

# Table of Contents

<b>Globus Basics and File Transfers</b>	<b>2</b>
Executive Summary . . . . .	2
Prerequisites . . . . .	2
Getting Started: Logging Into Globus . . . . .	3
Navigating the Globus Interface . . . . .	5
Transferring Files Between Collections . . . . .	8
Monitoring Transfer Progress . . . . .	14
Handling Authentication Requests . . . . .	15
Troubleshooting Common Issues . . . . .	16
Next Steps: Building on the Basics . . . . .	17
Quick Reference . . . . .	17
<b>Creating Guest Collections for External Collaboration</b>	<b>19</b>
Executive Summary . . . . .	19
Globus Access and Account Requirements . . . . .	19
Prerequisites . . . . .	19
Creating Your Guest Collection . . . . .	20
Managing Ongoing Collaborations . . . . .	27
Troubleshooting Sharing Issues . . . . .	27
Collaboration Best Practices . . . . .	28
For Your Collaborators: Accessing Shared Data . . . . .	28
Support and Additional Resources . . . . .	29
<b>Globus Connect Personal</b>	<b>30</b>
Executive Summary . . . . .	30
Prerequisites . . . . .	30
Installing Globus Connect Personal on a Smithsonian Managed Computer . . . . .	30
Installing Globus Connect Personal on a Personal Device . . . . .	32
Transferring Data To and From Your Device . . . . .	36
Managing Globus Connect Personal . . . . .	38
Working with External Drives . . . . .	38
Troubleshooting . . . . .	39
Best Practices for Large Transfer Projects . . . . .	39
Next Steps . . . . .	40
Quick Reference . . . . .	40
<b>Advanced Globus Features</b>	<b>41</b>
Executive Summary . . . . .	41
Sync vs. Transfer: Choosing the Right Tool . . . . .	41
Scheduling Transfers and Syncs . . . . .	43
Globus Flows: Automated Multi-Step Workflows . . . . .	47
Advanced Transfer Options . . . . .	49
Real-World Workflow Examples . . . . .	50
Troubleshooting Advanced Features . . . . .	51
Quick Reference . . . . .	52

# Globus Basics and File Transfers

## Executive Summary

This guide introduces Smithsonian staff to Globus, a secure platform for transferring large datasets between storage systems. You'll learn how to log in with your Smithsonian credentials, navigate the web interface, transfer files between collections, and monitor your transfer activity. This foundational guide prepares you for more advanced Globus capabilities covered in our companion guides.

### What is Globus?

Globus is a data management service that enables fast, reliable transfer of large files between storage systems. Unlike email attachments or consumer cloud services, Globus is designed for large-scale data—think gigabytes to terabytes of genomic sequences, archival video collections, high-resolution imagery, or simulation outputs. Transfers happen server-to-server, so you can start a transfer and close your laptop; Globus handles the rest.

### Smithsonian Globus Infrastructure

Smithsonian maintains dedicated **Globus Data Transfer Nodes** at several locations:

- **Smithsonian Data Center** — Includes access to Hydra (high-performance computing cluster) storage, DAMS NAS storage, and certain unit managed network drives
- **STRI** (Smithsonian Tropical Research Institute) — For tropical research data workflows
- **SAO** (Smithsonian Astrophysical Observatory) — For astrophysical research data

These institutional endpoints allow you to move data between Smithsonian systems and collaborate with partners worldwide.

## Prerequisites

### Account Requirements

What You Need	How to Get It
<b>Smithsonian network account</b>	Your standard SI credentials (same as email login)
<b>Server account</b> (for Smithsonian storage access)	Contact system administrators (see below)

**To access specific storage systems:** - **Hydra:** Request an HPC account on the SI Service Portal - **Hydra HPC Support:** SI-HPC@si.edu - **DAMS NAS Workflows:** SI-Globus@si.edu - **STRI users:** Contact STRIhelp@si.edu to request Globus Data Transfer Node Server access

Need more general support? Contact *SI-Globus@si.edu*

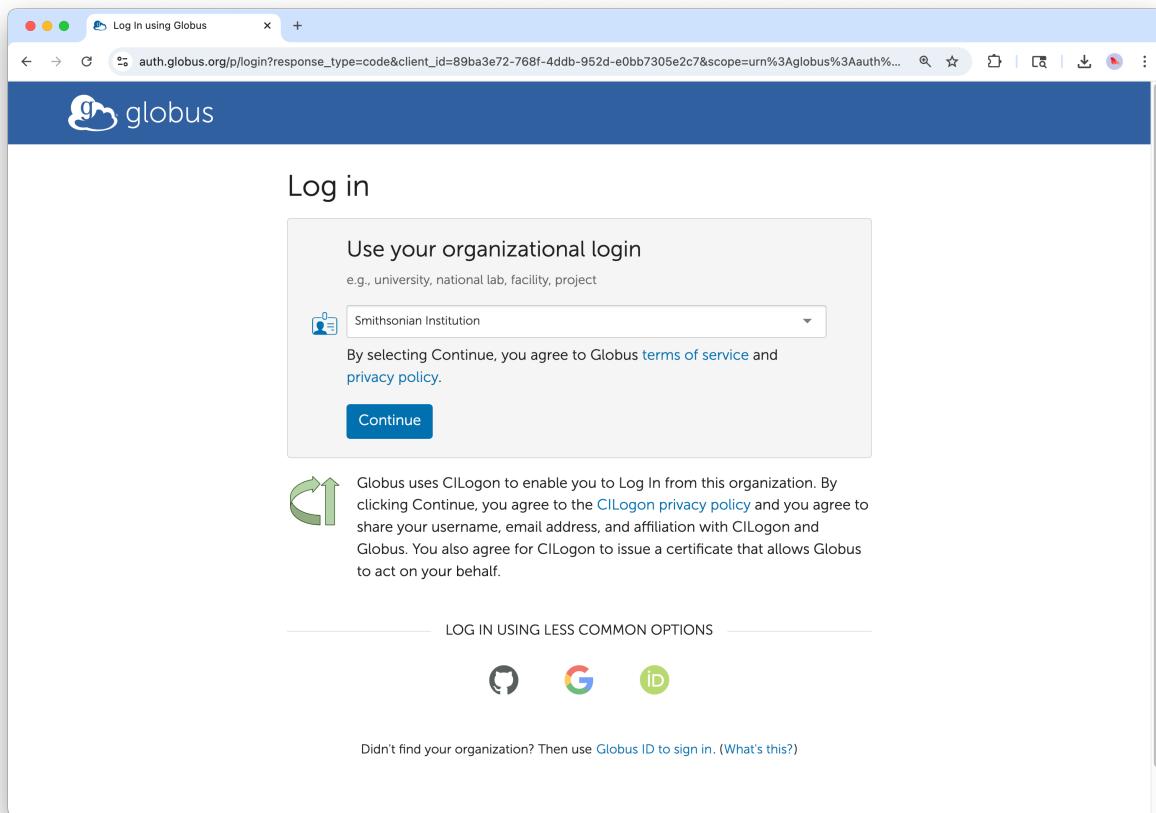
### Network Requirements

You can access Globus from anywhere with an internet connection—no VPN required for the web interface. However, you'll need appropriate server accounts to access data stored on Smithsonian systems.

## Getting Started: Logging Into Globus

### Step 1: Navigate to Globus

Open your web browser and go to [app.globus.org](https://app.globus.org). You'll see the Globus login page.



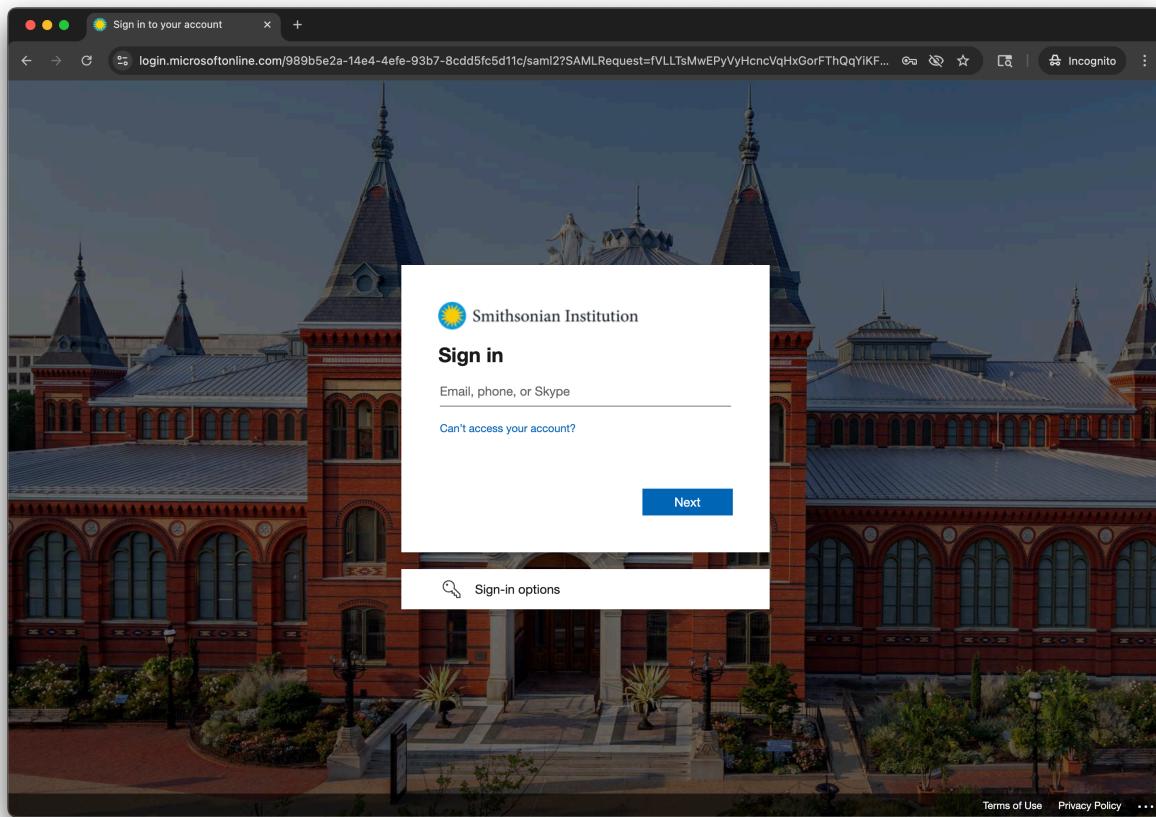
The *Globus login page*. Use the dropdown to search for *Smithsonian Institution*.

### Step 2: Select Your Institution

Click the organization dropdown and search for “Smithsonian Institution.” Select it from the list, then click **Continue**.

Globus uses the institution’s existing login system—you’ll authenticate with the same credentials you use for Smithsonian email and other institutional services.

## Step 3: Authenticate with Smithsonian Credentials



*Enter your Smithsonian network credentials when prompted.*

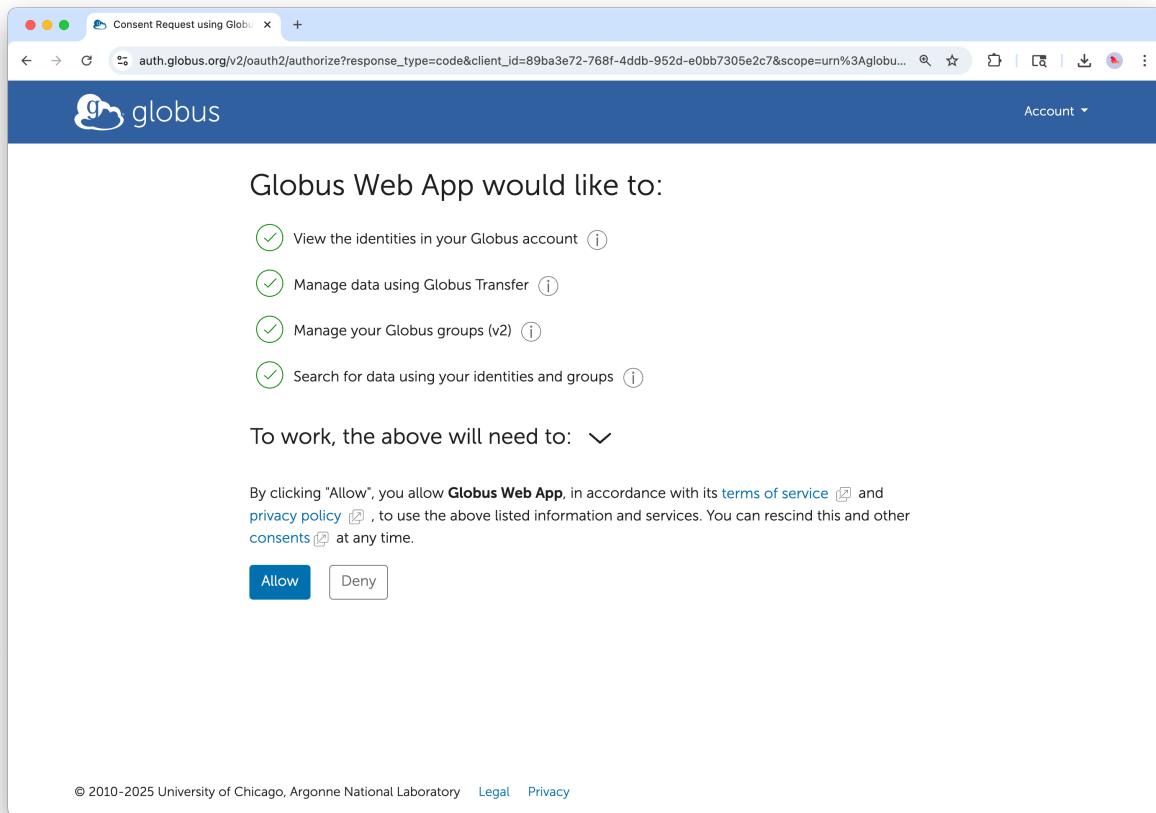
You'll be redirected to the Smithsonian login page. Enter your network username and password, and complete multi-factor authentication.

## Step 4: First-Time Login Setup

If this is your first time using Globus with your Smithsonian credentials:

1. You'll see a welcome message confirming successful authentication
2. If you've used Globus before with a different login (personal Google account, another institution), you can link accounts to maintain a unified history
3. For most users, simply click **Continue** to proceed

## Step 5: Grant Application Permissions



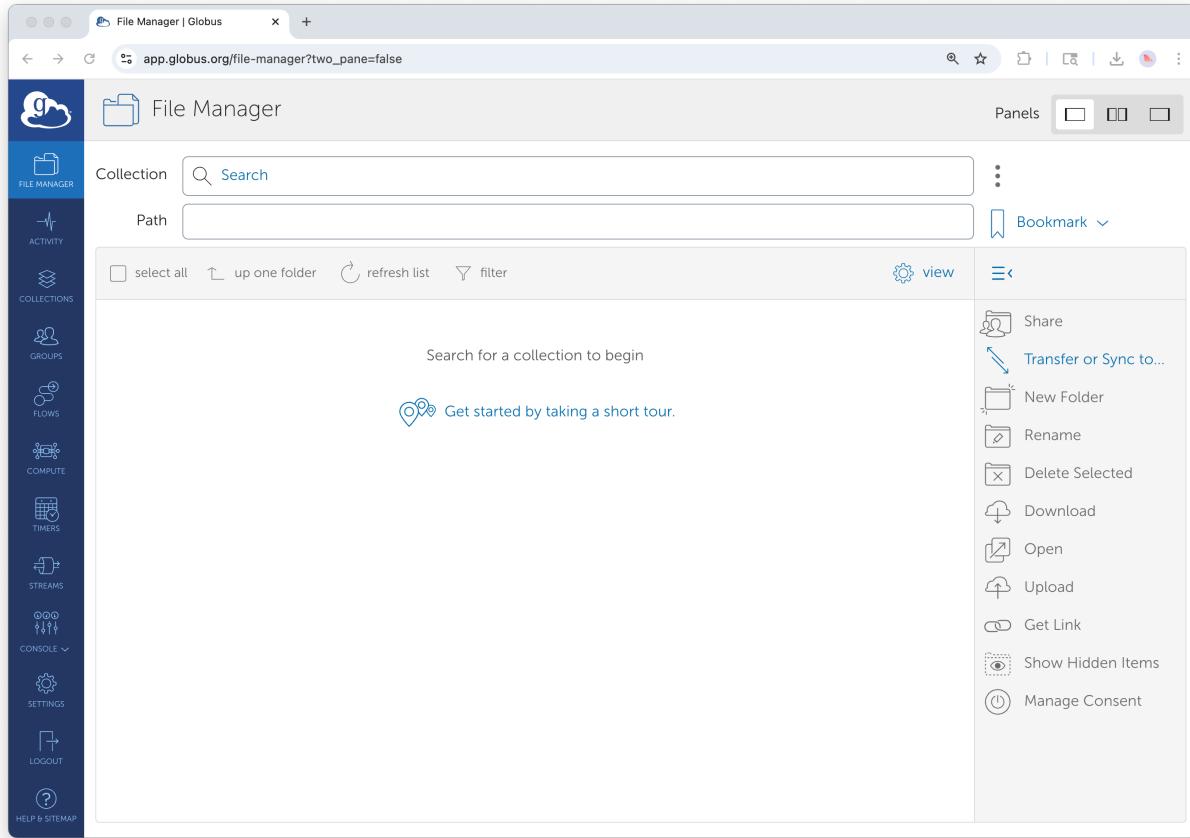
*Globus requests permission to access data on your behalf. Click Allow to proceed.*

Globus will ask for permission to perform actions on your behalf—searching collections, managing transfers, and accessing your data. Review the permissions and click **Allow** to continue to the File Manager.

