



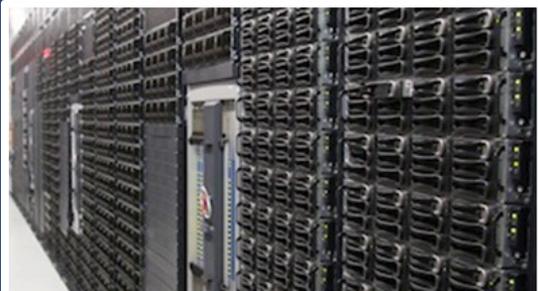
Beyond File Transfer

Steve Tuecke
tuecke@uchicago.edu

NCAR – September 5, 2018



Unify access to data across tiers



Research Computing HPC



Personal Resources



Desktop Workstations



Mass Storage



Instruments





Storage Connectors - globus.org/connectors

Current

IBM Spectrum Scale



ceph



Lustre®

HPSS



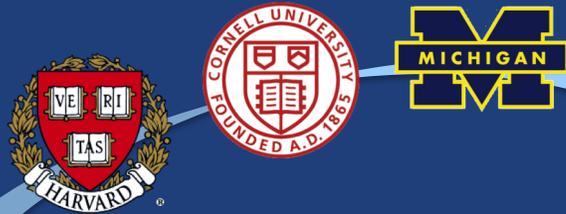
Planned

box



Microsoft Azure

Share with collaborators/community



XSEDE
Extreme Science and Engineering
Discovery Environment

EC2

Jetstream

compute canada | **calcul canada**



External
campus
storage



Project
repositories,
replication stores



Public repositories

Public / private cloud stores

red cloud **ceph** **openstack**

Google Drive **amazon web services™ S3**



Manage data from instruments



Next-Gen Sequencer



Advanced Light Source



Cryo-EM



MRI



Light Sheet Microscope



Analysis store



High-durability, low-cost store



Remote visualization



Personal system



Develop apps, services, and workflows

UCAR NCAR Closures/Emergencies Locations/Directions Find Pe

Hello tuecke@uchicago.edu [dashboard](#) [sign out](#)

NCAR UCAR Research Data Archive
Computational & Information Systems Lab *weather • data • climate*

Go to Dataset:

Home Find Data Ancillary Services About/Contact Data Citation Web Services For Staff

NCEP Climate Forecast System Version 2 (CFSv2) Monthly Products
ds094.2

For assistance, contact Bob Dattore (303-497-1825)

Description Data Access

Mouse over the table headings for detailed descriptions

Data Description	Data File Downloads		Customizable Data Requests	Other Access Methods
	Web Server Holdings	Globus Transfer Service (GridFTP)	Subsetting	THREDDS
Union of Available Products	Web File Listing	Request Globus Invitation	Get a Subset	Thumbnail Access
Diurnal monthly means	Web File Listing		Get a Subset	
Regular monthly means	Web File Listing		Get a Subset	

Data Selection Summary Signed in as ANANTHAKRISHNANR1

mergesonde1mace c1 @ fkb M1 [Generate Citation](#) 274 file(s) // 6014 MB

2007-04-01

2007-12-31

Order Complete Datastream Extract Specific Measurements

Note: All variables will be delivered for this datastream.

Measurement : Atmospheric temperature
Variable : Temperature // temp

Combine files by datastream

File format

Data Delivery Options

- FTP
- Globus
- THREDDS
- Dropbox

Suspect

"Extract Specific Measurements" is selected.

as part of all orders.

Sanger Imputation Service **Beta** Home About Instructions Resources Status

Sanger Imputation Service

This is a free genotype **imputation** and **phasing** service provided by the Wellcome Trust Sanger Institute. You can upload GWAS data in VCF or 23andMe format and receive imputed and phased genomes back. Click [here](#) to learn more and [follow us on Twitter](#).

Before you start

Be sure to [read through the instructions](#).

You will need to set up a free account with [Globus](#) and have [Globus Connect](#) running at your institute or on your computer to transfer files to and from the service.

Ready to start?

If you are ready to upload your data, please fill in the details below to **register an imputation and/or phasing job**. If you need more information, see the [about](#) page.

Full name

Organisation

Email address

What is this 

Globus user identity

News [@sangerimpute](#)

11/05/2016
Thanks to [EAGLE](#), we can now return **phased data**. The HRC panel has been updated to r1.1 to fix a [known issue](#). See [ChangeLog](#) for more details.

15/02/2016
Globus API changed, please see [updated instructions](#).

17/12/2015
New status page and reworked internals. See [ChangeLog](#).

09/11/2015
Pipeline updated to add some features requested by users. See [ChangeLog](#).

