

Introduction

The University of Michigan Advanced Genomics Core (AGC) generates terabytes of data every day, and delivers that genomic data to clients throughout the research community. The [University has chosen Globus as the preferred file transfer mechanism](#) both within campus and externally to collaborators outside of U of M. Globus is used by both the [ITS](#) and HITS organization on campus, and is used by many of our collaborators on other campuses.

Globus is not a *place*; it doesn't store files. It is a file transfer tool. It's advantage over other file movement tools is that it is built for the reliable movement of large quantities of large-sized files. If you are familiar with FTP or Filezilla, think of Globus as these tools on steroids.



In order to use Globus, we need to have two “endpoints”. **Endpoint** is simply the Globus term for a “file share”. That is, an “endpoint” is a location to store files that Globus can read and write. You may also know of these as “Network Drives”, “Shared Drives”, or “Mount Points”. You’ll need the help of an [IT organization](#) to help set up these drives, because they need to be accessible by Globus. Both ITS and HITS can help set up Globus Endpoints if you already are using them to provide storage (via “Turbo”, “Locker”, “Isilon” and many other storage locations).

If you take the time to read about Globus, you’ll see references to “Globus Personal Connect”. This is a piece of software that you install on your PC to make it accessible by Globus (that is, your PC becomes a globus endpoint). **We strongly recommend against using Globus Personal Connect.** Why? Because when you use Globus Personal Connect, you are copying

your precious genomic data files from the AGC to your PC, which likely does not auto-backup, is subject to getting coffee spilled on it, getting stolen at an airport, and having a hardware failure.

Even using an external USB drive (aka “USB RAID”) that you buy from Amazon or Best Buy isn’t good enough as it does not provide backup redundancy, disaster recovery, and eventually [will fail](#) (The “I” in “RAID” stands for “Inexpensive”).

Using data storage from UM’s ITS or HITS gives you three things:

- 1) **Resilience:** A single hard drive failure won’t impact your data at all.
- 2) **Redundancy:** Data on Locker and Data Den from ITS, and Isilon from HITS are stored in full at two different locations on campus. So a disaster at one data center will not impact your data.
- 3) **Recovery:** Most levels of service for these data storage options include “snapshots” of the data that get taken when the data changes, so if you accidentally modify or delete a file, you can recover an older version of the file with minimal interruption.

And most importantly, using campus-supported data storage gives you a person who you can talk to when you have a problem. Never underestimate having someone who can help you.

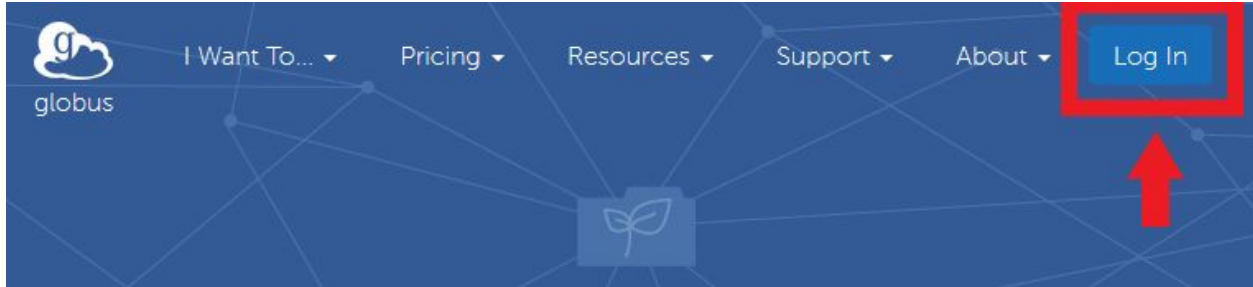
In a nutshell, we can’t stop you from using Globus Personal Connect to put your data on to your laptop. But it’s not good data stewardship, and you’re putting your data at tremendous risk. Best practice it to keep your data on a University file server, and copy data from your share to your work computer as needed to do analysis.

For Medical School researchers, you should [talk with HITS about your data storage needs](#). For the rest of campus, or for medical school researchers who already use Flux or the Great Lakes High-Performance Cluster computers, we suggest you [start with ARC-TS](#) to discuss Turbo, Locker and Data Den for your storage needs.

