVE YOUR DATA
DEEP BLUE DATA
deepblue.lib.umich.edu/data

DATA SHARING WORKFLOW FOR LARGE DATASETS (>1TB) WITH GLOBUS

Susan Borda - Data Workflows Specialist, University of Michigan - Library

Research Data Services

WHAT IS DEEP BLUE DATA?

Research data repository , which preserves and provides access to the intellectual output of the UM (e.g. published datasets)

Fedora and Samvera

Soft launch Feb 2016, public launch Sept 2016

Open access data – non-restricted datasets

Deep Blue Data accepts:

- Data from all disciplines from STEM to Humanities
- Data in any format. Open, nonproprietary and widely used formats are preferable
- Data in all sizes that are technically possible, from KBs → TBs

https://deepblue.lib.umich.edu/data/

THE NUMBERS

19/110 works > 10GB

40/8566 files > 10GB

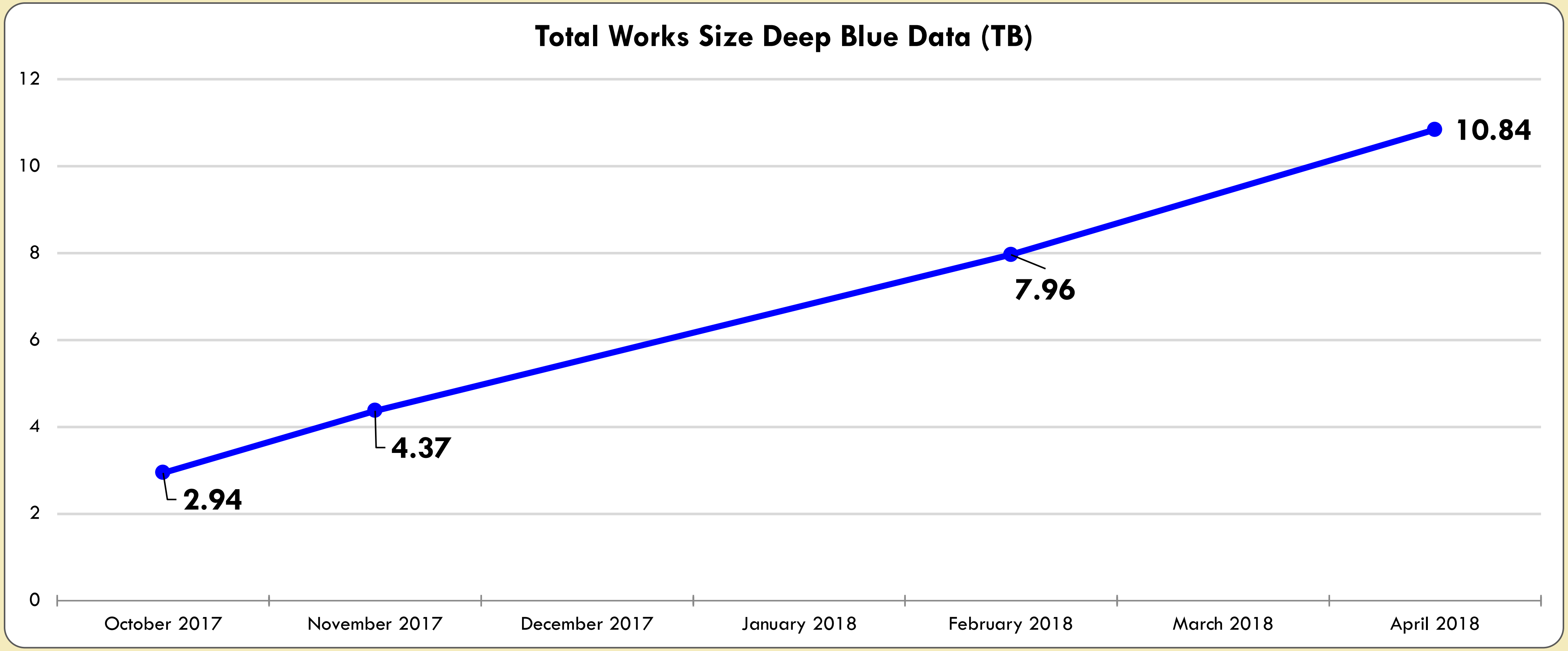
Largest single work = 3.9TB

Largest file = 754GB

All but 1 (> 10GB) work has been deposited since July 2017

The repository continues to grow…

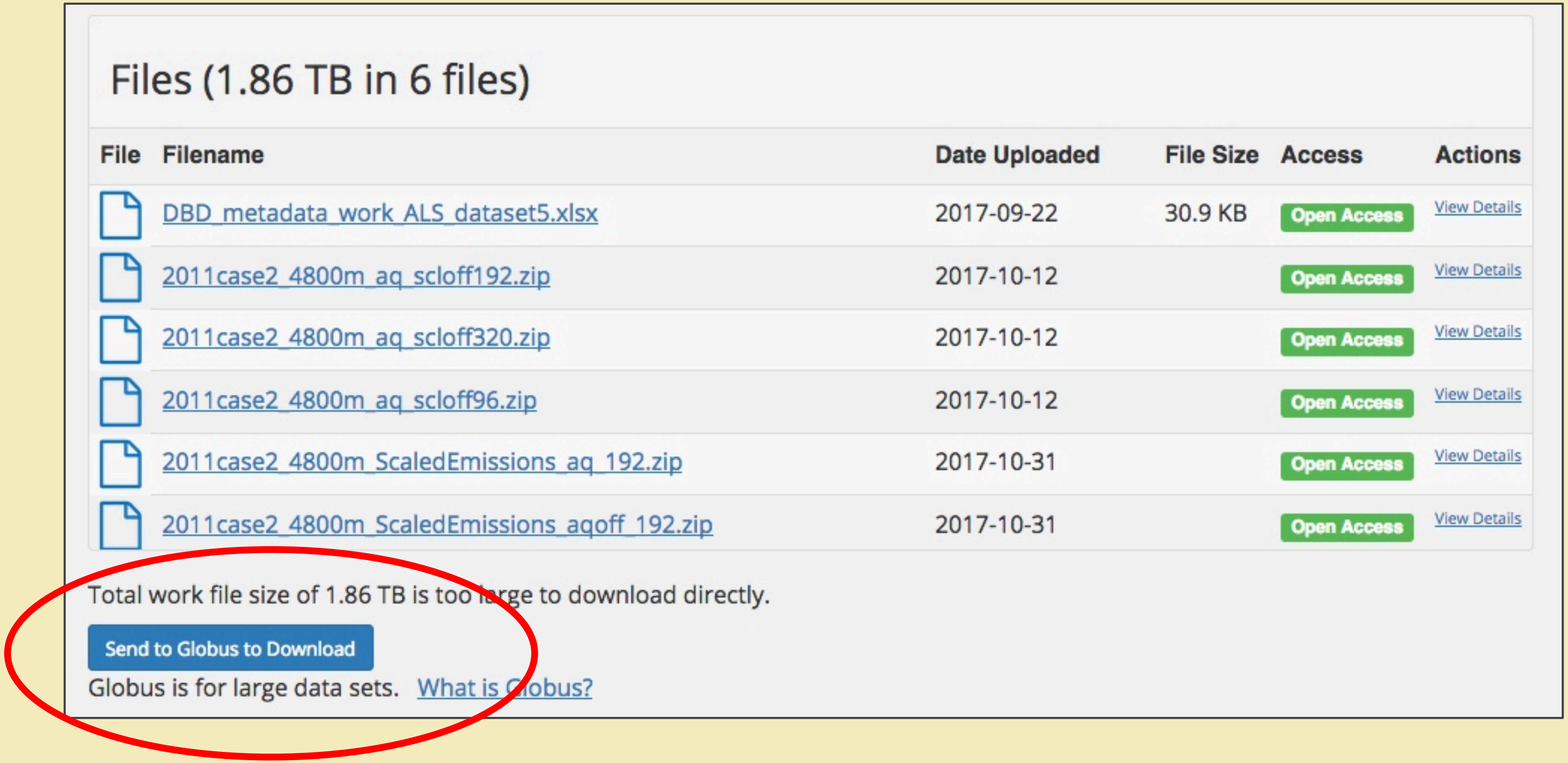
FROM KB, MB & GB → TB



DATA GOES IN BUT...

Hides Dead Bugs
ROACH MOTEL®
ROACHES CHECK IN... BUT THEY DON'T CHECK OUT®
GUARANTEED TO WORK OR YOUR MONEY BACK*
Traps Roaches, Waterbugs, Palmetto Bugs, Spiders, Crickets and Scorpions
FROM THE MAKERS OF BLACK FLAG
Contains 2 Traps
*For terms of guarantee, see back panel.
KEEP OUT OF REACH OF CHILDREN
CAUTION See back panel for additional precautionary statements and first aid.
ROACH MOTEL® ROACH TRAPS
FROM THE MAKERS OF BLACK FLAG

GLOBUS TO THE RESCUE!!



