



## **Implementation of the CoreTrustSeal**

The CoreTrustSeal board hereby confirms that the Trusted Digital repository FDAT complies with the guidelines version 2017-2019 set by the CoreTrustSeal Board.

The afore-mentioned repository has therefore acquired the CoreTrustSeal of 2016 on March 27, 2018.

The Trusted Digital repository is allowed to place an image of the CoreTrustSeal logo corresponding to the guidelines version date on their website. This image must link to this file which is hosted on the CoreTrustSeal website.

Yours sincerely,

The CoreTrustSeal Board

## Assessment Information

Guidelines Version: 2017-2019 | November 10, 2016  
Guidelines Information Booklet: [CTS Requirements 2017-2019](#)  
All Guidelines Documentation: [Documentation](#)

Repository: FDAT  
Seal Acquiry Date: Mar. 27, 2018

For the latest version of the awarded  
CoreTrustSeal for this repository: <https://www.coretrustseal.org/why-certification/certified-repositories/>

Previously Acquired Seals: None

This repository is owned by:

**Eberhard Karls Universität Tübingen**  
Main Building / Library  
Wilhelmstraße 32  
Raum 204a  
72074 Tübingen  
Germany

T +49 070712977848  
E [forschungsdaten@ikm.uni-tuebingen.de](mailto:forschungsdaten@ikm.uni-tuebingen.de)  
W  
<http://www.uni-tuebingen.de/einrichtungen/informations-kommunikations-und-medienzentrum-ikm.html>

# Assessment

## 0. Context

### Applicant Entry

*Self-assessment statement:*

**Repository Type:** Institutional repository

**Brief Description:** As a central and permanent infrastructure facility at the eScience-Center of the University of Tübingen (Germany), the research data repository FDAT offers various services as well as all the necessary technical equipment for the long-term archiving and reuse of research data to local researchers. Although this repository is primarily intended to support the local departments of the humanities and social sciences, it is fundamentally available to all scientific disciplines at the University of Tübingen, if no more research specific repository is available.

The development of FDAT started in late 2014 through the merger of 4 institutes of the University each providing different competencies and resources to the repository. As the most important facility, the Center for Information, Communication, and Media (IKM) consists of two units, the university library, and the computing center. Below the IKM is the eScience-Center with a focus on digital humanities, which provides human resources for the development and operation of the repository, as well as for the care of researchers.

As a general note, the FDAT repository itself is not a separate legal entity but part of the University of Tübingen, which is an institution governed by public law.

The web frontend of the repository is accessible via:

<https://fdat.escience.uni-tuebingen.de/portal/>.

The FDAT repository is registered at [re3data.org](http://re3data.org):

<http://www.re3data.org/repository/r3d100012296>.

**CoreTrustSeal Board**

W [www.coretrustseal.org](http://www.coretrustseal.org)

E [info@coretrustseal.org](mailto:info@coretrustseal.org)

In all essential aspects of the archiving workflow and the data structures used, we follow the conventions of an Open Archival Information System (OAIS).

**Repository's Designated Community:** The target community includes researchers from all fields of humanities and social sciences and cultural sciences, e.g. archaeology or ancient history. However, as FDAT is an institutional repository, the target group is expected to expand over time to address scientists of all main areas of research at the University of Tübingen.

The web frontend of the repository is also used in lectures in the field of digital humanities at the University of Tübingen, where young scientists get in touch with the infrastructure at an early stage of their education.

**Level of Curation Performed:** All the curation is based on technical aspects of research data at the moment. We do not systematically curate research data with regard to contents, mainly because of a lack of qualified staff in all research fields we support.

Different levels of technical curation occur depending on the type of research project. We support levels B and C. We check for completeness and consistency of metadata information, we help with the selection or creation of research-specific metadata schemas. In addition to scientific metadata content given by the researcher, we add technical metadata information (e.g. filename, file size, mime type, ...). We support and enforce file conversion to long-term preservation formats like PDF/A-1A(B) or XML. In the case of a data format conversion, the original unconverted records are also saved. Data format identification and validation are performed via standard tools like Droid and VeraPDF.

FDAT also offers tools for the generation of data management plans from interactive web forms. With such tools in charge, researchers are forced to deal with important general aspects of long-term archiving of their research data. The tool is accessible via:

[https://fdat.escience.uni-tuebingen.de/portal/#/service\\_downloads](https://fdat.escience.uni-tuebingen.de/portal/#/service_downloads).

**Outsource Partners:** The FDAT research repository is strongly supported by the university's computing center, offering geo-redundant and backed-up storage on demand as well as access to a virtualization infrastructure where the repository system is built up. System administrators at the computing center assure the operability of the repository and its web frontend.

In order to make archived digital data records citable, we get support from the Handle.Net Registry:

<https://www.handle.net/>

mediated by the University Library of Tübingen. Handle persistent identifiers exist for every data record. The FDAT research data repository has its own registry prefix 10900.1.

The authentication of users in FDAT is managed via the German National Research Net (DFN)

<https://www.aai.dfn.de/>

, where FDAT works as a so-called service provider to outsource the user authentication process via the help of the DFN using the SAML 2.0 protocol as implemented in shibboleth (<https://shibboleth.net/>) software package.

#### **Other Relevant Information:**

The repository currently considers the archiving and securing of large inventory data from the department of humanities and social sciences at the University of Tübingen as a priority. Due to a data collection which took place many years ago, nonstandard and customized processes to create the archival capability of these data records are necessary, requiring time and effort. As a first result, research data from archaeological excavations in the ancient city of Troy is currently available in the archive.

