Administering Globus Sharing in XSEDE

This guide provides the information necessary for an XSEDE Level-1 or Level-2 service provider (SP) to set up and properly configure a Globus sharing service as part of the XSEDE system. In addition to generic Globus features, it includes XSEDE-specific instructions and recommendations designed for the XSEDE community. Some of these instructions, as noted, are not appropriate for individuals or organizations that are not XSEDE Level-1 or Level-2 service providers.

Table of Contents

1. About Globus sharing

Sharing is a feature offered by the Globus services that XSEDE uses for data transfers between systems. What is Globus? Globus is the solution for XSEDE's managed data transfer use cases.[1] Globus provides high-performance, highly reliable file transfer services between Globus "endpoints."[2] Globus is an easy-to-use, thin-client, web-based service. XSEDE provides a downloadable server application, GridFTP, that allows XSEDE SPs, campus IT administrators, and laboratory IT administrators to turn their own storage systems into Globus endpoints. Once the endpoint is defined in Globus, the "endpoint administrator" can authorize local users to transfer data to or from the endpoint using Globus services.

In addition to these basic file transfer services, Globus offers a sharing feature that allows end users to create and manage their own "shared endpoints." Creating and using shared endpoints is what we mean by "Globus sharing." A shared endpoint provides access to a single file or a directory of files, owned by the end user, to users or groups defined by the end user. The end user creates the shared endpoint and defines the access control rules using Globus interfaces, and his or her colleagues then use Globus interfaces and the name of the shared endpoint to access the files or directories. Globus sharing provides the following benefits to end users.

1. Once authorized by the system administrator to do so, the owner of a file or directory (the "owner") can create shared endpoints on his or her own without further system administrator help.
2. The owner can offer access to his/her files to anyone with a Globus account. His/her colleagues can be added to access control rules (by name!) without having accounts on the local storage system. (Globus accounts and XSEDE accounts are essentially the same thing. There are no restrictions on who may create a Globus or XSEDE account, though having an account doesn't automatically provide access to anything.)
3. The owner can change access control rules at any time, instantly, without needing to wait for an administrator's help.

Globus's sharing feature was designed to provide end users the greatest possible convenience and flexibility. But it also offers important features for system administrators.

1. Administrators control which Globus endpoints allow sharing, which local users are allowed to create shared endpoints, and which portions of the endpoint can be shared.
2. Local logs track all accesses to data on the system, whether via normal or shared endpoints. The identity of the end user is recorded even if the user doesn't have an account on the local system.

[1] XSEDE Canonical Use Case 2.0: Managed File Transfer. (http://hdl.handle.net/2142/48673)
[2] Globus. (https://www.globus.org/)

3. The administrator has complete control over the endpoint itself. The administrator can shut down access to the endpoint (including any shared endpoints) at any time, either by everyone or by specific users (including the shared endpoints they created).
4. Globus provides endpoint administrators, particularly those within XSEDE, an interface for managing and monitoring their endpoints.

Globus sharing's role in XSEDE

The sharing features provided by Globus satisfy the user needs expressed in a wide variety of XSEDE use cases. The following XSEDE use cases are implemented using these features.[3]

Campus bridging:

- UCCB-4: Use of data resources from campus on XSEDE, or from XSEDE at a campus
- UCCB-5: Support for distributed workflows spanning XSEDE and campus-based data, computational, and/or visualization resources

Science gateways:

- UCSGW-2: Science Gateway community file transfers
- UCSGW-4: Large File Data Movement from/to users desktop/laptop to XSEDE resource mediated by the gateway

"Big Data" data management:

- UCDM-1: Share a common repository of data with a distributed user community
- UCDM-3: Shared use of large-scale/streaming sensor input data

Globus's sharing feature is available to XSEDE SPs as part of the XSEDE project. Campus IT operators and independent laboratories may access Globus sharing features via a Globus Provider Plan.[4] SPs who provide systems that require sharing features may enable sharing on any of their XSEDE endpoints at their discretion. XSEDE does not require that SPs enable sharing.