## Navigating the Globus Interface

After logging in, you'll land in the **File Manager**—your primary workspace for browsing and transferring files.

## The File Manager Overview



The File Manager interface. The left sidebar provides navigation; the main area displays your current collection.

**Left Sidebar Navigation:** - **File Manager** — Browse collections and initiate transfers (where you'll spend most of your time) - **Activity** — Monitor ongoing and completed transfers - **Collections** — View and manage collections you own or have access to - **Groups** — Manage group memberships for collaborative access - **Settings** — Configure preferences and manage your Globus identity

**Main Panel:** - **Collection** field — Search for and select data collections - **Path** field — Shows your current directory location within the collection - **File listing** — Browse files and folders in the selected collection

### Finding Collections

Collections are the fundamental unit in Globus. They represent accessible storage locations. To work with any data, you first need to find and select a collection.

The screenshot shows a web-based file manager interface titled "File Manager - Collection Search". On the left is a sidebar with various icons for FILE MANAGER, ACTIVITY, COLLECTIONS, GROUPS, FLOWS, COMPUTE, TIMERS, STREAMS, CONSOLE, SETTINGS, LOGOUT, and HELP & SITEMAP. The main area has a search bar with the text "Smithsonian" and a checked checkbox for "Search All Collections". Below the search bar is a table listing five collections:

Collection	Description	Actions
SI_DAMS_Staging_JPCARCHIVE	Subscribed Mapped Collection (GCS) on <b>Smithsonian Hydra Globus Endpoint</b> Owner: siglobus@globusid.org Domain: m-482676.9d21.8443.data.globus.org Description: Smithsonian DAMS Staging JPCARCHIVE Mount	Edit, Delete, More
SI_DAMS_Staging_NMAH-Video	Subscribed Mapped Collection (GCS) on <b>Smithsonian Hydra Globus Endpoint</b> Owner: siglobus@globusid.org Domain: m-244cea.9d21.8443.data.globus.org Description: Smithsonian DAMS Staging NMAH-Video Mount	Edit, Delete, More
SI_DAMS_Staging_NMNH-HSFA	Subscribed Mapped Collection (GCS) on <b>Smithsonian Hydra Globus Endpoint</b> Owner: siglobus@globusid.org Domain: m-d4b319.9d21.8443.data.globus.org Description: Smithsonian DAMS Staging NMNH-HSFA Mount	Edit, Delete, More
SI_DAMS_Staging_SLA-AVMPI	Subscribed Mapped Collection (GCS) on <b>Smithsonian Hydra Globus Endpoint</b> Owner: siglobus@globusid.org Domain: m-6509b9.9d21.8443.data.globus.org Description: Smithsonian DAMS Staging SLA-AVMPI Mount	Edit, Delete, More
SI_DAMS_Staging_TEST	Subscribed Mapped Collection (GCS) on <b>Smithsonian Hydra Globus Endpoint</b> Owner: siglobus@globusid.org	Edit, Delete, More

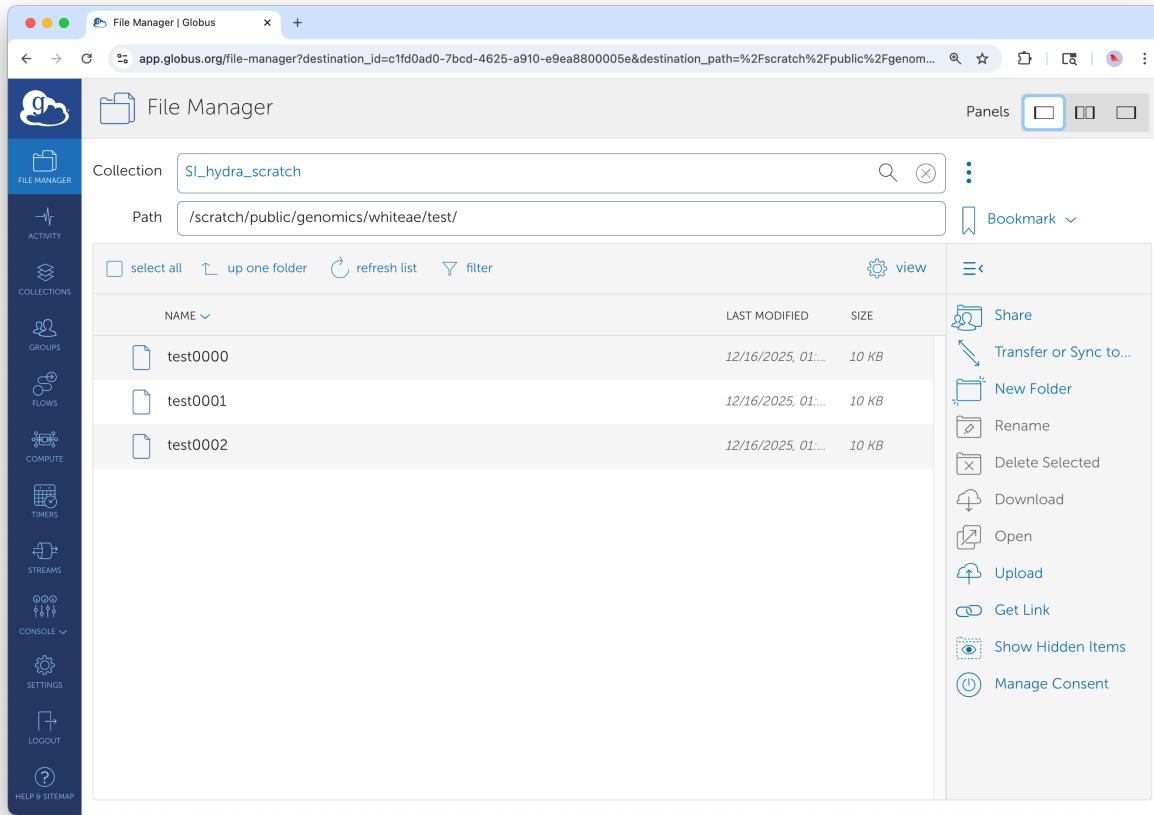
*Click in the Collection field and type to search. Results appear as you type.*

### To find a Smithsonian collection:

1. Click in the **Collection** search field
2. Search for "**Smithsonian**" to see all collections managed on Smithsonian Data Transfer Nodes
3. Browse the results to find the collection matching your storage system (Hydra, DAMS Staging NAS, STRI, SAO, etc.)
4. Select the appropriate collection from the search results

**Tip:** There are dozens of Smithsonian-managed collections. If you're unsure which collection corresponds to your storage system, contact SI-Globus@si.edu for guidance.

## Browsing Files and Folders



After selecting a collection, you can browse directories and files.

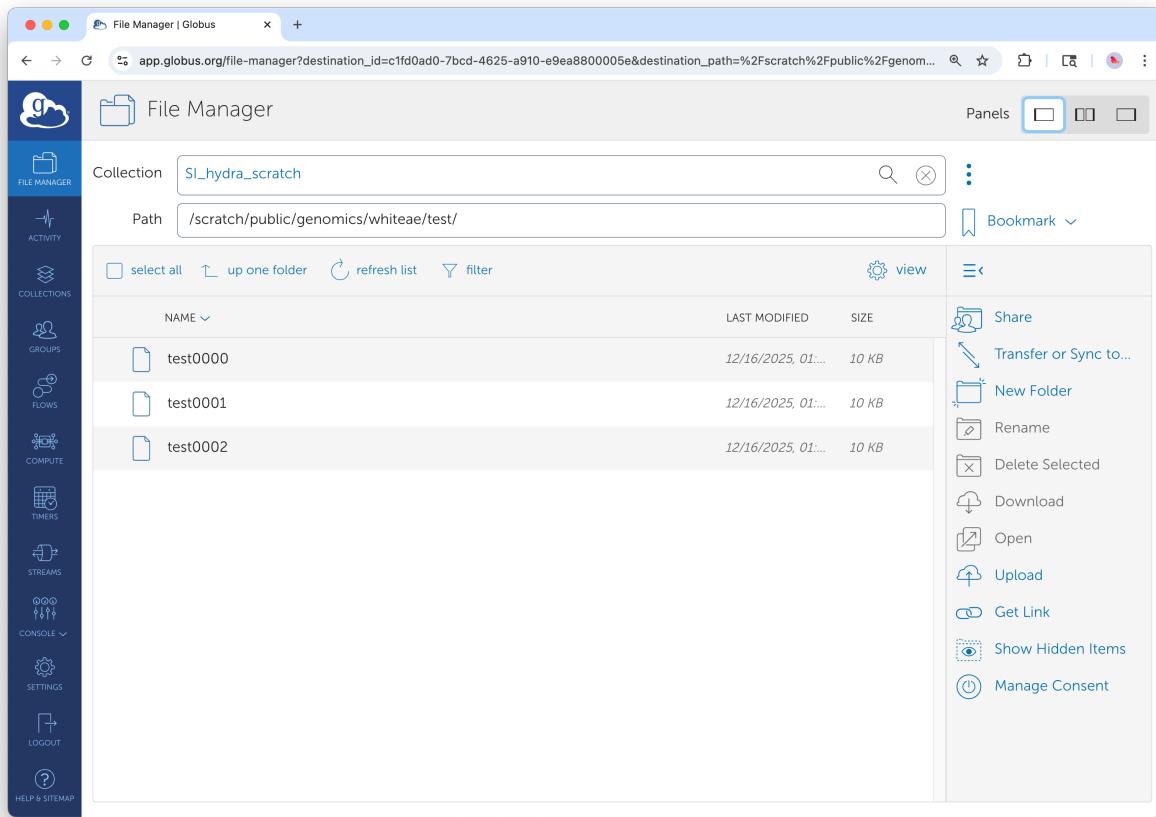
Once you've selected a collection:

- **Double-click folders** to navigate into them
- Use "**up one folder**" to move to the parent directory
- The **Path** field shows your current location—you can also type paths directly
- Click **refresh list** to update the view if files have changed

## Transferring Files Between Collections

The core Globus workflow involves transferring files from a **source** collection to a **destination** collection. Globus uses a two-panel interface that makes this intuitive.

## Step 1: Open the Two-Panel View

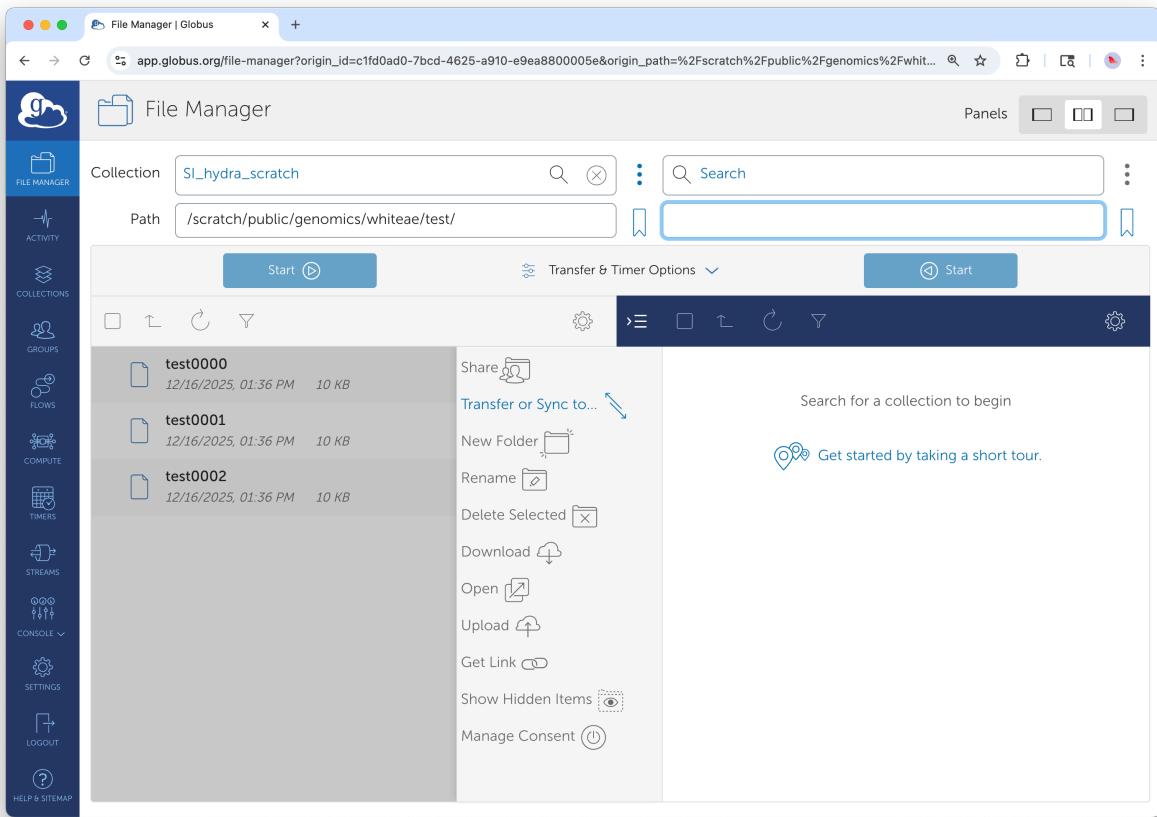


Click the **Panels** button in the top right to enable two-panel view for transfers.

In the top-right corner of the File Manager, you'll see **Panels** options. Click to enable the two-panel view, which displays source and destination side by side.

## Step 2: Select Your Source Collection

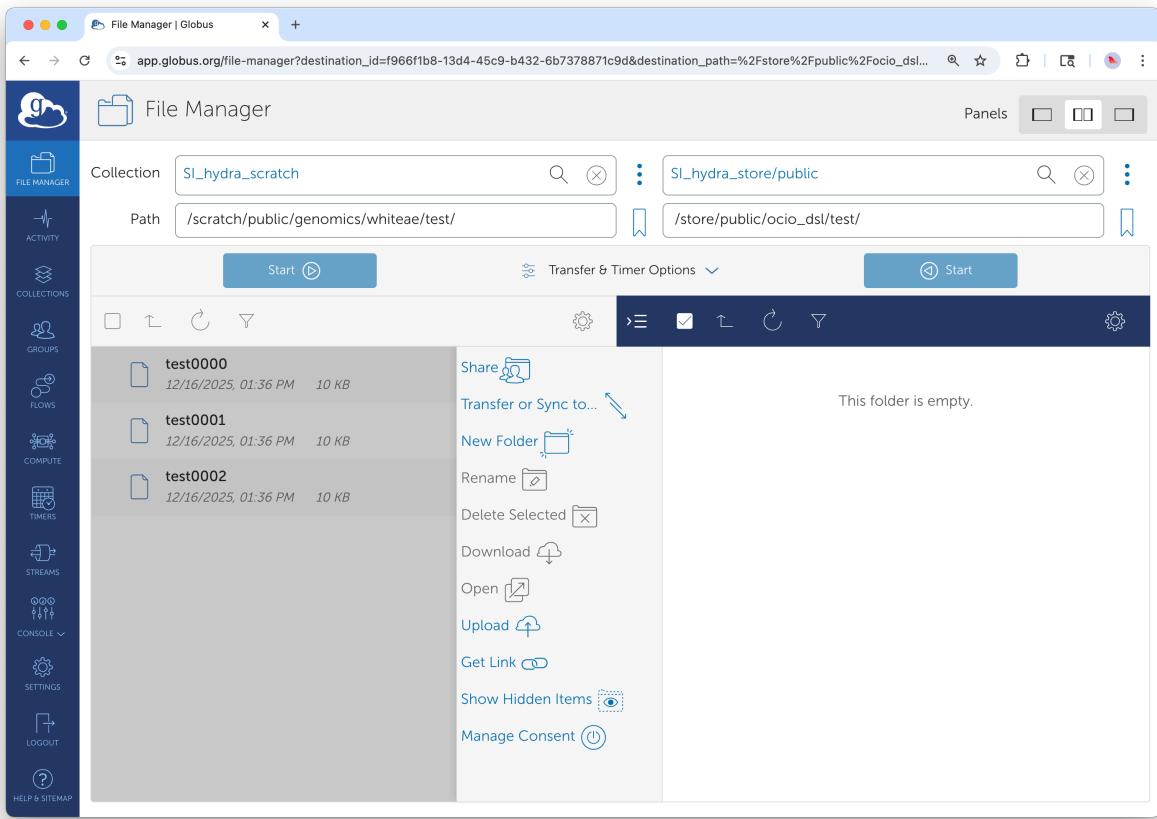
In the **left panel**, search for and select the collection containing the files you want to transfer. Navigate to the directory with your data.



The left panel shows your source collection. Browse to the files you want to transfer.