Moreover, about 65,000 documents from the area of Egyptology are currently being processed, which are expected to be recorded in the archive by the end of 2018. An estimated 50,000 more picture documents of archaeological findings in Troy will be processed for archival storage over a period of the next 2-3 years (2019-2020). At present, multimedia data records of about 1.3 TB in volume from the field of Indology is in preparation and expected to be archived in the first half of 2019. In addition, the FDAT repository has taken responsibility for the long-term preservation of research data for two large multidisciplinary research clusters (SFB 1253, SFB 1070) of the University of Tübingen. These projects will run for 4 to 10 more years providing research data records in a large measure to the repository over this period.

The web portal of the repository is online since 01.01.2017. Due to a currently existing temporary access restriction for most data records in the archive, no meaningful user statistics could be determined yet.

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

FDAT has only one dataset at time of review, but the explanation and plans given here seem reasonable and the statement is acceptable.

## 1. Mission/Scope

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

## Applicant Entry

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

The FDAT repository at the University of Tübingen pursues two main goals. Most importantly, it was developed to support local research projects from the fields of humanities and social sciences during the process of long-term archiving and reuse of their research data. Especially in those research fields, there is usually a lack of appropriate technical infrastructures and expertise available, therefore this problem needs to be addressed by a central institutional repository, which is FDAT.

The second goal of the repository is to strengthen the awareness of the changing conditions in modern sciences when it comes to scientific data and its management. The University of Tübingen has determined to make a public mission statement for the responsible and sustainable handling of research data. The full text can be found here:

<https://www.uni-tuebingen.de/forschung/service-fuer-forschende/leitlinien-zum-forschungsdatenmanagement.html>  
(only in German)

and here:

<https://fdat.escience.uni-tuebingen.de/portal/#/policies> -> see first panel Guidelines for research data management  
(in German and in English)

## Reviewer Entry

*Accept or send back to applicant for modification:*

Accept

*Comments:*

**CoreTrustSeal Board**

W [www.coretrustseal.org](http://www.coretrustseal.org)

E [info@coretrustseal.org](mailto:info@coretrustseal.org)

## 2. Licenses

### *Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

## Applicant Entry

### *Statement of Compliance:*

3. In progress: We are in the implementation phase.

### *Self-assessment statement:*

In the FDAT repository, every single research data record is strictly assigned a license determining the conditions of its reuse. The license is part of the metadata information of that data record and therefore persistently stored together with the record itself in the archive. Via the metadata, license information is distributed, e.g. via OAI-PMH protocol. Although the data producer (scientist) is basically free of choosing an appropriate license system for his data, the FDAT repository strongly encourages the use of Creative Commons licenses:

<https://fdat.escience.uni-tuebingen.de/portal/#/deposit> -> see tab: **Legal Aspects**

, as highly recommended by the German Research Foundation (DFG):

[http://www.dfg.de/foerderung/info\\_wissenschaft/2014/info\\_wissenschaft\\_14\\_68/index.html](http://www.dfg.de/foerderung/info_wissenschaft/2014/info_wissenschaft_14_68/index.html).

As a policy in FDAT, all metadata information given by the data producer is strictly published open access (license CC0 1.0 Universal).

The web frontend of FDAT emphasizes on license conditions for research data records at various stages, e.g. via automatic popups appearing at opening or downloading attempts on data records. On the web frontend, license information is always displayed as clickable links, pointing to the websites of, e.g. Creative Commons for further explanations.

While licenses specify the conditions of reproduction of data records in the archive, there is a higher level of access control implemented in FDAT which enables the data producer to temporarily make data records accessible only to a limited number of persons via the web portal. If access to a resource is restricted to the user, additional information is provided via clickable link 'No authorization' on the web portal. The conditions of access, including



the time span of restriction, the explicit data records in use and the authorized person groups are clearly laid down in a data contract.

In compliance with the Creative Commons, resources in this archive may be subject to a technical access restriction by the web portal, although they are subject to an open license in the aftermath provisions.

In FAT, the use of provided and disseminated research data is monitored on the level of collecting download rates. For data records under an access control, the user information is also collected on a download event. Non-compliant use of sensitive data by a registered user will be punished by a withdrawal of its access rights.

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

### 3. Continuity of access

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

#### Applicant Entry

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

The long-term availability and accessibility of research data in the FDAT archive is mainly due to the fact that this archive is maintained and supported by central and persistent institutions of the University of Tübingen, i.e. the computing center (ZDV), the university library (UB) as well as the eScience center directly hosting the repository. Therefore, no external funding is necessary for a continuous operation of the archive system.

Following the guidelines of good scientific practice from the German Research Foundation (DFG), we ensure archival storage for a standard time period of 10 years with the declared ambition, as stated in the standard data contract, to extend this to an indefinite time in the future, if required. By no means will data be removed from the repository after this time span without direct advice from the data owner. The operators of the repository including all institutions mentioned above are totally aware of the fact that research data, especially from the field of humanities and cultural sciences, usually need to be stored for an infinite period of time and agree to take all necessary technical and organizational steps to ensure this.

Therefore, data recovery measures that include format migration, normalization, or emulation will be performed. The steps to be taken in each case in order to permanently ensure the usability of a digital object is mainly determined by the nature of the file format. The repository commits itself to document the procedures of preserving data records and to render all subsequent modifications or extensions to the data required by the data provider transparently by means of a corresponding versioning strategy. To ensure long-term access and usability of the data, we also make recommendations to the data producers about necessary migrations into other formats as a result of newly established technical standards. The development of software components used to build up the archive system is also being pursued towards a necessary transition to more established or more stable components. The eScience-Center and its partners, the Computing Center (ZDV) and the University Library, are aware of the fact that constant monitoring of the sometimes rapidly changing technologies is indispensable and endeavors to take on this responsibility. But also the user's view and changes in the usage habits are taken into account and balanced.