Globus Basics

Logging in to the Globus Web Page

Visit www.globus.org and click "Login" at the top of the page. On the Globus login page, choose an organization you’re already registered with, such as your school or your employer. Additional instructions can be found in the [Globus documentation](#).



Creating a Public Globus Group

If more than one person needs access to the data, the best way to share is to set up a Globus Group. Step by step instructions on how to create and manage Globus Groups can be found [here](#).

Getting Your Data from the AGC

There are three ways that the AGC can deliver data to you. All of these use Globus to get the data to you, although at Delivery Method #1 you don't need to interact with Globus after an initial setup.

Delivery Method #3: Individual Submitter/PI Pull

This is the default if no other action is taken by your Lab.

1. Your lab will need a Globus endpoint already created (by [HITS](#), [ARC-TS](#), or some other [IT organization](#)).
2. Once your data is ready for delivery, **we put your data into our temporary storage area**, which is also a Globus endpoint.
3. We will grant read access to that temporary area to the submitter and their PI in Globus, and then notify those two people through MiCores that the data is ready for download.
4. Either the submitter or the PI will need to go in to Globus and pull the data from our server to a Globus endpoint that you have already created. **You will have 2 weeks to copy the data off of our endpoint before we delete your data from our file server.**

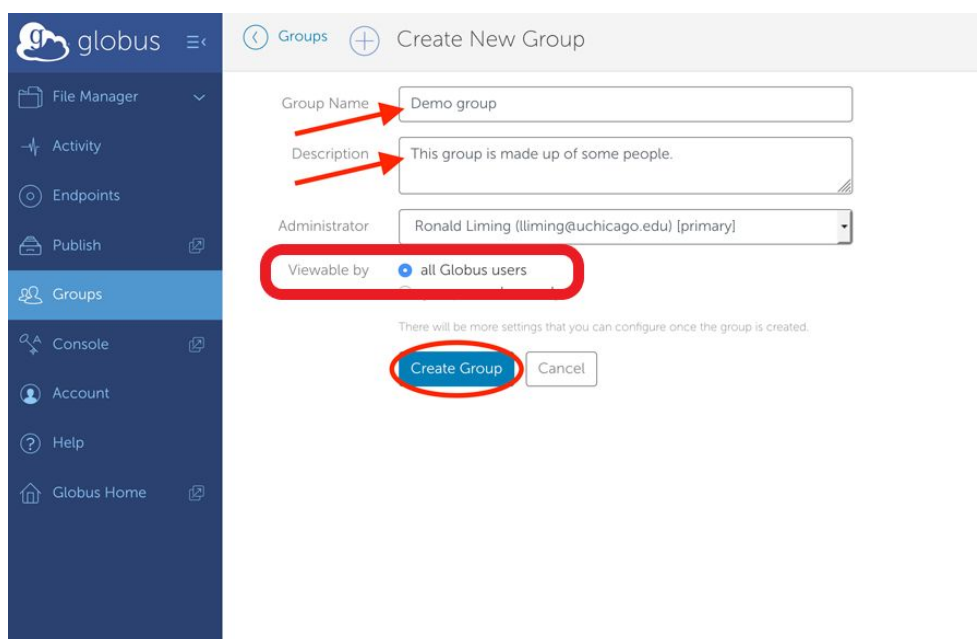
Even this basic level requires your lab to have file storage somewhere, and a "Globus endpoint" configured on that file storage. There are no obvious advantages to this data method, unless the PI or the submitter plan to do all of the processing of the data themselves.

Again this is not recommended, but If you choose to use Globus Personal Connect to put your precious data on your highly-susceptible PC, then this option will work for you, but *only* for the PI and the person who submitted the Service Request in MiCores.

Delivery Method #2: Group Pull

This model is similar to what the DNA Sequencing Core used to do, but with a slight Globus twist.

