THE TRR170-DB REPOSITORY: A DATA REPOSITORY FOR DIVERS PLANETARY DATA, Elfrun Lehmann[1], Harry Becker[1], [1]Freie Universität Berlin, Germany, elfrun.lehmann@fu-berlin.de.

Introduction: The TRR170-DB data repository (https://planetary-data-portal.org/) manages research data from the collaborative research center 'Late Accretion onto Terrestrial Planets' (TRR 170). The repository makes TRR 170 data and other planetary science data accessible to researchers. Data in the repository reflects the diverse methods and approaches applied in the planetary sciences, including astromaterials data, experimental studies, remote sensing data, images and geophysical modeling data. The repository provides new machine-readable data and resources to the planetary science community.

The TRR170-DB repository aligns its data policy and practice to Open Science and the FAIR principles [1] as promoted by various national [2] and international [3] data initiatives. TRR170-DB is permanently hosted on a server at Freie Universität Berlin to ensure long-term preservation and access to its published data. The long-term data preservation shall encourage usage by researchers and the interested public. The TRR170-DB is referenced by Re3data, a global registry of research data repositories [4].

The **TRR170-DB system**: The repository is operated on the open source data management software Dataverse [5]. Users access the repository directly through its interface that connects to the storage environment of the datasets. Alternatively, a web portal allows for repository access while also guiding users how to use TRR170-DB. The website provides additional information to the community related to data management and data publication and informs about legal conditions and embargo periods to safeguard the data publication process.

Data access, storage, and publication: While the TRR170-DB serves as a hub to exchange data for the TRR 170 user community through a password-guarded member area, published data are also accessible for other interested researchers.

The repository interface provides search tools to allow users to conduct general searches or specify queries for published data via filters. Filters can direct searches to specific published data types (i.e., planetary materials/geophysics, planetary surface data, astronomy) via their metadata information. Specified searches result in compilations of data and metadata information available in the repository that can be directly viewed, saved, and downloaded. Recently, we integrated an advanced search tool, the data explorer (Fig.1), into TRR170-DB. The graphical user interface (GUI) allows users to search for specific information associated with archives, datasets and files and provides a listing of summary statistics. The data explorer complements the TRR170-DB and is an integrated part of a dataset.

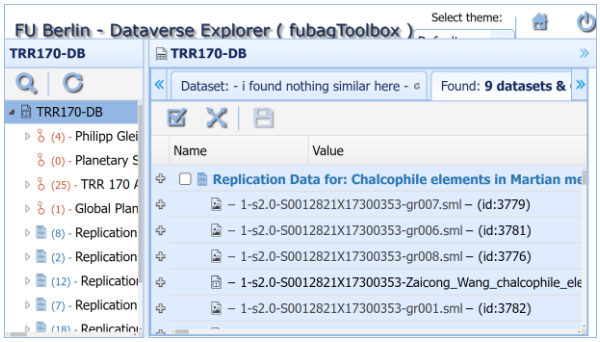

Fig. 1: TRR170-DB data explorer query results.

Any data supplier receives data curation services, a permanent archive and a digital object identifier (DOI) to make the dataset unique and findable. Our datasets obtain their DOIs via DataCite [6]. Data suppliers are requested to apply standardized ways to annotate, structure and organize data. This metadata information on the content, quality, origin, and other characteristics of the datasets ensures reliable data quality for future use in other research projects. In this way, data from different projects can be easier exchanged and be understood when used in different contexts.

Increasingly, journals, and funding agencies require researchers to provide replication datasets in a public repository. We support TRR 170 members and other users to meet this requirement by providing storage for replication datasets (Fig. 2). We implemented data publishing workflows to link publications to data.