By granting persistent identifiers (PID) by the repository and the local operation of the necessary infrastructure, the FDAT repository also guarantees the permanent citability of archived research data for the agreed duration of archiving.

There is a permanent exchange between the decision-makers about the present and future scope of FDAT, including a clear governance structure across all institutions involved.

ZDV UB -> <http://www.uni-tuebingen.de/einrichtungen/zentrum-fuer-datenverarbeitung/>

\\

IKM -> <https://www.uni-tuebingen.de/einrichtungen/informations-kommunikations-und-medienzentrum-ikm.html>

|

eScience center -> <http://www.escience.uni-tuebingen.de/>

|

FDAT repository

Long-term availability of all metadata information of data records in the FDAT archive is already ensured by its systematic distribution via standard protocols to other mainly more research field-specific data repositories. As an example, FDAT distributes all metadata information on archaeological data records to the IANUS data archive:

<https://www.ianus-fdz.de/>

, as a partner of the German archaeological institution.

To further enhance the sustainability of the FDAT repository, an infrastructure project called ORDP (Open Research Data Portal) is currently in process, funded by the Ministry of Art and Science in the federal state of Baden-Württemberg:

<https://fit.uni-tuebingen.de/Project/Details?id=4667>

The aim of this application is to merge and consolidate already existing approaches in research data management and archival at the University of Tübingen into a uniform infrastructural solution.

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## 4. Confidentiality/Ethics

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

### Applicant Entry

*Statement of Compliance:*

3. In progress: We are in the implementation phase.

*Self-assessment statement:*

Prior to archive storage in FDAT, data producers must sign a data contract stating that they have all the necessary rights to the submitted data and be themselves responsible for compliance with legal regulations especially concerning copyright and personal privacy. A template of this contract is available under <https://fdat.escience.uni-tuebingen.de/portal/#/deposit> -> see tab Legal Aspects and open/download contract template as a pdf document (only in German).

Further information for the data provider about data collection and distribution policies in FDAT including data protection is available via the policy statements:

<https://fdat.escience.uni-tuebingen.de/portal/#/policies>

The repository staff is actively advising data providers in this respect, if necessary. Via the information material presented on the FDAT website, the data provider will be informed about aspects of data protection including disclosure risk: <https://www.fdat.escience.uni-tuebingen.de/portal/#/deposit> -> see tab legal aspects

In general, data sets providing an obvious disclosure risk are not appropriate for archival storage in the FDAT repository since they can possibly never be published open access which is a central part of the policy of this repository. Therefore, the disclosure risk must be removed from submitted data sets prior to archival storage. As an interim solution, affected data sets can be access restricted to a very limited group of trusting persons for a certain time span.

The systematic identification of potential disclosure risk in large data sets is, however, a complex task requiring inside to specific mathematical methods. At the moment, the FDAT repository does not offer any expertise or support with regard to the application of such methods. However, the operators of the repository are actively searching for appropriate software and standard workflows to establish a procedure to support data providers in this respect.

The operators of the repository are aware of the EU's General Data Protection Regulation (GDPR) to handle privacy and data breaches and are preparing for the upcoming legal changes.

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

Acceptable as "in progress" phase.

## 5. Organizational infrastructure

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

### Applicant Entry

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

The sustainability of the FDAT repository is first of all given by the support of central and persistent institutions of the University of Tübingen, i.e. the computing center and the university library, which form as a merger the Information, Communication, and Media Center (IKM) at the university:

<https://www.uni-tuebingen.de/einrichtungen/informations-kommunikations-und-medienzentrum-ikm.html>

, as well as the eScience center of the university in the hierarchy below, which is actually hosting the FDAT repository:

<http://www.escience.uni-tuebingen.de/forschungsdatenarchiv-fdat.html>.

The eScience center is part of the so-called Core-Facilities of the University of Tübingen:

<https://www.uni-tuebingen.de/exzellenzinitiative/core-facilities.html>

, which denote permanent and central scientific service facilities. The computing center of the university is in charge of hosting and maintaining the hardware resources of the FDAT repository.

A current funding for further technical development of FDAT (3 years) is given by the Ministry of Art and Science in the federal state of Baden-Württemberg:

**CoreTrustSeal Board**

W [www.coretrustseal.org](http://www.coretrustseal.org)

E [info@coretrustseal.org](mailto:info@coretrustseal.org)

<https://fit.uni-tuebingen.de/Project/Details?id=4667>

Staff resources, both permanent and non-permanent, include IT employees (4 FTE / 50% temporary), system administrators (2 FTE / permanent), legal staff (0.5 FTE / permanent), as well as scientists from the field of humanities (2 FTE / 50% temporary). They are provided by either the e-Science center:

<http://www.escience.uni-tuebingen.de/mitarbeiter.html>

, or by the computing center as well as by the university library. The staff expertise is appropriate to the mission. However, a higher ratio of permanent staff is strongly sought. The staff in charge is regularly attending conferences and workshops referring to the topics of research data management and e-science in general.

All necessary infrastructural resources of the repository including technical facilities and system administrators from the computing center are permanent. Legal staff and software developers from the university library supporting FDAT are permanent too. A part of the staff from the eScience-Center, including software developers and scientists from the field of humanities and natural sciences, work project-based. There is a clear attempt to change their position from periodic renewal to permanent.

