DSS How Globus Data Transfer and Globus Sharing for DSS works

In the following, we want to give you an overview on how we integrate Globus Sharing into DSS.

What is Globus?

Globus is a sustainable, non-profit business within the University of Chicago and Argonne National Laboratory. It has its roots in the Grid Computing era and started out as a research project to provide reliable high performance data transfers between Grid Computing Sites. It was started in 1997 and since then has evolved into a powerful and easy to use toolkit for data intensive science.

Their main Product or Service is the Globus Data Transfer Portal, which lets you easily move data between GridFTP Servers or personal GridFTP endpoints, using an intuitive web-based File-browser.

What is Globus Sharing?

Globus Sharing is an extension to the Globus Transfer service. If a particular GridFTP Server has been enabled for Globus Sharing, local users are allowed to create so called Shared Endpoints on-top of the GridFTP server. Access to these shared endpoints can then be granted to arbitrary Globus Users. And since everyone can sign up for a user account at Globus (called a Globus ID), access can be granted to everyone, who has a valid E-Mail address. When an external user tries to access a shared endpoint, Globus will translate these accesses to the local system as if they were carried out by the local user, who created the Shared Endpoint.

For more details on how this works, please take a look at

Globus Security Deep Dive
Globus Sharing Effective File Permissions

How DSS Integrates with Globus Data Transfer

For our DSS Systems, we operate a dedicated GridFTP Gateway, that can be used to access the data stored in DSS from outside of LRZ. For more information on how to use this, please see the DSS documentation for users. Authentication and Authorisation in Globus/GridFTP is based on X.509 certificates. In order to hide that complexity from you, we use the InCommon CILogon Service, that issues short lived certificates for users, which can be authenticated via Shibboleth SSO. For more information about the authentication and authorisation and data flows, please look at this.

How DSS Integrates with Globus Sharing?

In our LRZ Data Science Management approach, only the data curators of a data project have the privilege to decide which users are able to access the data stored in the project's data containers. So basically for every container (or path), a certain set of users is allowed to control data access to this container (or path).

In the Globus Sharing approach, the sharing privilege can only be restricted to certain paths and certain users. However, it cannot be restricted to a combination of path and user. This means for example, if user Alice is data curator of Project A and Bop is data curator of Project B, we would need to give both of them data sharing permissions. However, we cannot tell Globus that Alice is only allowed to share Paths that belong to Project A's containers. That is no problem as long as Alice has only access to containers of Project A. But we run into a problem when Bop gives Alice access to one of his containers. As then Alice would be allowed to share this container via Globus, which may not be desired by Bop.

In order to work around this conceptual incompatibility we control the Globus Sharing functionality for DSS via the DSSWeb Self Service Portal. The solution we've come up, is the following:

The Globus Sharing privilege on our GridFTP Gateways is restricted to a special functional user, called dssglobus. So only this user is allowed to create shared endpoints in Globus. This user is "owned" by the DSSWeb Self Service Portal. When you enable Globus Sharing on a certain DSS Container via the DSSWeb Self Service Portal, the following will happen:

1. The user dssglobus is added to the respective DSS Container Group, so he has access to the data in the container.
2. As dssglobus user, we create a Globus Shared Endpoint in Globus.

After the Globus Shared Endpoint for your container has been created, you can now set, modify and delete Globus Sharing ACLs via the DSSWeb Self Service Portal. So when you tell DSSWeb to invite some external user to your container, DSSWeb will authenticate against Globus as the dssglobus user and then create an respective Globus Sharing ACL/Invitation for the Shared Endpoint of your container, using the Globus REST API.

Known Limitations
As we create the Globus Shared Endpoints as dssglobus user, all external accesses translate to accesses from dssglobus. While this should be no problem in normal situations, as all container users should have access to all data, it could be a problem when using Container Streamline Mode NONE or if users revoked the DSS default ACLs (on purpose or by accident). In this situation, dssglobus may not have access to all or even any of the files in the container and therefore the observed result of sharing the container via Globus may not be exactly as expected.

Please note that the only solution to this problem is to give dssglobus access to the files and directories, that should be shared via Globus.

While Globus Sharing is a really convenient and unique solution to share Petabytes of data with arbitrary users around the globe, you should also note that because of the technical nature of the Globus Sharing feature, once you enable the Shared Endpoint for your container, you technically give Globus the possibility to access your data. So you have to trust Globus to manage and protect access to your data in a well-behaved manner. For more details on Globus Sharing Security, please see Globus Security Deep Dive. Even though LRZ has a long personal history with the makers of Globus and we have decent trust in Globus, you should consider that Globus is based in the US and therefore is subject to US law and has to cooperate with US government and law enforcement officials to comply with the US laws. So there remains some residual risk for unauthorized data access.

We therefore by default do not enable Globus Sharing on DSS containers. It has to be explicitly enabled by the data curators for each container individually. For use cases, which require very high security or confidentiality, like personal-data, trade secrets, etc. we recommend that you either leave Globus Sharing for your container disabled or/and store your data using a secure encryption method.

### Related articles

- [DSS How to apply for storage on TUM-DSS](#)
- [DSS How to transfer data between SuperMUC and DSS](#)
- [DSS Understanding DSS on DGX-1](#)
- [DSS How an external invited user get's started with Globus Online](#)
- [DSS How DSS Container Auto Group Links work](#)