[See older news...](#)



Globus by the numbers

1,042

most shared
endpoints
at a single
institution

425 PB

transferred

70 billion

files processed

1,700

active GCS
endpoints

90

subscribers

100,000

users

3 months

longest running transfer

18,000

active GCP
endpoints

500

identity providers

1 PB

largest single
transfer to date

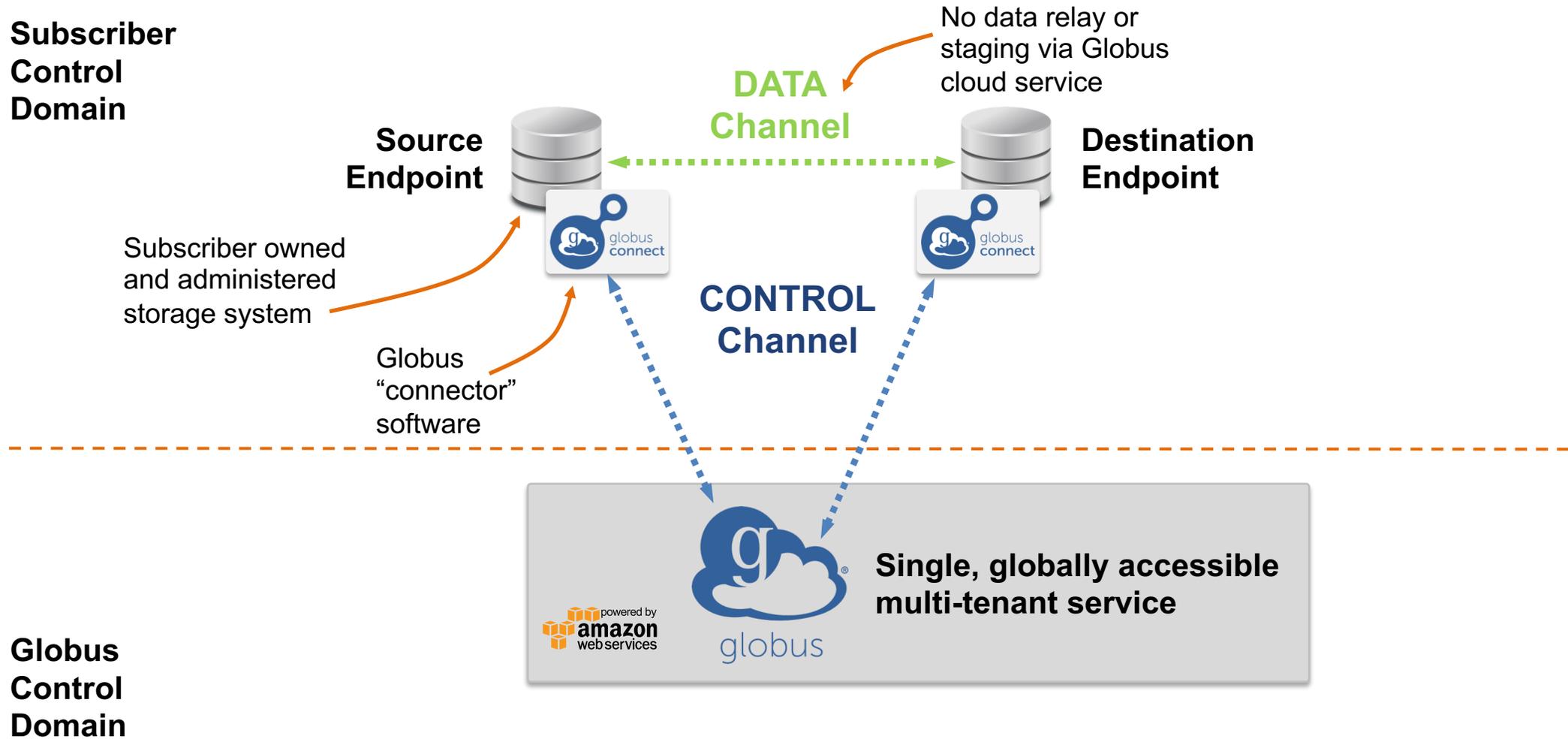
8,000

active shared
endpoints

99.9%

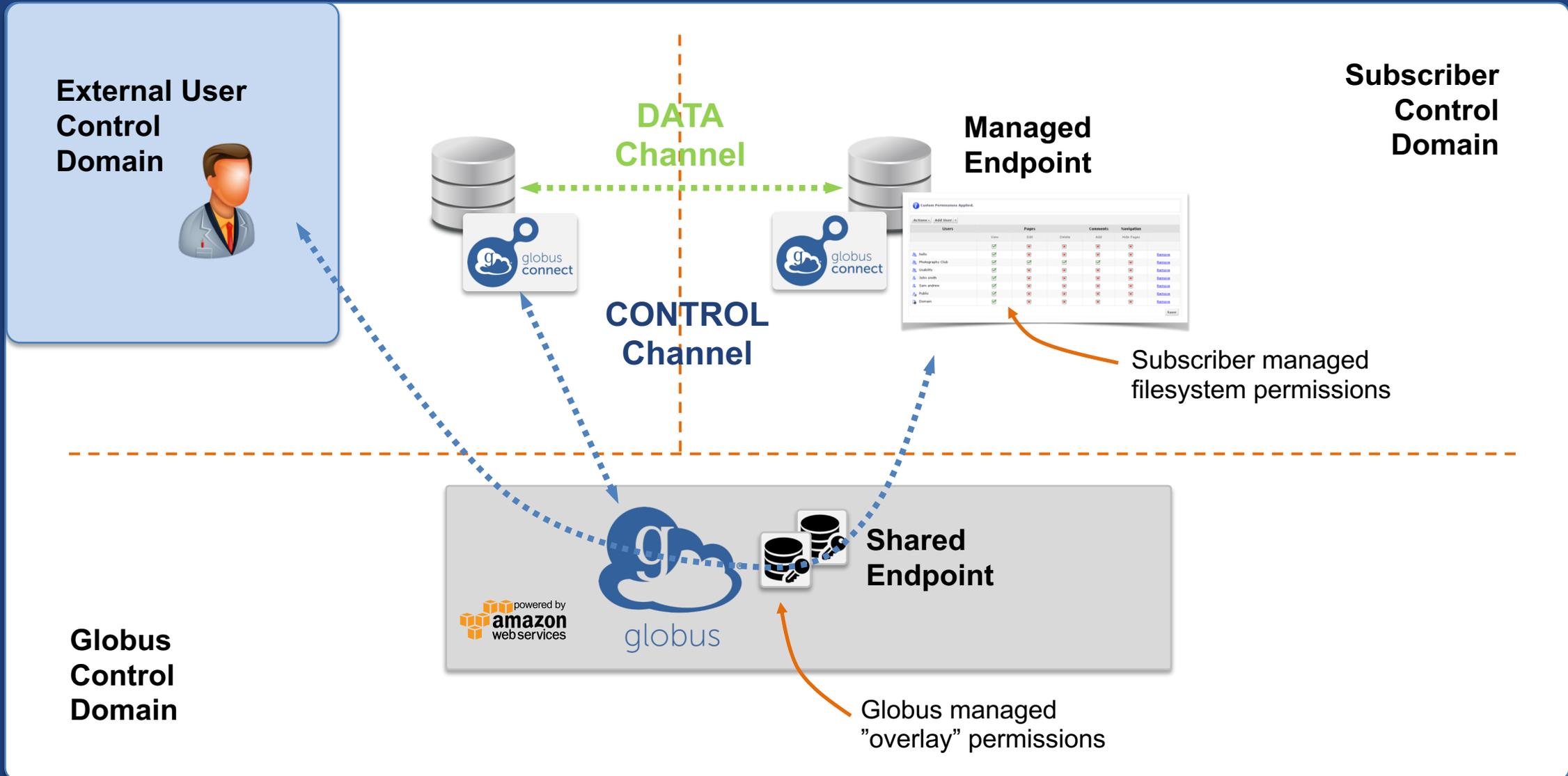
availability

Conceptual architecture: Hybrid SaaS





