



Data Policy

*Authors: Prof. Dr. Thomas Dandekar, Prof. Dr. Birte Kehr, Dr. Juan Prada, Dr. Meik Kunz.
Version: 1.0
Date: 17.01.2025*

This Data Policy was adapted from a document of the CRC1002 (Version 2.0, Date: 14.6.2017) for the requirements of the TR221 “Modulation of graft-versus-host and graft-versus-leukemia immune responses after allogeneic stem cell transplantation”.*

** Authors of the original document:*

*Prof. Dr. Blanche Schwappach (orcid.org/0000-0003-0225-6432),
PD Dr. Sara Y. Nußbeck (orcid.org/0000-0003-1223-6494),
Dr. Harald Kusch (orcid.org/0000-0002-9895-2469),
Bartłomiej Marzec (orcid.org/0000-0002-9640-8881)*

Contents

Preamble3

§1 Definitions.....3

§2 Coverage3

§4 Documentation of datasets4

§5 Access to data in the Research Data Platform4

§6 Use of RDP data by TR221 members4

§7 Access to project data by third parties.....5

§8 Delivery of data to the Research Data Platform and quality control5

§9 Data sharing.....5

§10 Authorship and copyright5

§11 Implementation and enforcement5

Preamble

The subprojects of the DFG funded TR221 are collecting datasets of high diversity. The TR221 results and publications arise from collaborations of different subprojects and even between different institutions (i.e. the university clinics in Regensburg, Erlangen and Würzburg, and others). The TR221 aims for international visibility and active support of the research community in the area. This requires high quality data documentation, clear and comprehensive metadata annotation, storage capacity, and curation of data sets shared within and outside the consortium. This data policy document is created as a guideline for this task.

§1 Definitions

- (1) Project data means any information in analogue or digital format including (not necessarily all) raw data obtained from measurements, processed data, images, descriptions, software as well as any publications, which are produced, used, acquired, derived from data or stored for the TR221 in the course of conducting research by any of the TR221 members.
- (2) Metadata means all data describing and documenting the project data.
- (3) TR221 data means all project data and metadata related to the TR221 and/or produced within the TR221. The Research Data Platform (RDP, §1 (7)) will store selected TR221 data.
- (4) TR221 members are principal investigators (PIs), doctoral and postdoctoral researchers (irrespective of their individual funding source), student research assistants, and associate fellows as well as associated scientists who contribute directly and specifically to TR221 subprojects.
- (5) Subproject refers to any subproject of the TR221. PI's are the leaders of the respective subproject.
- (6) INF team refers to all staff involved in digital infrastructure development in TR221.
- (7) Research Data Platform (RDP) refers to all electronic data capturing systems containing TR221 data, which are provided or developed and managed by the INF team.
- (8) Data owners are those who produce the data, normally in wet-lab experiments. Data owners are also those who produce data from clinical work.
- (9) Proprietary data of TR221 refers to data produced by any of the TR221 subprojects. This data can also become public after publication. Public data refers to any data relevant to the TR221 that has already been published and can be found in some database.
- (10) The Published Data Registry is the compilation of publications resulting from the TR221. The compilation is done and updated by the INF team.

§2 Coverage

- (1) This policy applies to all TR221 members as far as research activities are carried out under the TR221 or when TR221 data is used by other scientists or projects.

§3 Rights and responsibilities

- (1) Primary research data will be acquired and stored by TR221 members according to Good Scientific Practice. TR221 members are responsible for the organization of primary research data on their research group servers.
- (2) Data specifically relevant to the project shall be submitted to the INF project and eventually transferred to the RDP for sharing with other researchers in accordance with the RDP data access policy (§5).
- (3) Supported by the INF team, submitting scientists provide metadata necessary for the understanding of the entered subproject data.
- (4) All data and metadata will be – as far as available and in accordance with national and international standards – provided in formats suitable for sustainable data storage and documentation.

- (5) Supported by the INF team, data selected by the PI for public sharing will be prepared for the transfer into public repositories (e.g. NCBI, EGA, ENA, STRING, etc.).
- (6) If a TR221 member leaves a subproject, e.g. finishing of PhD thesis, the entire body of the original electronic, analog and written information including the metadata for all research projects from the beginning to the end of employment must be handed over to the responsible PI. If a PI leaves the TR221, data and metadata transferred to the RDP remain stored there under the same hierarchy of use and access rules. Briefly, all data previously accessible only to the PI who has left will be accessible only to the Speaker of the TR221 or representatives of the PI's former institution (while a member of TR221) to support any putative investigation of scientific conduct. All data and metadata made accessible by the PI to TR221 members or the public via the RDP will retain the same accessibility status unless the PI specifically requests an access status change.
- (7) The INF team is responsible for the maintenance and management of the RDP, which is hosted at the University of Würzburg.
- (8) The TR221 steering committee serves to adjudicate possible disputes relating to this data policy and data sharing.