## Reviewer Entry

*Accept or send back to applicant for modification:*

Accept

*Comments:*



## 6. Expert guidance

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

### Applicant Entry

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

The eScience-Center hosting the FDAT repository closely works together with the university library which, throughout regular meetings, give advise in terms of, e.g. metadata standards, publishing policies and standards for sharing metadata sets with other institutions.

The computing center of the university permanently provides the eScience center with valuable information on the current hardware infrastructure status. In regular meetings, future hardware requirements and procurements for the FDAT repository are discussed and planned. This includes storage and virtualization technologies as well as security technologies.

The German Research Network (DFN -> <https://www.aai.dfn.de/>) acts as an external advisor and infrastructure provider for data security with its federated identity management framework. The FDAT repository receives regular feedback from the DFN in terms of developments in the framework and is officially registered as a user identity service provider at the DFN:

<https://www.aai.dfn.de/fileadmin/metadata/dfn-aai-metadata.xml> ->  
entityID="https://fdat.escience.uni-tuebingen.de"

In-house software developers from the eScience center permanently monitor new developments in archive software solutions and peripheral technologies to keep the overall system up to date and therefore efficient and capable.

The FDAT repository primarily holds data records from the field of humanities and social sciences. For these scientific fields, disciplinary experts exist in the eScience staff providing scientific advisory if required.

The repository staff together with employees of the university library collect feedback from designated communities directly via the web portal:

<https://www.fdat.escience.uni-tuebingen.de/portal/#/contact>

, or via the repositories email address: **forschungsdaten@ikm.uni-tuebingen.de**

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## 7. Data integrity and authenticity

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

### Applicant Entry

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

The core element of the FDAT repository is the open source repository software Fedora Commons 4.x:

<http://fedorarepository.org/>

Fedora automatically adds checksums to data records ingested into the repository. Via Fedoras build-in fixity-check functionality, data provenance and related issues are regularly monitored via automated fixity checks:

<https://wiki.duraspace.org/display/FEDORA40/RESTful+HTTP+API+-+Fixity>

Several fields of technical metadata information are created on ingest, ie.g. about the checksum and latest updates of a record. The FDAT repository utilizes several metadata schemes to capture all relevant general, technical and research specific metadata of a digital object. On the FDAT web portal there is comprehensive further information available:

<https://fdat.escience.uni-tuebingen.de/portal/#/deposit> -> see tab **Metadata/Vocabularies**

Fedora 4 natively supports versioning of digital objects:

<https://wiki.duraspace.org/display/FEDORA473/Versioning>

Whenever changes have to be made to a data record or to its related metadata set (both are stored in Fedora), a new version of this record is created by default including a new persistent identifier. Fedora keeps track of the membership of versions to the original record, and the web portal displays by default all available versions of a data record. However, Fedora does not yet offer functionalities to quickly evaluate and display the difference between versions of the same digital object.

All changes applied to existing data and metadata in the repository are logged via repository internal versioning as well as via log protocols, automatically generated and also archived by an in-house developed ingest software. All write operations on the repository are carried out via the FDA manager software, always producing log files which are archived as well.

-----

The FDAT repository provides optional metadata fields to store links, via persistent identifiers, to other data sets in the repository. Persistent links from or to other repositories can be used as well.

-----

The recording of data and metadata information of a digital object is performed via the open source software docuteam packer:

<https://wiki.docuteam.ch/doku.php?id=docuteam:packer>

This software provides capabilities in terms of checking for completeness and correct usage of data types for metadata and is also logging various events like creation, update or deletion of records. The data provider uses the docuteam packer to transform his research data into a standardized archive package format (OAIS: Submission Information Package), realized via the structural metadata standard METS:

<http://www.loc.gov/standards/mets/mets-profiles.html>

The METS data structure of each research project is, after evaluation for completeness and integrity, also stored in the archive.

-----

Data providers are always local scientists at the university to which the repository staff always talks personally in order to understand their specific requirements in terms data processing, archival storage, and reuse. This personal contact also ensures the authenticity and trustworthiness of involved persons and provided data sets.

-----

The archive package is ingested via an in-house developed software which is done by the repository staff only and not by the data providers themselves, which have strict read-only access to the repository. In other words, changes to the repository state can only be done by a very limited number of the repository staff. The integrity of data and metadata information is checked prior to the final ingest via the FDA manager. Furthermore, after a complete ingest, the new data records in the archive remain hidden from the public until the data provider has successfully inspected and approved its records.

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

FDAT needs to seek to ensure that there will continue to be enough staff for the personal contacts with the data providers.

## 8. Appraisal

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

### Applicant Entry

*Statement of Compliance:*

3. In progress: We are in the implementation phase.

*Self-assessment statement:*

The FDAT collection policy defines the conditions for accepting data into the repository:

<https://fdat.escience.uni-tuebingen.de/portal/#/policies> -> see panel **Data Collection and Transfer Policy**

For the research areas of humanities and social sciences, the repository has a scientific expert council at the eScience-Center to support the selection of data to be archived. In its role as an institutional repository, FDAT is also responsible for research data from other disciplines that do not have access to subject-specific repositories. In this case, scientific advice from experts is requested on demand from local research groups at the university.