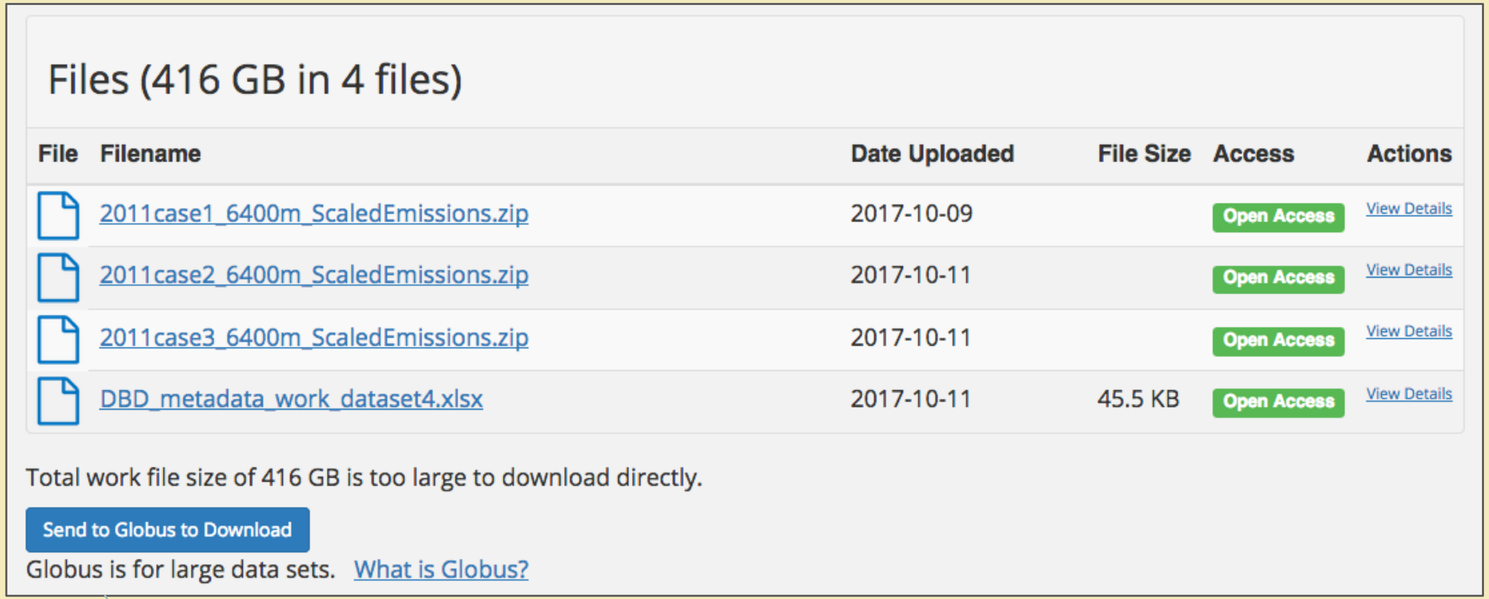
Files (1.86 TB in 6 files)

File	Filename	Date Uploaded	File Size	Access	Actions
🗋	DBD_metadata_work_ALS_dataset5.xlsx	2017-09-22	30.9 KB	Open Access	View Details
🗋	2011case2_4800m_aq_scloff192.zip	2017-10-12		Open Access	View Details
🗋	2011case2_4800m_aq_scloff320.zip	2017-10-12		Open Access	View Details
🗋	2011case2_4800m_aq_scloff96.zip	2017-10-12		Open Access	View Details
🗋	2011case2_4800m_ScaledEmissions_aq_192.zip	2017-10-31		Open Access	View Details
🗋	2011case2_4800m_ScaledEmissions_aqoff_192.zip	2017-10-31		Open Access	View Details

Total work file size of 1.86 TB is too large to download directly.

Send to Globus to Download

Globus is for large data sets. What is Globus?

