



4TU.Centre for Research Data

Notes Before Completing the Application

We have read and understood the notes concerning our application submission.

True

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments:

CORE TRUSTWORTHY DATA REPOSITORIES REQUIREMENTS

Background & General Guidance

Glossary of Terms

BACKGROUND INFORMATION

Context

R0. Please provide context for your repository.

Repository Type. Select all relevant types from:

Domain or subject-based repository, National repository system; including governmental

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

Comments

4TU.ResearchData (previously known as 4TU.Centre for Research Data) was started in 2008 as an initiative of the three technical universities in the Netherlands – Delft University of Technology, Eindhoven University of Technology, and the University of Twente. The ambition was, and still is, to create and maintain a national state-of-the-art facility for storing and preserving science and engineering research data and for making those data openly accessible. The data repository has been fully operational since 2010 and it has evolved to become a trusted and certified repository for science and engineering. Currently, the repository holds just over 6800 datasets, corresponding to about 42 TB of data.

4TU.ResearchData is engaged in Research Data Netherlands, a national coalition of data archives with a mission to promote long-term archiving and reuse of research data.

The 4TU.ResearchData repository can be accessed via <https://data.4tu.nl>.

Besides the repository website at <https://data.4tu.nl>, 4TU.ResearchData maintains three other websites:

Portal website at <https://data.4tu.nl/portal>, which we consider to be the 'home' for our repository and from where users can search and access the repository.

Info website at <https://data.4tu.nl/info>, providing information about the repository service including its policies and guidelines.

Community website at <https://community.data.4tu.nl/>, which is an online platform to connect, engage and collaborate with our community around 4TU.ResearchData.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

Brief Description of the Repository's Designated Community.

The core business of 4TU.ResearchData is to provide a repository for long-term access of curated research datasets, with a focus on data from science, engineering and design. The repository went live in 2010 and since then has been managed as a service for researchers (from research institutions around the world) to deposit and share their data, and for other researchers to download and use data in their research.

Every researcher, both in the Netherlands and abroad, can upload data to the data repository or access and download data for use in their research. Nevertheless, the majority of the data in the 4TU.ResearchData repository comes from the Dutch technical universities.

Datasets deposited by researchers from the Netherlands as well as from abroad undergo the same data quality checks. Each dataset submitted is checked for completeness and accuracy of the metadata, for the sustainability of the file format, the presence of a readme file, for privacy issues in case of confidential data, etc. The (meta)data quality review is an essential part of our curation process and ensures that the quality requirements of the repository are met. (See also guideline R11)

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

Level of Curation Performed. Select all relevant types from:

B. Basic curation – e.g. brief checking; addition of basic metadata or documentation, C. Enhanced curation – e.g. conversion to new formats; enhancement of documentation

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

Comments

4TU.ResearchData adds value to the content of the repository by performing quality assurance routine checks on the deposited data, the addition of basic metadata and documentation, if applicable, and check on preferred file formats to improve the sustainability of the dataset. By performing these checks we comply to the Findable, Accessible, Interoperable and Reusable (FAIR) data principles and help researchers to increase the impact of their research.

4TU.ResearchData offers more than one level of curation. Integrity of original copies is always maintained. In all cases there will be a thorough checking and addition of metadata and documentation (B). For the vast majority of datasets curation level C is being performed where curators help amend provided files to address issues with their formatting and content.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

Outsource Partners. If applicable, please list them.

The technology underpinning the 4TU.ResearchData repository is provided by Figshare, using the 'Figshare for Institutions' service, which includes the intuitive interface, search functionalities and visualisation tools.

Although the technology is provided at present by Figshare, the data repository itself is an independent service designed for the long term, of which its content will remain controlled by 4TU.ResearchData.

4TU.ResearchData (with TU Delft as the contract party) and Figshare have signed three agreements:

1. Licence agreement; in which are written down the agreed functionalities, as well as the roles, responsibilities and liabilities of all parties involved.
2. A processor agreement, in which the (IT) security measures and the technical and legal obligations of the parties involved are laid down. Through this agreement, 4TU.ResearchData and Figshare comply with the European General Data Protection Regulation (GDPR);
3. A Service Level Agreement, in which the mutual obligations in service level are agreed, e.g. updates, downtime, customer support, technical infrastructure and development.

The data themselves are stored locally on TU Delft servers for which the ICT department of Delft University of Technology is responsible.

A Service Level Agreement between 4TU.ResearchData and TU Delft ICT is in place.

All agreements are available upon request.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

Other Relevant Information.