The data producer is by all means responsible for the understandability of provided data and metadata information:

<https://fdat.escience.uni-tuebingen.de/portal/#/policies> -> see panel **Data Collection and Transfer Policy**

However, the repository staff does monitor completeness of data and metadata information. This is done via the use of data collection tools specifically developed for the needs of modern digital archives:

<https://wiki.docuteam.ch/doku.php?id=docuteam:packer>

To increase the understandability of metadata information provided, the use of controlled vocabularies is required for several mandatory metadata fields in the repository:

<https://fdat.escience.uni-tuebingen.de/portal/#/deposit> -> see tab **Metadata/Vocabularies**

The information provided here helps the data provider to understand the meaning of the metadata fields based on textual descriptions.

The FDAT repository requires a set of mandatory metadata fields to be filled out which is also automatically checked for by specific software tools. The decision whether this minimum field set required denotes a sufficient description of the respective data is up to the data provider. FDAT provides metadata fields for general use only. Also supported are research specific metadata fields which can be defined additionally. This allows data providers to exhaustively describe their scientific data contentwise.

There is currently no procedure implemented to figure out whether provided metadata is, with regard to its content, insufficient for long-term preservation neither have related criteria been defined yet. Generally, we strongly advise the data provider not to include any variable personal data, like addresses, telephone numbers and so forth in the metadata. From the technical point of view we consider metadata information in the FDAT repository to be suitable for long-term storage since it is stored and disseminated in well established and long-lasting data formats only, i.e. EAD, Dublin Core and Marc21.

A list of recommended data formats is given here:

<https://fdat.escience.uni-tuebingen.de/portal/#/deposit> -> see tab **Data Formats/Validation**

We make use of in-house developed tools to ensure only data of preferred formats are ingested into the repository. Generally, file formats of all data files provided will be determined and validated via open source tools like Droid:

<http://www.nationalarchives.gov.uk/information-management/manage-information/preserving-digital-records/droid/>

See also explanations at the end of the page of

<https://fdat.escience.uni-tuebingen.de/portal/#/deposit> -> see tab **Data Formats/Validation**

The data provider is instructed by the repository staff about data formats suitable for archive storage. The format conversion will be done by either the data provider or the repository staff. If no suitable and lossless format conversion is possible for the data records provided, a non-proprietary open source software able to properly open the file content is being searched for and archived together with those files, if in line with the software license.

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*



## 9. Documented storage procedures

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

### Applicant Entry

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

The basic steps of the overall procedure of data acquisition up to ingest and access are coarsely described on the web portal of the FDAT repository:

<https://fdat.escience.uni-tuebingen.de/portal/#/deposit> -> see tab Workflow **Archive Storage**

Here, we make use of OAIS terms extensively to put emphasis on the fact, that we try to follow the OAIS framework in terms of functional units, procedures, and data structures. The workflow describes the basic steps of collaboration between the data provider and the repository operators in terms of data collection, augmentation with metadata, the ingest process and so forth.

The first relevant process is the data and metadata collection done by the data provider. In the default case, he uses the open source software docuteam packer:

<https://www.docuteam.ch/angebot/archivinformatik/software/>

, adapted and configured by the repository staff for his needs. The packer software comprehensively documents all write processes on the data via the PREMIS metadata framework. As a result, a submission information package (SIP) with PREMIS metadata in standard METS format is generated which is also ingested into the repository, therefore all information about the data collection processes are kept permanently.

In terms of data handling and management, a PDF -> PDF-A/1a(b) conversion is a common process. Afterward, the VeraPDF tool is used to validate the quality of conversion. This tool offers comprehensive documentation of the conversion process (PDF/A compliance) in XML format, which is also permanently stored in the repository.

<http://verapdf.org/home/>

Furthermore, an in-house developed ingest software logs all processes of the ingest procedure of data records into the repository, including also the generation and storage of persistent identifiers for data records and the copy of metadata into search engines. Also, this log file is permanently stored in the repository in order to be able to trace all processes at a later point in time.

Since all write operations on the repository are strictly performed by the repository staff only, security requirements are small compared to systems where the data providers are allowed to deposit data on their own. We routinely check for malware of any kind in research data records handed over to us via the clam software before ingest:

<https://www.clamav.net/>.

Information about data preservation policies are given here:

<https://fdat.escience.uni-tuebingen.de/portal/#/policies> -> see panel **Preservation Policy and Citability**

The Repository takes responsibility for recovery measures of the data which can include format migration, normalization, or emulation. All procedures of data preservation on data records will be documented by the operators of FDAT. The repository operators constantly monitor changes in technical standards such as suitable archive data formats, archive software components as well as workflows of preservation to be able to adapt to more accepted standards.

The strategy of the local computing center established for the FDAT repository is for a fully mirrored recovery site (data center I to data center II). The strategy includes maintenance of a fully mirrored duplicate site, which will enable manual switching between the life system and the backup site.

The activities of data recovery for the repository are taken over by the local computing center. The recovery strategy is built on a fully mirrored duplicate site, which will enable manual switching between the life system and

the backup site.

The risk management for FDAT is taken over by the local computing center. The potential disruptive threats which can occur at any time and affect the normal business process are listed below. Each potential environmental disaster or emergency situation has been examined by the computing center. The focus here is on the level of business disruption which could arise from each type of disaster.

Potential disaster	Probability Rating	Impact Rating	Consequences and Actions
--------------------	--------------------	---------------	--------------------------

-----	-----	-----	-----
-------	-------	-------	-------

Flood	4	4	All critical equipment is located on 1st or 3rd floor
-------	---	---	---

Fire detectors on all floors	3	4	Fire suppression systems installed, fire and smoke
---------------------------------	---	---	--

Gale	5	5	-
------	---	---	---

Electric power failure Monitoring: 24/7	3	4	Redundant UPS systems with standby generators.
--	---	---	--

Communication Network loss	4	4	Redundant Connection to DFN via BelWue
----------------------------	---	---	--

Sabotage	5	-	-
----------	---	---	---

Terrorism	5	-	-
-----------	---	---	---

Archival copies of the FDAT repository include timestamps in order to trace the chronological sequence and find the most recent one to be restored in case of an accident. Archival copies of the file system are strictly write protected to circumvent any consistency issues over time.