The user experience

Globus's sharing features were designed for end users. This section briefly describes the experience users have when sharing with Globus. This description is provided so that you--as the administrator--know what your users will do when using Globus sharing. Documentation specifically for users is also available. [ref]

In order to use the sharing features of a Globus endpoint, one must first obtain: (1) an XSEDE identity (by registering with the XSEDE User Portal, or XUP[5]), and (2) an account on the system that provides the Globus endpoint. The latter is usually accomplished by applying for and

[3] XSEDE Use Case Registry. (https://software.xsede.org/registry-dev/index.php)
[4] Globus Provider Plans. (https://www.globus.org/providers/provider-plans)
[5] XSEDE User Portal. (https://www.xsede.org/portal)

obtaining an XSEDE allocation via the XSEDE Resource Allocation System (XRAS)[6], but a local account could also be obtained directly from the Service Provider organization.

Open a web browser, login to XUP, navigate to XSEDE's Globus services, and click "Transfer Files" to view the endpoint selection screen, shown in Figure 1. All XSEDE endpoints have names beginning with xsede#, so it's easy to find the right XSEDE endpoint by typing "xsede#" in the Endpoint field. The interface will respond by providing a pop-up list of all XSEDE endpoints.

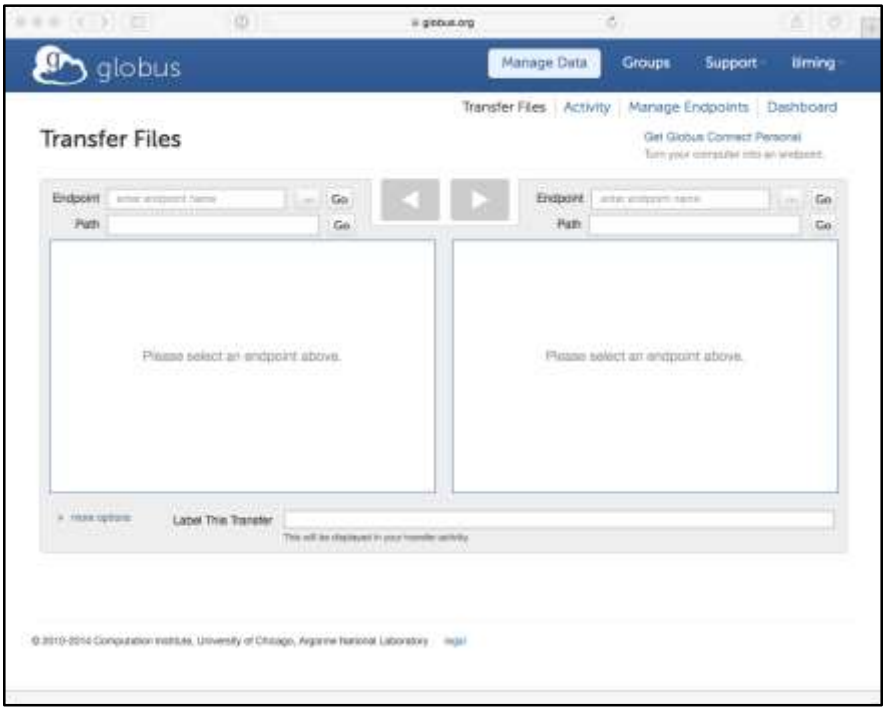

Figure 1. Globus endpoint selection interface

Once the endpoint is selected, the endpoint activation interface might require you to authenticate to the endpoint. (This usually isn't necessary if the endpoint is an XSEDE endpoint and the user obtained access via XRAS, but it may be in some situations.) If authentication is required, follow the prompts to authenticate.

Once authenticated, the contents of the endpoint will be displayed in the Transfer Files interface. This is usually (not always) your home directory.