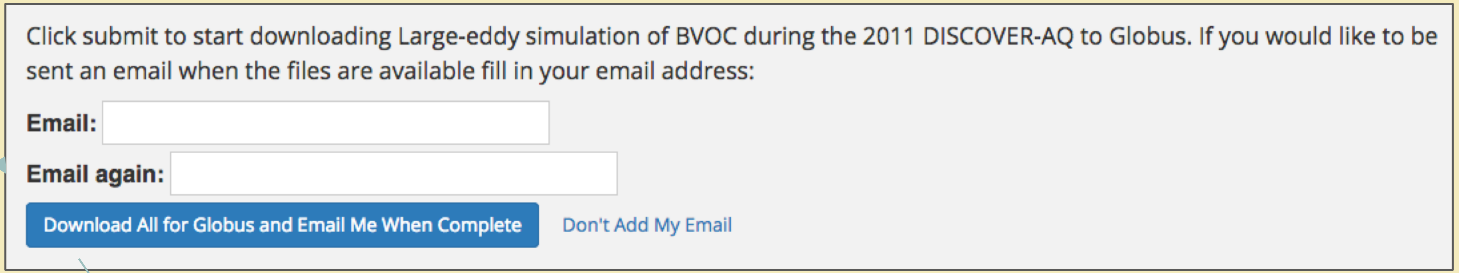


Files (416 GB in 4 files)

File	Filename	Date Uploaded	File Size	Access	Actions
	2011case1_6400m_ScaledEmissions.zip	2017-10-09		Open Access	View Details
	2011case2_6400m_ScaledEmissions.zip	2017-10-11		Open Access	View Details
	2011case3_6400m_ScaledEmissions.zip	2017-10-11		Open Access	View Details
	DBD_metadata_work_dataset4.xlsx	2017-10-11	45.5 KB	Open Access	View Details

Total work file size of 416 GB is too large to download directly.

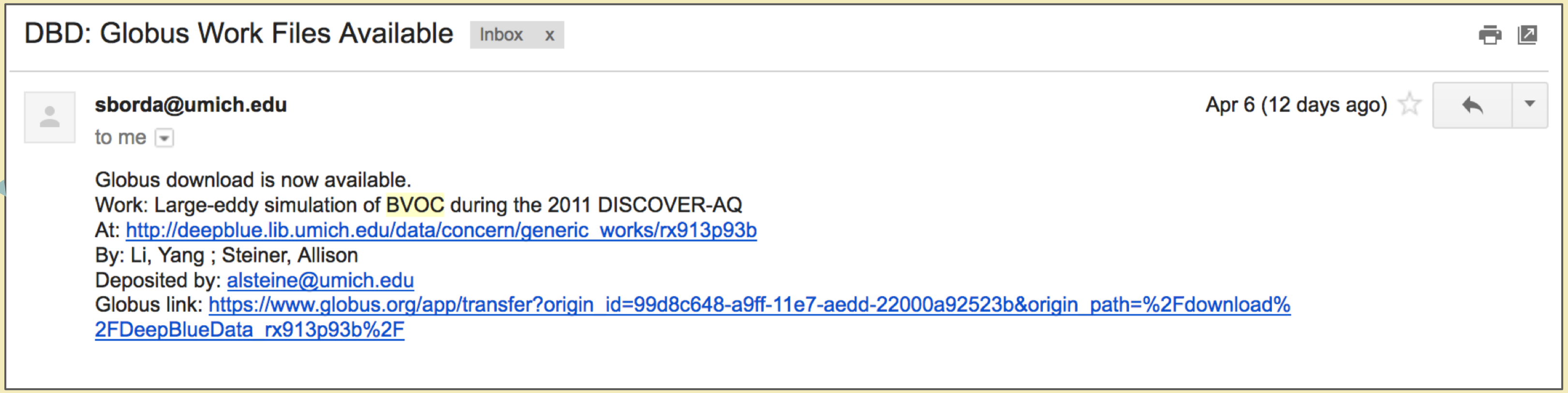
Globus is for large data sets. What is Globus?