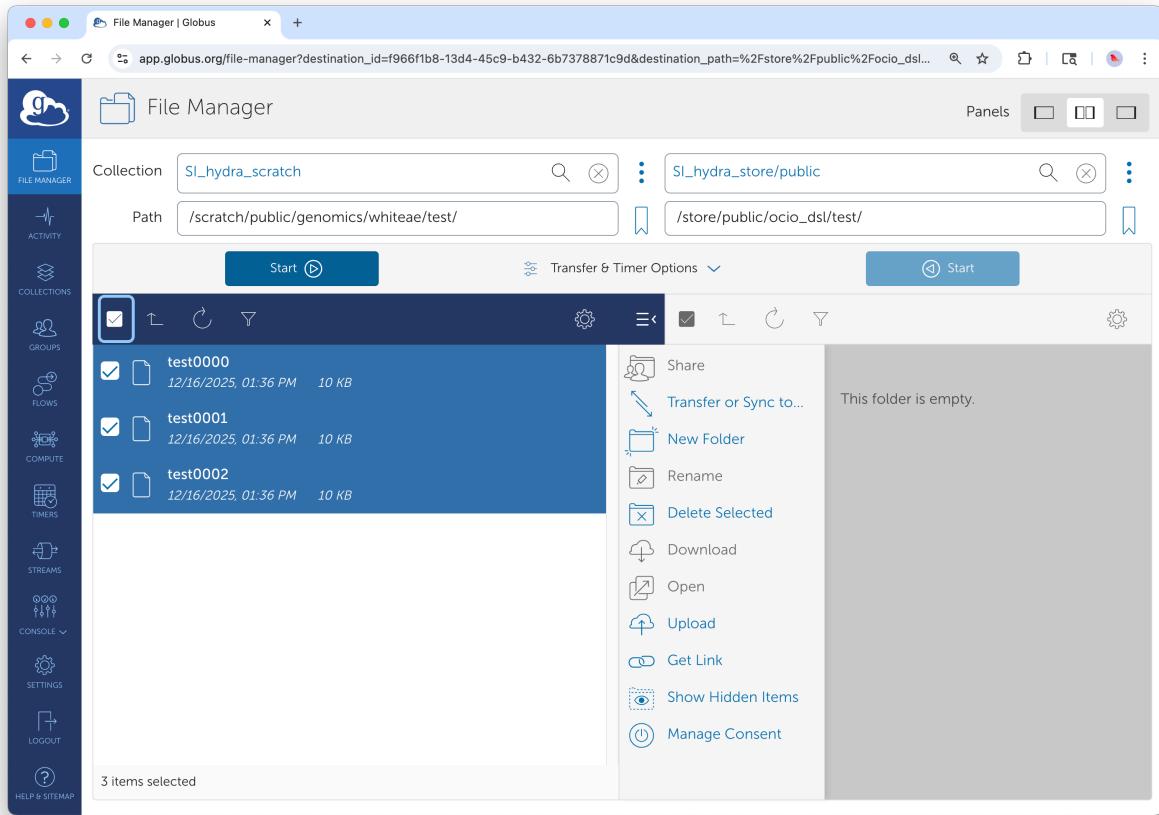
### Step 3: Select Your Destination Collection

In the **right panel**, search for and select where you want the files to go. Navigate to the target directory.



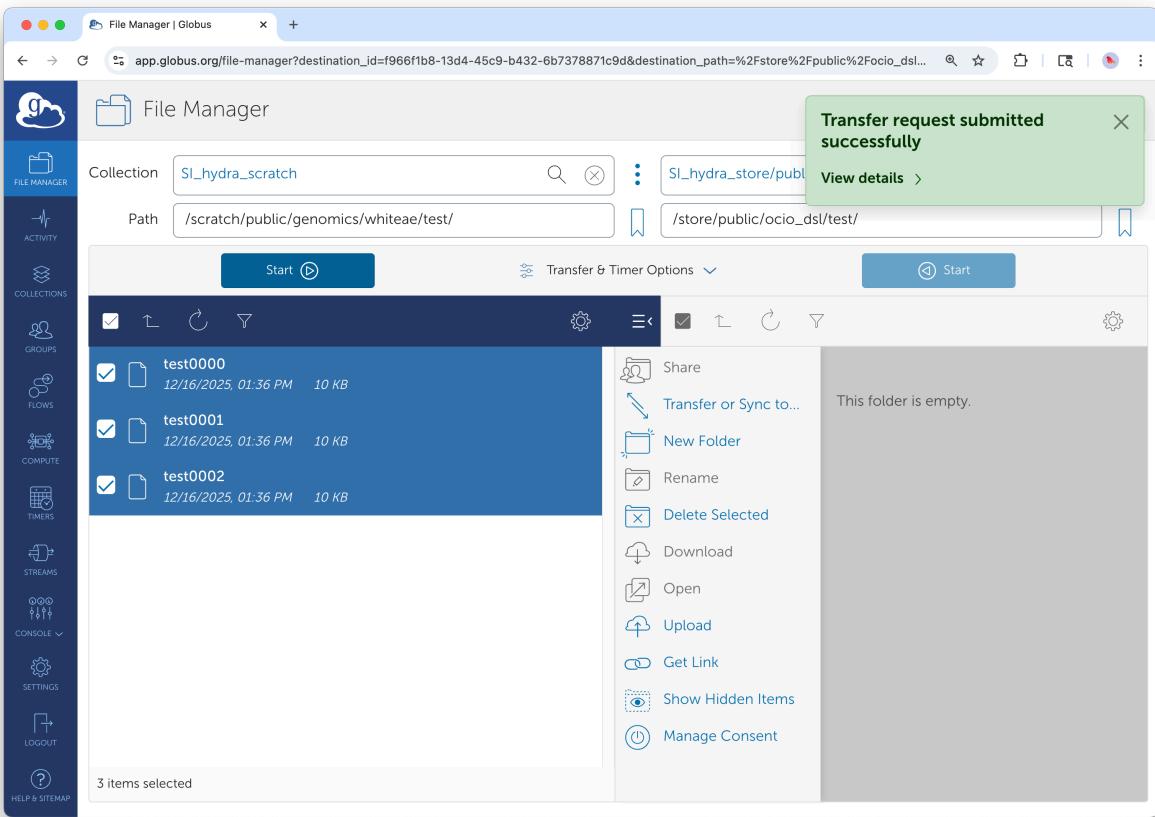
*Set up both panels: source on the left, destination on the right.*

## Step 4: Select Files and Start Transfer



Select files using checkboxes, then click Start to initiate the transfer.

1. In the source panel (left), **check the boxes** next to files or folders you want to transfer
2. Selected items will be highlighted
3. Click the **Start** button (blue button with arrow) above the source panel
4. You'll see a confirmation that your transfer request was submitted



Transfer submitted confirmation. Click “View details” to monitor progress.

**Important:** You can now close your browser or shut down your laptop. Globus handles the transfer server-to-server, you don’t need to stay connected.

## Understanding Transfer Direction

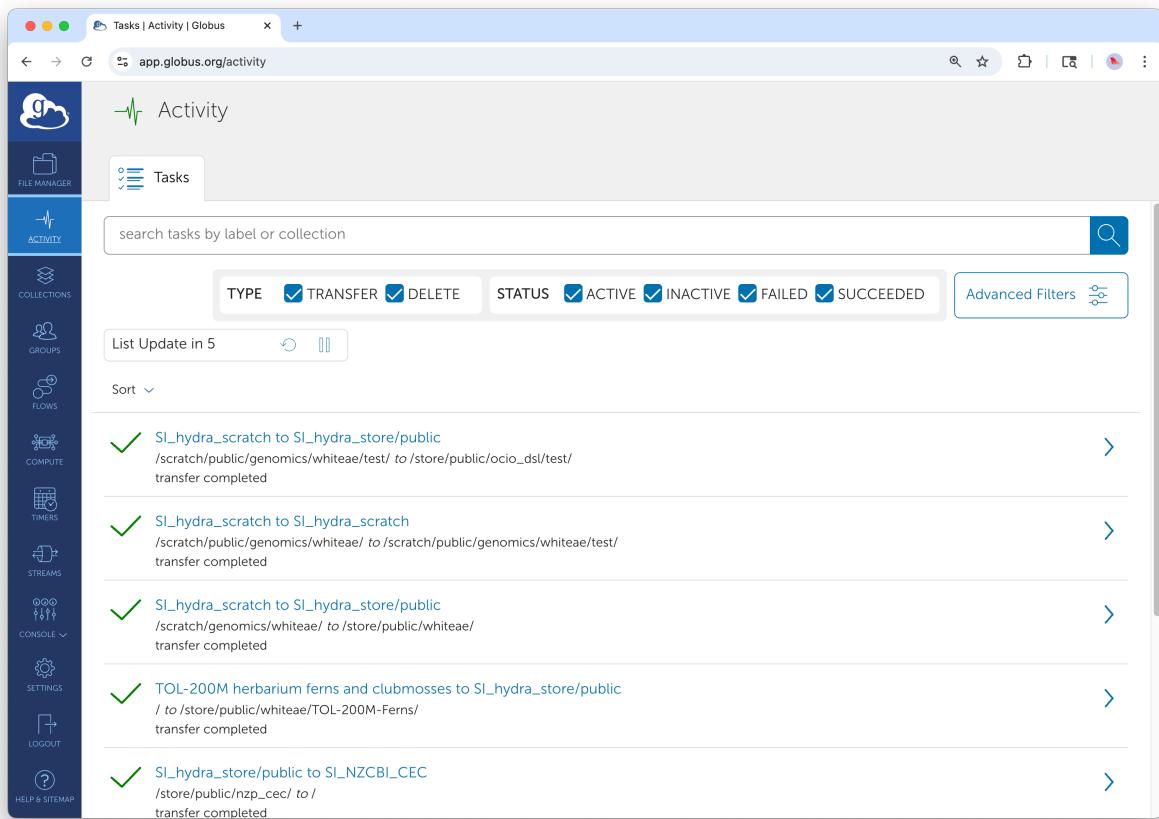
The **Start** button appears on both panels. The button you click determines the transfer direction:

- Click **Start** on the **left panel** → Transfer from left to right
- Click **Start** on the **right panel** → Transfer from right to left

The blue “Start” button with the play icon initiates the transfer; the teal button with the clock icon is for scheduling (covered in Guide 4: Advanced Features).

## Monitoring Transfer Progress

### Viewing Active and Completed Transfers



The screenshot shows the 'Activity' page of the Globus web interface. The left sidebar has 'ACTIVITY' selected. The main area is titled 'Activity' with a 'Tasks' tab selected. A search bar says 'search tasks by label or collection'. Filter buttons for 'TYPE' (TRANSFER, DELETE), 'STATUS' (ACTIVE, INACTIVE, FAILED, SUCCEEDED), and 'Advanced Filters' are present. A 'List Update in 5' button and a 'Sort' dropdown are also visible. Below is a list of completed transfers:

- SI\_hydra\_scratch to SI\_hydra\_store/public  
/scratch/public/genomics/whiteae/test/ to /store/public/ocio\_dsl/test/  
transfer completed
- SI\_hydra\_scratch to SI\_hydra\_scratch  
/scratch/public/genomics/whiteae/ to /scratch/public/genomics/whiteae/test/  
transfer completed
- SI\_hydra\_scratch to SI\_hydra\_store/public  
/scratch/genomics/whiteae/ to /store/public/whiteae/  
transfer completed
- TOL-200M herbarium ferns and clubmosses to SI\_hydra\_store/public  
/ to /store/public/whiteae/TOL-200M-Ferns/  
transfer completed
- SI\_hydra\_store/public to SI\_NZCBL\_CEC  
/store/public/nzp\_cec/ to /  
transfer completed

The Activity page shows all your transfers. Click the arrow to view details.

Click **Activity** in the left sidebar to see your transfer history. This page shows:

- **Active transfers** — Currently in progress
- **Completed transfers** — Successfully finished
- **Failed transfers** — Transfers that encountered errors

You can filter by status (Active, Succeeded, Failed) and search by collection name.

## Transfer Details

The screenshot shows the Globus Transfer Details interface. On the left is a sidebar with icons for FILE MANAGER, ACTIVITY, COLLECTIONS, GROUPS, FLOWS, COMPUTE, TIMERS, STREAMS, CONSOLE, SETTINGS, LOGOUT, and HELP & SITEMAP. The main area has tabs for Overview (selected) and Event Log. A message at the top says "SI\_hydra\_scratch to SI\_hydra\_store/public transfer completed". The Overview tab displays detailed transfer information:

Task Label	SI_hydra_scratch to SI_hydra_store/public
Source	▶ SI_hydra_scratch
Source Local User	whiteae
Destination	▶ SI_hydra_store/public
Destination Local User	whiteae
Task ID	750f9aa9-dab7-11f0-b03d-0213754b0ca1
Owner	▶ Alexander White (whiteae@si.edu)
Condition	SUCCEEDED
Requested	12/16/2025, 01:43 PM
Completed	12/16/2025, 01:43 PM
Duration	4 seconds
Base Paths	Source /scratch/public/genomics/whiteae/test/ Destination /store/public/ocio_dsl/test/
Transfer Settings	<ul style="list-style-type: none"><li>verify file integrity after transfer</li><li>transfer is not encrypted</li><li>overwriting all files on destination</li></ul>

A summary box on the right lists:

3	Files
0	Directories
3	Files Transferred
30 KB	Bytes Transferred
9.66 KB/s	Effective Speed
n/a	Skipped files on sync
n/a	Skipped files on error

[View debug data](#)

Transfer details show complete information: files transferred, speed, and any errors.

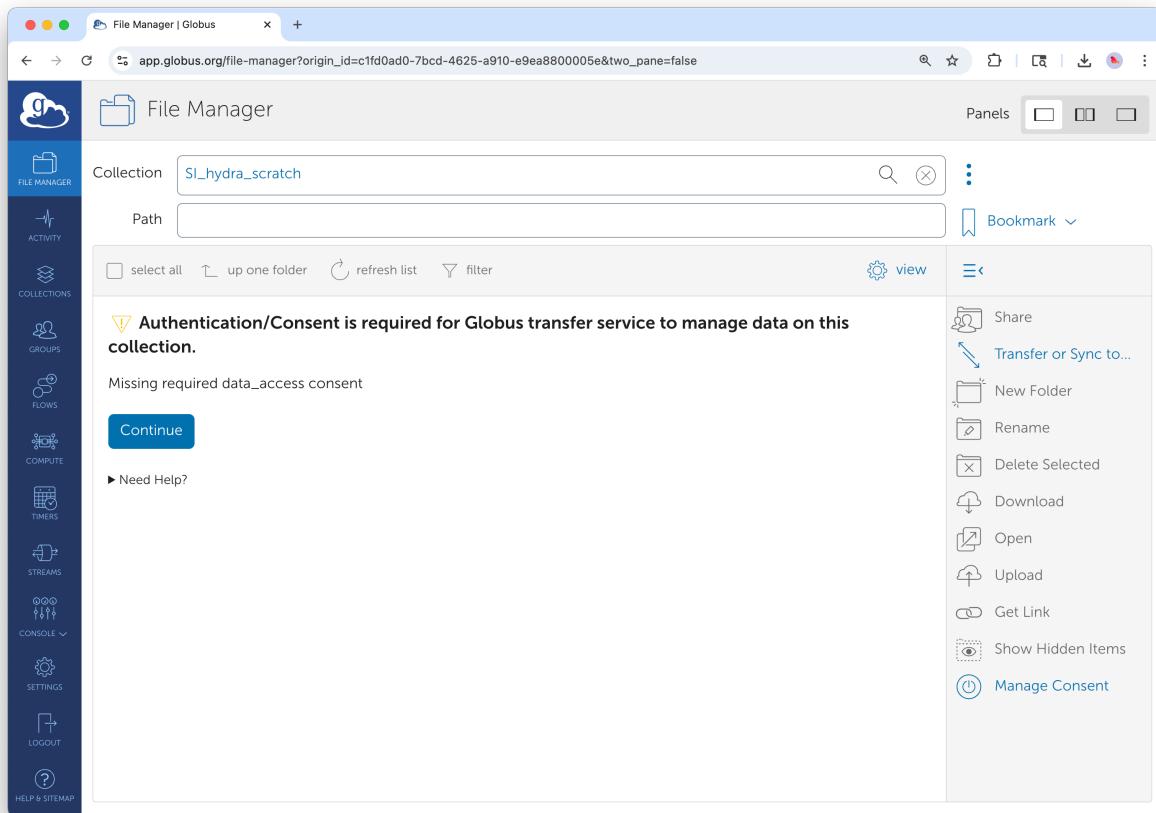
Click on any transfer to see detailed information:

- **Task Label** — Description of the transfer
- **Source and Destination** — The collections involved
- **Condition** — Current status (Active, Succeeded, Failed)
- **Files/Bytes Transferred** — Progress information
- **Duration** — How long the transfer took
- **Effective Speed** — Transfer rate achieved

The **Event Log** tab provides a timeline of transfer events, useful for troubleshooting.

## Handling Authentication Requests

When accessing certain collections, you may see an “Authentication/Consent Required” message.



*Some collections require additional authentication. Click Continue to proceed.*

This happens when: - You're accessing a collection for the first time - Your authentication has expired - The collection requires specific credentials

Click **Continue** and follow the prompts to authenticate. This is normal and ensures secure access to protected data.