To share a set of files or directories, start by creating a new directory with an appropriate name (e.g., "Shared"). Select the new directory by clicking on its name, and pull down the menu in the upper right of the window and select "share." Then click the "+Add Shared Endpoint" button to initiate the sharing interface.

[6] XSEDE Allocations. (https://www.xsede.org/allocations)

The first thing the sharing interface will ask for is the name of the new shared endpoint. This is the name that your collaborators will use to access your shared files. The name will begin with your XSEDE identity and a hash symbol (e.g., yourid#) Enter something appropriate in the "New Endpoint Name" field (e.g., ProjectData). This will create a shared endpoint named yourid#ProjectData that provides access the contents of your newly created directory. Click the "Create and Manage Access" button to proceed.

The next panel will allow you to set the access permissions for the shared endpoint. Permissions may be set for any of the following: one or more specific XSEDE or Globus users, one or more XSEDE or Globus groups, all users, or an email address. If you enter an email address, Globus will send email inviting the recipient to access the files, including registering with XSEDE if they aren't already registered. Once you've finished setting the access permissions, you may close the interface panel by clicking the "x" in the upper right corner of the panel.

At this point, you've created a shared endpoint, given it a name, and set access controls for the endpoint. You may now put any data you'd like to share into this directory, send your collaborators the endpoint name (e.g., yourid#ProjectData), and they can use XSEDE's Globus interface to access the contents of your shared endpoint.

If you want to stop sharing this data at any time, you can do this by returning to the Transfer Files interface and clicking "Manage Endpoints" at the top of the page. You will see a list of the endpoints you have used or created. Find your shared endpoint in the list and click the triangle symbol at the right side of the list entry. Then click the "Delete Endpoint" button. None of the data in your directory will be changed, but the shared endpoint name you created will no longer be recognized by Globus, so no one else will be able to access the data.

2. Administering Globus sharing

This section explains how you, as a service provider, can enable and disable sharing on your Globus endpoints, control who can and who cannot create shared endpoints, control the parts of your system that may be shared, monitor usage, enable and disable XSEDE automated testing for your endpoint, and prepare for the possibility of security incidents.

Creating a Globus endpoint

Although this document focuses on enabling sharing on existing endpoints, it's worth noting that there are two common methods of creating an endpoint on your system.

The method assumed in this document is to use XSEDE's GridFTP server distribution. XSEDE provides installation packages and configuration instructions[7] for GridFTP servers that are geared for Level-1 or Level-2 service providers (SPs). The packages are identical to those available from the Globus Toolkit (http://toolkit.globus.org/toolkit/), but the instructions include settings that enable XSEDE-specific features. ***The remainder of this document assumes that you used XSEDE's GridFTP server distribution.***

An alternate method for campus administrators, science gateway administrators, and anyone else who hasn't already used GridFTP servers is to use the Globus Connect Server (GCS) distribution, available from the Globus website (https://www.globus.org/globus-connect-server). This is the distribution used by most Globus endpoint providers, but it lacks XSEDE-specific configuration settings and is not specifically configured to satisfy all of XSEDE's managed file transfer use cases.

Enable sharing on a Globus endpoint

Register an XSEDE endpoint

This section is specific to XSEDE Level-1 and Level-2 service providers. Other service providers may access Globus sharing features via a Globus Provider Plan.

Globus's sharing features are available at no additional cost for all XSEDE endpoints. XSEDE endpoints are defined as endpoints with names beginning with xsede#. The first step of enabling sharing is making sure that the endpoint is registered as an XSEDE endpoint.

A new XSEDE endpoint is created by an automated process when the endpoint administrator (you) registers the endpoint in XSEDE's information service. Assuming that you already have the software installed and configured as a normal Globus endpoint, you can register it in the information service by editing configuration files on your local system. Follow XSEDE's documentation for deploying and registering GridFTP servers to register your service with XSEDE. Once registered, an XSEDE endpoint will be defined in Globus automatically after a short period of time, as described in the documentation.