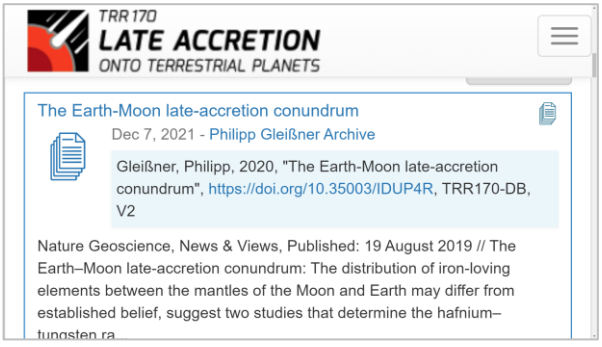

Fig 2: Replication datasets in TRR170-DB linked to published research articles.

Replication datasets make published data more easily findable and usable for other researchers. At present, most replication datasets were provided and used by TRR 170 authors of articles appearing in international journals since 2016. These replication datasets are freely available, and no special permission is required for reuse and verification of a study without having to contact the study's authors.

TRR170-DB provides links to some datasets at external sites and archives. They can be retrieved by entering metadata information while the datasets are stored in the formats of their external locations.

The metadata model: TRR170-DB has a flexible data-driven metadata system that uses tailored "metadata blocks" for specific data communities. The metadata blocks are based on standards that are compliant with several international metadata schemata (i.e. DDI, DataCite, Dublin Core, VOResource Schema, etc.) and controlled vocabularies. Once a dataset has been published, its metadata and files can be exported in various other metadata standards and file formats (Fig. 3). It allows for easy data and information transfer of published data among various geoscience subdisciplines so that they can be easily integrated into international repositories. This ensures that all data are generally accessible and interoperable between various external databases and repositories.

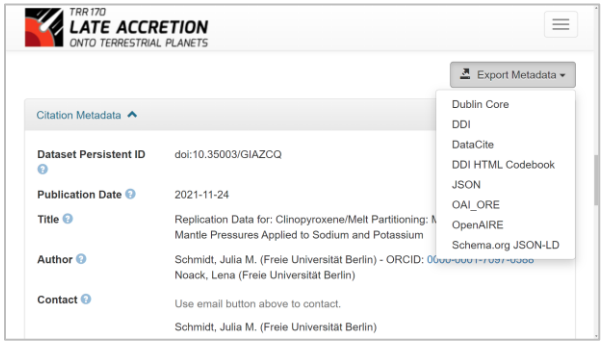

Fig. 3: Extract of TRR170-DB metadata and controlled vocabulary to metadata formats, upper right corner.

User and community services: We inform the TRR 170 network regularly about news in the data community and provide training on data related topics through webinars and workshops. So far, our on-demand training resources focused on managing research data using TRR170-DB and building data management plans. We will extend our learning resources to inform on policies and regulations and how to extract efficiently knowledge from our data collections.

Future Work: For effective management and structuring of data from heterogeneous datasets, we will adapt metadata in the repository to common international ontologies. We are currently expanding the metadata model to support interoperability, findability, preservation, and reuse of geochemical data in TRR170-DB. The new geochemical metadata blocks will be templates that incorporate standardized metadata information on samples and materials, analytical methods and additional experimental data. Advancing metadata will be an ongoing process in which the international scientific community and various initiatives (OneGeochemistry, Astromaterials Data System [7] etc.) need to interact and discuss what is required. .

All of TRR170's metadata content will be mapped to Freie Universität's central library system. This provides the data to a larger community of researchers, teachers, and students.

Acknowledgements: The TRR170-DB repository is managed and funded by subproject TRR 170-INF (German Research Foundation (DFG), 263649064–TRR 170). Freie Universität Berlin provides server infrastructure and a web content management system for managing the TRR170-DB website. Dataverse software is maintained by the IQSS Dataverse team, Harvard University. We thank A. Balduin, R. Schleuer, João Cardoso, and Archium for technical support.

References: [1] DFG, www.dfg.de/research_funding/programmes/nfdi/index.html. [2] Nationale Forschungsdaten Infrastruktur, www.nfdi.de. [3] Coalition for Publishing Data in the Earth and Space Sciences, copdess.org. [4] r3data, www.re3data.org. [5] Dataverse, dataverse.org. [6] DataCite, datacite.org. [7] Astromaterials Data System, https://www.astromat.org/.