## Troubleshooting Common Issues

### Transfer Failures

**Issue: Transfer fails with “Permission Denied”** - **Cause:** You don’t have appropriate access to the source or destination - **Solution:** Verify you have an account on the relevant storage system; contact system administrators if needed - **For Hydra/DAMS:** SI-HPC@si.edu - **For DAMS Staging:** SI-Globus@si.edu - **For STRI:** STRIhelp@si.edu

**Issue: Transfer stalls or times out** - **Cause:** Network interruption or endpoint temporarily unavailable - **Solution:** Globus automatically retries failed transfers. Check the Activity page—if the transfer is still “Active,” it’s working. If it failed, you can restart it.

**Issue: “Endpoint not found” or collection not visible** - **Cause:** The endpoint may be offline, or you may not have access - **Solution:** Try searching for the collection again; contact SI-Globus@si.edu if the issue persists

## **Login Problems**

**Issue: Can't find Smithsonian Institution in the organization list - Solution:** Try searching for "Smithsonian" rather than specific unit names; ensure you're on app.globus.org

**Issue: Login succeeds but collections show “Access Denied” - Cause:** You're authenticated to Globus but don't have an account on the underlying storage system - **Solution:** Request appropriate server access (Hydra, STRI, etc.) from system administrators

## **Performance Issues**

**Issue: Transfers are slow - Cause:** Large number of small files, network congestion, or endpoint load - **Solution:** For many small files, consider compressing into an archive first; schedule large transfers during off-peak hours

## **Next Steps: Building on the Basics**

Now that you understand Globus fundamentals, explore these companion guides based on your needs:

### **Guide 2: Creating Guest Collections for External Collaboration**

**For:** Staff who need to share data with collaborators outside Smithsonian

Learn how to create Guest Collections that give external partners controlled access to specific directories without requiring them to have Smithsonian accounts.

### **Guide 3: Globus Connect Personal**

**For:** Staff with data on laptops, workstations, or external hard drives

Install Globus Connect Personal to create a collection on your personal device, enabling transfers between your laptop and institutional storage. This is essential for users who want to transfer data from workstations or external drives to institutionally managed storage like DAMS, Hydra, or STRI's NAS.

### **Guide 4: Advanced Globus Features**

**For:** Power users ready to automate their data workflows

Explore advanced capabilities: Sync (instead of Transfer) for keeping directories synchronized, scheduled and recurring transfers, and Globus Flows for automated multi-step workflows.

## **Quick Reference**

### **Key URLs**

- **Globus Web App:** [app.globus.org](http://app.globus.org)
- **Globus Documentation:** [docs.globus.org](http://docs.globus.org)

### **Smithsonian Support Contacts**

- **General Globus Support:** SI-Globus@si.edu
- **Hydra HPC Support:** SI-HPC@si.edu
- **DAMS NAS Workflows:** SI-Globus@si.edu

- **STRI Data Transfer Server:** STRIhelp@si.edu

## Glossary

Term	Definition
<b>Collection</b>	A storage location accessible through Globus (formerly called “endpoint”)
<b>Endpoint</b>	The server or system hosting one or more collections
<b>Guest Collection</b>	A shared folder you create to give others access to your data
<b>Mapped Collection</b>	An institutional collection that maps to underlying storage
<b>Transfer</b>	A one-time copy of files from source to destination
<b>Sync</b>	A transfer that only copies new or changed files

---

*For technical support with Globus, contact SI-Globus@si.edu. For questions about specific storage system access, contact the appropriate system administrators listed above.*

# Creating Guest Collections for External Collaboration

## Executive Summary

This guide provides step-by-step instructions for Smithsonian staff to create **Guest Collections** in Globus, enabling secure data sharing with external collaborators who don't have access to Smithsonian storage systems. Guest Collections act as a bridge, using the data owner's credentials to provide controlled access to specific directories.

## Example Use Cases

Many Smithsonian staff use Globus to share large datasets with external collaborators, particularly those requiring access to files that are too large for email or cloud storage:

- A collaborative team at NMNH and STRI uses Globus to transfer pollen images generated on microscopes at NMNH in DC to STRI collaborators in Panama, where they are analyzed and shared with external collaborators who provide taxonomic expertise by annotating the images according to their specialized knowledge of pollen identification.
- The Center for Conservation Genomics at NZCBI uses Globus to share genomic data externally with research partners around the world.

## Globus Access and Account Requirements

**TL;DR: Your Smithsonian account gives you access to Globus. To share data, you also need an account on the server where your data lives. When you create a Guest Collection, it uses YOUR server permissions to let external collaborators access your data - they don't need their own server accounts.**

Access Type	What It Provides	Requirements
<b>Globus Web Application Creating Guest Collections</b>	Interface to manage transfers, view activity, browse collections Ability to share your data with external collaborators	Smithsonian network account Smithsonian network account + user account on the specific server hosting your data
<b>Accessing Shared Guest Collections</b>	Read/write data in collections shared with you	Globus account only (uses collection owner's server permissions)

**Important Note:** While you can access the Globus web application interface with just your Smithsonian network account, you need user accounts on the specific storage systems (Hydra, STRI servers, etc.) to view and work with Collections containing data stored on those systems. Without the underlying storage access, Collections will appear as inaccessible in the Globus interface.

## Prerequisites

### For Data Owners (Sharing Data)

- Active Smithsonian network account
- User account on the server hosting your data:
  - **Hydra:** [Request an HPC account on the SI Service Portal](#)

- **DAMS NAS Workflows:** SI-Globus@si.edu
- **STRI users:** Contact STRIhelp@si.edu to request Globus Data Transfer Node Server access
- Data organized in a shareable directory structure
- Evaluate your data's sensitivity (note: PII should never be transferred using Globus)

### For Data Recipients (Accessing Shared Data)

- Globus account (free institutional or personal account)
- Valid email address for sharing notifications

**Note:** Recipients do not need accounts on the underlying storage systems. The Guest Collection uses the data owner's permissions to provide limited access.

### Where Is Your Data Right Now?

Before creating a Guest Collection, your data must be on Smithsonian institutional storage, not on your personal device.

If your data is...	What to do
<b>Already on Smithsonian storage</b> (Hydra, STRI servers, DAMS NAS)	You're ready to proceed with this guide
<b>On your laptop, workstation, or external hard drive</b>	First transfer it to institutional storage using <b>Guide 3: Globus Connect Personal</b> , then return here

**Why this matters:** Smithsonian Data Transfer Nodes are optimized for high-bandwidth transfers. Sharing directly from a personal device would tie up your machine for the duration of every collaborator's download and severely limit transfer speeds. By hosting data on institutional storage, your collaborators get fast downloads and you can close your laptop.

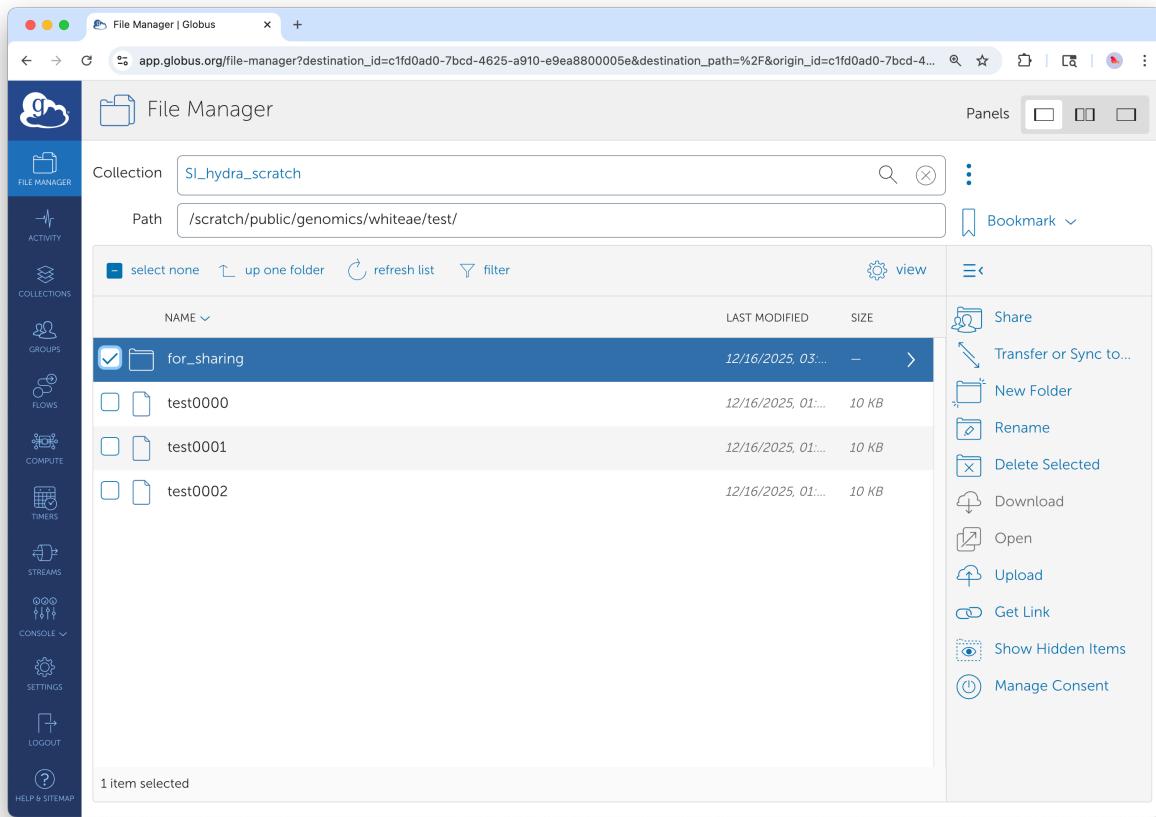
### Network Requirements

Once data is deposited in a Smithsonian-managed collection, all Globus actions can be managed from the web application ([app.globus.org](http://app.globus.org)) without being required to be on the Smithsonian network or VPN.

### Creating Your Guest Collection

Think of a Guest Collection as creating a secure "window" into your data. You maintain full control over what's shared and who can access it, while your collaborators can use their existing Globus accounts to access, download, or even upload files to your shared space.

## Step 1: Locate Your Data

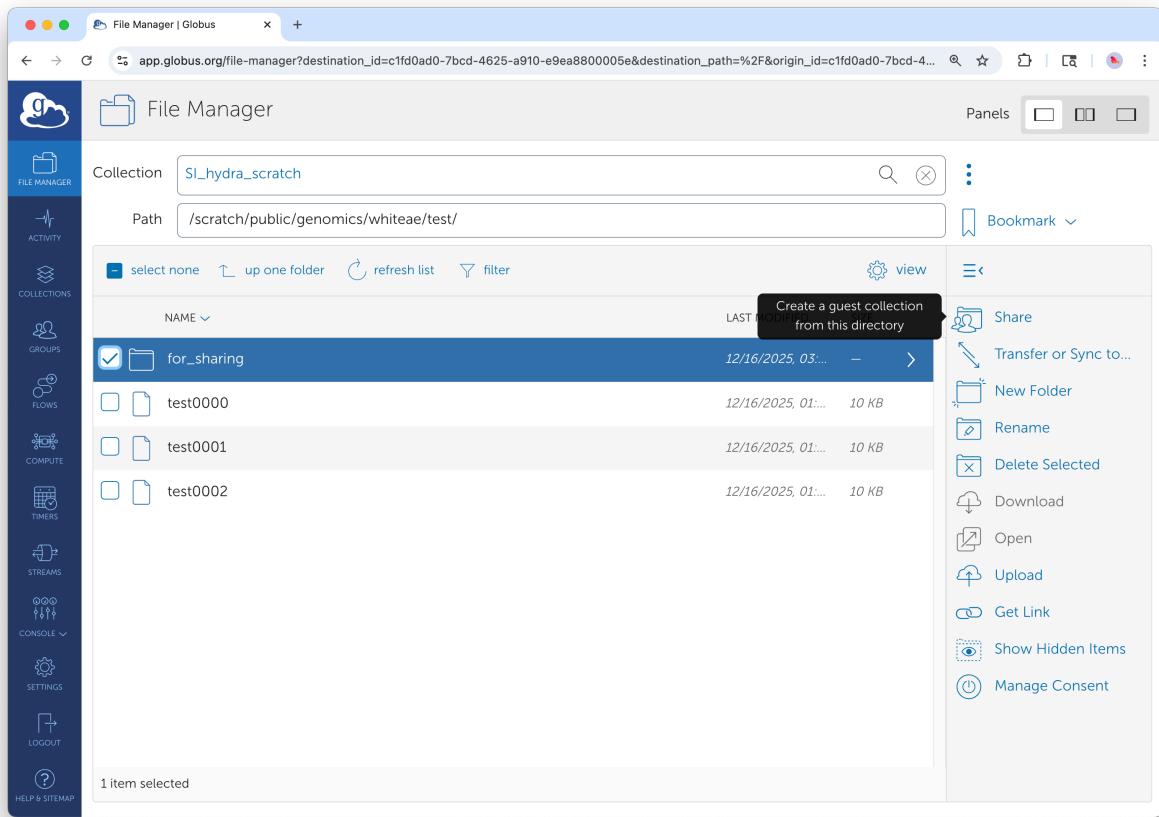


Access your *SI* collection through the File Manager interface. Note the “*for\_sharing*” folder selection.

Log into the Globus web interface and navigate to the File Manager. Search for the collection where your data are housed - for example, “*SI\_hydra\_scratch*” shown in the interface. Browse to find the directory containing the data you want to share.

The key decision here is choosing the right directory level. You can share an entire project folder or drill down to specific subdirectories. Consider what your collaborators actually need access to - sharing at too high a level might expose data they don't need to see, while sharing too narrowly might require creating multiple Guest Collections later.

## Step 2: Initiate Guest Collection Creation



Select the folder you want to share (blue highlighting shows selection) and click the Share button from the action menu on the right.

Once you've located the directory to share, select it using the checkbox. You'll see it highlighted in blue. Click the "Share" button from the action menu on the right side of the interface.

From the sharing options, select "Add Guest Collection." This creates a new access point that external users can connect to, separate from your main Smithsonian collection. Think of this as creating a dedicated "front door" for your collaborators that only opens to the specific data you want to share.

Guest Collections on SI\_hydra\_scratch

FILE MANAGER

ACTIVITY

COLLECTIONS

GROUPS

FLOWS

COMPUTE

TIMERS

STREAMS

CONSOLE

SETTINGS

LOGOUT

?

SI\_hydra\_scratch

Overview Roles Collections Credentials

Root Path: /

Add Guest Collection

No collections match that filter.

DISPLAY NAME CREATED LAST ACCESS

https://app.globus.org/file-manager/collections/c1fd0ad0-7bcd-4625-a910-e9ea8800005e/shares/create?host\_path=%2Fscratch%2Fpublic%2Fgenomics%2Fwhiteae%2Ftest%2Ffor\_sharing%2F

*Alternative method: From the collection overview, click 'Add Guest Collection' to begin sharing process.*

### Step 3: Configure Collection Identity

You are creating a guest collection on **SI\_hydra\_scratch** to share data

User Credential: [Alexander White \(whiteae\)](#)

Globus Identity: Alexander White (whiteae@si.edu)

Directory: /scratch/public/genomics/whiteae/test/for\_sharii  [Browse](#) [?](#)

Display Name:

Description:

Keywords:

Preserve ModTime: Not set - users can choose to preserve file modification times. [▼](#)

Activity Notification Policy: Use:  Source  Destination  Either  None (Default)

Force encryption on transfers to and from this collection

[view more fields ▾](#)

[Create Collection](#) [Cancel](#)

Configure your guest collection with descriptive information and security settings.

When creating the Guest Collection, you'll configure several important settings:

**Display Name:** Choose something descriptive. “DemoGuest” works for testing, but for real collaborations, use names like “PollenAnalysisProject2025” or “GenomicsData\_SmithLab.” Your collaborators will see this name, and it helps them understand what they’re accessing.

**Description:** Explain what the data contains, any important usage guidelines, and perhaps how it relates to your collaboration. For example: “High-resolution pollen microscopy images from Costa Rica field sites. Please download images, add taxonomic annotations to the provided spreadsheet, and return completed annotations via email.”

**Directory Path:** This confirms exactly which directory you’re sharing. Double-check this carefully - you can restrict access further later, but you can’t expand beyond what you select here without creating a new Guest Collection.

**Keywords:** Add searchable terms that help with discoverability if you’re managing multiple Guest Collections. Terms like your project name, institution names, or data types work well.

Click “Create Collection” to establish the new endpoint.

