TRR 170 Research Data Management Plan

Administrative details of the project

Project name	Late accretion onto terrestrial planets
Speaker	Prof. Harry Becker (hbecker@zedat.fu-berlin.de)
Contact data management	Elfrun Lehmann (elfrun.lehmann@fu-berlin.de)
Project runtime	2016 – 2019, proposed 2020-2023
Funded by	DFG
Version and date of DMP	August 1 (V2)

Data and formats

The data collected by the TRR 170 represents a significant effort in research with emphasis on providing new data and resources to the planetary science community. The strong interdisciplinary alignment of the TRR 170 research network comprises heterogeneous data and different data formats. Accordingly, research data generated in TRR 170 subprojects are diverse, reflecting the range of methods used, including laboratory and other instrumental data on planetary samples, remote sensing data, geological maps and model simulations.

After collection, quantitative data are converted to plain-text (ASCII), EXCEL/CVS or imaging (JPEG 2000, png) files (with lossless compression) or uncompressed TIFF files depending what format is the most appropriate for the specific data set. These formats are fully supported by our **TRR170-DB** that operates on *Dataverse*. *Dataverse* is a public repository that performs archival format migration, metadata extraction, and validity checks. Documentation is deposited in PDF or plain-text formats, to ensure long-term accessibility. All figures, images and tables are separately deposited from the documentation in pdf format.

Access and Sharing

All TRR 170 data collected or generated in the TRR 170 projects will be deposited in the **TRR170-DB** hosted by a virtual machine at the central computing center ZEDAT of Freie Universität Berlin. Currently, **TRR170-DB** is accessed directly via the *Dataverse* software. The **TRR170-DB** facilitates data access by providing descriptive searches and topical browsing. However, we are in the process of building a web-based Planetary Data Portal that will be advanced to provide enhanced user-friendly functionalities to find and retrieve data in the TRR 170-DB. The future functionalities will provide on-line analysis, variable/question-level search, data extraction and re-formatting, and other enhanced access capabilities.

Data are continuously deposited over the course of a funding period. Data deposit will be completed up to 90 days after the expiration of the funding. Data may be embargoed until the publication of research based on the data or until 2 years after the expiration of the funding, whichever is sooner. Users will be required to agree to terms that prohibit unlawful uses and intentional violations of privacy and require attribution (see TRR 170 data policy). Use of the data will be otherwise unrestricted and free of charge.

Submitting and Retrieving Data

We have established a standardized approach to submit research data to **TRR170-DB** to make it available online for the TRR 170 community. Through the **TRR170-DB** interface, the user is guided to a metadata template that provides a choice of parameters that users can select from to best describe their research data-related information. In such way, we ensure that research data will be found via defined ontology vocabulary.

The process of submitting to and/or retrieving research data from **TRR 170-DB** requires the user to

(1) comply with the TRR 170 <u>policy</u> on research data. This policy informs users on requested quality, storage, accessibility, security and long-term sustainability of TRR 170 research data;

(2) fill in a <u>request form</u> if the user wishes to obtain research data generated in TRR 170;

(3) read and acknowledge a contract form for <u>liability</u> reasons; and

(4) if the researcher wishes to provide feedback on the retrieval process s/he fills in a <u>comment</u> form. In this form the researcher comments on how user-friendly the retrieval process is and whether information in the data is sufficient for trustworthy use, such as the methodologies used to generate the data, etc. In this form, data users do not evaluate their downloaded data. Data are evaluated in a published peer-reviewed scientific paper. We collect and store this paper in TRR170-DB as part of the data information available to all users.

All forms are available at https://www.trr170-lateaccretion.de/trr170-db-information.

Documentation, Metadata and Bibliographic Information

We will create documentation detailing the sources of all data in sufficient detail to enable other researchers to replicate them from original sources. Descriptive metadata for each dataset includes a title, author, description, descriptive keywords, and file descriptions. We include bibliographic information for each dataset based on that data.

For data entries, **TRR170-DB** provides standard-compliant metadata ('templating') for consistency of information across datasets. In the last quarter of 2019, the **TRR170-DB** repository will be associated with *DataCite* to automatically generate persistent identifiers for datasets. **TRR170-DB** facilitates open distribution of metadata with a variety of standard formats (Data Cite, DDI, Dublin Core, VO Resource) and protocols (OAI-PMH, SWORD).

