Data Policy

Authors: *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), Bartlomiej Marzec (orcid.org/0000-0002-9640-8881)*
Version: *2.0*
Date: *14.06.2017*

Contents

Preamble

The subprojects of the DFG funded Collaborative Research Center (CRC) 1002 “Modulatory Units in Heart Failure” are collecting extensive heterogeneous datasets. Many of the results and publications will arise as collaborations of different subprojects, also between different institutions (University Medical Center Göttingen and Max-Planck Institutes for Biophysical Chemistry and for Dynamics and Self-Organization). The CRC 1002 aims at international visibility and broad use of its generated data. The latter requires careful data documentation, high-quality metadata annotation, storage, and curation of data sets shared within the CRC or publicly with the scientific community. These aspects are subject of this data policy.

§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 is produced, used, acquired, derived from data or stored for the CRC 1002 in the course of conducting research by any of the CRC 1002 project members.
(2) Metadata means all data describing and documenting the project data.
(3) CRC 1002 data means all project data and metadata. The Research Data Platform (RDP, §1 (9)) will store selected CRC 1002 data.
(4) CRC 1002 project 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 CRC 1002 projects.
(5) Project refers to the whole CRC 1002.
(6) Subproject refers to any subproject of the CRC 1002.
(7) PIs are the leaders of the respective subproject.
(8) Administrator refers to the staff in charge of maintenance and management of the IT infrastructure used to store and access project data.
(9) INF team refers to all staff involved in digital infrastructure development in CRC 1002 and CRC 1190.
(10) Research Data Platform (RDP) refers to all electronic data capturing systems containing project data, which are provided or developed and managed by the INF team, e.g. the Antibody Catalogue, the Lab Notebook Registry, or the Electronic Laboratory Notebook.

§2 Coverage

(1) This policy applies to all CRC 1002 project members as far as research activities are carried out under the CRC 1002 project or when CRC 1002 data is used by other scientists or projects. It also covers interactions with CRC 1190 “Compartmental Gates and Contact Sites in Cells” and International Research Training Group (IRTG) 1816 “Phosphorylation- and Redox-mediated Signaling Mechanisms in the Failing Heart”, i.e. via the joint development and consented use of RDP modules.

§3 Data Management Committee

(1) The CRC 1002 steering committee (“Vorstand”, see §6 CRC 1002 Bylaws, 10.03.2016) also serves as the CRC 1002 Data Management Committee (DMC). The DMC is advised by the INF team.
(2) The DMC serves to adjudicate possible disputes relating to this data policy.

§4 Use and Access Committee for biomaterials and data

(1) The CRC 1002 steering committee ("Vorstand", see §6 CRC 1002 Bylaws, 10.03.2016) also serves as the CRC 1002 Use and Access Committee biomaterial and data of the "Transition to failure" study.
(2) Request for biomaterial and data have to be addressed to this committee and will be dealt with in the regular face to face meetings or by written consent in lieu of a live meeting.
(3) The request has to cover aspects for the planned use of samples (project idea and research question), a detailed description of the samples and data requested (in order for the UMG Biobank to check for the feasibility of the study) and contact details of the requesting scientist.
(4) The Use and Access Committee will make a decision based on the scientific aspects and the expected outcome.

§5 Rights and responsibilities

(1) According to Good Scientific Practice, primary research data will be acquired and stored by the respective research groups using established services of the Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen (GWDG), UMG IT service or Max Planck Society. The research group's and institution's data management plans determine the organization of the group servers.
(2) Ownership of project data is according to the respective legal framework pertaining to each institution involved in the CRC 1002. The copyrights and intellectual property rights on these data and datasets apply irrespective of the storage location.
(3) If a CRC project 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 have to be handed over to the responsible PI. If a PI leaves the CRC, data and metadata entered into the CRC 1002 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 CRC 1002 or representatives of the PI's former institution (while a member of CRC 1002) to support any putative investigation of scientific conduct. All data and metadata made accessible by the PI to CRC 1002 members or the public via the RDP will retain the same accessibility status unless the PI specifically requests access changes from the DMC.
(4) The INF team is responsible for the maintenance and management of the RDP, which is hosted at the GWDG.
(5) Data specifically relevant to the CRC 1002 RDP will be copied to the respective server. Supported by the INF team, project members provide metadata necessary for the understanding of the entered subproject data. 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. Data selected by the PI for public sharing will be prepared for the transfer into public repositories (e.g. Antibodypedia, Mouse Phenome Database).