4TU.ResearchData is engaged in Research Data Netherlands, a national coalition of research data organisations with a mission to promote long-term archiving and reuse of research data.

4TU.ResearchData is gold member of The Carpentries (<https://carpentries.org>), a non-profit organization that teaches computer programming and data science skills to researchers through hands-on workshops. The Carpentries membership allows 4TU.ResearchData to request centrally organized workshops on software source code management and arrange for researchers, or research support staff, to receive training as instructors. These benefits are fundamental to build our community around software and computational aspects of reproducible working with research data. We provide the training on a regular basis at the member institutions of 4TU.ResearchData (TU Delft, Eindhoven University of Technology, University of Twente).

The 4TU.ResearchData repository is run by the 4TU.ResearchData Consortium, which consists of Delft University of Technology (TU Delft), Eindhoven University of Technology and University of Twente. 4TU.ResearchData is managed by the TU Delft Library, of which the datasets are locally stored on TU Delft servers, and is governed by the legal statements of its host institution, the TU Delft.

4TU.ResearchData staff, all located at the TU Delft Library, is responsible for managing the data repository, including its preservation policy to guarantee long-term usability and comprehension of its archive's data. The 4TU.ResearchData team cooperates closely with the TU Delft ICT Department, which is responsible for the storage infrastructure (storage on redundant servers, backup strategy, and data recovery).

TU Delft is responsible for managing the data repository and the storage infrastructure. These tasks are not distributed among partners of the 4TU.ResearchData Consortium. More details on how continuous access to the data files are ensured are provided in guideline R16 on Security.

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

ORGANIZATIONAL INFRASTRUCTURE

I. Mission/Scope

R1. The repository has an explicit mission to provide access to and preserve data in its domain.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

The mission of 4TU.ResearchData is to ensure the accessibility of technical scientific research data to give a quality boost to contemporary and future research. Mission statement is published on the website:

<https://data.4tu.nl/info/en/about/organisation/>

4TU.ResearchData promotes its services as well as best practices to data management to the scientific community in a variety of ways, including a monthly newsletter, exhibits at scientific conferences and meetings of professional associations, print and electronic brochures, training and workshops, and dedicated efforts in specific areas (e.g. community of practice for data stewards and thematic working groups).

An annual Highlights report is published to update our stakeholders and users on our results.

Furthermore, 4TU.ResearchData collaborates nationally and internationally with various organisations, including Research Data Netherlands (RDNL), DataCite, the Dutch National Coordination Point Research Data Management (LCRDM), and actively contributes to the CESAER Group on Research Data Management.

References:

Monthly newsletters: <https://data.4tu.nl/info/en/news-events/monthly-newsletters>

Highlights report 2020: <https://data.4tu.nl/info/en/about-4turesearchdata/highlights>

Research Data Netherlands: <https://researchdata.nl/en/>

DataCite: <https://datacite.org/>

National Coordination Point Research Data Management (LCRDM): <https://www.lcrdm.nl/en>

CESAER: <https://www.cesaer.org/>

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

II. Licenses

R2. The repository maintains all applicable licenses covering data access and use and monitors compliance.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

When depositing data in 4TU.ResearchData depositors are required to select a licence from a predefined list.

4TU.ResearchData offers the full range of Creative Commons licences for datasets, as well as dedicated licences for software and code (the most commonly used software licenses are supported).

4TU.ResearchData has adopted CC0 (Creative Commons Zero) as the default means for researchers to share their datasets and software to make its reuse as easy as possible without any legal barrier. If there are reasons or circumstances when data can't be shared with a CC0 licence, depositors can choose another, more appropriate licence for their data/software. Guidance on all licence types offered can be found on the website:

<https://data.4tu.nl/info/en/use/publish-cite/upload-your-data-in-our-data-repository/licencing/>

These licences govern the usage, distribution and author attribution requirements of the datasets.

The licence selected by the data depositor is displayed on the public dataset landing page, with a link to the relevant licence web page for full information on the licence's requirements.

4TU.ResearchData prefers data that can reside in the public domain. Data that is personal, confidential or sensitive in

nature, is only accepted for archiving when the data has been anonymized so that individuals, organisations or businesses cannot be identified. (See also R4, Confidentiality/Ethics).

When depositing data in 4TU.ResearchData, the depositor is asked to consent to a Deposit Agreement. The Deposit Agreement gives permission to 4TU.ResearchData to disseminate the contents under the access conditions and use specified by the depositor at the time of deposit. It also allows 4TU.ResearchData to store, copy and modify the dataset, without changing the content, to ensure that it can be preserved and made available in the future.

