Preservation Policy

1. Purpose

This preservation policy describes the digital preservation strategies and principles of the 4TU.ResearchData repository and provides transparency on the procedures involved in ensuring adequate preservation of and access to the data and software items within the repository.

As a trusted digital repository that has received the CoreTrustSeal certification, 4TU.ResearchData demonstrates its long-standing good archival practice by drafting this preservation policy. The policy is addressed to our designated community of data producers and users, as well as research funders, 4TU.ResearchData staff, and other interested audiences.

2. File formats and preservation

One of the goals of digital preservation is to prevent loss of access to files due to file format obsolescence. To enhance the chance of future interpretability of the data, 4TU.ResearchData has created a list of preferred formats for full preservation for which it guarantees long-term usability. In general, the preferred file formats used for full preservation of data, are standard, exchangeable or open file formats. Data depositors are strongly recommended to provide their data in the preferred format, which is most suitable to the type of data. The 4TU.ResearchData repository also accepts other formats, but informs depositors that only bit-level preservation will be provided.

4TU.ResearchData applies two levels of file support:

- Full preservation: All reasonable actions to maintain usability will be taken. Actions may include migration, normalization or conversion.
- Bit-level preservation: Only access to the object in its submission file format is provided.

The current accepted and preferred file formats are reviewed periodically to see whether they are still suitable for preservation. The list of preferred formats is available on the website.

3. Archival workflow

The 4TU.ResearchData preservation policy adheres to the terminology and preservation practices outlined by the Open Archival Information System (OAIS) Reference Model as well as the FAIR Data principles. The OAIS model provides both a functional model – the specific tasks performed by the archive, such as storage or access – and a corresponding information model, to support long-term maintenance and access to digital material. All processes are organised according to this model.

3.1 Pre-ingest:
Data producers wishing to deposit data are offered guidance and assistance. This can include technical support, but also consultancy about issues such as data formats, the preferred data model, and legal/privacy aspects. Data depositors are provided with clear instructions on how to prepare, document and deposit their data.

3.2 Ingest:
4TU.ResearchData collects data from all fields and subjects in the science, engineering and design domains as outlined in the Data Collection Policy. Once a data depositor has completed the upload, the data are submitted as a Submission Information Package (SIP) to the data repository.

To ensure the completeness, accuracy and usability of the metadata, all deposited datasets undergo the following metadata quality checks:

- completeness of required and optional metadata,
- documentation,
- links to related materials,
- file format,
- privacy issues.

Once the SIP is accepted into the system, an Archival Information Package (AIP) is created. The AIP is the information package the system stores, preserves and sustains.

3.3 Archival storage:
The AIP is saved in our repository system. All metadata and data files are saved on the TU Delft S3 storage service, except for data in NetCDF format. These files are saved on our OPeNDAP server.

4TU.ResearchData uses the DataCite service to assign Digital Object Identifiers (DOIs) to ensure the accessibility and authenticity of the data. DOIs are reserved when items are deposited into the repository. Once a dataset is published, its DOI is displayed at the top of the dataset's landing page, along with a suggested citation. When a new version of a dataset is published, a new metadata landing page is created and a new (versioned) DOI is created. This way, the already existing persistent identifier (DOI) will continue to refer uniquely to the earlier version of the dataset.

3.4 Access:
All datasets are findable and accessible from the repository's Home page. The data and their associated metadata and documentation are made available to the user as a Dissemination Information Package (DIP). The DIP is the information package created to distribute the digital content. Users are able to download the complete DIP, as well as separate files within the package.

Access to NetCDF data (and HDF5) is further enhanced through the OPeNDAP protocol. A major advantage of using OPeNDAP is the ability to retrieve subsets of files without the need to download the whole dataset, and the ability to aggregate series of data files, e.g. a time series, into one 'virtual' dataset.

4. FAIR Data Principles

4TU.ResearchData aims to operate according to the FAIR Data Principles in making its data *Findable*, *Accessible, Interoperable* and *Reusable:*

- A DOI (Digital Object Identifier) is assigned to each published dataset or software item to make it *Findable.*
- Support for open and standard file formats making data *Interoperable*.
- All datasets, except for those that are restricted, are *Accessible* to everyone and there is no need for special protocols to download and obtain the data.
- Metadata are always openly *Accessible*, even if the data are restricted or under embargo.
- The metadata of datasets and research software published in 4TU.ResearchData follows the DataCite metadata schema, one of the most used metadata standards to guarantee its *Interoperability*. The metadata model contains required, recommended and optional fields to make your dataset more *Findable* and *Reusable*. All the data and software items have a clear and accessible data usage licence to facilitate its *reuse*.
- Controlled vocabularies, community standards, or ontologies are supported where possible, making data *Interoperable* and exchangeable with other data and systems.

5. Back-up, integrity, and authenticity

Data storage of 4TU.ResearchData is managed by the ICT department of Delft University of Technology according to their procedures. The stored research data are backed up (and stored) on hard disks (RAID6) and synchronized (one way) daily. Two times a month a backup is made on disks at another location and retained for one year. Data are stored on two mirror sites.

In order to ensure restore procedures, the root-filesystems are backed up incrementally on a daily basis, and once a week full backups are made. These backups are saved on tapes and will be kept for three months on another location. A restore can be carried out upon request. Security updates and patches are installed on a regular basis.

These preservation procedures are outsourced to the ICT department of Delft University of Technology and recorded in a service level agreement.

To ensure the integrity of the datasets, for every deposited file a checksum (md5 type) is made which allows 4TU.ResearchData staff to check the files for defects. In case file degradation is discovered, the corrupted data will be removed and replaced with its uncorrupted counterpart from mirror sites. All changes are logged in an audit trail.

6. Retention

A minimum retention period for all objects stored in the repository is 15 years. After this time, 4TU.ResearchData continues to retain the objects until a significant change occurs which affects the conditions of preservation. For example, this could include file format obsolescence of objects deposited in a non-preferred file format for which no long-term preservation could be guaranteed.

In these cases, the published dataset may be removed from access, resulting in the deaccessioning or ‘unpublishing’ of the dataset with a statement on the landing page explaining why the files are no longer available.

7. Financial sustainability

Long-term sustainability of 4TU.ResearchData requires adequate and reliable sources of funding so that data is preserved properly. To ensure the data repository is able to fulfil its mission, structural funding is received from the members of the 4TU.ResearchData Consortium.

Should a situation arise which threatens the continued existence of the data repository, these organisations are committed to taking responsibility for the future availability of the data entrusted to the repository.

8. Responsibilities

All 4TU.ResearchData staff work together to ensure that the data stored remains accessible and understandable over the long term. They closely cooperate with the TU Delft ICT department, which is responsible for providing the storage infrastructure.

4TU.ResearchData is responsible for the maintenance, review and revision of all its policies and documentation, including the Preservation Policy. If you have any comments or questions regarding this policy, please contact us at researchdata@4tu.nl.