

Using Data Science Storage via TUM Workbench

A How-to Guide

1. Data Science Storage

Data Science Storage (DSS) is a storage solution for data intensive research offered by the Leibniz Supercomputing Centre. The Technical University of Munich (TUM) has purchased a share of DSS, called TUM-DSS, which is organised in data projects and managed by data curators. TUM researchers who need vast amounts of storage can apply for a DSS container via the Information Officer (IO) of their faculty or school.¹

Data stored in a DSS container can be accessed and shared within the LRZ computing ecosystem. A DSS container can also be mounted on the TUM Workbench². This allows you to manage your DSS files via the TUM Workbench, for example by describing them with metadata, assigning them to projects or linking them to other elements. For general questions on DSS, please consult the LRZ documentation³. This guide explains how to use DSS via TUM Workbench.

Using DSS via TUM Workbench has several advantages. A decisive point is that data can be accessed from anywhere simply by using a web browser, which makes the evaluation of data more flexible. The permissions for accessing data can be set to “read only”, which ensures that data remains unchanged and cannot be manipulated. Furthermore, it is very easy to grant access permissions to data in a DSS container to both TUM members and external users. The users can be added to a project to which the DSS files are assigned. Using a DSS container within a group or chair also supports uniform workflows and enriches the organisation.

¹ IO Contact information: <https://www.it.tum.de/en/it/committee/members/>,
<https://doku.lrz.de/display/PUBLIC/DSS+How+to+apply+for+storage+on+TUM-DSS>

² TUM Workbench: <https://www.ub.tum.de/workbench>

³ LRZ documentation: <https://doku.lrz.de/display/PUBLIC/Data+Science+Storage>

2. Mounting a DSS container on the TUM Workbench

To mount a DSS container on the TUM Workbench, the LRZ data curator has to activate the DSS container configuration via the following steps: ⁴

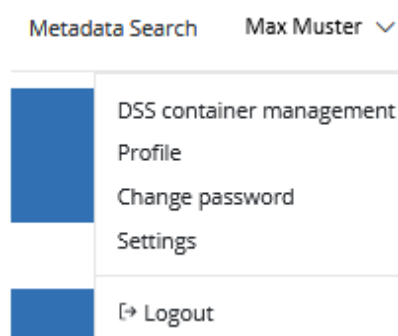
1. Navigate to the container's overview page in the DSSWeb⁵.
2. Scroll down to the section "Container Configs".
3. Click on the "+Add new Config" button.
4. Select the "TUM Workbench (Beta)" and click on "Apply".



Config	Choose config
Description	Choose config
	TUM Workbench (Beta)
	LRZ Remote Visualisation
	Apply

A DSS container can be managed via the TUM Workbench by TUM Workbench DSS curators. Please contact TUM Workbench support⁶ to assign you the role of a TUM Workbench DSS curator. This allows you to connect DSS containers to the TUM Workbench and configure that connection. The TUM Workbench DSS curator can be identical with the LRZ container curator, but need not be.

After you have been granted this role, select the option "DSS container management" in the drop-down menu below your user name in the top right corner of the TUM Workbench.



Metadata Search Max Muster ▾

- DSS container management
- Profile
- Change password
- Settings
- Logout

⁴ LRZ documentation: <https://doku.lrz.de/display/PUBLIC/DSS+documentation+for+data+curators#DSSdocumentationfordatacurators-ManagingDSSContainerConfigs>

⁵ DSSWeb: <https://dssweb.dss.lrz.de>

⁶ TUM Workbench support: workbench@ub.tum.de

Click on the “+ DSS Container” and fill in the form as follows:

1. Name

This is used as a reference for you. Please enter a text not longer than 255 characters.

2. Path

The path must consist of four parts: the DSS file system, the name of the directory (e.g. “tumdss”), an identifier such as a TUM User ID, and the rest of the container mount path, which usually also includes the identifier. There should be no slashes at the beginning or the end of the path, e.g. dssfs03/tumdss/pr53ve/pr53ve-dss-0000.