References:

Deposit Agreement:

https://data.4tu.nl/info/fileadmin/user_upload/Documenten/4TU.Research_Data_-_Deposit_agreement.pdf

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

III. Continuity of access

R3. The repository has a continuity plan to ensure ongoing access to and preservation of its holdings.

Compliance Level:

3 – The repository is in the implementation phase

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

3 – The repository is in the implementation phase

Response:

According to Article 2.4 of the Deposit Agreement "The repository shall ensure, to the best of its ability and resources, that the deposited dataset is archived in a sustainable manner and remains legible and accessible". See:

https://data.4tu.nl/info/fileadmin/user_upload/Documenten/4TU.Research_Data_-_Deposit_agreement.pdf

Published datasets are guaranteed for a minimum of 15 years.

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 partners of the 4TU.ResearchData Consortium (currently: TU Delft, Eindhoven University of Technology, University of Twente). 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.

The rights and obligations of the partner institutions are laid down in a Consortium Agreement: https://data.4tu.nl/info/fileadmin/user_upload/Documenten/Signed_Consortium_Agreement_4TU_Centre_for_Research_Data_20190220.pdf

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
accept

IV. Confidentiality/Ethics

R4. The repository ensures, to the extent possible, that data are created, curated, accessed, and used in compliance with disciplinary and ethical norms.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

4TU.ResearchData prefers data that can reside in the public domain.

Data that is personal, confidential or sensitive in nature, is only accepted for archiving when the data has been anonymised so that individuals cannot be identified. Pseudonymised data (with the exception of sensitive personal data and special categories of personal data which cannot be deposited in 4TU.ResearchData) can be shared under restricted access conditions. In the Data Collection Policy a matrix is included that shows the different type of data and risk classifications, with examples of data that fit into each classification.

The following access levels can be applied in the 4TU.ResearchData repository:

- 1) Open access; the default setting by which all datasets will be directly accessible to others, free of charge.
- 2) Temporary embargo; 4TU.ResearchData will restrict access to the data until the end of the embargo period; at which time, the content will automatically become publicly available.
- 3) Permanent embargo (restricted access); whereby the depositor can set conditions on when and how access is granted, allowing the data to be securely reused, protecting the data for commercial, ethical or legal reasons. Recommended when data are sensitive, and the sensitive information can't be removed without the dataset losing value.
- 4) Metadata only; when the data is published elsewhere, but there is a need to publish a record also in 4TU.ResearchData to fulfil funder or institutional requirements. A link to the location of the dataset should be provided.
- 5) Private link/URL; to share a dataset (e.g. for peer review purposes) before it is actually published.

While providing open access to data, 4TU.ResearchData is at the same time committed to protecting the confidentiality of research participants and the rights of the data depositors.

As a way of ensuring confidentiality, 4TU.ResearchData relies on a combination of anonymization, access eligibility criteria, informed consent among study participants, and the terms of the licence mentioned on the dataset description. These terms make explicit how the dataset can be used. (see also R2, Licences).

4TU.ResearchData provides guidance on the responsible use of data with disclosure risk (consultancy in the pre-ingest phase and deposit guidelines on our website). The deposited data are furthermore screened by staff for disclosure risk and, where appropriate, returned to the depositor with the request to fully anonymize the data. As a rule, datasets are not anonymised by 4TU.ResearchData itself. Anonymised datasets are accepted.

As stated in the Terms of Use, depositors are responsible for ensuring that any dataset submitted to 4TU.ResearchData is compiled with due observance of the Netherlands Code of Conduct for Research Integrity, the General Data Protection Regulation (GDPR) and other applicable laws and regulations.

A Privacy Policy is available on the website that explains how 4TU.ResearchData collects, stores, protects, and utilizes personal information provided by its users. In addition to the Privacy Policy, a separate Cookie Policy is in place to inform our website's visitors and users about which cookies our website uses, what type of data they collect and for what purposes this data is processed.