Enable sharing

Once your endpoint is registered as an XSEDE endpoint or is included in another Globus Provider Plan, you can enable your users to create shared endpoints. In this section we explain how to turn the sharing feature on. In the following sections, we add information on how to restrict the feature to match your use policies and to avoid security vulnerabilities. ***Please review all of these sections, including the recommendations below for XSEDE administrators, before enabling sharing on your endpoints.***

Using the XSEDE GridFTP server distribution, the configuration change to enable sharing must be made in the GridFTP server configuration file. To enable file sharing using Globus Sharing,

[7] XSEDE GridFTP server deployment and registration instructions. (http://software.xsede.org/production/gridftp/latest/XSEDE-GridFTP-install.html)

you have to add the Globus Sharing CA certificates to your trusted certificates directory (/etc/grid-security/certificates) and use -sharing-dn option in the server as follows:

```
globus-gridftp-server -sharing-dn "/C=US/O=Globus Consortium/OU=Globus
Online/OU=Transfer User/CN=__transfer__"
```

and use -sharing-rp option to restrict the file paths allowed for sharing:

```
globus-gridftp-server -sharing-rp <path>
```

Now any user who can authenticate to your managed endpoint can create a shared endpoint.

Restrict users who can create shared endpoints

You may find that you need to restrict the set of users who are able to create shared endpoints on your system.

Use the sharing_users_allow and sharing_users_deny directives in the GridFTP server configuration file to explicitly allow or deny specific users access to the sharing features. Both directives accept a comma-separated list of user IDs (local account names). If sharing_users_allow is used, only the users explicitly listed will be allowed to create shared endpoints. If sharing_users_deny is used, any users explicitly listed will not be allowed to create shared endpoints even if they are listed in the sharing_users_allow directive.

Restrict the portions of an endpoint that can be shared

You will most likely need to designate certain parts of your endpoint as “unshareable” in order to prevent your users from accidentally exposing security credentials or other critical data. It also might be useful to make the “sharable” parts of the endpoint explicit by name so that your users are sure to be aware that any data there is sharable. It is particularly important that Globus services should not have access to dot files (or dot directories) in a user’s home directory, as these can contain critical security information and even control security configuration.

Use the restrict_paths and sharing_rp in the GridFTP server configuration file to explicitly define and narrow the portions of your filesystem that can be accessed by your local users via Globus and that can be shared with others, respectively.

For example, the following configuration change will allow your local users to both read and write files in your system’s /shared and /scratch directories (and all sub-directories), subject to the Unix permissions that you have set. They will also be able to read and write files in their home directories. But they *will not* be able to use Globus to read or write files or directories in their home directories whose names begin with a dot. They will be able to create shared endpoints in /shared (but not in /scratch) and in their home directories (but not any file or directory whose name begins with a dot.)

- Edit the GridFTP server configuration file and add the following configuration lines:

```
restrict_paths /shared,/scratch,~,N~/.*
sharing_rp /shared,~,N~/.*
```

- Restart your GridFTP server to activate the new settings.

Obviously, you should replace the specific path restrictions in the example above to match your site's access policies.

NOTE: *A bug in Globus Toolkit 5.x software allowed wildcards (*) in restrict_paths to match the '/' character in a path. (Intuitively, the '/' character in a path should not be matched by a wildcard.) If you are using a Globus Toolkit 5.x GridFTP server, you must be very careful how you use wildcards to avoid unintended path matches. GridFTP servers that use GT 6.x software follow the intuitive behavior, thus we strongly recommend that you use a distribution based on Globus Toolkit 6.x software.*

Restrict sharing to read-only

In some cases you might want to enable your users to create shared endpoints, but restrict the sharing to "read-only" mode. In read-only mode, users who create shared endpoints can give their collaborators permission to view data but they can't give their collaborators permission to change it.