## Step 4: Add Your Collaborators

Shared With

Mapped Collection: SI\_hydra\_scratch  
Base Path: /scratch/public/genomics/whiteae/test/for\_sharing/

USER OR GROUP	CREATED	EXPIRATION	READ	WRITE
hydra-globus02.si.edu (76f9ff3b-d495-4416-abd7-eb11a434815e@clients.auth.globus.org)	-	never expires	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Alexander White (whiteae@si.edu)	-	never expires	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

Add Permissions — Share With

*Access permission management from your newly created guest collection using the “Add Permissions — Share With” button.*

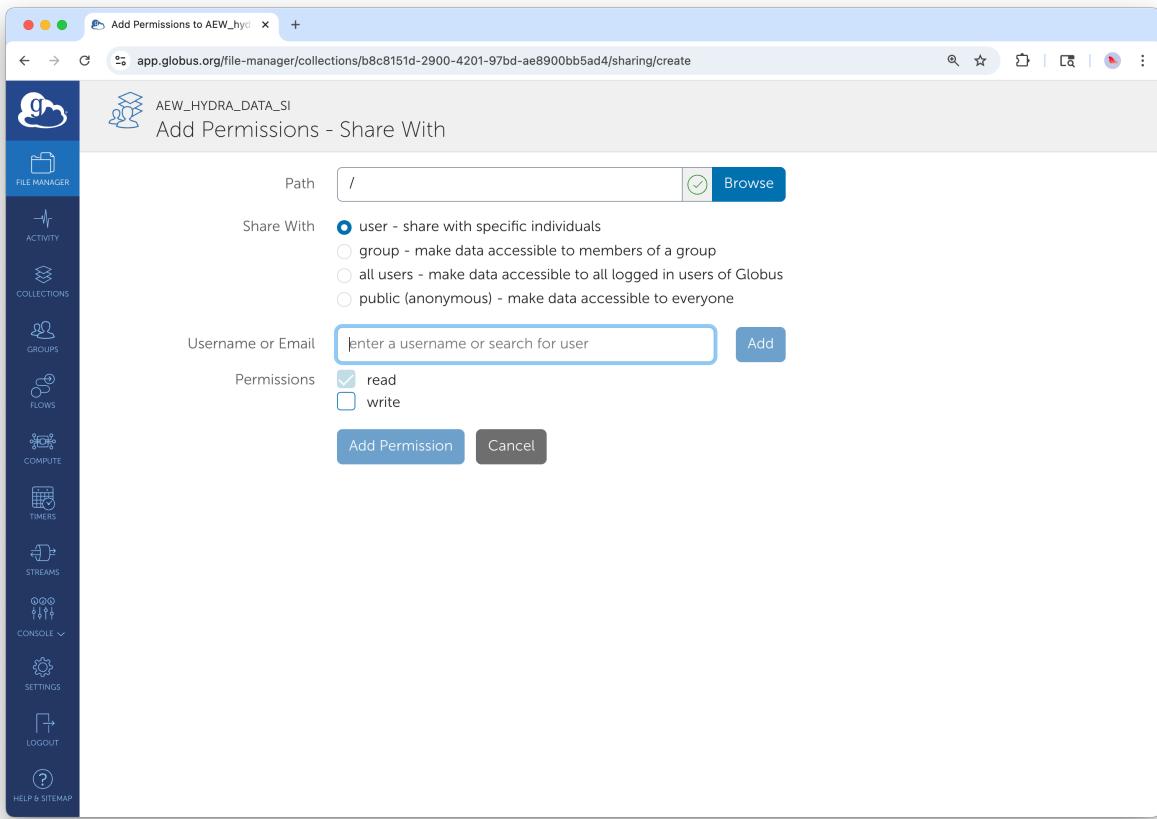
Now comes the crucial step of defining who can access your data and what they can do with it. Click “Add Permissions — Share With” to begin configuring access.

### Choosing Your Sharing Scope:

Most research collaborations use “User” sharing, where you specify individuals by their email addresses or institutional identifiers. This gives you precise control over who has access.

“Group” sharing works well if your collaborator’s institution has pre-defined research groups in Globus, but this is less common for external collaborations.

Do not select “All Users” or “Public”.



Configure basic user permissions for root directory access. Note the “user” button selection, Email Notification checkbox, and “Add Permission” button.

### Configuring User Permissions:

For each collaborator, you can specify several settings:

**Path Access:** The default is “/” which gives access to everything in your Guest Collection. However, you can restrict access to specific subdirectories. For example, if your Guest Collection contains both raw data and preliminary analysis, you might give collaborators access only to “/rawdata/” initially, then later grant access to “/analysis/” as the project progresses.

**Read vs. Write Permissions:** - **Read:** Allows downloading and viewing files. This is appropriate when you’re sharing datasets for analysis. - **Write:** Allows uploading, modifying, and deleting files. DO NOT ENABLE WRITE ACCESS.

**Email Notifications:** Enable this and customize the message if desired. Your collaborators will receive an email with instructions on how to access the shared data.

### Advanced Access Control:

For enhanced customization, you can limit access to specific subdirectories by changing the Path field from / to /subdir1/ or other specific paths. This is particularly useful when your Guest Collection contains multiple types of data but you only want to share specific portions with certain collaborators.

## Managing Ongoing Collaborations

### Monitoring Collection Usage

The Activity section shows you what's happening with your shared data - who's downloading what, when transfers occur, and if there are any access problems. This is valuable for understanding how actively your collaboration is proceeding and for troubleshooting if collaborators report problems.

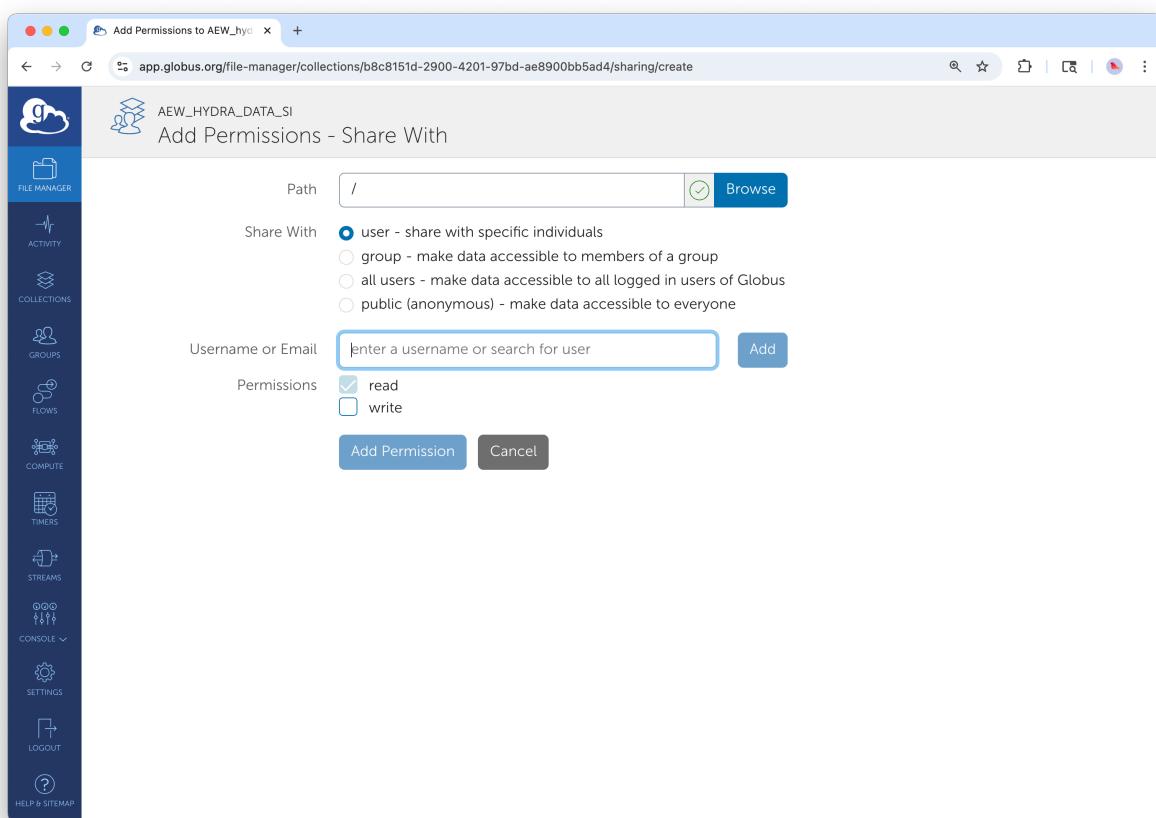
### Lifecycle Management

Research collaborations have natural endpoints. When a project concludes, or when certain collaborators no longer need access, remove their permissions through the Permissions tab. This isn't just good security practice - it also helps keep your Guest Collection organized and ensures you're only supporting active collaborations.

## Troubleshooting Sharing Issues

### Access Problems

#### Issue: “Access Denied” when accessing shared collection



Verify that read permissions are enabled (checkboxes) and that the collaborator has been properly notified.

- **Cause:** Permission not properly configured in Guest Collection, or data owner lacks sufficient permissions on underlying storage
- **Solution:** Verify permission settings in Guest Collection; ensure data owner has appropriate access to shared directories
- **Prevention:** Test Guest Collection access with known collaborators; verify data owner permissions before sharing

**Issue: Collection not visible to collaborator - Cause:** Permission not properly configured or email notification not received - **Solution:** Re-send invitation through permission management; verify user identity - **Prevention:** Always enable email notifications; test sharing with known users first

## Collaboration Best Practices

### Data Organization

- **Share Only What's Necessary:** Limit access to specific directories required for collaboration rather than providing broad access
- **Clear Directory Structure:** Use intuitive folder hierarchies like shown in the file listings
- **Access Boundaries:** Use path restrictions to prevent unnecessary directory access

### Security Considerations

- **Principle of Least Privilege:** Grant minimum required access
- **Review Collaborator Access:** Remove permissions when no longer needed
- **Data Sensitivity:** Do not share personally identifiable information (PII) or any export controlled data with Globus

## Using Globus for Internal Smithsonian Collaboration

While this guide focuses on external collaboration, these same tools and techniques can be used effectively for sharing data between Smithsonian departments, research units, centers, and institutes. Internal collaborations may have simplified account management since all parties already have Smithsonian network access, but the same principles of organized sharing, clear permissions, and proper access management apply.

## For Your Collaborators: Accessing Shared Data

*Share this section with external collaborators who receive your Guest Collection invitation.*

---

When a Smithsonian staff member shares data with you through Globus, you'll receive an email notification with a link to access the shared files. Here's how to get started:

### Step 1: Create a Globus Account (If You Don't Have One)

If you've never used Globus before:

1. Go to [app.globus.org](https://app.globus.org)
2. Look for your institution in the organization dropdown—many universities and research institutions are already connected
3. If your institution isn't listed, click "Sign in with Google" or "Sign in with ORCID iD" to create a free account
4. Complete the brief registration process

If you already have a Globus account through your institution or a previous collaboration, simply log in with those credentials.

## **Step 2: Access the Shared Collection**

1. Click the link in the email invitation you received, OR
2. Log into [app.globus.org](https://app.globus.org), go to **File Manager**, and search for the collection name provided by your collaborator

You should see the shared files and folders. If you see “Access Denied,” contact the person who shared the data—they may need to verify your permissions.

## **Step 3: Download the Data**

### **Option A: Download to your computer (small files)**

For files under a few gigabytes, you can download directly through your browser: 1. Select the files you want 2. Click **Download** in the right-hand menu

### **Option B: Transfer to your institution’s storage (large files)**

For larger datasets, transfer server-to-server for faster, more reliable results: 1. In File Manager, enable the two-panel view (click **Panels** in the top right) 2. The shared collection should be in one panel 3. In the other panel, search for your institution’s Globus collection 4. Select files in the shared collection, then click **Start** to transfer

### **Option C: Transfer to your personal computer**

If you need large files on your laptop or workstation: 1. Install **Globus Connect Personal** from [globus.org/globus-connect-personal](https://globus.org/globus-connect-personal) 2. This creates a collection on your personal device 3. Transfer from the shared collection to your personal collection

## **Need Help?**

If you have trouble accessing shared data, contact the Smithsonian staff member who sent the invitation. They can verify your permissions and resend the invitation if needed.

For general Globus questions, visit [docs.globus.org](https://docs.globus.org) or contact your institution’s research computing support.

---

*End of external collaborator instructions section.*

## **Support and Additional Resources**

### **Smithsonian Support Contacts**

- **General Globus Support:** SI-Globus@si.edu
- **Hydra HPC Support:** SI-HPC@si.edu
- **DAMS NAS Workflows:** SI-Globus@si.edu
- **STRI Data Transfer Server:** STRIhelp@si.edu

**Globus Documentation:** Comprehensive platform [guides and tutorials](https://docs.globus.org)

---

*For technical support with Globus collections and data access, contact SI-Globus@si.edu. For questions about account creation or specific systems access, contact the appropriate system administrators listed above.*

# Globus Connect Personal

## Executive Summary

This guide shows Smithsonian staff how to install and configure **Globus Connect Personal**, a lightweight application that turns your laptop, workstation, or external hard drive into a Globus collection. Once set up, you can transfer files directly between your personal device and Smithsonian storage systems. This workflow is ideal for uploading fieldwork data, migrating archival collections, syncing project outputs, or backing up local files to institutional storage.

## When to Use Globus Connect Personal

Globus Connect Personal bridges the gap between personal devices and institutional infrastructure. Common scenarios include:

- **Returning from fieldwork** with terabytes of images, audio recordings, or sensor data on an external drive
- **Migrating archival collections** from external hard drives to DAMS or other institutional storage
- **Working on a laptop** and needing to move outputs to Hydra for further processing
- **Downloading shared datasets** from collaborators directly to your workstation

Unlike browser-based uploads (which are limited and unreliable for large files), Globus Connect Personal uses the same robust transfer engine as server-to-server transfers, with automatic retry, integrity verification, and no size limits.

## Prerequisites

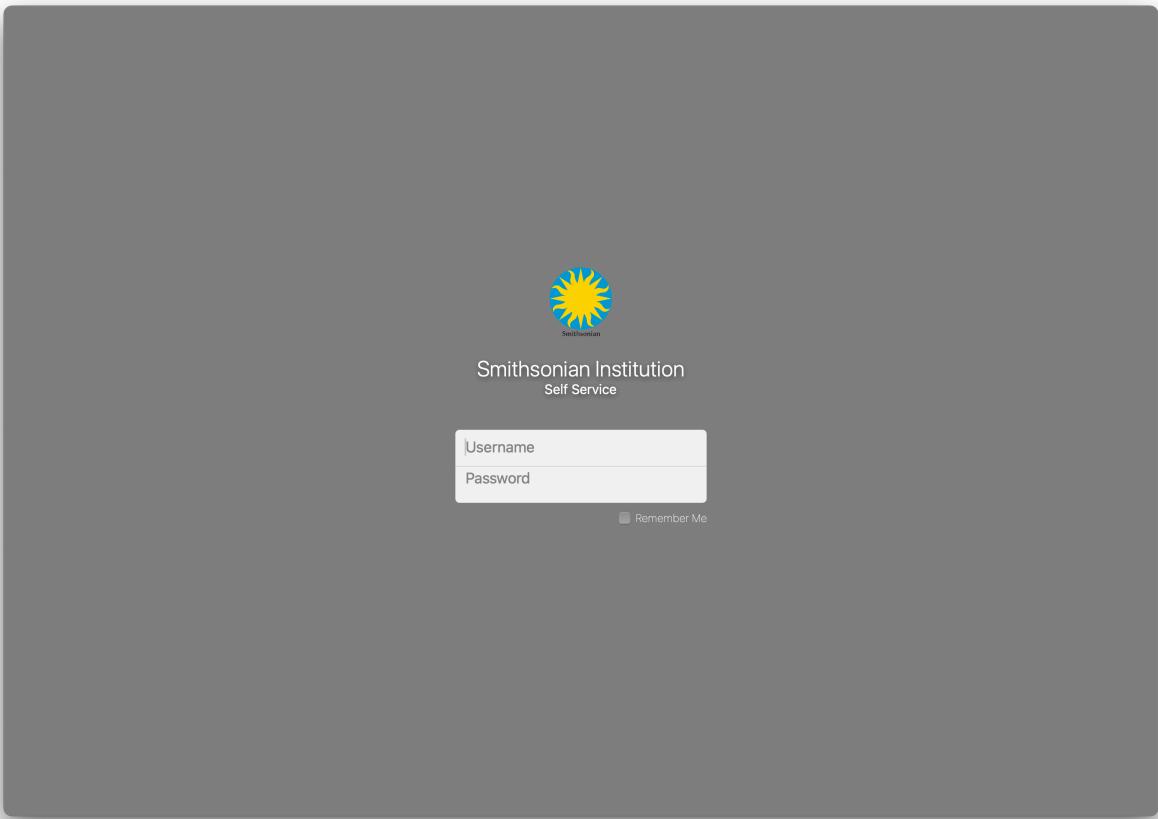
### What You Need

Requirement	Details
<b>Globus account</b>	Log in at least once via <a href="https://app.globus.org">app.globus.org</a> with your Smithsonian credentials
<b>Personal device</b>	Windows, macOS, or Linux computer where you want to install Globus Connect Personal
<b>Storage to connect</b>	Internal drive, external hard drive, or network-attached storage