References:

Data Collection Policy: https://data.4tu.nl/info/fileadmin/user_upload/Documenten/Data_collection_policy_2020.pdf

Deposit guidelines: https://data.4tu.nl/info/fileadmin/user_upload/Documenten/Deposit_Guidelines_2019.pdf

Terms of Use: https://data.4tu.nl/info/fileadmin/user_upload/Documenten/Terms_of_Use_4TU.ResearchData.pdf

Netherlands Code of Conduct for Research Integrity: <http://www.vsnu.nl/files/documents/Netherlands%20Code%20of%20Conduct%20for%20Research%20Integrity%202018.pdf>

The General Data Protection Regulation (GDPR):

https://ec.europa.eu/info/law/law-topic/data-protection/data-protection-eu_en

Privacy Policy: https://data.4tu.nl/info/fileadmin/user_upload/Documenten/4TU-Privacy-policy.pdf

Cookie Policy: https://data.4tu.nl/info/fileadmin/user_upload/Documenten/Cookie-Policy.pdf

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

V. Organizational infrastructure

R5. The repository has adequate funding and sufficient numbers of qualified staff managed through a clear system of governance to effectively carry out the mission.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

Within the 4TU.Federation foundation, 4TU.ResearchData was established in 2008. It started as an initiative of three technical universities, namely Eindhoven, Delft and Twente, despite the fact that the name 4TU.ResearchData might

suggest four technical universities. The duties and responsibilities of each member of 4TU.ResearchData are laid down in a Consortium Agreement. The Board of the 4TU.ResearchData Consortium sets down the 4TU.ResearchData's long-term plan, annual plan, approves budgets, and takes decisions on its strategy.

The 4TU.ResearchData repository is managed by the TU Delft Library, datasets are locally stored on TU Delft servers, and governed by the legal statements of its host institution, the TU Delft. The technology underpinning 4TU.ResearchData is provided by figshare, using the 'Figshare for Institutions' service. It provides the intuitive interfaces, search, and visualisation tools.

4TU.ResearchData staff, all located at the TU Delft Library, is responsible for managing the data repository including its preservation policy to guarantee long-term usability and comprehension of its repository's data. Staff includes a repository manager, data engineer, developers and data curators. The 4TU.ResearchData team cooperates closely with the TU Delft ICT Department, which is responsible for the storage infrastructure (storage on redundant servers, backup strategy, and data recovery).

Although the back office is located at TU Delft, each member institute has a so-called Front Office with staff (usually data stewards) who provide direct support to researchers of their institution. Periodically a Front Office meeting is organised where staff members from all member institutions that are in contact with customers/users discuss new or current developments.

References:

Consortium Agreement: https://data.4tu.nl/info/fileadmin/user_upload/Documenten/Signed_Consortium_Agreement_4TU_Centre_for_Research_Data_20190220.pdf

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

VI. Expert guidance

R6. The repository adopts mechanism(s) to secure ongoing expert guidance and feedback (either inhouse or external, including scientific guidance, if relevant).

Compliance Level:

3 – The repository is in the implementation phase

Reviewer Entry

Reviewer 1

Comments:

3 – The repository is in the implementation phase

Reviewer 2

Comments:

3 – The repository is in the implementation phase

Response:

Around the technical infrastructure of the repository, an online community has been established which comprises currently more than 70 members who are researchers, research supporters and technical experts. See:

<https://community.data.4tu.nl/>

The overarching goal of the community is to provide an inclusive space for members to connect and exchange knowledge about best practices for the creation and reuse of FAIR data within the technical sciences. Programming currently includes adhoc one-to-one engagement with members through interviews, monthly community-led working group meetings, and biannual community calls. Communication channels include Slack for direct messaging, the online platform (to showcase member profiles, community groups, a blog, news and events) and a monthly newsletter.

Furthermore, the repository encourages contact with the designated community by having links to its email address or physical contact on the website. There is an open line of communication for all data depositors regarding ingest and use of their data. Data depositors are also asked for feedback on the deposit process when their dataset has been published.

In the near future, a sounding board with actual users will be established to fulfil the function of raising issues and giving advice on the repository's usability and functionalities.

4TU.ResearchData staff regularly participate in community events, seminars and conferences.

With regard to the deposit and archiving of research data, the team works closely with research support staff of member- and other institutions (so called 'front offices').

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

DIGITAL OBJECT MANAGEMENT

VII. Data integrity and authenticity

R7. The repository guarantees the integrity and authenticity of the data.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

4TU.ResearchData uses the DataCite service to assign Digital Object Identifiers (DOIs) to ensure accessibility and authenticity of the data. DOIs are minted when items are deposited and published, and are included with a suggested citation. 4TU.ResearchData prefers data that can reside in the public domain for the purpose of scientific validation. Once published, depositors are allowed to update their dataset. Modifying the dataset title, authors, or the file(s) associated with the dataset will automatically create a new version of the item and its DOI. Updating any other metadata field (description, categories, keywords, licence, item type, references, funding information, etc.) will not generate a new version. The original DOI of the published dataset will always lead to the latest version of the item. Versions are listed and accessible in the drop down menu under the item title, each is timestamped.