1. Your lab will need a Globus endpoint already created (by HITS, ITS, or some other IT organization).
2. Your Lab will [create a Globus Group](#) with the identities of all Globus users who will be able to copy your data from the AGC's temporary file server.
3. **That Globus Group will need to be viewable by "all Globus users".**



The screenshot shows the 'Create New Group' form in the Globus interface. The 'Group Name' field contains 'Demo group'. The 'Description' field contains 'This group is made up of some people.'. The 'Administrator' field shows 'Ronald Liming (rliming@uchicago.edu) [primary]'. The 'Viewable by' dropdown is set to 'all Globus users' and is circled in red. Below the form, there is a note: 'There will be more settings that you can configure once the group is created.' At the bottom, there are two buttons: 'Create Group' (circled in red) and 'Cancel'.

4. When you submit your request to the AGC in MiCores, you will include the name of your **public** Globus Group in the submission.
5. When your data is ready, **we will put your data in our temporary storage area**, and grant that group read-only access, and notify you via a comment in your MiCores submission.
6. Someone from that group must log in to Globus and transfer the data to your Globus endpoint. **Your team will have 2 weeks to copy the data off of our endpoint before we delete your data from our file server.**

The advantage of Delivery Method #2 over #3 is that you can delegate the actual copying of the data to someone else. Once you complete steps 1-3, you just re-use the same public Globus group for all future submissions. The disadvantage of this delivery method is that you must create a public Globus Group.

If you choose to use Globus Personal Connect to put your precious data on your highly-susceptible PC, then this option will work for all people who are in the Globus Group.

Delivery Method #1: Push (good for frequent AGC user-groups)

If your lab has worked with ITS or HITS to set up storage, we can push data directly to that storage location. Your IT contacts can work with the AGC to make sure that your storage location is shared on Globus, and that the AGC's Globus Group (UMICH-AGC-Data) has write access.

1. When you submit your request to the AGC through MiCores, you can tell us your group's endpoint name
2. When your data is ready for delivery, the **AGC will write your data directly to the endpoint** your specified in your submission on MiCores.
3. Once the copy is complete, we will comment on the request in MiCores, and you can begin using your data.

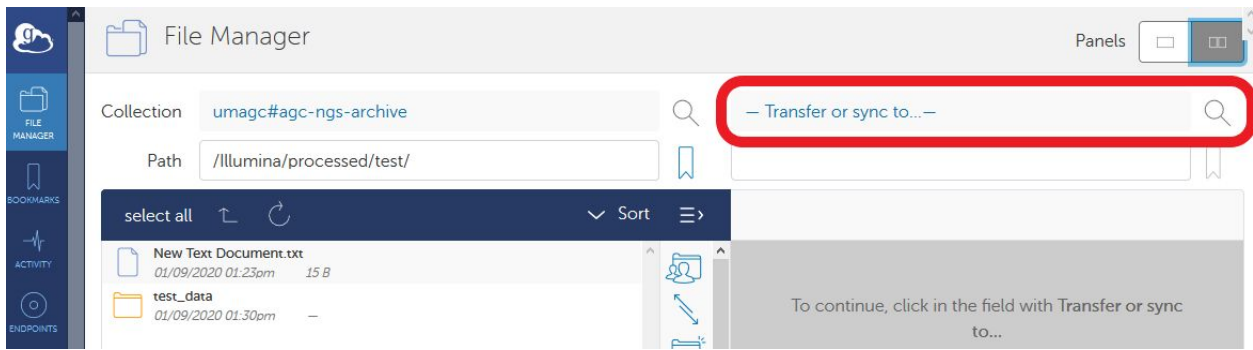
The advantage of Delivery Method #1 over #2 is that your data will get to your file server with no effort on the part of your Lab. The disadvantage is that the AGC will need to have write access to your file server, and may fill up your storage area if you are not monitoring it.

We believe that this delivery method is what most of our "power users" will want to use. After the initial set-up, this method requires minimal effort on the part of your Lab.