§4 Documentation of datasets

- (1) All datasets in the RDP must be documented with descriptive metadata and should be linked to related publications in the Published Data Registry (§1 (12)). All datasets will be stored with information on who collected them, version of the data, date of the latest update, rights holders and use and access rules. As far as possible, information on the datasets will be accessible via persistent resolvable identifiers.
- (2) The INF team provides templates to facilitate metadata entry and trains and supports data entry into the RDP.

§5 Access to data in the Research Data Platform

- (1) The RDP access model consists of three levels:
 - (I) Public access: open to the global scientific community,
 - (II) TR221 access: open to all registered TR221 members,
 - (III) TR221 subproject access: open to all registered TR221 members of one particular subproject and
 - (IV) Selected access: open to a selected group of researchers which might or might not be TR221 members.
- (2) The RDP user model builds on five user roles:
 - (I) Unregistered user: member of global scientific community,
 - (II) Principal investigator (PI): head of a TR221 subproject,
 - (III) Registrar: responsible for registering and managing RDP datasets (nominated by PI),
 - (IV) Researcher: member of a TR221 research group, and
 - (V) Administrator: responsible for maintenance and management of the RDP. Administrators may not share RDP data with anyone without PI's consent.
- (3) The responsible PIs authorize the visibility of and access to all TR221 data entered by the respective subproject. PIs may delegate this responsibility to registrars.
- (4) The RDP will not contain any automatic features that make data accessible without consent.

§6 Use of RDP data by TR221 members

- (1) Data use must always be based on an agreement between original TR221 data supplier and data user. Original data suppliers are the scientists originally acquiring the TR221 data and the PIs of the respective subprojects.
- (2) Data accessed by a scientist must only be used for purposes necessary to carry out his/her own

research in the TR221. It is prohibited to distribute other scientist's data to a third party without written consent of the owner of the data.

§7 Access to project data by third parties

- (1) Full access will be provided as required by Good Scientific Practice in case the university or funding agencies need to investigate any case of scientific misconduct.
- (2) In any other case, if a third party external to the TR221 needs access to TR221 data, a written permission of the data owner is necessary.

§8 Delivery of data to the Research Data Platform and quality control

- (1) All key data related to publications must be transferred to the RDP by the time of manuscript acceptance at the latest.
- (2) The INF team helps during data submission to the RDP to ensure documentation quality by requiring the submission of suitable metadata.
- (3) Every TR221 member submitting TR221 data is responsible for the quality of the submitted data.

§9 Data sharing

- (1) During scientific discussions and collaborations, some TR221 data will be shared between subprojects to foster the TR221-wide implementation of cutting-edge technology and synergy between individual subprojects. Unpublished data discussed during TR221 activities will not be communicated to third parties unless the written consent of the data owner has been given.

§10 Authorship and copyright

- (1) This data policy does not affect any rights of TR221 members under applicable copyright legislation. As far as projects are subject to third party's copyright, such right must be respected.
- (2) The provisions contained in this data policy do not in any way affect the legitimate rights and interests of TR221 members concerning their authorship regarding any scientific publication resulting from the use of TR221 data.
- (3) Authorship shall be determined by reference to national and international good practice guidelines in the sciences and relevant standards regarding Good Academic Practice.
- (4) In any published or unpublished writing, TR221 data should be referred to in the form author, TR221, subproject number, date of access if appropriate.

§11 Implementation and enforcement

- (1) During development of data handling and management procedures of the TR221, mandatory contributions of individual PIs are limited to the requirements of Good Scientific Practice as defined by the participating institutions and the DFG.
- (2) The INF team creates and provides the necessary tools for efficient and synergistic data handling and long-term storage by continuous extensions and enhancements of the TR221 RDP.
- (3) The aim of the INF team is to support interoperability between TR221 subprojects and create added value regarding re-usability of research data generated within TR221.
- (4) This data policy has been accepted by all PIs of the TR221, will be subject to constant development and replaced by future improved versions as required.