## Installing Globus Connect Personal on a Smithsonian Managed Computer

### Step 1: Navigate to the Self Service Application

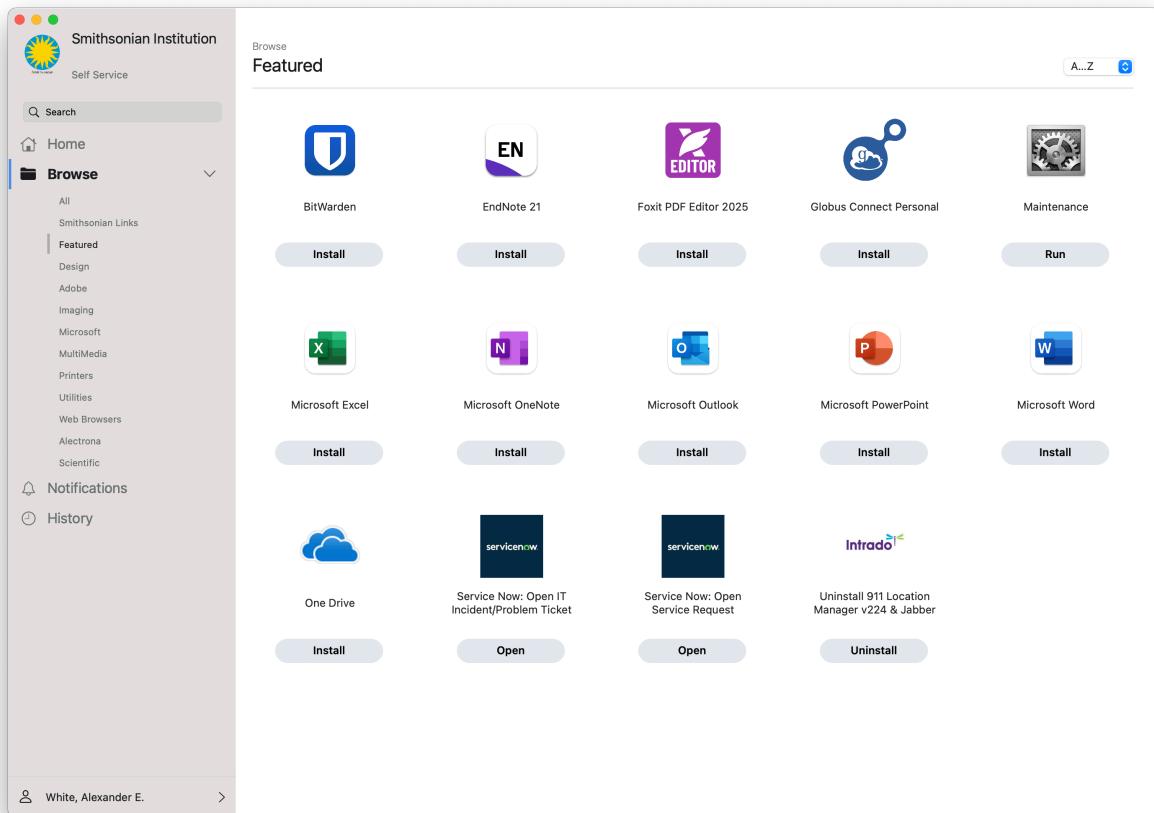
Open the “Self Service” or “Software Center” application on your machine. This application comes pre-installed on Smithsonian managed computers.



*The Smithsonian “Self Service” login. Authenticate with your SI credentials.*

### **Step 2: Run the Installer**

In the main self service page you will likely see the Globus Connect Personal application. Search for “globus” if it is not immediately available. Select the application to install it on your machine.



*Globus Connect Personal is available via the self service software center on both Windows and macOS.*

## **Installing Globus Connect Personal on a Personal Device**

### **Step 1: Download the Installer**

Go to [globus.org/globus-connect-personal](https://globus.org/globus-connect-personal) and download the installer for your operating system.

The screenshot shows a web browser displaying the Globus Connect Personal download page at [globus.org/globus-connect-personal](https://globus.org/globus-connect-personal). The page has a blue header with the Globus logo and navigation links for Solutions, Resources, Pricing, Newsroom, Developers, and About. On the right side of the header are 'GET STARTED' and 'LOG IN' buttons. The main content area is titled 'Install Globus Connect Personal' and includes a sub-instruction: 'Create a Globus collection on your laptop. Globus Connect Personal is available for all major operating systems.' Below this, there are three download options:

- Globus Connect Personal for Mac**: Available for Mac OS X 10.13 or higher. Includes an 'INSTALL NOW' button.
- Globus Connect Personal for Windows**: Currently supported Windows versions. Includes an 'INSTALL NOW' button.
- Globus Connect Personal for Linux**: For common x86 distributions. Includes an 'INSTALL NOW' button.

At the bottom of the page, there is a link to 'frequently asked questions' and an email address 'support@globus.org'. A 'Related Content' section follows, featuring two cards: 'PAGE' (Globus Connect Server) and 'USER STORY' (Simplifying Research with Globus for Box).

*The Globus Connect Personal download page. Select your operating system.*

## **Step 2: Run the Installer**

**On Windows:** 1. Run the downloaded .exe file 2. Follow the installation wizard prompts 3. The application will launch automatically when complete

**On macOS:** 1. Open the downloaded .dmg file 2. Drag Globus Connect Personal to your Applications folder 3. Launch from Applications (you may need to approve it in Security & Privacy settings)

**On Linux:** 1. Extract the downloaded .tgz archive 2. Run the `globusconnectpersonal` script 3. Follow terminal prompts to complete setup

## **Step 3: Log In and Create Your Collection**

Launch the application. When Globus Connect Personal launches, it will open a browser window for authentication.



## Globus Connect Personal Setup



Log In

Advanced Options ▾

*Globus Connect Personal prompts you to log in and authorize the application.*

1. Log in with your Smithsonian credentials (same as the web interface)
2. Grant the requested permissions
3. You'll be prompted to name your collection

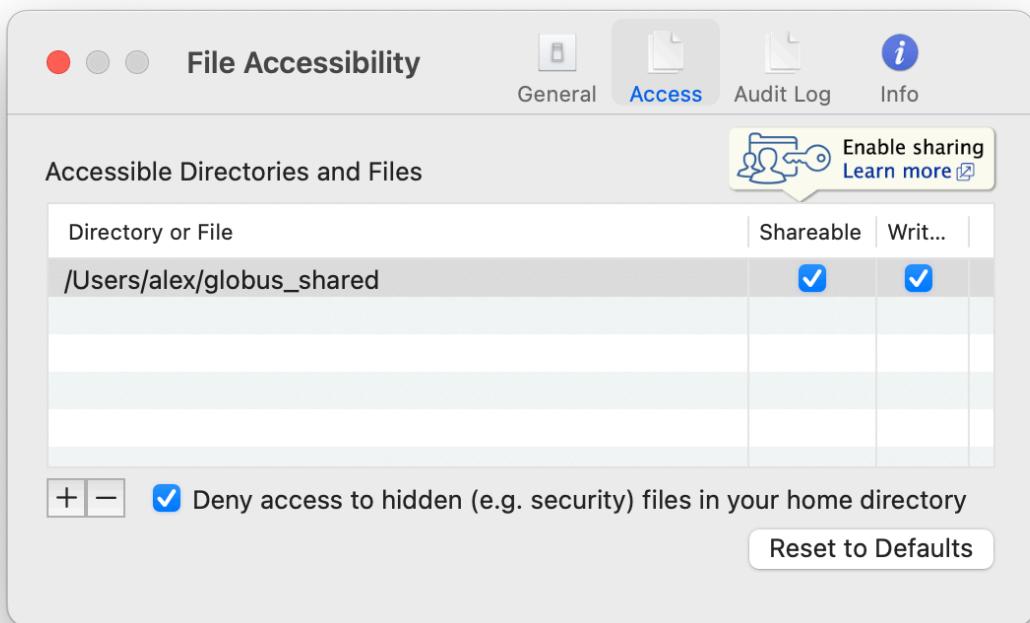
### **Choosing a Collection Name:**

Pick something descriptive that you'll recognize later. Good examples: - "JSmith-MacBook-Pro" - "FieldLaptop-2025" - "Lab-Workstation-NMNH"

Avoid generic names like "My Computer"—if you set up multiple devices, you'll want to distinguish them.

### **Step 4: Configure Accessible Folders**

By default, Globus Connect Personal only exposes your home directory (see security note below). You'll want to instead add other locations, especially external drives.



The Preferences panel lets you control which folders are accessible via Globus.

**To add folders:**

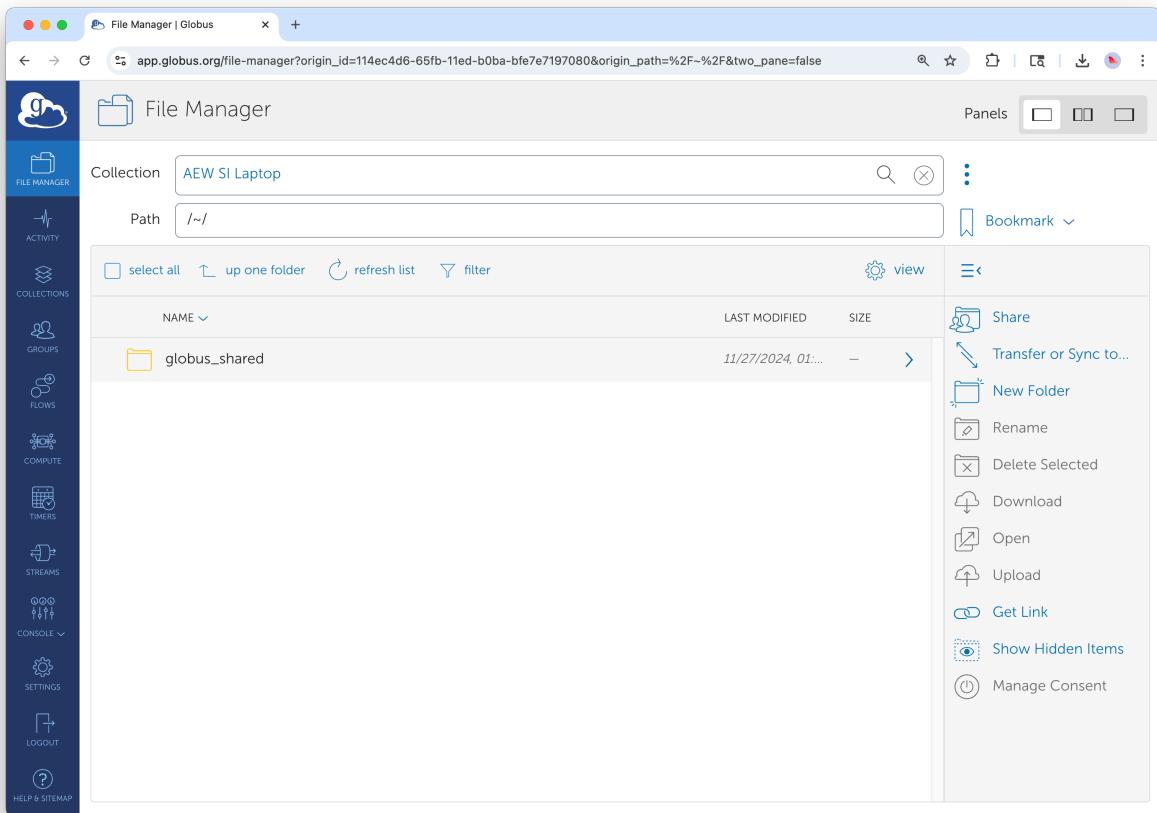
1. Open Globus Connect Personal preferences (right-click the system tray/menu bar icon)
2. Go to the **Access** tab
3. Click the + button to add folders
4. Navigate to the folder or drive you want to share
5. Choose whether to allow **Writable** access (allows uploads to this location)

**Common folders to add:** - External hard drives (e.g., /Volumes/FieldData on Mac, E:\ on Windows) - Project folders outside your home directory - Network drives mounted on your computer

**Security note:** Only add folders you actually need to transfer. The principle of “least access required” applies. To remove a folder, select it and click the - button. It is recommended to remove your home directory and only add specific folders you need.

**Step 5: Verify Your Collection is Online**

Once configured, your collection should appear in the Globus web interface.



Your personal collection now appears in the File Manager, ready for transfers.

1. Go to [app.globus.org](https://app.globus.org)
2. In the File Manager, search for your collection name
3. You should see your configured folders

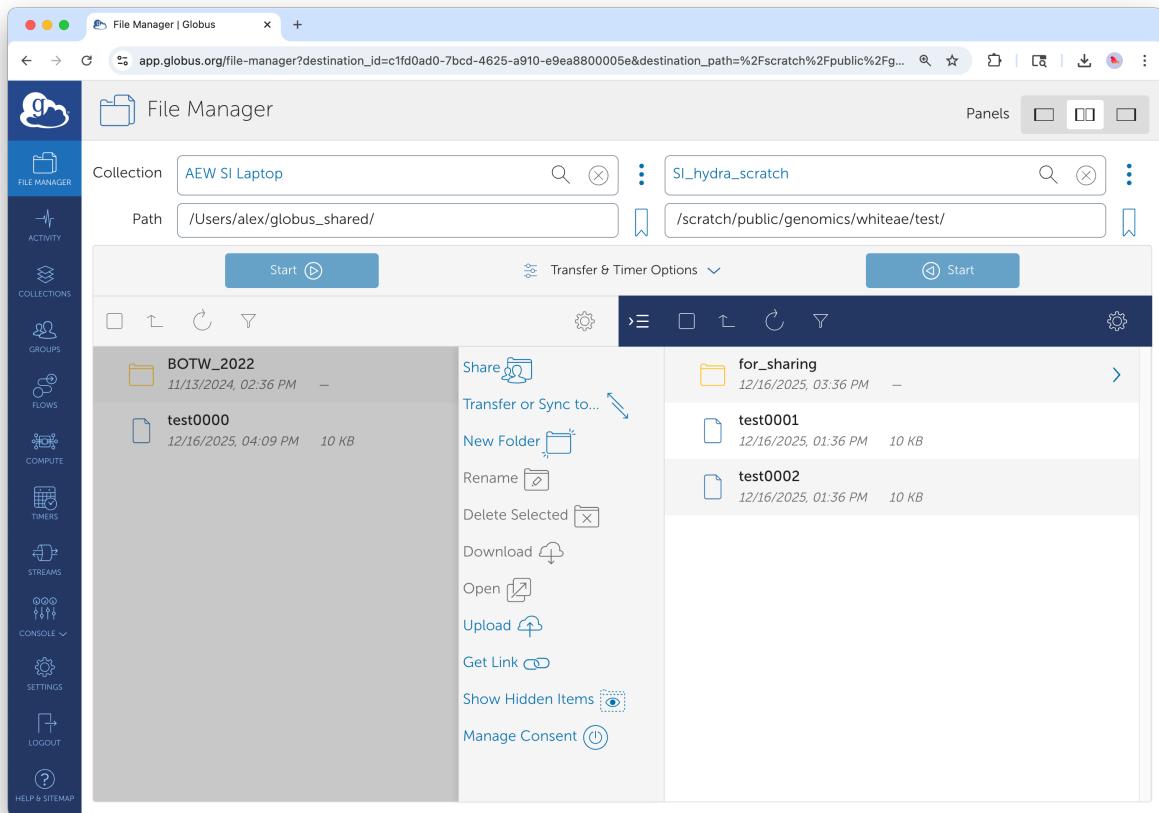
If your collection doesn't appear, check that Globus Connect Personal is running (look for the icon in your system tray or menu bar, the same location where you might find your wifi connection).

## Transferring Data To and From Your Device

With Globus Connect Personal running, transfers work exactly like server-to-server transfers covered in Guide 1.

### Use Case: Uploading Data to Storage at the Smithsonian Data Center

A common workflow: you've returned from the field with an external drive full of data and need to upload it to Hydra for processing. Another scenario: you have project files on an external drive that need to be uploaded to DAMS Staging.



*Two-panel view: personal collection on the left, in this case, Hydra on the right.*

1. Open the two-panel view in File Manager
2. **Left panel:** Select your personal collection, navigate to the external drive
3. **Right panel:** Select the Smithsonian managed collection, navigate to your destination directory
4. Select files/folders on the left, click **Start**

The transfer will proceed even if you close the browser—but **keep Globus Connect Personal running** until the transfer completes.

## Downloading Shared Data to Your Laptop

If a collaborator has shared a Guest Collection with you, you can download directly to your laptop:

1. **Left panel:** Navigate to the shared Guest Collection
2. **Right panel:** Select your personal collection
3. Select files on the left, click **Start**

This is much more reliable than browser downloads for large datasets.

## Managing Globus Connect Personal

### Understanding the System Tray Icon

Globus Connect Personal runs in the background and shows a small icon in your system tray (Windows) or menu bar (macOS).

Icon State	Meaning
<b>Solid/connected</b>	Your collection is online and accessible
<b>Greyed out/paused</b>	Transfers are paused; collection is offline
<b>Spinning/active</b>	A transfer is currently in progress

Right-click the icon for options: Pause, Preferences, Web Dashboard, Quit.

### Pausing and Resuming

You can pause Globus Connect Personal when you don't need it—for example, when on a slow connection or to preserve battery life.

**To pause:** Right-click the icon and select **Pause**

**To resume:** Right-click and select **Unpause** or **Resume**

While paused, your collection is offline and transfers cannot proceed. Any pending transfers will resume automatically when you unpause.