§5 Documentation of datasets

(1) All datasets in the RDP have to be documented with descriptive metadata and should be linked to related publications in the Published Data Registry. 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.

§6 Access to data in the Research Data Platform

(1) The RDP access model consists of three levels: (I) Public (open to the global scientific community), (II) CRC 1002 (open to all registered CRC 1002 members, and (III) CRC 1002 Research Group (open to all registered CRC 1002 members of a particular research group).
(2) The RDP user model builds on five user roles: (I) Unregistered user (member of global scientific community, interested in published CRC 1002 results), (II) Principal investigator (PI, head of a CRC 1002 research group), (III) Registrar (responsible of registration and management of RDP Datasets, nominated by PI), (IV) Researcher (member of a CRC 1002 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 CRC 1002 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.
(5) PIs control data access to entries in the Electronic Laboratory Notebook (ELN) module, entered by their respective research group members.
(6) Access to RDP data for non-CRC 1002 members needs to be approved by the DMC (§3).

§7 Use of RDP data by CRC 1002 members

(1) Data use must always be based on an agreement between original CRC 1002 data supplier and data user. Original data suppliers are the scientists originally obtaining the CRC 1002 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 CRC 1002. It is prohibited to distribute other scientist's data to a third party without written consent of the DMC (§3).

§8 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 have to investigate any case of scientific misconduct.
(2) In any other case, if a third party external to the CRC 1002 needs access to project data, a written permission of the DMC (§3) is necessary.

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

(1) All key data related to publications should be delivered to the RDP at latest at the moment of acceptance of a manuscript.
(2) Every project member submitting CRC 1002 data is responsible for the quality of the submitted data.
(3) Assisted dialogues during data submission to the RDP ensure documentation quality by requiring the submission of suitable metadata.

§10 Data sharing

(1) During scientific discussions and collaborations, some CRC 1002 project data will be shared between subprojects to foster the CRC 1002-wide implementation of cutting-edge technology and synergy between individual subprojects. Unpublished data discussed during CRC 1002 activities will not be communicated to third parties unless the written consent of the DMC (§3) has been given.

(2) Data of the Antibody, Mouse Line and Cell Model Catalogue will be shared with all members of CRC 1190 and IRTG 1816 to further enhance local exchange of data and productivity.

§11 Authorship and copyright

(1) This data policy does not affect any rights of CRC 1002 project members under applicable copyright legislation. As far as projects are subject to third party's copyright, such right has to be respected.
(2) The provisions contained in this data policy do not in any way affect the legitimate rights and interests of CRC 1002 project members concerning their authorship regarding any scientific publication resulting from the use of CRC 1002 data.
(3) Such 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, CRC 1002 project data should be referred to in the form author, CRC 1002, subproject number, date of access if appropriate.

§12 Implementation and enforcement

(1) During development of data handling and management procedures of the CRC 1002, 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 CRC 1002 RDP.
(3) The aim of the INF team is to support interoperability between CRC 1002 subprojects and create added value regarding re-usability of research data generated within CRC 1002.
(4) This data policy has been accepted by all PIs of the CRC 1002, will be subject to constant development and replaced by future improved versions as required.
(5) This data policy replaces the data policy version 0.7 (final amendment, 14.04.2014) developed within the first funding period of CRC 1002.