Click submit to start downloading Large-eddy simulation of BVOC during the 2011 DISCOVER-AQ to Globus. If you would like to be sent an email when the files are available fill in your email address:

Email:

Email again:



DBD: Globus Work Files Available

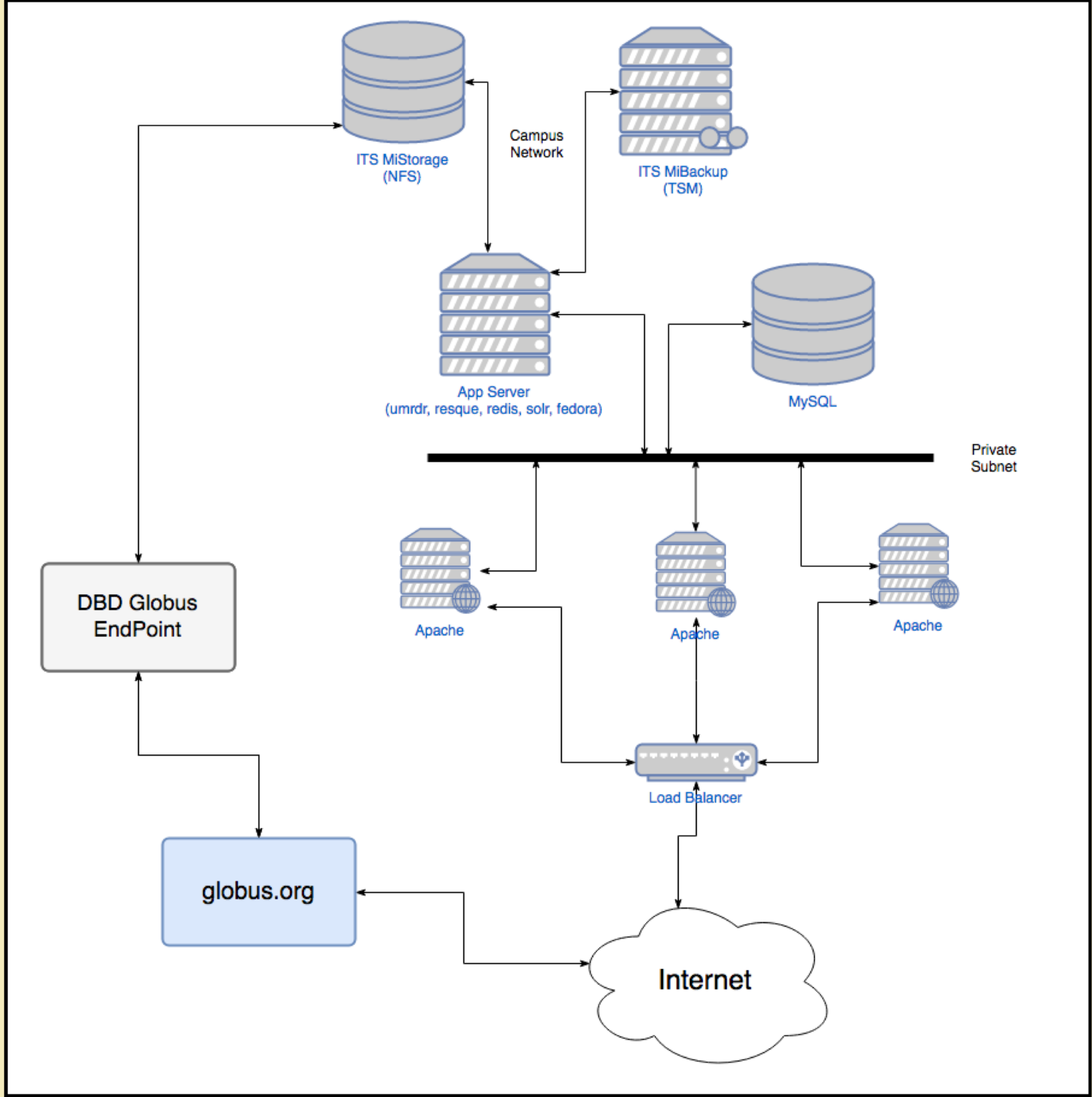
sborda@umich.edu

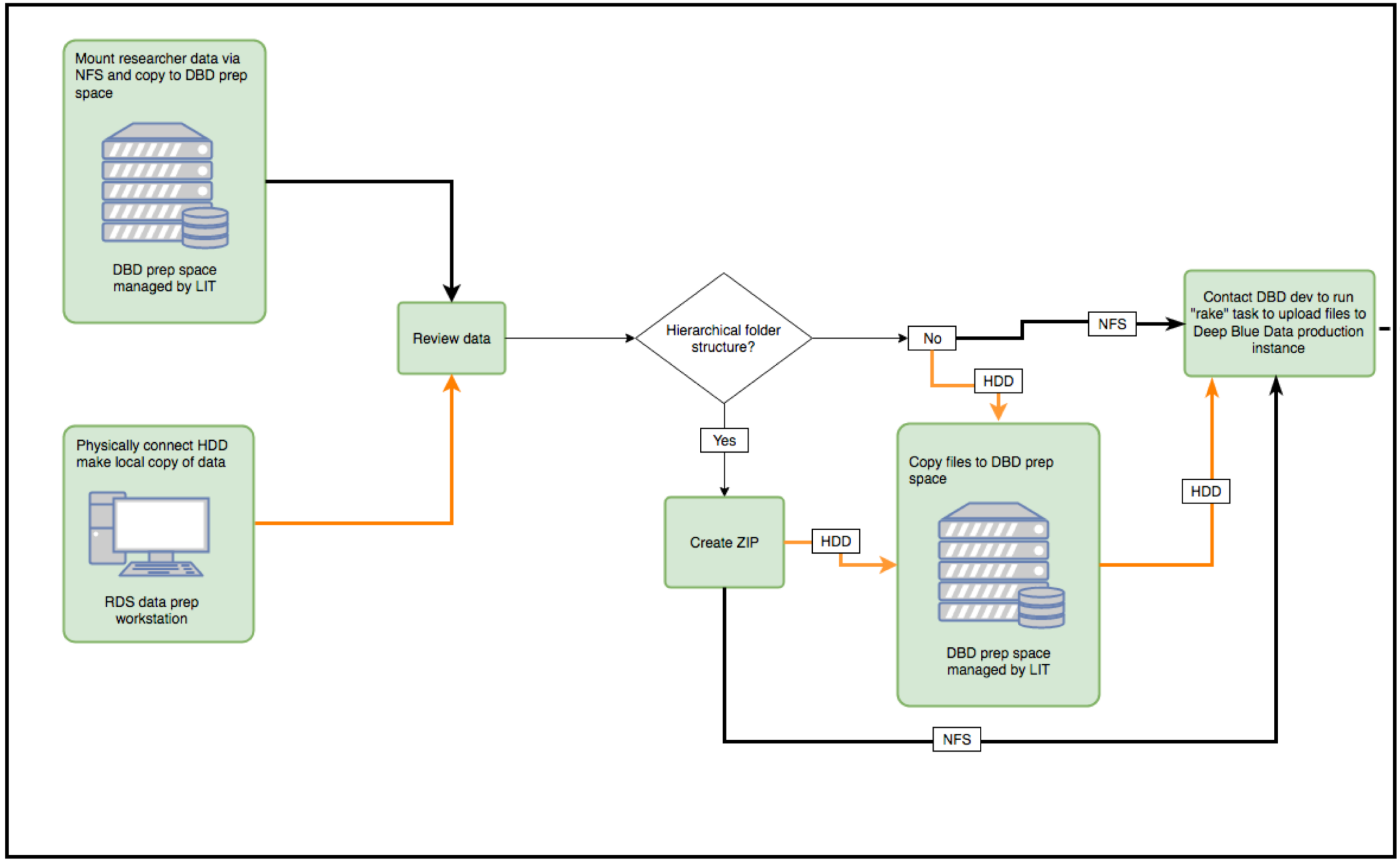
to me

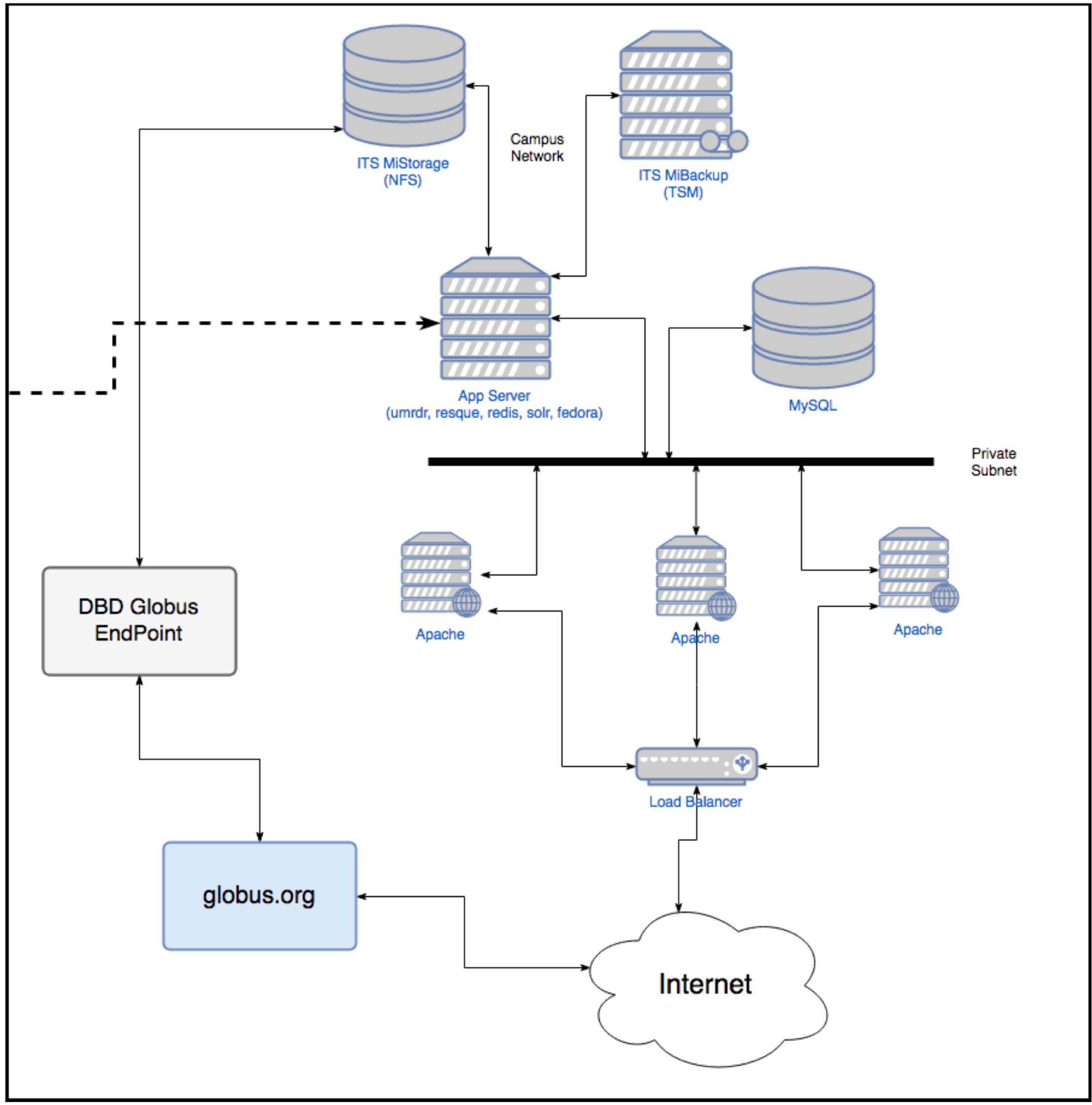
Apr 6 (12 days ago)

Globus download is now available.
Work: Large-eddy simulation of BVOC during the 2011 DISCOVER-AQ
At: http://deepblue.lib.umich.edu/data/concern/generic_works/rx913p93b
By: Li, Yang ; Steiner, Allison
Deposited by: alsteine@umich.edu
Globus link: https://www.globus.org/app/transfer?origin_id=99d8c648-a9ff-11e7-aedd-22000a92523b&origin_path=%2Fdownload%2FDeepBlueData_rx913p93b%2F