Because new versions are created when crucial metadata and/or data files are changed, each new version will go through the metadata review again to ensure the quality requirements of the repository are still met.

If applicable, we create collections which provide a context for the data within the collection.

To ensure the integrity of the data files, the repository runs virus checks and calculates MD5 checksums for all versions of a data file, and stores the size (in bytes).

The data repository maintains links to other relevant materials (e.g. article, thesis, documentation, other datasets, etc.) and to metadata of measuring instruments, locations, etc. whenever applicable.

The unique identity of a depositor is ensured by the required login using institutional credentials via SSO (Single Sign-On) or eduID, both provided through SURFconext.

Reviewer Entry

Reviewer 1

Comments:
Accept.

It's good that the library has a doi minting function.

Reviewer 2

Comments:
Accept.

VIII. Appraisal

R8. The repository accepts data and metadata based on defined criteria to ensure relevance and understandability for data users.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

The principles and criteria by which the repository develops its data collection are indicated in our Data Collection Policy. This policy defines the scope and nature of the datasets published in the data repository as well as the supported model for access and licensing. The Data Collection Policy is flexible and responds to future developments and shifting requirements that will influence the repository's data collections (technology, scientific standards, thematic focus, etc.). By making the Data Collection Policy available, we want to demonstrate 4TU.ResearchData's commitment in this area, and inform the research community of the guiding principles and practices, and challenges we face in our data collection development.

We perform quality assurance checks on the data files, metadata and documentation files of the SIP. The extensive metadata record created by depositors is checked for completeness and comprehensibility (see R7, Data integrity and authenticity and R11, Data quality). If that screening reveals major problems (disclosure risk, completeness, plausibility, etc.), the SIP is usually returned to the depositor with instructions on how to address the standards and to produce another SIP to deposit. In case of minor problems, the metadata, and more rarely the data files, are modified by the repository.

More details about our review process can be found in our Metadata review guidelines.

Our preferred and accepted formats are communicated transparently in the Data Collection Policy as well as in our Preservation Policy. In all cases we ask the data depositor to convert the data and documentation files into a format accepted by 4TU.ResearchData. Currently the formats for data files that we receive most often are in netCDF. These are also the dissemination formats most commonly requested by users (see R14, Data reuse).

References:

Data Collection Policy: https://data.4tu.nl/info/fileadmin/user_upload/Documenten/Data_collection_policy_2020.pdf

Metadata review guidelines: https://data.4tu.nl/info/fileadmin/user_upload/Metadata_review_guidelines_June_2021.pdf

Preservation Policy: https://data.4tu.nl/info/fileadmin/user_upload/Documenten/4TU.Preservation_Policy.pdf

Preferred File Formats: https://data.4tu.nl/info/fileadmin/user_upload/Documenten/Preferred_File_Formats_2019.pdf

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

IX. Documented storage procedures

R9. The repository applies documented processes and procedures in managing archival storage of the data.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

The main principles of long-term preservation and permanent storage facilities are outlined in our Preservation Policy.

See: https://data.4tu.nl/info/fileadmin/user_upload/Documenten/4TU.Preservation_Policy.pdf

The 4TU.ResearchData preservation policy adheres to the terminology and preservation practices outlined by the Open Archival Information System (OAIS) Reference Model.

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 dataset has been approved, the dataset is accepted and will be subsequently published and archived for long-term preservation.

To ensure the integrity of the datasets, for every deposited file a checksum (MD5 type) is generated. Once the deposit has been completed, the MD5 checksum value is calculated again and compared to the previously generated value to ensure the file is not corrupted and modified during the upload process. MD5 checksums allow 4TU.ResearchData staff to check the files for defects and restore correct versions if necessary.

Data storage of 4TU.ResearchData is managed by the ICT department of the TU Delft according to their procedures. Files are stored on 3 locations (3 separate copies): 2 locations at TU Delft and 1 at Leiden University. If for any reason a file gets corrupted it will be detected by the system regularly checking all hashes of the files. In this case the storage admins will get a notification and the file will be repaired using the other copies. (See also R16. Security)

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

X. Preservation plan

R10. The repository assumes responsibility for long-term preservation and manages this function in a planned and documented way.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

In our Preservation Policy we describe in more detail the specific preservation actions taken on the data we hold in order to maintain its integrity and ensure its accessibility over time. Preservation Policy:

https://data.4tu.nl/info/fileadmin/user_upload/Documenten/4TU.Preservation_Policy.pdf

4TU.ResearchData has selected a number of preferred formats for full preservation for which it guarantees longterm usability. In general, the preferred file formats used for full preservation of data, are nonproprietary, well documented, and well understood by 4TU.ResearchData staff.