### Updating the Application

Globus Connect Personal occasionally requires updates. You'll see a notification when an update is available. It's good practice to keep the application current for security and compatibility.

## Working with External Drives

### Connecting a New Drive

When you connect an external drive for the first time:

1. Mount the drive normally on your computer
2. Open Globus Connect Personal preferences
3. Add the drive's mount point to your accessible folders
4. The drive will now appear as a subfolder in your collection

**On macOS:** External drives typically mount at /Volumes/DriveName

**On Windows:** External drives appear as drive letters like D:\ or E:\

### Drive Naming Considerations

If you regularly swap between multiple external drives, the folder structure in Globus will change based on which drive is connected. Consider:

- Using consistent drive names when formatting
- Creating a dedicated “incoming data” folder structure on each drive
- Documenting your drive naming conventions for your colleagues or lab group

## Safely Disconnecting

Before physically disconnecting an external drive:

1. Ensure no transfers are in progress (check the Activity page)
2. Pause Globus Connect Personal (optional but recommended)
3. Eject the drive using your operating system's normal process

## Troubleshooting

### Connection Issues

**Issue: Collection shows as “offline” in the web interface - Cause:** Globus Connect Personal isn't running, or is paused - **Solution:** Check for the application icon; restart if needed; verify you're connected to the internet

**Issue: Collection is online but folders are empty - Cause:** Folder permissions not configured in Globus Connect Personal - **Solution:** Open Preferences → Access tab → verify the folders you need are listed

**Issue: “Permission denied” when transferring to your device - Cause:** The destination folder isn't marked as Writable in preferences - **Solution:** Edit the folder in Access preferences and enable Writable

### Transfer Issues

**Issue: Transfer fails when laptop goes to sleep - Cause:** Globus Connect Personal stops when the computer sleeps - **Solution:** Adjust your power settings to prevent sleep during transfers, or use a desktop workstation for large transfers

**Issue: Transfer is very slow - Cause:** Home internet connection, WiFi limitations, or VPN overhead - **Solution:** Use a wired connection if possible; avoid VPN for Globus transfers (it's not required); consider transferring large datasets from your office rather than home

**Issue: External drive not visible in Globus - Cause:** Drive not added to accessible folders, or not mounted - **Solution:** Verify the drive is mounted and visible in your file explorer; add the mount point in Globus Connect Personal preferences

### Authentication Issues

**Issue: “Login required” error when collection was working before - Cause:** Authentication credentials have expired - **Solution:** Open the Globus Connect Personal menu and select “Log In” or visit the Web Dashboard to re-authenticate

## Best Practices for Large Transfer Projects

Whether you're uploading fieldwork data, migrating archival collections, or consolidating project files, these practices will help ensure successful transfers.

### Before You Start

- Install and test Globus Connect Personal on your laptop or workstation
- Verify you can successfully transfer a small test file to your destination
- Add your external drive mount points in advance
- Confirm you have the necessary account access on destination systems

## **During Transfer**

- Connect to a reliable, fast network before starting large transfers
- Keep your computer plugged in and prevent sleep
- Start the transfer and monitor initial progress before walking away
- Use the Activity page to verify completion

## **Next Steps**

Now that you can transfer data to and from your personal devices, explore these related guides:

**Guide 2: Creating Guest Collections** — Share data with external collaborators by creating Guest Collections on Smithsonian storage systems.

**Guide 4: Advanced Features** — Set up recurring transfers to automatically sync data between your laptop and institutional storage.

## **Quick Reference**

### **Key URLs**

- **Download:** [globus.org/globus-connect-personal](http://globus.org/globus-connect-personal)
- **Web Interface:** [app.globus.org](http://app.globus.org)
- **Documentation:** [docs.globus.org/how-to/globus-connect-personal](http://docs.globus.org/how-to/globus-connect-personal)

### **Smithsonian Support Contacts**

- **General Globus Support:** SI-Globus@si.edu
- **Hydra HPC Support:** SI-HPC@si.edu
- **DAMS NAS Workflows:** SI-Globus@si.edu
- **STRI Data Transfer Server:** STRIhelp@si.edu

---

*For technical support with Globus Connect Personal, contact SI-Globus@si.edu. For issues with destination storage systems, contact the appropriate system administrators listed above.*

# Advanced Globus Features

## Executive Summary

This guide covers Globus power-user features that automate and optimize data workflows. You'll learn when to use **Sync** instead of Transfer, how to schedule **recurring transfers**, and how to build **Globus Flows** for multi-step automated workflows. These capabilities transform Globus from a manual file-copying tool into an automated data pipeline.

## Who This Guide Is For

This guide assumes you're comfortable with basic Globus operations (covered in Guide 1) and are ready to:

- Automate repetitive transfer tasks
- Keep directories synchronized across systems
- Build workflows that run without manual intervention
- Optimize transfer performance for your specific needs

## Sync vs. Transfer: Choosing the Right Tool

### Understanding the Difference

Both Transfer and Sync copy files from source to destination, but they handle existing files differently.

Operation	What It Does	Best For
<b>Transfer</b>	Copies all selected files, overwriting any existing files at the destination	One-time data moves; initial uploads; replacing outdated datasets
<b>Sync</b>	Copies only new or modified files; leaves unchanged files alone	Ongoing synchronization; incremental backups; collaborative folders

Think of it this way: **Transfer** is “copy everything I selected.” **Sync** is “make the destination match the source.”

### When to Use Transfer

Use standard Transfer when:

- Moving data for the first time to a new location
- Replacing an entire dataset with a new version
- You want to overwrite everything regardless of what's already there
- You're moving data to an empty destination

### When to Use Sync

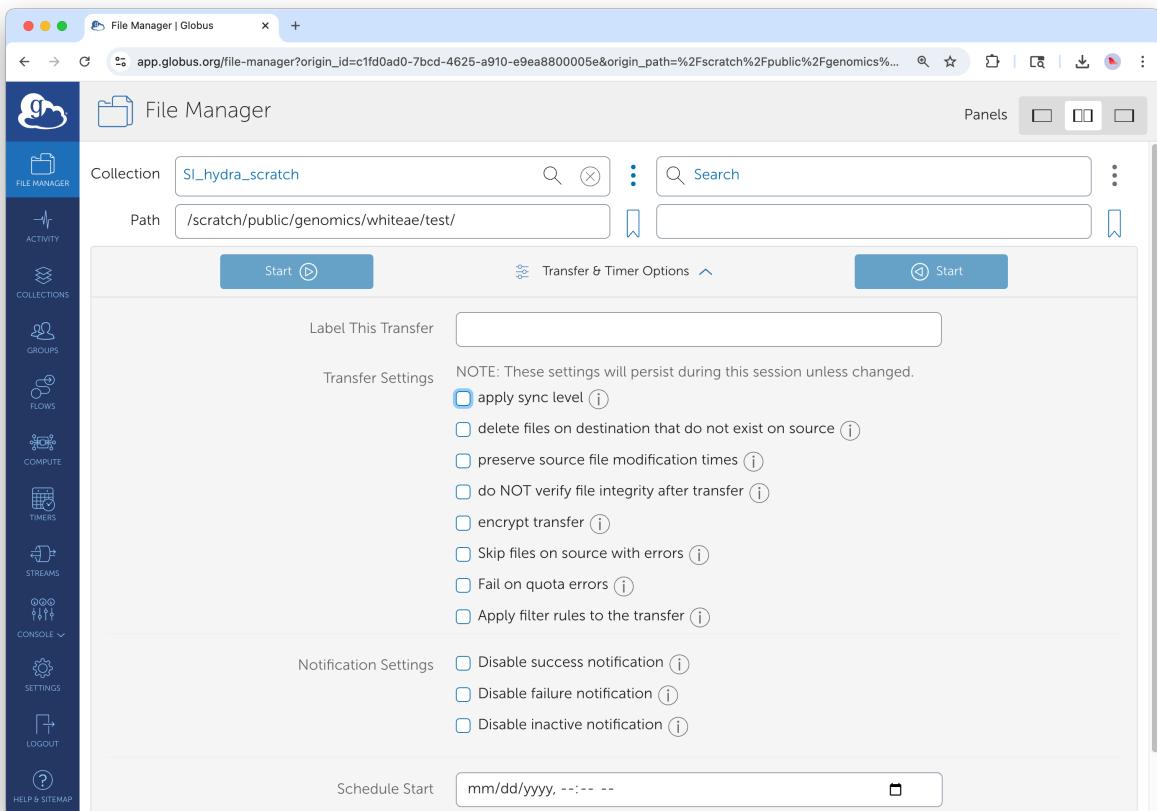
Use Sync when:

- You're updating a destination that already has some of the files
- Running regular backups where most files haven't changed
- Collaborating on a shared folder that multiple people update

- Resuming an interrupted transfer (Sync will skip already-completed files)

**Example scenario:** You maintain a project folder on your laptop and periodically sync it to Hydra. Each week, only a few files change. Using Sync means only those changed files transfer—saving time and bandwidth.

## How to Initiate a Sync



*The Transfer & Timer Options panel. Select “sync” to enable synchronization mode.*

1. Set up your source and destination as you would for a normal transfer
2. Click **Transfer & Timer Options** (between the two panels)
3. Under **Transfer Options**, select **apply sync level**
4. Choose your desired sync behavior (see next section)
5. Click **Start**

## Sync Behavior Options

When you enable Sync, you can fine-tune how Globus detects changes:

Option	Behavior
<b>Modification time</b>	Compare timestamps; transfer if source is newer (default)
<b>Checksum</b>	Calculate file hashes; transfer if content differs (slower but more accurate)

Option	Behavior
<b>File size</b>	Compare sizes; transfer if different (fast but less reliable)

For most workflows, the default (modification time) works well. Use checksum verification for critical data where you need to guarantee integrity.

### Sync Limitations

Sync has an important limitation: **it only adds and updates files at the destination; it doesn't delete files that no longer exist at the source.** If you delete a file from your source folder, that file will remain at the destination.

For true two-way synchronization or deletion handling, you'll need Globus Flows (covered below).

## Scheduling Transfers and Syncs

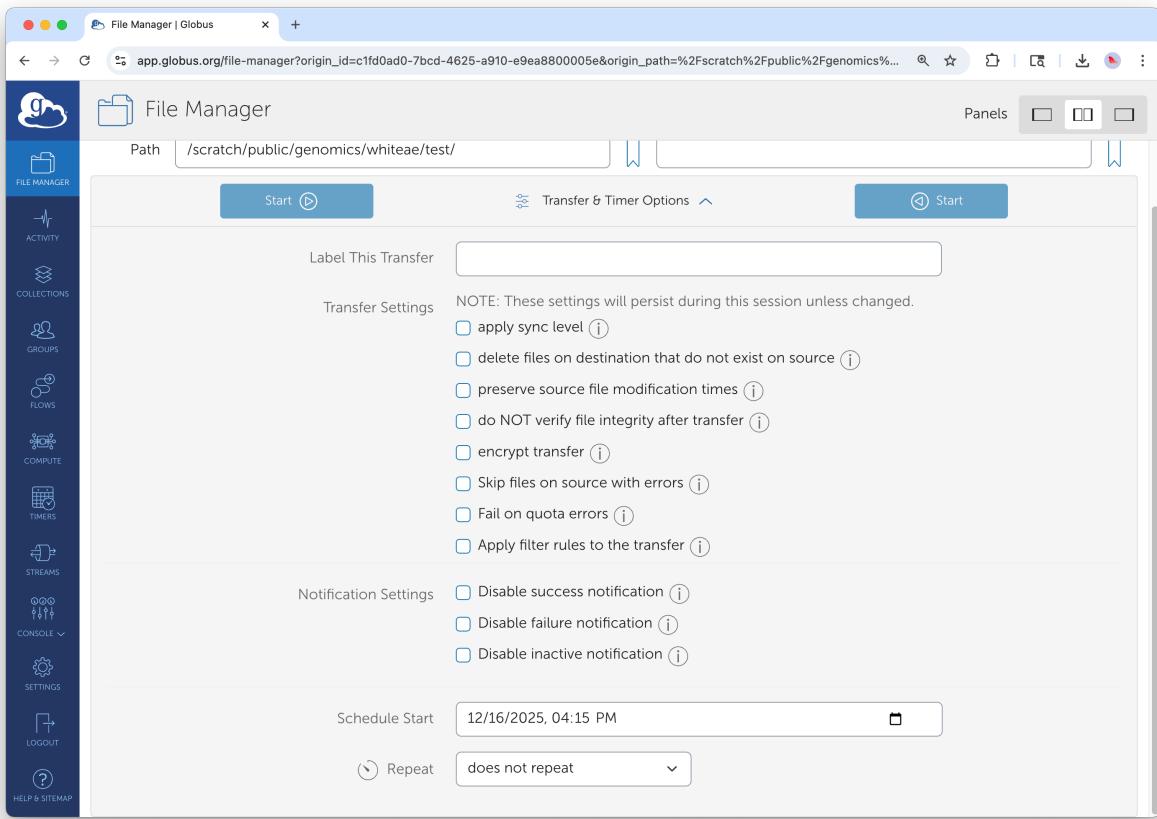
### Why Schedule Transfers?

Scheduling lets you:

- Run large transfers during off-peak hours when networks are faster
- Set up regular backups that run automatically
- Coordinate data movement around processing schedules
- Avoid tying up bandwidth during working hours

### One-Time Scheduled Transfers

To run a transfer at a specific future time:



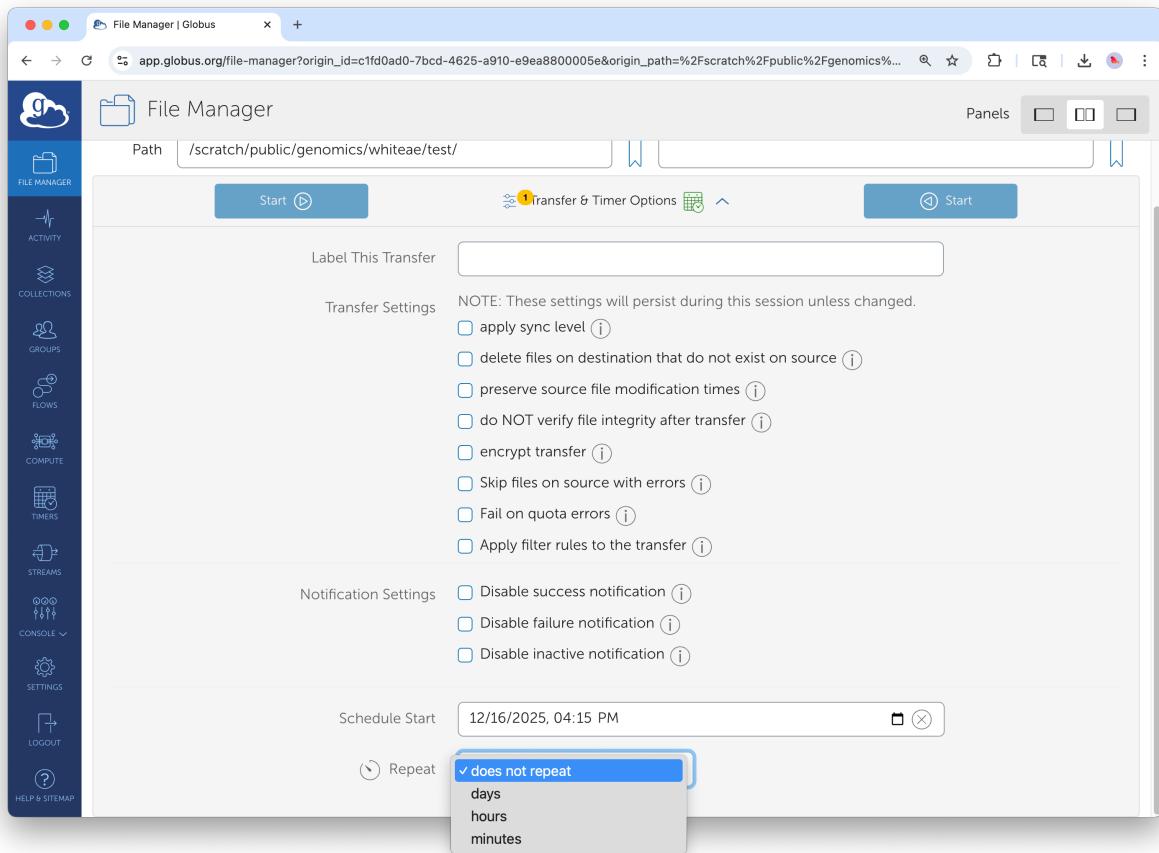
*Schedule a transfer to run at a specific date and time.*

1. Set up your transfer as normal
2. Click **Transfer & Timer Options**
3. Under **Schedule Start**, select the date and time
4. Click the **teal clock button** (not the blue Start button) to schedule

The transfer will start automatically at the specified time. You'll receive an email notification when it begins and when it completes.

## Recurring Transfers (Timers)

For automated, repeating transfers like nightly backups or weekly syncs use Globus Timers.