Globus and the Deep Blue Data Infrastructure



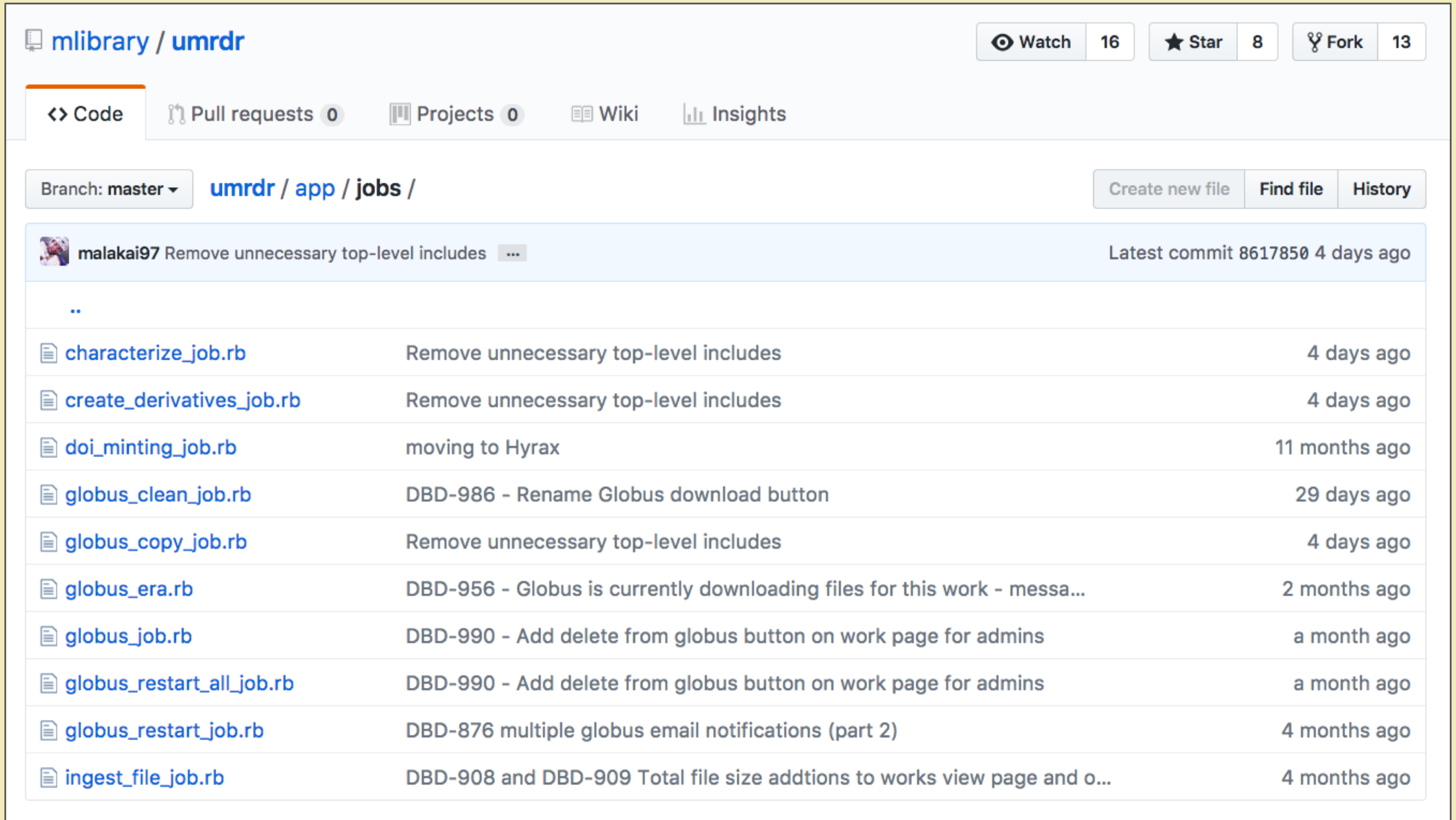

Mount researcher data via NFS and copy to DBD prep space
DBD prep space managed by LIT
Physically connect HDD make local copy of data
RDS data prep workstation
Review data
Hierarchical folder structure?
No
Yes
NFS
HDD
Contact DBD dev to run "rake" task to upload files to Deep Blue Data production instance
Create ZIP
HDD
Copy files to DBD prep space
DBD prep space managed by LIT
HDD
NFS



Future Plans for Globus…
Data Ingest!!

SHARE THE LOVE!! – GITHUB (MLIBRARY/UMRDR)

https://github.com/mlibrary/umrdr/tree/master/app/jobs



malakai97 Remove unnecessary top-level includes … Latest commit 8617850 4 days ago

..

characterize_job.rb	Remove unnecessary top-level includes	4 days ago
create_derivatives_job.rb	Remove unnecessary top-level includes	4 days ago
doi_minting_job.rb	moving to Hyrax	11 months ago
globus_clean_job.rb	DBD-986 - Rename Globus download button	29 days ago
globus_copy_job.rb	Remove unnecessary top-level includes	4 days ago
globus_era.rb	DBD-956 - Globus is currently downloading files for this work - messa…	2 months ago
globus_job.rb	DBD-990 - Add delete from globus button on work page for admins	a month ago
globus_restart_all_job.rb	DBD-990 - Add delete from globus button on work page for admins	a month ago
globus_restart_job.rb	DBD-876 multiple globus email notifications (part 2)	4 months ago
ingest_file_job.rb	DBD-908 and DBD-909 Total file size addtions to works view page and o…	4 months ago