

About Tallinn University of Technology Data Repository

Tallinn University of Technology Data Repository (TalTechData) is a storage space for researchers to deposit data sets associated with their research.

The main goal of TalTechData is to gather all fields of research data and other research outputs and stands for responsible management of research data and encouraging **open science** and **FAIR** principles (**F**indable, **A**ccessible, **I**nteroperable and **R**eusable).

Through preserving open research data TalTechData enriches academic quality and collaboration, supports innovative developments and supports overall use of scientific materials.

FAIR research output management, including the management of research data, and other outputs such as software, algorithms, code, protocols, workflows, data management plans and project reports, makes it possible to maximize the impact of research.

Data **stored** in TalTechData can be open access, embargoed, restricted, or closed. Open access means that the dataset is freely accessible and downloadable. Other options, embargoed, restricted or closed, have some specific limitations. After the embargo period has ended, the dataset will be automatically made public. Closed or restricted datasets need access rights given by the dataset owner.

It is the responsibility of the submitters to guarantee the rights and permissions for uploading and sharing the data. Datasets should contain **documentation** (e.g., README.txt file) describing the purpose and terms of reuse and explaining how research data was collected and how to use this specific dataset. The maximum upload size of one dataset is 10GB. For larger datasets, please contact the repository administrator.

The service provider (TalTech) makes 1 TB of data capacity available at no charge per institute, for larger amount the price list of the HPC Center is followed. See price list <https://taltech.ee/en/itcollege/hpc-centre>

TalTechData supports all file formats, consistent with the goal of accommodating all research objects. We are working on guidelines and features that will help people deposit in preservation friendly formats.

Submitted datasets should have standard **metadata** to ensure that data is findable, accessible, interoperable, reusable, and effectively curable. Metadata is openly accessible, and TalTechData's metadata is reachable via [OAI-PMH](#) and is available for harvesting.

Metadata may be reused in any medium without prior permission for not-for-profit purposes. Metadata must not be reused in any medium for commercial purposes without formal permission.

Metadata is exported in several standard formats: JSON, CSL, DataCite JSON, DataCite XML, Dublin Core XML.

If there is a new and improved **version** of a dataset, it should be uploaded next to the previous version(s). It is recommended that each new version is reflected in the file name (e.g., dataset_name_v1, dataset_name_v2 etc.). When adding a new file, the database gives each version its own DOI.

When a dataset is submitted, a **license** must be included to determine what terms and conditions data can be (re)used.

The proprietary rights of the data are transferred to the university by a contract of employment (academic staff) or to another umbrella organization by a written document (the act of handing over the intellectual property).

The metadata in the repository can be used under the license Creative Commons Zero v1.0 Universal, the data and other uploaded files in the repository can be used according to the license specified in the record.

A third party whose data has been used for the grant results may set restrictions for the usage of the data. In this case, these limitations must be considered when licensing the data, i.e. the use of the data can only be licensed to the extent of the rights granted by the third party (i.e. the number of rights received by the university from third parties).

Once a dataset is published and given a DOI, it won't be removed from the repository because it may have been cited and reused. However, there can be exceptions and exceptional cases, such as data falsification and corruption. A dataset can be removed from the repository wholly, partly or with only some parts being restricted. Access can be restricted temporarily or permanently. DOI remains publicly available. Any copyright violations are entirely the responsibility of the datasets' authors. If TalTechData administrators receive proof of copyright violation, the relevant dataset will be removed immediately. The uploaded content should not consist of private, confidential, sensitive or restricted information that might harm others.

Dataset replicas are stored in TalTech HPC centre servers and are backed up on a weekly basis.

TalTechData provides access to all submitted datasets and their metadata.

Metadata of datasets are retained for the lifetime of the repository.

In closing the repository, all efforts will be made to find a replacement institution and/or repository.

The datasets will be retained for at least 10 years. After 10 years, the datasets will go through periodic review and the datasets deemed no longer necessary can be removed from repository. If longer than 10 year retention periods are needed, please contact TalTech HPC Centre.

TalTechData administrative staff has a right to remove or restrict datasets that violate the copyright or consist of private, confidential, sensitive or restricted information that might harm others.

TalTechData gives itself a right to change the terms of this policy according to the best practices of open research data.

TalTechData ensures that submitted datasets follow open access rules and FAIR principles.

FAIR Principles

FAIR Principles definition as referenced from *Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3:160018 doi: [10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18) (2016).*

To be Findable:

F1: (meta)data are assigned a globally unique and persistent identifier.

A DOI is issued to every published record on TalTechData

F2: data are described with rich metadata (defined by R1 below)

TalTechData metadata is compliant with [DataCite's Metadata Schema](#) minimum and recommended terms.

F3: metadata clearly and explicitly include the identifier of the data it describes

The DOI is a top-level and mandatory field in the metadata of each record.

F4: (meta)data are registered or indexed in a searchable resource

The metadata of each record is indexed and searchable directly in the TalTechData search engine immediately after publishing.

Metadata of each record is sent to DataCite servers during DOI registration and indexed there.

To be Accessible:

A1: (meta)data are retrievable by their identifier using a standardized communications protocol

Metadata for individual records and record collections are harvestable using the [OAI-PMH](#) protocol by the record identifier and the collection name.

Metadata is also retrievable through the public [REST API](#).

A1.1: the protocol is open, free, and universally implementable

A1.2: the protocol allows for an authentication and authorization procedure, where necessary

Metadata are publicly accessible and licensed under the public domain. No authorization is ever necessary to retrieve it.

A2: metadata are accessible, even when the data are no longer available

Metadata will be retained for the lifetime of the repository.

To be Interoperable:

I1: (meta)data use a formal, accessible, shared and broadly applicable language for knowledge representation.

TalTechData uses JSON Schema as the internal representation of metadata and offers export to several standard formats such as CSL, DataCite JSON, DataCite XML, Dublin Core XML

I2: (meta)data use vocabularies that follow FAIR principles

For specific terms, we refer to open, external vocabularies, e.g., license

Dataset replicas are stored in TalTech HPC centre servers and are backed up on a weekly basis.

I3: (meta)data include qualified references to other (meta)data.

To be Reusable:

R1: (meta)data are richly described with a plurality of accurate and relevant attributes.

Each record contains a minimum of DataCite's mandatory terms, with optionally additional DataCite, recommended terms.

R1.1: (meta)data are released with a clear and accessible data usage license

License is one of the recommended terms in TalTechData metadata.

Data downloaded by the users are subject to the license specified in the metadata by the uploader.

R1.2: (meta)data are associated with detailed provenance

All data and metadata uploaded are traceable to a registered TalTechData user.

Metadata can optionally describe the original authors of the published work.

R1.3: (Meta)data meet domain-relevant community standards

Through compliance with DataCite's Metadata Schema, TalTechData metadata meets one of the broadest standards available.

Technical

TalTechData repository is hosted by TalTech HPC Centre and uses Invenio RDM as the repository software.

Physically, the technical infrastructure is located in TalTech Campus in Tallinn. The data is stored on a Lenovo DSS-G storage system.

Copy of backups in encrypted form may be stored in tape library in HPC Centre in University of Tartu.

Dataset replicas are stored in TalTech HPC centre servers and are backed up on a weekly basis.

Security

The TalTechData repository is hosted in a server room. The physical access is restricted to authorized personnel only.

The HPC Centre follows the best security practices.

In case of any questions, please contact us: data@taltech.ee