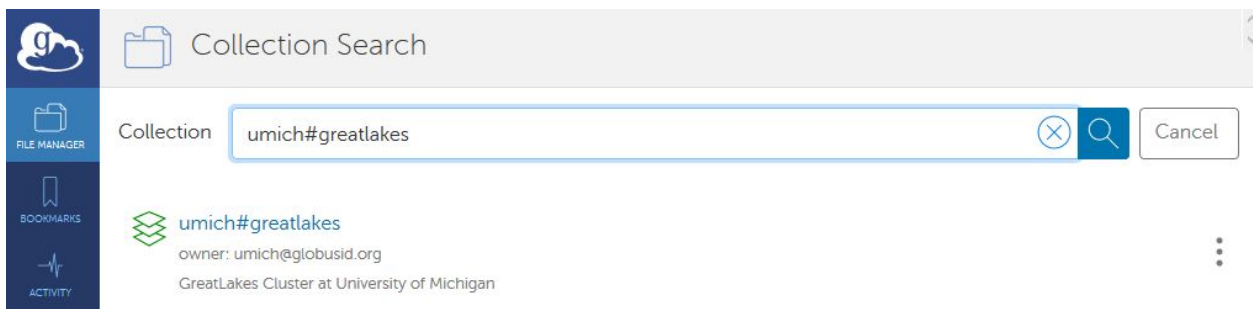
You cannot use this option if you are using Globus Personal Connect to store your precious data on a computer in your Lab, because pushes are not supported by Globus Personal Connect endpoint.

How to Transfer Files with Globus (Methods #2 or #3)

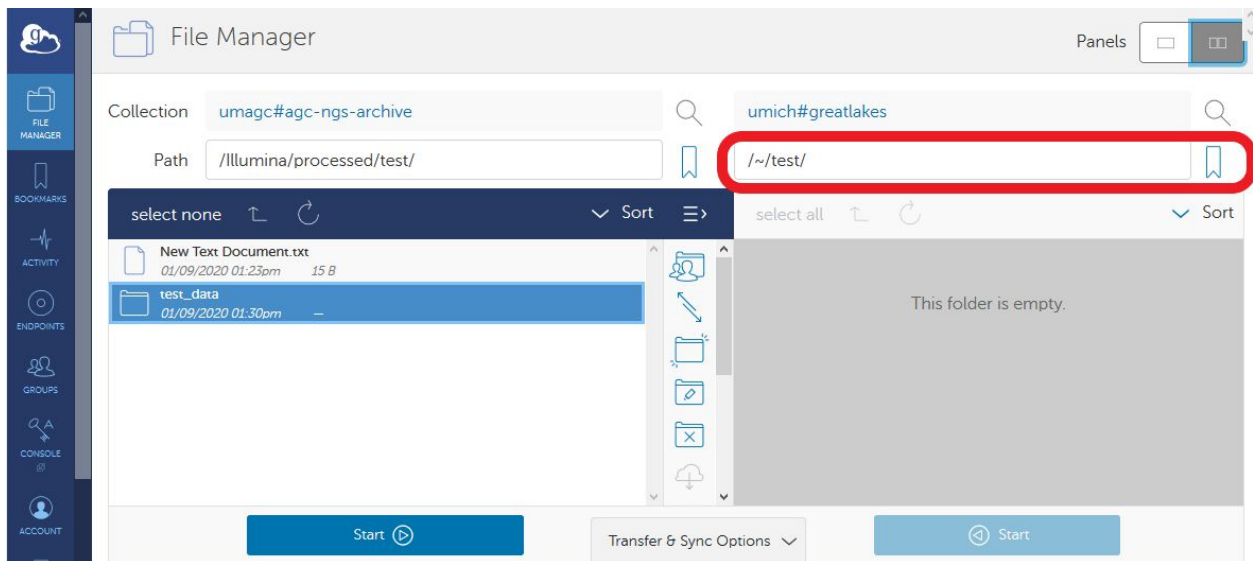
When the AGC has made your data accessible, you will be sent a link from Globus to the shared folder. Clicking this link will send you to the Globus login. Once you are logged-in and ready to transfer files, you will be taken to the "File Manager" page, as seen below. Click "[-Transfer or sync to...-](#)".