Like the example above, this is also accomplished using the sharing_rp configuration setting.

- Edit the GridFTP server configuration file and add the following configuration lines:

```
restrict_paths /shared,R/scratch,~,N~/.*
sharing_rp /shared,R~,N~/.*
```

- Restart your GridFTP server to activate the new settings.

Note the R before /scratch in the restrict_paths and before the ~ in sharing_rp. This explicitly restricts local user access to read-only in the /scratch directory (even if their local Unix permissions give them write access) and restricts shared endpoints created in user home directories to no more than read-only access.

Please see the note in the preceding section about wildcard matching in Globus Toolkit 5.x-based GridFTP servers.

Use Unix file permissions to control sharing-enabled endpoints

Globus gives local Unix file permissions precedence in all operations. The greatest access that sharing can offer to an end user is the access held by the creator of the shared endpoint. The creator's access is determined by the mapping of the creator's Globus identity to his or her local Unix account and the file permissions granted to that account. As the system administrator, you can "cap" access to shared endpoints by limiting the local access of the users who can create shared endpoints.

For example, if my Globus identity is lliming and the local system is configured to map Globus lliming to the local lee account, then shared endpoints created by lliming (endpoints beginning

with lliming#) on this system can never have more access than the local lee account itself. You can set the maximum access to local files granted via lliming# shared endpoints by restricting the access granted to the local lee account. If you set the Unix permissions on a directory to r---r------ with owner lee and group staff, and I create a shared endpoint lliming#secrets that includes that directory, then Globus will be unable to give anyone write access to the directory via my shared endpoint--regardless of the permissions I set in Globus.

Disable sharing on a Globus endpoint

After enabling sharing on an endpoint, you may need to disable it. This will disable the ability to create new shares and also disable any shared endpoints that have already been created by your users. The shared endpoints will no longer be accessible via Globus services. (File owners will still be able to access their files and directories using the non-shared endpoints that you created for them before sharing was enabled.)

Disable file sharing on a GridFTP server endpoint by removing the -sharing-dn flag on your GridFTP server startup command and the sharing_dn setting in the GridFTP configuration file.

Now, all previously created shared endpoints are no longer accessible on your server, and no new shares may be created.

Monitor a Globus endpoint

As a resource administrator, you are most likely responsible for ensuring appropriate and efficient use of your IT services. Being able to monitor your services and see how they are being used is an important part of your job, and Globus provides features to help you with it.

Administrators of XSEDE endpoints are given access to the Globus Management Console[8] feature at no additional cost. Contact support@globus.org or help@xsede.org if you do not already know how to access the Globus Management Console.

Prepare for security incidents

Globus sharing adds little danger to security incidents at your site, but there are a few basic things you need to do to ensure that an incident doesn't become larger or persist longer than necessary because of the sharing features.

Disabling an end user's account

Disabling an end user's account is no different with Globus sharing than with Globus without sharing. The most important things to do are to: (1) disable the user's local Unix account, and (2) remove the user's entry in the local Globus grid-mapfile.

When sharing is enabled on a system, the administrator can take the extra step of checking the sharing_state_dir and making certain that the user has no active shares defined on the system. (Removing any files owned by the user in the sharing_state_dir will disable any active shares.)

[8] https://www.globus.org/blog/globus-management-console

Responding to security compromises

If a vulnerability, intrusion or exploit of the Globus services is detected, the Globus team will report the issue to XSEDE security officer, who in turn will alert all XSEDE service providers using a pre-established communication mechanism.

Your response to this alert will depend on the nature of the exploit that has been reported. If the issue has been confirmed to be limited to a specific set of users, you may choose to disable those user accounts until the accounts are certain to be free of intrusion. If the issue is not known to be limited in scope, you may wish to follow the instructions above to "Disable sharing on a Globus endpoint," or even shut down the GridFTP service entirely to prevent all Globus-based access to the system.