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 list of the preferred formats is available on our website:

https://data.4tu.nl/info/fileadmin/user_upload/Documenten/Preferred_File_Formats_2019.pdf

The Deposit Agreement clearly highlights the responsibilities and the rights of 4TU.ResearchData and the data depositor (see R2, Licences). The deposit agreement, as well as guidelines for deposit, are available on our website and in the data upload form. The transfer of custody of data, the actual deposit in 4TU.ResearchData, is clearly marked in our system, since the data can only be deposited if the deposit agreement is accepted by the data depositor. The depositor maintains the copyright on the deposited data as well as on the corresponding documentation. 4TU.ResearchData has the rights to copy, transform and store the items, as well as to provide access to them as stated in the Deposit Agreement (see R2, Licences).

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

XI. Data quality

R11. The repository has appropriate expertise to address technical data and metadata quality and ensures that sufficient information is available for end users to make quality-related evaluations.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

Upon ingest the dedicated moderator checks the submitted data, metadata and other documentation. This could result in going back to the data depositor/producer to improve the metadata and/or documentation. This could be the case when metadata are not sufficient, a readme file is missing or data are incomplete or do not match the description. More details about our review process can be found in our Metadata review guidelines. (See also R8 Appraisal, and R12 Workflows) Deposit guidelines and Guidelines for creating a readme file are provided on the website and are part of the pre-ingest phase.

Our preferred formats are communicated transparently in the list of Preferred file formats. In all cases we ask the data depositor to convert the data and documentation files into a format accepted by 4TU.ResearchData. Currently the formats for data files that we receive most often are in netCDF. These are also the dissemination formats most commonly requested by users (see also R8 Appraisal, and R14 Data reuse).

Basically, providing a minimum set of metadata is obligatory for all deposits. Some metadata fields are mandatory, while others are optional.

Citations to related works are provided as standard field in the data upload form. 4TU.ResearchData automatically generates the citation to the dataset concerned, following the DataCite format.

Our metadata standard is DataCite compliant.

The performed metadata checks includes automatic schema validation and automatic creation of rich DataCite metadata.

Currently only data depositors and 4TU.ResearchData staff are able to comment on the quality of the metadata or data in

4TU.ResearchData, no other members of the designated community.

References:

Metadata review guidelines: https://data.4tu.nl/info/fileadmin/user_upload/Metadata_review_guidelines_June_2021.pdf

Guidelines for creating a README file:

https://data.4tu.nl/info/fileadmin/user_upload/Documenten/Guidelines_for_creating_a_README_file.pdf

Deposit guidelines: https://data.4tu.nl/info/fileadmin/user_upload/Documenten/Deposit_Guidelines_2020.pdf

DataCite Metadata schema: <https://schema.datacite.org/>

Preferred file formats: https://data.4tu.nl/info/fileadmin/user_upload/Documenten/Preferred_File_Formats_2019.pdf

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

accept

XII. Workflows

R12. Archiving takes place according to defined workflows from ingest to dissemination.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

The archival workflow of 4TU.ResearchData is described in our Preservation Policy and is documented in internal workflow and processing guides. It is oriented towards the OAIS reference model (Ingest, Data management, Access, Archival storage, Administration, Preservation planning). 4TU.ResearchData has added the content and activities of the

pre-ingest function to the original OAIS framework in order to ensure the quality of the data and determine problematic issues, such as confidentiality, prior to the official deposit. The pre-Ingest function includes negotiation of data acquisitions, checking rights and access criteria, licencing, and guidance and technical support for data producers wishing to deposit data.

Data deposit guidelines provide data depositors with clear instructions on how to properly prepare, document and submit their data (see R11, Data quality). The principles and criteria by which the repository develops its data collection are indicated in the Collection Policy.

In short, the archiving workflow is as follows:

- Once a dataset Submission Information Package (SIP) has been created, it is reviewed by a qualified curator.
- The curator verifies if the dataset meets the requirements; both the requirements and the appraisal process are described in R8, Appraisal.
- Datasets which do not pass the review process, are rejected and the depositor is contacted.
- Upon approval by the curator, the dataset will be accepted and published online. At the same time a DOI for the dataset will be registered and attached.
- The curator informs the depositor that the dataset is published online.