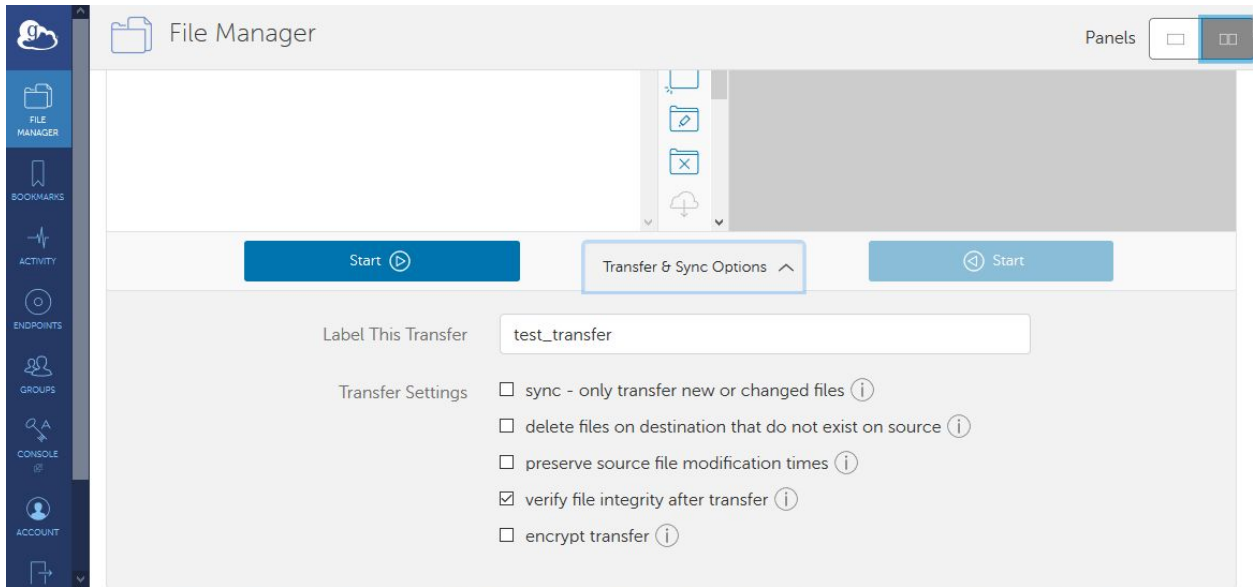
Which allows you to search for your group's endpoint (or personal endpoint).



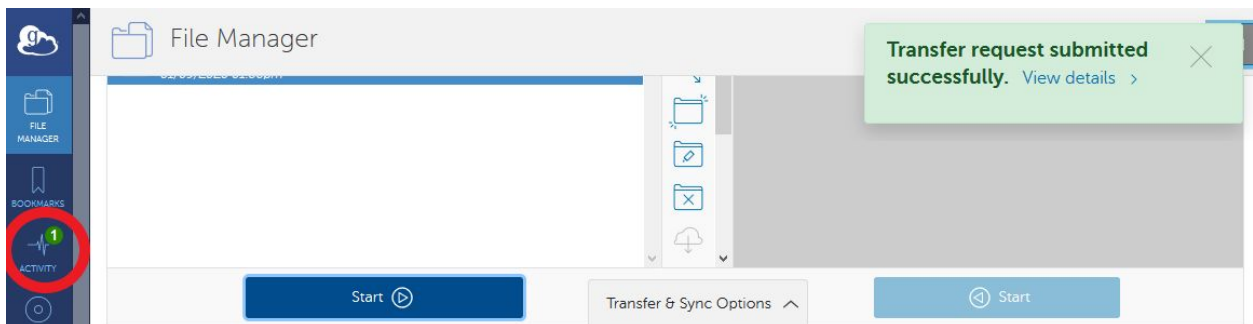
Clicking on your endpoint will take you back to the File Manager. If you get an error, you may need to set up a [Great Lakes account](#) or you may need to set up your or your group's endpoint. Then, enter the path (red in below image) where you want to transfer your files on your endpoint and select the files/folders you want to transfer (folder now in blue on left in below image).