3. Recommendations for XSEDE administrators

The following recommendations are the result of reviews and discussions held by XSEDE personnel when considering the sharing feature, and are intended to help administrators avoid pitfalls foreseen by others. They are not XSEDE community requirements.

Highlighting the fact that shared endpoints are shared with others

Some end users may not realize (or may forget) that a shared endpoint is, in fact, shared with other people. There are two classes of people to consider: (1) those who create shared endpoints, and (2) those who use endpoints shared by someone else.

The second class is easiest, because the sharing mechanism itself provides an unavoidable clue that the endpoint is shared. The name of a shared endpoint will be different. Specifically, the prefix **will not** be xsede#. It will instead be the identity of the user who created the shared endpoint. Thus, anyone who is using an endpoint created by someone else will necessarily need to specify the other person's identity when accessing the endpoint. This should be sufficient to remind them that they are using someone else's shared data.

The first class (those who create shared endpoints) is a little trickier. These users will be able to access the shared files via the shared endpoints they created, but they will also be able to access the data via the original, non-shared, endpoint. When they do so, there will be no obvious clue that the data they are accessing is visible to others or might have been changed by others.

The simplest way for an endpoint administrator to remind his/her users that the data might be shared would be to require that the path to the data include a clue that the data might be shared. This can be accomplished using Unix file system naming and Globus's restrict_paths directive. (See "Restrict the portions of an endpoint that can be shared" above.) For example, if sharing is limited to paths in the /shareable namespace, and that namespace is only available via a local endpoint named xsede#sdsc-shareable, all end users, including those who can create shared endpoints, will be reminded whenever they access the data that they might have shared it with others.

Users who share inappropriately

In general, scientific end users are responsible for their own data management. However, because XSEDE resources may be used for research in sensitive fields, XSEDE resource administrators must be careful to do everything reasonably possible to ensure that end users do not misuse the system in a way that violates law. Specifically, some of the data used on XSEDE may be subject to national export controls or similar regulations. Because sharing is initiated by end users--and generally configured by end users--there is no surefire way for the administrator to prevent an end user from intentionally or unintentionally sharing data in violation of such regulations, short of disabling sharing entirely.

On systems with sharing enabled, an administrator might be notified by law enforcement, management, or other users that some of the dtaa shared by a particular user is in violation of these regulations. When that happens, the administrator has a range of actions he/she can take to correct the situation.

1. The administrator can disable sharing, system-wide. (See "Disable sharing on a Globus endpoint" above.) This will, of course, affect all users with shared endpoints on the system and anyone to whom they have given access to their shared endpoints.
2. The administrator can change the local permissions on the specific data in question, preventing read access by the data owner. Data in a shared endpoint is subject to the owner's access, so if the owner can't access it on the local system, no user of the shared endpoint can access it, either.
3. The administrator may revoke the data owner's ability to use the sharing feature. (See "Restrict users who can create shared endpoints" above. In addition to the action described there, the administrator should also remove any files owned by the owner in the sharing_state_dir.) This will allow the data owner to continue using the system but will prevent him/her from sharing any data.
4. The administrator may, of course, deactivate the owner's account. (See "Disabling an end user's account" above.) This will prevent access to any shared endpoints created by the owner.

Managing the sharing_state_dir

The security of the sharing_state_dir is important to the overall security of the Globus sharing mechanism. GridFTP itself does not prevent administrators from applying settings that are not 100% safe. The XSEDE service providers at San Diego Supercomputing Center (SDSC) have developed a local mechanism[9] for managing the Globus sharing feature for end users that ensures safe practices for the sharing_state_dir and that also allows per-user control over the sharing feature. We recommend that XSEDE service providers make use of SDSC's mechanism or provide one of their own that closely matches the practices demonstrated by SDSC.

[9] https://security.sdsc.edu/resources/sharectl