References:

Preservation Policy: https://data.4tu.nl/info/fileadmin/user_upload/Documenten/4TU.Preservation_Policy.pdf

Deposit guidelines: https://data.4tu.nl/info/fileadmin/user_upload/Documenten/Deposit_Guidelines_2020.pdf

Data Collection Policy: https://data.4tu.nl/info/fileadmin/user_upload/Documenten/Data_collection_policy_2020.pdf

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

XIII. Data discovery and identification

R13. The repository enables users to discover the data and refer to them in a persistent way through proper citation.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

4TU.ResearchData supports data discovery in various ways. The datasets can always be found through the Data Portal, <https://data.4tu.nl/portal>, browsing by subject or institution, or by searching the metadata. To enhance the discoverability of the datasets, 4TU.ResearchData stimulates adding as many metadata and other relevant documentation as possible. All visible metadata and some metadata extracted from data files are indexed and searchable. Queries can contain special operators, fieldnames, wildcards, etc. and results can be refined using facets by the user.

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 also the ability to aggregate series of data files, e.g. a time series, into one 'virtual' dataset.

Research data are currently available in formats suggested, required and frequently used by the data consumers or in formats in which 4TU.ResearchData has the highest confidence with regard to durability. However, over 80% of all data are in netCDF format.

4TU.ResearchData uses the DataCite service to assign Digital Object Identifiers (DOIs) to ensure the accessibility and authenticity of the data. DOIs are minted when items are deposited and published, and are included with a suggested citation. When a new version of a dataset is published, a new metadata landing page is created and a new DOI is minted. Each of the versions have their own DOI and can be cited independently. Versions are listed and accessible in the drop down menu under the item title, each is timestamped.

The version DOI is created from the base DOI by adding the v<x> suffix, where x is the version number. The base DOI always points to the latest public version.

All metadata can be harvested via the OAI-PMH protocol and are findable through portals like NARCIS, the gateway to scholarly information in The Netherlands, Google Dataset Search, and DataCite Search.

We've embedded the schema.org standard in the dataset landing pages to improve the discoverability of our datasets and follow our community recommendations for embedding machine-readable metadata in landing pages.

Every dataset published in the 4TU.ResearchData repository, which has an associated article or project, becomes automatically aggregated to the OpenAIRE portal, where it can be found alongside other research.

4TU.ResearchData is also represented in the Registry of Research Data Repositories (re3data.org) and in FAIRsharing.org.

Reviewer Entry

Reviewer 1

Comments:

Accept.

The DataCite service adopted by this repository seems effective.

Reviewer 2

Comments:

Accept

XIV. Data reuse

R14. The repository enables reuse of the data over time, ensuring that appropriate metadata are available to support the understanding and use of the data.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

The 4TU.ResearchData repository aims to ensure that sufficient metadata of high quality are captured to support understanding and reuse of data.

As described in R8, mandatory and optional metadata are provided by the data depositor when submitting a dataset.

To help ensure understandability of the data to consumers, data depositors must provide a clear title and description; they are also encouraged to indicate the steps to reproduce the research which led to the data, for instance methods, workflow and/or software used; and to provide links to any software or other datasets used in generating the data.

In addition, data depositors are strongly recommended to provide a README file to help ensure that the data can be correctly interpreted when sharing the data with others. Guidelines for creating a README file are available on the website: https://data.4tu.nl/info/fileadmin/user_upload/Documenten/Guidelines_for_creating_a_README_file.pdf

In order to be discoverable and usable, the metadata added to the data repository needs to contain a minimum set of

attributes. A summary of metadata fields which are mandatory or seen as desirable is provided on the website: <https://data.4tu.nl/info/en/use/publish-cite/upload-your-data-in-our-data-repository> (tab "Metadata").

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

TECHNOLOGY

XV. Technical infrastructure

R15. The repository functions on well-supported operating systems and other core infrastructural software and is using hardware and software technologies appropriate to the services it provides to its Designated Community.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

The 4TU.ResearchData repository's core infrastructural software runs on Figshare servers using the 'Figshare for Institutions' service. Data files are locally stored on TU Delft servers. The connection between TU Delft servers and Figshare is provided through the S3 protocol.

Agreements on the repository's availability, maintenance windows and support services, are laid down in a Service Level Agreement.

4TU.ResearchData also maintains an OPeNDAP server for the storage and querying of data in netCDF format. For data residing on the OPeNDAP server, a link to the server is provided on the dataset landing page allowing users to access the data files. From the OPeNDAP server there is a link back via the DOI to the dataset landing page in the repository.