At the moment, there is no established procedure available for a systematic monitoring and handling of deteriorating digital data records.

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## 10. Preservation plan

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

### Applicant Entry

*Statement of Compliance:*

3. In progress: We are in the implementation phase.

*Self-assessment statement:*

Information about data preservation strategies for the FDAT repository are given here:

<https://fdat.escience.uni-tuebingen.de/portal/#/deposit> -> see tab **Data Preservation Strategies**

The preservation strategy covers 5 key aspects of a repository system, that is storage and geographic location, file fixity and data integrity, information security, metadata and file formats. For each of those areas, four different levels of preservation are defined. FDAT is strived to reach the highest level in all areas to provide the best possible preservation for the data entrusted to it. The following list includes current actions performed for data preservation.

storage and geographic location:

- Two complete copies that are not collocated
- For data on heterogeneous media (optical discs, hard drives, etc.) get the content off the medium and into your storage system
- Document your storage system(s) and storage media and what you need to use them

- At least three complete copies
- At least one copy in a geographic location with a different disaster threat
- Obsolescence monitoring process for your storage system(s) and media

file fixity and data integrity:

- Check file fixity on ingesting if it has been provided with the content
- Create fixity info if it wasn't provided with the content
- Check fixity on all ingests
- Use write-blockers when working with original media
- Virus-check high-risk content
- Ability to detect corrupt data

- Virus-check all content
- Ability to replace/repair corrupted data
- Check fixity of all response to specific events or activities
- Ability to replace/repair corrupted data
- Ensure no one person has write access to all copies

information security:

- Identify who has read, write, move and delete authorization to individual files
- Restrict who has those authorizations to individual files
- Document access restrictions for content
- Maintain logs of who performed what actions on files, including deletions and preservation actions

metadata:

- Inventory of content and its storage location
- Ensure backup and non-collocation of inventory
- Store administrative metadata
- Store transformative metadata and log events
- Store standard technical and descriptive metadata
- Store standard preservation metadata

file formats:

- When you can give input into the creation of digital files encourage the use of a limited set of known open formats and codecs



- Inventory of file formats in use
- Monitor file format obsolescence issues
- Perform format migrations, emulation and similar activities as needed

Following the definitions of the National Digital Stewardship Alliance (NDSA), preservation levels are defined as follows in the FDAT repository:

<http://www.digitalpreservation.gov:8081/ndsa/activities/levels.html>

We refer to this scheme and give an estimation of the current level (1-4) of implementation for the FDAT repository.

The current contract version between a depositor and the FDAT repository contains general statements referring to preservation activities:

1. "Technical, non-content processing of the data stock, in particular conversions to archival data formats"
2. "Migration of the data stock into technical successor systems or other archive systems"

The data contract clarifies the transfer of custody and responsibility handover.

Via a standard data contract, the repository requests authorization from the depositor to copy, transform and store the items, as well as provide access to them. There is further information for depositors on the web portal of the FDAT archive:

<https://fdat.escience.uni-tuebingen.de/portal/#/policies> -> see panel **Usage and Distribution Policies**

All information relevant to data preservation are documented either in the data contract template:

<https://fdat.escience.uni-tuebingen.de/portal/#/deposit> -> see tab **Legal Aspects**

or in the statement for data preservation strategies:

<https://fdat.escience.uni-tuebingen.de/portal/#/deposit> -> see tab **Data Preservation Strategies**

Such measures are not fully established yet. The repository operators are actively working on a comprehensive documentation system. For some areas of preservation, documentation is already available.

For metadata, the results of preservation can partially be viewed via the repository portal. An inventory of all metadata fields, their meaning and usage are stored and preserved.

For the preservation area of file formats, an inventory of accepted file formats is available online, the format conversion is actively done by the repository operators and can be viewed from the repository portal.

For the field of information security, the repository backend automatically logs of who performed what actions on files, including deletions and preservation actions. Those logs are permanently stored in the repository database. Furthermore, document access restrictions for content are stored in a standard data contract as well as in the repository database.

For the area of file fixity and data integrity, the repository backend automatically performs fixity checks on all data records in the repository on regular bases and stores the results permanently in the repository database.

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## 11. Data quality

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

### Applicant Entry

*Statement of Compliance:*

3. In progress: We are in the implementation phase.

*Self-assessment statement:*

Repository employees work closely with the data providers to assess the quality of the data and metadata that is passed on.

Since this repository is open to a wide range of research disciplines, there is generally a shortage of qualified personnel at the repository capable of curating the data and metadata from a content perspective.

However, we strongly advise and support the data provider in selecting a suitable, research specific metadata schema, possibly linked to a controlled vocabulary, so that the data can be understood and interpreted by the community of interest.

This information is then checked by the repository for completeness and technical integrity. Furthermore, in the FDAT repository, metadata information is mandatory down to the bitstream level (single documents), i.e a high granularity of data description exists in the repository.

If metadata is backed by a controlled vocabulary, as provided by the depositor, configured dropdown lists in data collection tools, as provided by the repository, ensure that metadata fields can only be filled with appropriate content, understandable by the community.