Conceptual architecture: Sharing

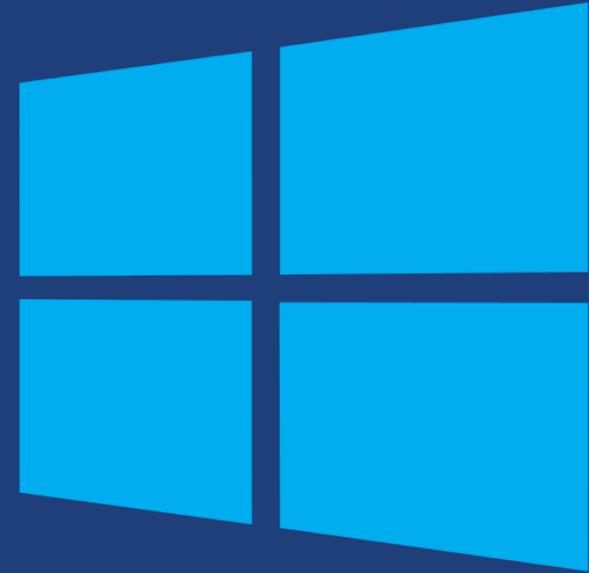


Endpoints (Collections)

- **Storage abstraction**
 - All transfers happen between two endpoints
 - Globus Connect instantiates endpoints
- **Collection \approx Endpoint**
- **Test / Demo Endpoints**
 - Globus Tutorial Endpoint 1
 - Globus Tutorial Endpoint 2
 - ESnet Test Endpoints
 - Contain file samples of various sizes
- **Globus Connect Personal**
 - Now your laptop is an endpoint
 - <https://www.globus.org/globus-connect-personal>