An API is provided to allow every user to programmatically move content to and from the repository and to automate their workflow. The API documentation is available at: <https://docs.figshare.com/>

To support the sharing and discovery of code and software, the 4TU.ResearchData repository allows users to integrate with GitHub, GitLab and Bitbucket repositories. The integration results in the assignment of a DOI to each version/release of a code repository which enables users to refer to and cite specific versions of code/software. DOIs are registered through the DataCite service.

Development of 4TU.ResearchData's core infrastructure is a continuously and ongoing process and led by Figshare.

Although the production server is controlled by Figshare, 4TU.ResearchData maintains a development server where new functionality can be tested before releasing on production.

Development and maintenance of TU Delft servers is taken care of by qualified staff in narrow collaboration with the TU Delft ICT Department. A software inventory is regularly updated and system documentation is maintained.

References:

NetCDF and OPeNDAP: <https://data.4tu.nl/info/en/about-your-data/netcdf-and-opendap>

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

accept

XVI. Security

R16. The technical infrastructure of the repository provides for protection of the facility and its data, products, services, and users.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

The 4TU.ResearchData repository's core infrastructural software runs on Figshare servers using the 'Figshare for Institutions' service. As part of this service, a storage service (hosted by Hetzner Online GmbH) is provided for storing metadata and through which the data finally ends up on the servers of the TU Delft. The connection between the Figshare servers and the TU Delft storage servers is provided through the S3 protocol. The repository platform is deployed with AWS in the EU Ireland region. All data transfer, process and storage inside the Figshare platform is limited to the AWS EU region. The recommended AWS TLS Security Policies are deployed.

Information security

As part of the ISMS IS027001 Figshare maintains an Information security policy. The objective of Information Security is to prevent any loss of, or threat to, the confidentiality, integrity and availability of the proprietary information of the business and of the information provided by clients (customers, suppliers, contractors/subcontractors) under the applicable statutory rules and requirements. This policy, along with the entire Information Security Management System, must observe the relevant legal regulations in Information Security that apply to the organization, as well as the contractual requirements.

In regards to user and personal data, Figshare is fully GDPR compliant. Figshare only processes a limited amount of personal information on behalf of 4TU.ResearchData, which may include details of individual users such as names and basic contact information (e.g. emails) and professional information (e.g. job title, place of work / research / department).

Secure authentication: 4TU.ResearchData users log in via a single authentication mechanism, provided through SURFconext. The authentication service is employed with the Security Assertion Markup Language 2.0 (SAML2.0).

4TU.ResearchData uses the central storage facility of TU Delft. The storage capacity can be increased on demand. To ensure the integrity of the datasets, for every deposited file a checksum (MD5 type) is generated. Once the deposit has been completed, the MD5 checksum value is calculated again and compared to the previously generated value to ensure the file is not corrupted and modified during the upload process. In case a file gets corrupted "in transit", so between the application server and the storage, it will be immediately recognized and the transfer will be performed again.

Files are stored on 3 locations (3 separate copies): 2 locations at TU Delft and 1 at Leiden University. If for any reason a file gets corrupted it will be detected by the system regularly checking all hashes of the files. In this case the storage admins will get a notification and the file will be repaired using the other copies.

If a file is corrupted for some reason on our storage, this corrupted version is stored as a new version. The original file

remains stored on our server. The storage works with version control. The storage servers are using the object based storage system and are configured to replicate content. If a disk within a cluster fails, a duplicate disk is available, ensuring that the system continues running with no interruption or performance degradation. Data can be replicated within nodes and clusters and among distributed data centers for additional back-up off-site.

Backup and restore of the OpenDAP server: A backup of the entire server is made on working days. Within 14 days, a backup of every working day can be restored. After these 14 days, weekly backups can be restored during a period of 1 year.

A restore can be made on request. This is possible for a complete restore of the server or individual files / folders. Data files on the OpenDAP server are stored on 3 locations (3 copies) of which 1 location (backup) at Leiden University.

The 4TU.ResearchData archive is available through a secure https connection and a "SSL" encryption layer on top of it.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

APPLICANT FEEDBACK

Comments/feedback

These requirements are not seen as final, and we value your input to improve the core certification procedure. To this end, please leave any comments you wish to make on both the quality of the Catalogue and its relevance to your organization, as well as any other related thoughts.

Response:

Since our repository was migrated to Figshare last year, we have revised almost all guidelines. We have addressed the comments we received last time. Apologies for the long delay in resubmitting the application, but this is mainly due to the major migration performed last year and consequent review of our procedures.

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments:

This looks great. Fully approve!