*Configure a recurring transfer with customizable frequency.*

1. Set up your source, destination, and any options (like Sync)
2. Click **Transfer & Timer Options**
3. Enable **Repeat** and configure the schedule:
  - Frequency: hourly, daily, weekly, or custom interval
  - Start time: when the first transfer should run
  - End date: when to stop recurring (optional)
4. Click the **teal clock button** to create the Timer

## Managing Timers

View and manage your scheduled recurring transfers from the **Timers** page:

The screenshot shows the Globus Timers interface. On the left is a sidebar with icons for File Manager, Activity, Collections, Groups, Flows, Compute, Timers (which is selected and highlighted in blue), Streams, Console, Settings, Logout, and Help & Sitemap. The main content area has a title 'Timers' and a sub-instruction: 'Use Timers to schedule and repeat data transfers or flow runs. Timers can be created either while setting up a transfer in File Manager using settings under the Transfer & Timer Options tab or while starting a flow run using the options under the start button.' Below this is a search bar with filters for ACTIVE, INACTIVE, and COMPLETED, and a 'filter by timer name' input field with a search icon. A table lists one timer entry:

TYPE	LATEST RUN	INTERVAL	STOP AFTER
TRANSFER	-	every 7 days	3 Runs

At the top right of the table are icons for pause, delete, and more. The table also has a 'Sort' dropdown and a note '• AWAITING NEXT RUN'.

*The Timers tab shows all your scheduled recurring transfers.*

1. Go to **Activity** in the left sidebar
2. Click the **Timers** tab
3. Here you can:
  - View upcoming scheduled runs
  - Pause or resume timers
  - Delete timers you no longer need
  - See the history of past runs

## Practical Scheduling Patterns

**Nightly backup:** Sync your project folder from a workstation to the SI Data Center every night at 2 AM when the network is quiet.

**Weekly archive:** Transfer completed datasets to long-term storage every Friday afternoon.

**Hourly sync:** Keep a shared collaboration folder synchronized between STRI and the Data Center.

**Field season automation:** Schedule daily syncs from a field station laptop during your research season, with an end date when the season concludes.

## Globus Flows: Automated Multi-Step Workflows

### What Are Flows?

Globus Flows chain multiple actions into automated workflows. Instead of manually running a transfer, then cleaning up old files, then notifying collaborators. You define a Flow that does all of this automatically.

Flows can include:

- Transfer or Sync operations
- Delete operations
- Conditional logic
- Email notifications
- Integration with external services

### Accessing Flows

The screenshot shows the 'Library | Flows | Globus' interface. On the left is a sidebar with icons for FILE MANAGER, ACTIVITY, COLLECTIONS, GROUPS, FLOWS (which is selected), COMPUTE, TIMERS, STREAMS, CONSOLE, SETTINGS, LOGOUT, and HELP & SITEMAP. The main area is titled 'Flows' and contains tabs for 'Runs', 'Library' (selected), and 'Deploy a Flow'. A search bar says 'search flow library'. Below it, a message says '26 flows available to you in the library'. There are three flow cards:

- Two-Stage Transfer** by Globus Team: 'Transfer from a source collection to a destination collection using an intermediate endpoint'. It has 27 steps, was created on 7/27/2023 at 12:10 PM, last modified on 10/21/2024 at 09:58 AM, and keywords: Globus-Provided,Two Stage,Two Hop,Intermediate,Globus Transfer,Transfer,Production. It has a 'Start' button and a 'View' button.
- Move (Copy and Delete) Files** by Globus Team: 'Transfers data from a source collection to a destination collection, then deletes the data from the source collection'. It has 20 steps, was created on 7/27/2023 at 12:09 PM, last modified on 10/21/2024 at 09:56 AM, and keywords: Globus-Provided,Move,Data Move,Globus Transfer,Transfer,Production. It has a 'Start' button and a 'View' button.
- Transfer and Set Permissions** by Globus Team: 'Transfer to a guest collection that you own or manage and grant an identity or group access to the data'. It has 0 steps, was created on 7/27/2023 at 12:09 PM, last modified on 10/21/2024 at 09:56 AM, and keywords: Globus-Provided,Move,Data Move,Globus Transfer,Transfer,Production. It has a 'Start' button and a 'View' button.

*The Flows section in the left sidebar gives access to workflow automation.*

Click **Flows** in the left sidebar to access the Flows interface. Here you can:

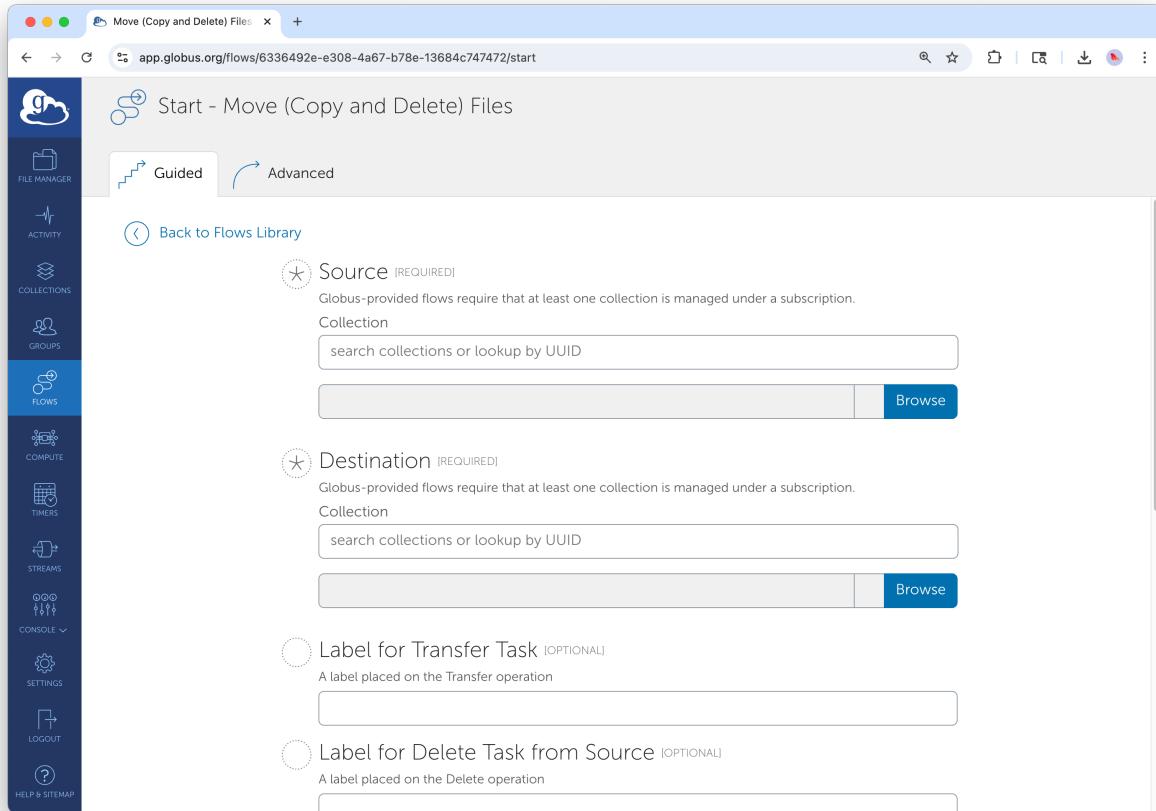
- Browse available Flow templates
- Run Flows that have been shared with you
- Create new Flows (requires additional permissions)
- Monitor running and completed Flow executions

### Example Flow: Transfer and Delete

A common workflow: transfer files from an instrument staging area or incoming folder to permanent storage, then delete the originals to free up space.

**Without Flows:** You manually transfer, wait for completion, verify success, then manually delete the source files.

**With a Flow:** Define a “Transfer and Delete” Flow that automatically: 1. Transfers all files from source to destination 2. Verifies the transfer completed successfully 3. Deletes the original files from the source 4. Sends you an email confirmation



*Running a Transfer and Delete Flow. Specify source, destination, and notification preferences.*

#### To run this Flow:

1. Go to **Flows** in the sidebar
2. Find the “Transfer and Delete” template (or one shared by your administrator)
3. Click **Run**
4. Specify your source collection and path
5. Specify your destination collection and path
6. Configure notification preferences
7. Click **Start Flow**

The Flow runs asynchronously. You'll receive email notifications as each step completes.

#### Flow Safety Considerations

Flows involving deletion are powerful but require caution:

- **Test first:** Run your Flow on a small test dataset before applying to critical data
- **Verify destinations:** Double-check that your destination path is correct—deletions from source only make sense if the data safely arrived at the destination

- **Understand the order:** The Flow won't delete source files unless the transfer succeeds
- **Keep notifications on:** Email alerts let you know if something goes wrong

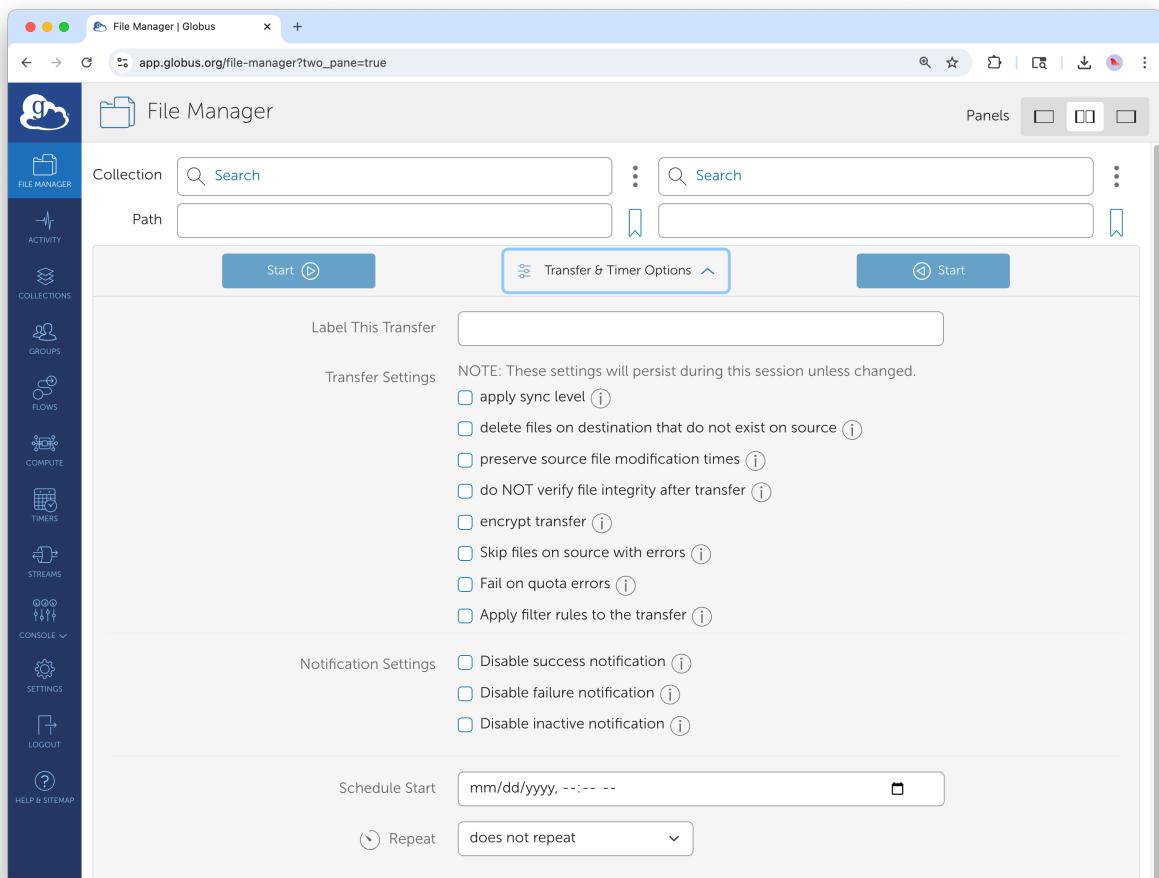
## Custom Flows

The ability to create custom Flows is available for all Smithsonian staff. For complex automation needs, contact SI-Globus@si.edu to discuss options.

## Advanced Transfer Options

Beyond Sync and scheduling, Globus offers several options to optimize and control transfers.

### Accessing Transfer Options



*The full Transfer & Timer Options panel with advanced settings.*

Click **Transfer & Timer Options** between the panels to access all settings.

## Verification and Integrity

Option	Purpose
<b>Skip file integrity verification after transfer</b>	Skip calculating checksums to confirm files transferred correctly (NOT recommended)
<b>Encrypt transfer</b>	Forces encryption in transit (enabled by default for most institutional endpoints)
<b>Preserve source file modification times</b>	Keeps original timestamps on copied files

## Handling Existing Files

Option	Purpose
<b>Overwrite files</b>	Replace existing files at destination (default behavior)
<b>Skip files that already exist</b>	Don't transfer files if a file with the same name exists
<b>Delete files on destination not in source</b>	Make destination exactly match source (use carefully)

## Performance Tuning

Option	Purpose
<b>Fail on quota errors</b>	Stop immediately if destination storage is full
<b>Skip source errors</b>	Continue transfer even if some files can't be read (useful for partially corrupted sources)

## Transfer Labels

Adding a descriptive label to your transfer makes it easier to find in your Activity history:

- “2025-03 Pollen Images - Costa Rica Site A”
- “Weekly backup - NMAH workstation”
- “Genomics data for collaborator review”

Labels are especially helpful when you run many transfers or need to find a specific historical transfer.

## Real-World Workflow Examples

### Example 1: Instrument Data Pipeline

**Scenario:** A microscope generates large image files that need to move to Hydra for processing, then to DAMS for archiving.

**Solution:** 1. Set up a **nightly Timer** that syncs from the instrument staging folder to Hydra 2. After processing completes, use a **second Timer** to transfer results to DAMS storage 3. Use a **Transfer and Delete Flow** to clean up the staging folder weekly

## **Example 2: Collaborative Research Sync**

**Scenario:** Teams at NMNH and STRI work on shared datasets that need to stay synchronized.

**Solution:** 1. Create a **recurring Sync** from NMNH to STRI (daily at midnight) 2. Create a **second recurring Sync** from STRI to NMNH (daily at 6 AM) 3. Both syncs use “modification time” to detect changes

**Note:** This pattern works for files that are only edited on one side at a time. If the same file might be edited at both locations simultaneously, you’ll need a more sophisticated conflict-resolution strategy.

## **Example 3: Field Season Automation**

**Scenario:** Researchers collect data during a 3-month field season and need regular backups to institutional storage.

**Solution:** 1. Install **Globus Connect Personal** on the field laptop (see Guide 3) 2. Create a **daily Timer** that syncs the data collection folder to SI Data Center storage (e.g., Hydra) 3. Set the Timer end date for when the field season concludes 4. Enable email notifications to confirm each sync completes

# **Troubleshooting Advanced Features**

## **Timer Issues**

**Issue: Timer runs but no files transfer - Cause:** Source folder is empty, or Sync detects no changes - **Solution:** Check that new/modified files exist in the source; verify the Timer is pointed at the correct path

**Issue: Timer stopped running - Cause:** Timer expired (reached end date) or was paused - **Solution:** Check the Timers tab in Activity; create a new Timer if needed

**Issue: Timer runs at unexpected time - Cause:** Timezone mismatch - **Solution:** Verify the timezone setting in your Globus account preferences

## **Sync Issues**

**Issue: Sync transfers files that haven't changed - Cause:** Modification times differ between source and destination (common when copying from certain filesystems) - **Solution:** Use checksum-based sync for more accurate change detection

**Issue: Deleted files reappear at destination - Cause:** Standard Sync doesn't delete; a subsequent Sync re-copies the files - **Solution:** Use “delete files not at source” option carefully, or use a Transfer and Delete Flow

## **Flow Issues**

**Issue: Flow fails partway through - Cause:** Permissions issue, endpoint offline, or network interruption - **Solution:** Check the Flow execution details for specific error messages; verify all endpoints are accessible; re-run the Flow

**Issue: Can't find expected Flow templates - Cause:** Flows may need to be shared with you or require additional permissions - **Solution:** Contact SI-Globus@si.edu to resolve access issues

## Quick Reference

### Sync vs. Transfer Decision Tree

- Do files already exist at the destination?
  - No → Use Transfer
  - Yes → Have most files changed since the last copy?
    - Yes → Use Transfer (overwrites everything)
    - No → Use Sync (only updates changed files)

### Timer Scheduling Cheat Sheet

Use Case	Frequency	Good Start Time
Nightly backup	Daily	2:00 AM
Weekly archive	Weekly	Friday 5:00 PM
Hourly sync	Hourly	Top of the hour
End-of-day sync	Daily	6:00 PM

### Key URLs

- **Globus Web App:** [app.globus.org](http://app.globus.org)
- **Flows Documentation:** [docs.globus.org/guides/flows/](http://docs.globus.org/guides/flows/)
- **Timer Documentation:** [docs.globus.org/guides/timers/](http://docs.globus.org/guides/timers/)

### Smithsonian Support Contacts

- **General Globus Support:** SI-Globus@si.edu
- **Custom Flows Requests:** SI-Globus@si.edu
- **Hydra Globus Support:** SI-HPC@si.edu
- **STRI Data Transfer Server:** STRIhelp@si.edu

---

*For help with advanced Globus features, contact SI-Globus@si.edu. For complex workflow automation needs, the Globus team can discuss custom Flow development options.*