Momentarily, there is no systematic rating or comment system available for archived data provided by the repository infrastructure. However, the FDAT web portal includes a general messaging functionality where visitors can give any kind of feedback related to the web portal or single data records.

Every digital object in the repository can contain specific metadata information pointing to other resources, e.g. related works inside and outside the repository.

### Reviewer Entry

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## 12. Workflows

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

### Applicant Entry

*Statement of Compliance:*

3. In progress: We are in the implementation phase.

*Self-assessment statement:*

The overall workflow of archive storage and reuse, following the OAIS system (ISO 14721:2012) is documented and communicated to the depositors on the web portal of the FDAT repository:

<https://fdat.escience.uni-tuebingen.de/portal/#/deposit> -> see tab **Workflow Archive Storage**

The handling of data in the pre-ingest phase, in particular regarding data formats, is communicated to the depositors via the web portal:

<https://fdat.escience.uni-tuebingen.de/portal/#/deposit> -> see tab **Data Formats/Validation**

Provided technical metadata information helps the users to handle data records from a technical perspective, while license information on each data record informs users about the appropriate data handling from a legal perspective.

Upon the process of archiving research data, data providers will be informed about handling security-relevant data, and how to achieve anonymous or pseudonymized data appropriate for archive storage:

<https://fdat.escience.uni-tuebingen.de/portal/#/deposit> -> see tab **Legal Aspects**

The data provider is informed that the repository is only accepting such anonymous or pseudonymized data records for storage. However, there is no specific workflow provided to handle security-relevant data via the FDAT repository.

No contentwise data curation is carried out by the repository staff, appraisal and selection of data are only done by the data which is part of the FDAT official policy:

<https://fdat.escience.uni-tuebingen.de/portal/#/policies> -> see panel **Data Collection and Transfer Policy**

No data records which do not fall within the mission/collection profile of the repository will be stored. In this case, the data provider will be asked to search for a more research specific repository.

Any kind of digital data records will be managed by the FDAT repository. The specific type of data determines the possibilities and requirements of data format conversion to archive appropriate formats and has, therefore, an impact on the pre-ingest part of the archive workflow.

A change management for the archival workflow is based on the documentation of any practical problems arising from the archival process for research data projects. On regular bases, the repository operators meet to discuss the adaption or extension of the currently implemented workflow to meet practical requirements.

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

### 13. Data discovery and identification

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

#### Applicant Entry

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

The FDAT repository contains all metadata information in the open source search engine Apache Lucene/Solr. Using the web frontend of the repository, free text search, and facet-based search can be carried out to find the resources of interest in the repository.

FDAT provides all necessary information on metadata fields, related schema and vocabularies in use via the repository web front end:

<https://fdat.escience.uni-tuebingen.de/portal/#/deposit> -> see tab **Data Metadata/Vocabularies**

FDAT provides comprehensive access to metadata information in the repository via the REST based OAI-PMH protocol:

<http://fdat.escience.uni-tuebingen.de/portal/rest/oai>

The repository is furthermore registered and validated by the Open Archives Initiative as an OAI-PMH data provider:

<http://www.openarchives.org/Register/BrowseSites?viewRecord=http://fdat.escience.uni-tuebingen.de/portal/rest/oai>

FDAT is listed in the generic repository registry re3Data:

<http://www.re3data.org/repository/r3d100012296>

In FDAT, there is a metadata field citation, where the data provider can determine how every single data record should be cited. From this field, a complete citation entity is generated. The repository follows a proposed standard as described in the following article:

<http://www.dlib.org/dlib/march07/altman/03altman.html>

More detailed information about data citations is provided via the repository web portal:

<https://fdat.escience.uni-tuebingen.de/portal/#/deposit> -> see tab Data Citation

In order to make archived digital data records citable, FDAT gets support from the Handle.Net registry:

<https://www.handle.net/>

, mediated by the university library of Tübingen. Via Handle persistent identifiers are automatically created for every data record. The FDAT repository has its own registry prefix (10900.1).

## Reviewer Entry

*Accept or send back to applicant for modification:*

Accept

*Comments:*



## 14. Data reuse

### *Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

## Applicant Entry

### *Statement of Compliance:*

3. In progress: We are in the implementation phase.

### *Self-assessment statement:*

Metadata information from the depositor can be provided in a free-form pair of keys and values. The open source data collection tool in use, docuteam packer, maps the provided information into the Encoded Archival Description (EAD) standard. In-house developed software tools do further, potentially lossy, mappings into Dublin Core (DC) and Marc21 which are used as standard dissemination formats via the OAI-PMH protocol.

It is a declared policy of the FDAT repository to store and provide data in formats which are considered to be long-term readable. A list of these recommended file formats can be found on the repository website:

<https://fdat.escience.uni-tuebingen.de/portal/#/deposit> -> see tab Data Formats/Validation

Therefore, data is not stored in the formats preferred by a designated community but in the formats recommended for long-term archive storage.

Our repository staff constantly monitors developments in the evolution of data formats and standards relevant to the repository. This effort is part of the FDAT preservation policy:

<https://fdat.escience.uni-tuebingen.de/portal/#/policies> -> Preservation Policy and Citability

A general data migration plan is currently under development and includes the following preliminary steps:

1. Permanent monitoring for changes in technical standards and future support of current data formats

2. Evaluation of possible information loss and other risks towards migration to the new data format
3. Discussion with researchers/data providers about current acceptance of the new data format in the scientific community
4. Search for software tools appropriate for the concrete format migration
6. Execution of format migration and subsequent data verification by researchers and repository operators
7. Systematic data update in the repository system

In terms of a timetable, the first point is already in permanent execution. All subsequent points, with the exception of the last one, are difficult to estimate in terms of their time and effort so far. However, the staff responsibilities have been clarified for this purpose. The last point denotes the comprehensive migration of all relevant documents in the repository to the new format. The preparation of a standardized workflow with test runs for this purpose is planned for the end of 2018.