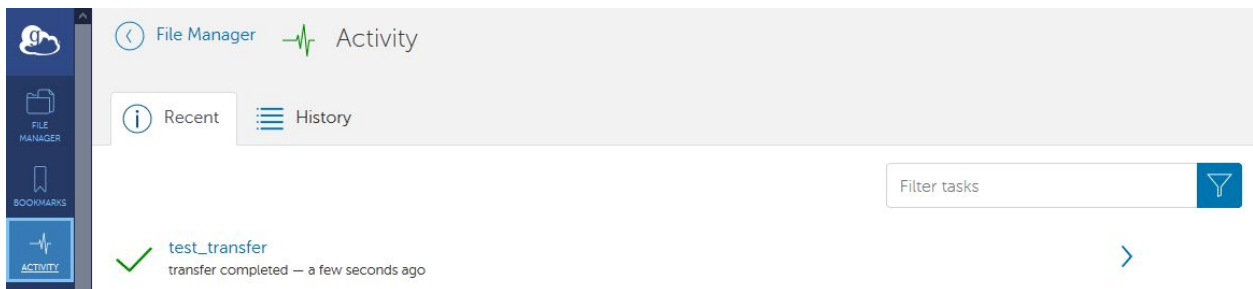
Click the “Transfer & Sync Options” at the bottom of the screen to change transfer settings and label the transfer for your records.



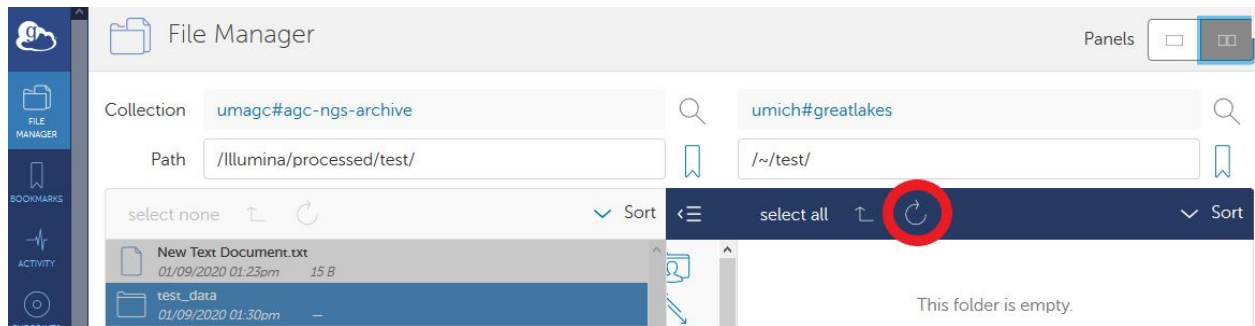
Click the “Start” button and you should see a message saying you submitted successfully. You can also check the status by clicking the “Activity” button on the side menu.



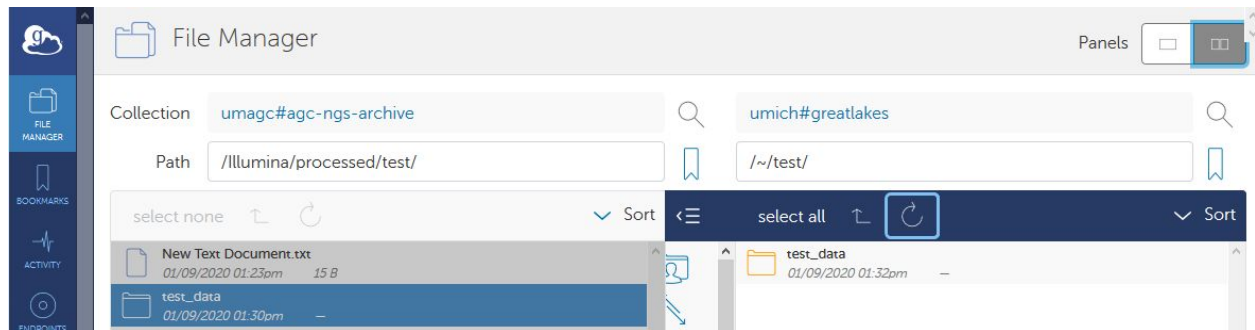
Which takes you to the activity page where you can see the details of your transfer.



On the File Manager, your files will not appear on your endpoint until you refresh the page by clicking “refresh” button, as seen below.



And voila! You have your data on your endpoint for your use.



Additional Support Documentation from Globus

[Instructional Videos](#)

[Getting Started - How To Log In and Transfer Files with Globus](#)

[How To Share Data Using Globus](#)

[How to Create and Manage Globus Groups](#)

[FAQs: Transfer and Sharing](#)

[FAQs: Globus Connect and Endpoints](#)