Globus Connect Personal



- **Installers do not require admin access**
- **Zero configuration; auto updating**
- **Handles NATs**
- **Installs in seconds – easy to delete - I'll prove it!**

The Globus Web App - Accounts

- **A Globus Account is**
 - A Primary Identity
 - Possible Linked Identities
- **Linking Identities**
- **Managing Identities**
- **Consents**



The Globus Web App - Hidden in Plain Sight

- **The Hamburger Menu**
 - Not always the same – based on the type of endpoint and the storage behind it
 - A great place to get the link to a share
- **Transfer Settings**
 - label – when using the activity monitor it's nice to see a recognizable name
 - sync - only transfer new or changed files
 - delete files on destination that do not exist on source
 - preserve source file modification times
 - verify file integrity after transfer
 - encrypt transfer
- **Search**
 - Look for the magnifying glass
 - Search for: Endpoints / Users / Groups



Activity Monitoring

- **Recent / History / Filter**
- **Drilling Down**
 - File transfer statistics
 - Overview
 - Event Log
 - Cancelling an active task

Groups

- **What can they be used for?**
 - Sharing: Access permissions for more than one person
 - Roles: Endpoint management and monitoring
- **Groups**
 - Creating groups and setting the visibility
 - Members (invitations), Subgroups, Settings
 - Settings
 - Policies / Membership Fields / Terms & Conditions
 - Roles
 - Giving others authority over your groups

Endpoint Sharing and Roles

- **Sharing**

- Select the directory and create the “share”
- A “share” is another type of endpoint
- Share with: Users / Groups / All Globus Users

- **Roles**

- Give others (users or groups) rights on your endpoint to:
 - Monitor activity on the endpoint
 - Manage activity on the endpoint (e.g., cancel, pause)
 - Manage access control permissions on the endpoint



Bookmarks

- **Just like browser bookmarks – frequently used, or maybe not used frequently enough!**
- **Creating a bookmark**
- **Using a bookmark**
- **Sorting and Filtering**
- **Editing and Deleting**



Globus Command Line Interface

```
(globus-cli) jupiter:~ vas$ globus
Usage: globus [OPTIONS] COMMAND [ARGS]...

Options:
  -v, --verbose          Control level of output
  -h, --help            Show this message and exit.
  -F, --format [json|text] Output format for stdout. Defaults to text
  --map-http-status TEXT Map HTTP statuses to any of these exit codes:
                        0,1,50-99. e.g. "404=50,403=51"

Commands:
  bookmark      Manage Endpoint Bookmarks
  config        Modify, view, and manage your Globus CLI config.
  delete        Submit a Delete Task
  endpoint      Manage Globus Endpoint definitions
  get-identities Lookup Globus Auth Identities
  list-commands List all CLI Commands
  login         Login to Globus to get credentials for the Globus CLI
  logout        Logout of the Globus CLI
  ls            List Endpoint directory contents
  mkdir         Make a directory on an Endpoint
  rename        Rename a file or directory on an Endpoint
  task         Manage asynchronous Tasks
  transfer      Submit a Transfer Task
  version       Show the version and exit
  whoami        Show the currently logged-in identity.
```

Open source, uses
Python SDK

docs.globus.org/cli

[github.com/globus/
globus-cli](https://github.com/globus/globus-cli)



The Globus CLI

- **Installation**
 - docs.globus.org/cli/installation
 - Prerequisites
- **Logging On (remember the consents?)**
 - globus login / logout
- **Getting help / list of commands**
 - globus –help
 - globus list-commands
- **Doing something**
 - It all about the UUIDs
 - Don't forget the file paths!

The Globus CLI – Let's do a few things...

- **Find endpoints**
 - `globus endpoint search Midway`
 - `globus endpoint search ESNNet`
 - `globus endpoint search --filter-scope=recently-used`
- **Find endpoint contents**
 - `globus ls af7bda53-6d04-11e5-ba46-22000b92c6ec`
 - `globus ls af7bda53-6d04-11e5-ba46-22000b92c6ec:RMACC2018`
- **Transfer a file**
 - From ESnet Read-Only Test DTN at CERN to Midway
 - Note the specific paths
 - `globus transfer d8eb36b6-6d04-11e5-ba46-22000b92c6ec:/~/data1/1M.dat af7bda53-6d04-11e5-ba46-22000b92c6ec:/~/1M.dat`
- **Transfer a directory**
 - From Globus Tutorial Endpoint 2 to Midway (create directory and contents)
 - `globus transfer --recursive ddb59af0-6d04-11e5-ba46-22000b92c6ec:/~/sync-demo af7bda53-6d04-11e5-ba46-22000b92c6ec:/~/syncDemo`
- <https://docs.globus.org/cli/examples/>