3. Read/Write setting

Select one of the following options:

- Read Only: Imported files cannot be updated and new files cannot be written to the container via the TUM Workbench.
- Read Write No New: Imported files can be updated but no new files can be added to the container via the TUM Workbench.
- Read Write Only New: Imported files cannot be updated but new files can be added via the TUM Workbench.
- Read Write All: Imported files can be updated and new files can be added via the TUM Workbench.

Regardless of the chosen option, the metadata of the DSS files can always be edited in the TUM Workbench.

4. Import option

Select one of the following options:

- Only new Globus imports: The TUM Workbench will be informed about new files added through Globus Online and the corresponding files will be imported.
- Import list: A .json file with the following structure can be uploaded within the detail view of a container and the files will then be imported:

```
[
  { "path": "/dss/dssfs01/pr53ve/pr53ve-dss-0000/envelope-1/storage-1/example-1-1-1.txt" },
  { "path": "/dss/dssfs01/pr53ve/pr53ve-dss-0000/envelope-1/storage-1/example-1-1-2.txt" },
  { "path": "/dss/dssfs01/pr53ve/pr53ve-dss-0000/envelope-1/storage-1/subdir-1/example-1-1.txt" },
  { "path": "/dss/dssfs01/pr53ve/pr53ve-dss-0000/envelope-2/storage-2/subdir-2/subdir-3/example-2-2.txt" },
  { "path": "/dss/dssfs01/pr53ve/pr53ve-dss-0000/envelope-3/storage-3/example-3-1.txt" }
]
```

- Import all: The TUM Workbench will periodically crawl the whole container mount path and will import all files that are found within envelopes (see below Section 3) that have a valid metadata.json.

Please inform the TUM Workbench administrators as soon as you have filled in the form for connecting DSS containers into the TUM Workbench. The DSS mount will then be finalised by the TUM Workbench administrators and the LRZ.

New DSS Container x

Name *

Path *

Read/Write setting *
 Read only v

Import option *
 v

3. Handling DSS files via TUM Workbench

Every container path should have one or many envelope folders underneath containing a metadata.json file, which should contain the following three elements:

```
{
  "tum_id": "pr53ve",
  "projects": [],
  "metadata_fields": []
}
```

All keys are required, but projects and *metadata_fields* can be empty lists.

- The *tum_id* should be a string containing the username of a TUM Workbench user. All files and storages within the envelope generated on the TUM Workbench

will be assigned to that user by setting the *created_by* attribute. Now, the user has access to all the files and storages within that envelope folder.

- The projects list can contain strings with primary keys (PKs) of projects. The *tum_id* user must have editing rights within the projects, which means the role must be at least *Project Member*. The *Observer* role does not grant editing rights. All members of the project can access the files. The access type depends on the role the users have in this project.
- The curator will receive failure notifications via email. PKs of projects can be found in their URLs, e.g. workbench.ub.tum.de/projects/4cea3b46-9f83-47ce-9683-bbf1f3150d92
PK: 4cea3b46-9f83-47ce-9683-bbf1f3150d92
- The *metadata_fields* list can contain .json objects with the ID and values of existing metadata fields in the TUM Workbench. Please ask workbench@ub.tum.de for the ID of relevant metadata fields.

tum_id users have access to all storages and files imported for them. If the container allows, the user can also add new files, or update them from the TUM Workbench. TUM Workbench container curators have access to all storages and files imported into their containers. Other access settings should be handled through TUM Workbench projects and their roles. Projects can be set from metadata.json files or directly on the TUM Workbench.

If a project is set on a storage, all TUM Workbench files within that storage receive automatic access from the project roles. This means *Observers* can view the storage and files, *Project Members* can view, edit, etc.

Technical University of Munich
University Library
Service Engineering and Research Support

Arcisstraße 21, 80333 München
Tel. +49 89 289 25323
workbench@ub.tum.de
www.ub.tum.de
www.tum.de

Last update: July 2022