TRR170-DB metadata also ensure that metadata in the database can be mapped easily to domain-specific metadata schemas of other repositories and be exported into JSON (Java Script Object Notation) format to exchange data for preservation and interoperability. **TRR170-DB** supports metadata formats for citation and geospatial metadata (DataCite, DDI, Codebook, Dublin Core) and also for astronomy and astrophysics metadata (VO) that can be mapped/exported to the International Virtual Observatory Alliance (IVOA). However, **TRR170-DB** also offers the option to adjust existing metadata templates to any other metadata standard.

Storage, backup, replication

We host the RDM system on a virtual machine (VM) at FU Berlin's central computing center ZEDAT. The VM emulates the most current Debian Linux 64 Bit version. Fast data access and handling are ensured by 3 GB RAM and 2 processors. **TRR170-DB** uses two hard drives. The root partition (hard drive 1, 64GB) stores all metadata information. The second partition uses a 2 TB hard drive to store research data and all related information. Data archiving capacities can easily be extended if necessary.

For all published papers we deposit replication datasets. Since **TRR170-DB** is run on *Dataverse* we use an integrated functionality that allows that this type of data is easy to find for researchers who want to reuse and to verify a study without having to contact the study's author.

The *Dataverse* software provides automatic version (revision) control over all deposited materials and no versions of deposited material are destroyed except where such destruction is legally required. All systems providing on-line storage for the *Dataverse* are contained in a physically secured facility of ZEDAT that is continually monitored. System backups are made on a daily basis.

Security

The **TRR170-DB** complies with Freie Universität Berlin-ZEDAT requirements for good computer use practices. ZEDAT has developed extensive technical and administrative procedures to ensure consistent and systematic information security. "Good practice" requirements include system security and operational requirements and regular auditing and review. The full Freie Universität Berlin-ZEDAT security policy can be found at https://www.fu-berlin.de/sites/it-sicherheit/downloads/IT-Sicherheitsrichtlinie.pdf.

Budget

The cost of preparing data and documentation will be covered by the project and is already reflected in the personnel costs included in the current budget. Since securing coverage of the incremental cost of permanent archiving activities, we will pursue opportunities immediately after the project has received funding.

Privacy, Intellectual Property, Other Legal Requirements

Information collected can be released without privacy restrictions because informed consent for full public release of the data will be obtained. Depending of the providing party, the data can be encumbered with intellectual property rights (copyright, database rights, license restrictions, trade secret, patent or trademark). Depositing with the **TRR170-DB** does not require a transfer of copyright, but instead grant permission for the **TRR170-DB** to re-disseminate the data and to transform the data as necessary for preservation and access.

Archiving, Preservation, Long-term Access

The TRR170-DB commits to good archival practice and regular content migration. All data will also be made available for replication by any party under the CC-attribution license.

The central computing center (ZEDAT) of Freie Universität Berlin operates an automated back-up service to ensure back-up of servers of the campus network, long-term archiving of research data and data recovery.

ZEDAT will carry out updates of the hard- and software system on a regular basis. In addition to ZEDAT backups, we use a *Dataverse* functionality that creates extra archival copies of all data files stored in **TRR 170-DB**. If needed, the extra archival copies are easier and faster to retrieve from our storage area than ZEDAT's centrally stored **TRR 170-DB** backups.

Adherence

The current coordinating scientist of TRR 170, Elfrun Lehmann, is in charge of managing data for TRR 170. She is also responsible for all amendments, updates, etc. documentation, and communication with the TRR 170 user community.

In the current funding period, adherence to this RDM will be checked by the speaker of TRR 170 at least ninety-days prior to the end of the funded year. Adherence checks will include review of the **TRR170-DB** content, number of datasets stored and published, availability for each dataset of preservation friendly data formats (possibly embargoed but listed), availability of documentation (public), and correctness of data citation.

This research data management plan was designed using examples by the *Dataverse Project* [1] *and the Göttingen eResearch Alliance* [2].

References

[1] http://best-practices.dataverse.org/data-management/index.html

[2] https://www.eresearch.uni-goettingen.de/de/knowledge-base/howto/how-to-data-management-planning/

Berlin, August 2019