

SOP Model Data Submission to THREDDS at GEOMAR

How to submit and generate FAIR data on GEOMAR THREDDS

Version 1.3

Authors:

Klaus Getzlaff (ORCID: [0000-0002-0347-7838](https://orcid.org/0000-0002-0347-7838))

Research Data Management

Information, Data and Computing Centre

GEOMAR Helmholtz Centre for Ocean Research Kiel <https://ror.org/02h2x0161>

Date: August 30th 2022

Contact: datamanagement@geomar.de

Purpose

The Purpose of this **Standard Operating Procedure (SOP)** is to unify the workflow of submitting model data and its import into the GEOMAR THREDDS System, in order to provide model data and metadata for exchange and long term storage following the [F.A.I.R. principles](#) and the [Research-Data-Policy at GEOMAR](#)

Scope

Description of applicable data

This SOP applies to numerical model data submissions covered by one of the following types of submission requests:

1. **Supplementary Data Submission** (permanent storage and access) as supplement to a scientific publication to fulfill requirements of journals about "data & code availability"
 - `aim`
permanent digital imprint of figures/tables with the documentation/scripts on post-processing steps and plotting ideally in combination with reference to **Data Publication** via [DataCite](#)
 - `storage size`
O ~100GB
 - `metadata`
minimum requirement such as creator, institute, title, project, UUID/PID; also included in NetCDF files
 - `files`

additionally README, LICENSE.txt and Checksums.sha256sum

- `UUID`

persistent identifier (PID) handle

2. **Stand-alone dataset** (permanent storage and access) as open-access distribution of model data; central experiments and their large-volume output should be published as **Data Publication** via [DataCite](#) at WDCC (World Data Centre for Climate) - CERA at DKRZ

- `aim`

permanent exchange/supply and publication of model data, citable

- `storage size`

less than 1TB

- `metadata`

enriched

- `files`

additionally README, LICENSE.txt and Checksums.sha256sum

- `UUID`

persistent identifier (PID) handle

3. **Data Exchange** (temporal storage only) as exchange of model data/output with external project partners with defined end date of storage

- `aim`

temporal/short-term exchange of model data

- `storage size`

O ~1TB

- `metadata`

minimum requirement such as creator, institute, title, project, UUID/PID; also included in NetCDF files

- `files`

additionally README, LICENSE.txt and Checksums.sha256sum

- `UUID`

URL (only temporal --> no PID handle)

If you are in doubt if your model data submission request fit into one of the three types above please contact datamanagement@geomar.de for advise

Applicable Roles

Involved / intended roles are:

1. **User:**

Users are those that request model data submission to the system. They can be scientists, technicians, lab staff, PIs, ...

2. **Curator:**

The **curator** takes care of the model data import to the THREDDS System, i.e. check completeness of data, metadata and minimum documentation control, licence, access rights. Could be the same person as the **storage manager**.

3. **Storage manager:**

The **storage manager** has information on the availability of storage on the system and transfers the submitted model data to the storage space. Could be the same person as the **curator**.

Instructions

Main steps:

1. **User** fills in [data submission form](#) with details on planned storage of model data, this triggers automated ticket creation and notification of **curator**

Ticket is created

please note: In each of the following steps please provide the ticket number received in step 1 for ALL communication between the above named user groups.

2. **Curator** checks completeness of information provided and creates a UUID connected with the above request to identify storage location for model data, metadata, documentation and licence
3. **Storage manager** creates storage location (target directory) on system and transfers data into target directory
4. **Curator** checks/updates necessary informations, metadata, checksums and creates PID handle if required
5. **Curator** provides link to data and sends confirmation email about successful upload to the THREDDS System

Ticket is closed

Step-by-Step

- 1.1 **User** fills in [data submission form](#) with details on planned storage of model data

Ticket is created

please note: In each of the following steps please provide the ticket number received in *step 1.1* for ALL communication between the above named user groups.

- 1.2 **Curator** is informed from ticket system by email about data submission request

- 1.3 **Curator** checks ticket information and modifies title, customerID according to information given and unlock ticket in system

- 2.1 **Curator** checks completeness of information provided, if needed **curator** contacts **user** for missing information and questions on data provided in *step 1.1*

2.2 **Curator** contacts appropriate **storage manager** to check availability of necessary resources

2.2 **Curator** creates a UUID connected with the above request to identify storage location for model data, metadata, documentation and licence

2.3 **Curator** moves ticket to queue: *DataSubmission: 01 Storage*

3.1 **Storage manager** creates storage location using UUID (target directory) on system

3.2 **Storage manager** transfers data into target directory

3.3 **Storage manager** moves ticket to queue: *DataSubmission: 02 Einbinden*

4.1 **Curator** creates LOG directory to store necessary scripts and informations for submission request

4.2 **Curator** creates landing page for data submission

4.3 **Curator** checks/updates metadata in NetCDF files (minimum requirement: creator, institute, title, project, UUID/PID)

4.4 **Curator** creates/updates README and LICENSE (CC BY 4.0) file

4.5 **Curator** verifies/updates Checksum

4.6 **Curator** moves ticket to queue: *DataSubmission: 03 Handle*

4.7 **Curator** creates PID handle link if required

5.1 **Curator** provides link to data and confirmation e-mail about successful upload to the THREDDS System

5.2 **User** may forward confirmation email with link to relevant 3rd party such as scientific journal, funding agency or project partners. This includes data not freely accessible due to embargos.

5.3 For *Supplementary Data Submission* under review the **curator** moves ticket to queue: *DataSubmission: 04 Review* for duration of review process with a reminder (3 month) enabled in the OTRS

5.4 For *Supplementary Data Submission* under review after successful revision the **curator** links dataset with publication in OceanRep and **curator** updates the information on landingpage of dataset

5.5 **Curator** closes ticket

█ Ticket is closed

Glossary

Term	Definition
Portal	GEOMAR data management portal: portal.geomar.de
THREDDS	GEOMAR in-house system to store, access and exchange data: data.geomar.de
UUID	Universally Unique Identifier (UUID) to be connected with each submission request
PID	Persistent Identifier: The PID binds a metadata object to a digital object – before and after processing. (Peter Wittenburg (Max Planck Computing & Data Facility, MPCDF) (2016) – Usage of PIDs from a Lab Perspective, S. 17)
OceanRep	OceanRep is an open access digital collection containing the research output of GEOMAR staff and students (GEOMAR Archive): oceanrep.geomar.de

Annex

Description of document: URL to document

Revisions

Version no.	Date	Author	Comment
1.0	2021-07-10	Lisa Paglialonga	New edition "Data Submission Guide" on portal.geomar.de
1.1	2022-03-08	Klaus Getzlaff	Documentation on internal documentation platform
1.2	2022-08-29	Klaus Getzlaff	Rewritten internal documents as SOP
1.2.1	2022-08-30	Klaus Getzlaff	Review comments and feedback integrated
1.3	2022-08-30	Klaus Getzlaff	final release