Understandability of the data is ensured by comprehensive metadata information FDAT provides down to the bitstream / single record level. We provide descriptive metadata for understandability in terms of scientific content, augmented by research specific custom metadata. For the latter, the repository staff intensively works together with the data provider to find a suitable set of research specific metadata schema for an appropriate description of the respective research data. Automatically generated technical metadata helps to understand the specifics of the respective electronic document.

## Reviewer Entry

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## 15. Technical infrastructure

### *Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

## Applicant Entry

### *Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

### *Self-assessment statement:*

The FDAT repository follows the OAIS (ISO 14721:2012) reference model for all relevant processes of archive storage. For further information refer to the FDAT web portal:

<https://fdat.escience.uni-tuebingen.de/portal/#/deposit> -> see tab **Workflow Archive Storage**

The XML Schema standard of the W3C is used for the declaration of data types in the repository.

The FDAT repository uses open source tools for data collection generating a defined data structure (OAIS: Submission Information Package) following the METS standard.

<https://wiki.docuteam.ch/doku.php?id=docuteam:packer>

The core element of FDAT is the open source repository software Fedora Commons 4.x:

<http://fedorarepository.org/>

Fedora holds the data objects (OAIS: Archival Storage) and also the metadata information. The FDAT repository also contains all the metadata information in the open source search engine Apache Lucene/Solr (OAIS: Descriptive Info). The data transfer from the provider to the repository (OAIS: Ingest/Archival Information Package) is done via an in-house developed software. The OAIS: Data Management is realized via a database system including all user accounts, access control lists (ACL) for the data records and so forth. The OAIS: Access

component is realized via an in-house developed web-portal.

FDAT is based on the Core-Facilities infrastructure hosted at the computing center of the University of Tübingen:

<https://www.uni-tuebingen.de/exzellenzinitiative/forschung/core-facilities.html>

The hardware infrastructure is regularly expanded and modernized to meet future technical requirements of FDAT, e.g. with respect to storage and compute capacity. Clear personal responsibilities at the computing center were defined for this purpose. Joint decisions are made in meetings on a regular base together with staff from the eScience-Center and the university library.

A current infrastructural development of the FDAT repository is the Open Research Data Portal (ORDP) project, funded by the federal state of Baden-Württemberg in Germany:

1.

<http://www.uni-tuebingen.de/einrichtungen/informations-kommunikations-und-medienzentrum-ikm/escience-center/landesprojekte-zum-1>

2. <https://fit.uni-tuebingen.de/Project/Details?id=4667>

The two in-house developed software components (web-portal and ingest software) are actively maintained via a cloud-based code repository systems (<https://bitbucket.org>). We are extensively working on a system documentation which will be available online as a doku-wiki in early 2019.

The following community-supported software is currently in use:

1. <https://wiki.docuteam.ch/doku.php?id=docuteam:packer> -> data & metadata collection tool

2. <http://fedorarepository.org/> -> data & metadata repository

3. <http://lucene.apache.org/solr/> -> metadata search engine

4. <https://www.handle.net/> -> persistent identifiers for data records

5. <https://www.postgresql.org/> -> various data

6. <https://www.shibboleth.net/> -> distributed user authentication system, actively supported by the German research network organization (<https://www.aai.dfn.de/>)

We did not yet consider real-time data streams and related necessities.

## Reviewer Entry

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## 16. Security

### *Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

## Applicant Entry

### *Statement of Compliance:*

3. In progress: We are in the implementation phase.

### *Self-assessment statement:*

The technical infrastructure of the repository is provided and maintained particularly by the Universities computing center. The servers and storage systems that build the repository both rely on virtualization techniques for scalability, ease of hardware migration and long-term availability of services.

The protection of all components of the repository is ensured by the following means:

The computer center provides two geographically (2km distance) separated server buildings that are secured against unauthorized access by means of an electronic access control system.

1.

Each server room provides monitored power and cooling and is protected against fire and power loss, utilizing a UPS.

2.

Servers and storage for the repository are located in both server rooms and setup to provide redundant services and redundant data storage.

3.

Data is protected by replication and frequent backup.

4.

Replication and backup ensure fast and complete recovery of repository data and services (server operating system and software) in case of any data loss or disaster.

5.

Maintenance and administrative access to the repository are strictly limited to a small number of system engineers.

6.

Migration of services (servers) and data (storage systems) is feasible without service disruption.

7.

Servers and storage systems are continuously monitored.

8.

IT security setup follows the recommendations of BSI-Grundschutz and are supervised by the IT security officer (<https://www.uni-tuebingen.de/en/facilities/zentrum-fuer-datenverarbeitung/aktuelles/it-sicherheit.html>)

9.

The Universities computing center infrastructure thus guarantees the physical safety of the digital archive.

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

Acceptable as "in progress" phase.



## 17. Comments/feedback

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

### Applicant Entry

*Statement of Compliance:*

0. N/A: Not Applicable.

*Self-assessment statement:*

-

### Reviewer Entry

*Accept or send back to applicant for modification:*

Accept

*Comments:*

FDAT is a new repository and many developments are ongoing. There is an expectation of progress if/when the certification is renewed. FDAT will need to seek to ensure that there will continue to be enough staff as their data collection and activities grow.