

WIAS Research Data Policy

2022

1 Preamble

1.1 Purpose and meaning

Research data are the sum of all data and facts within the research process that are required to achieve a scientific result. They also occur to a large extent in mathematics and are of great value to both the process of acquiring knowledge in general and scientific institutions. The generation and processing of research data is indispensable to the scientific work of the WIAS. Together with scientific software, it is a central component of the transfer of research results from the WIAS to industry or to other areas of expertise. Research data can be from both external partners as well as research.

In this context, adequate research data management (RDM)¹ plays an important role. The research data policy of the WIAS aims to support researchers in RDM in all phases of the research process (Research Data Lifecycle, see **4.1**) and provide a framework for the collection, generation, documentation, storage and reusability of research data in accordance with the FAIR principles².

The different types of research data management purposes can be distinguished as follows:

- for use as a working copy for scientific work, — for example at the WIAS, this includes data from external partners from industry or other scientific disciplines that cannot be processed correctly in mathematical research without sufficient metadata and may be subject to special rights of use.
- the reuse of research data for later research, — this often concerns use by new employees in the research group, but also use of data in later projects or by researchers outside of the WIAS.
- storage as documentation of correct scientific work — such a requirement results not only for reproducibility but also from the rules of good scientific practice and increasingly from the conditions for the publication of research results in scientific journals, — and

¹DFG, "Förderung von Informationsinfrastrukturen für die Wissenschaft. Ein Positionspapier der Deutschen Forschungsgemeinschaft" (2018), 38 http://www.dfg.de/download/pdf/foerderung/programme/lis/positionspapier_informationsinfrastrukturen.pdf.

²<https://www.nature.com/articles/sdata201618>.

- retention to comply with legal or other non-research requirements, — such cases can occur, for example, in connection with confidentiality obligations or with medical data and the necessary special protection of the personal rights of the subjects.

Research data management is intended to support researchers in dealing with research data in situations such as the above, and to improve the quality of the scientific process. The implementation of the FAIR principles provides suitable methods for this.

The research data guidelines³ of the WIAS are based on the guidelines of the Alliance Initiative Digital Information from the German scientific organizations⁴.

1.2 (Meta-)Data Standards and principles

The documentation and provision of research data is an aspect of good scientific practice. This ensures the traceability and replication of the research results and the research process. This increases the connectivity of research and the reusability of its results⁵. All members and associates of the WIAS are required to prepare the research data generated in their scientific work in accordance with the established subject-specific regulations and standards. Well-documented and structured metadata, which contain standardized subject-specific information about the data, play a central role in findability, traceability and reusability of research data.

1.3 Research data and research software at the WIAS

In principle, research data is understood to be all digital and electronically storable data that is generated during scientific work at the WIAS within the scope of various tasks and objectives. Primarily, these are *data from external partners*, *input and output* when using scientific software, aggregated data that is collected as part of the *analysis of literature sources* and *experimental data* generated. In a broader sense, the *mathematical results* of research at the WIAS (definitions, theorems, proofs) can be understood as mathematical research data. Examples of mathematical research data are:

1. Mathematical documents in PDF, LaTeX, XML, MathML etc.
2. Notebooks e.g. in Jupyter or Mathematica
3. Domain-specific scientific software packages and their libraries like R for statistics, Octave, NumPy/SciPy or Julia and Matlab for matrix calculations, and SCIP for integer programming, FEniCS, and Trilinos for numerical simulations, Paraview for data analysis and visualization
4. Computer algebra systems like SageMath, Mathematica and the data collections they contain

³Hiemenz, B.; Kuberek, M. 2018. Empfehlungen zur Erstellung institutioneller Forschungsdaten-Policies. Das Forschungsdaten-Policy-Kit als generischer Baukasten mit Leitfragen und Textbausteinen für Hochschulen in Deutschland. <http://dx.doi.org/10.14279/depositonce-7521>

⁴<http://doi.org/10.2312/allianz0a.029>.

⁵German Research Foundation. (2019). Guidelines for Safeguarding Good Research Practice. Code of Conduct <https://doi.org/10.5281/zenodo.3923601>

5. Programs and scripts in the above systems or in systems not developed in the math community; also input data for these systems such as algorithm parameters, grids, mathematical objects in collections, the definition of a neural network as a graph in machine learning, etc.
6. Data frames and simulation data, mostly series of states of a system, discretized fields, more generally very large but structured data as output of a simulation or as an experimental result, stored in established data formats (e.g. HDF5). This also includes geometry descriptions and calculation grids as well as output formats for the visualization (e.g. VTK).
7. Formalized mathematics: Coq, HOL, Isabelle, Mizar, NASA PVS library
8. Collections of mathematical objects, e.g. L-Functions and modular forms database (LMFDB), Online Encyclopedia of Integer Sequences (OEIS), Class Group Database, ATLAS of Finite Group Representations, Manifold Atlas, GAP Small Groups Library
9. Description of mathematical models in mathematical modeling languages, e.g. Modelica for component-based modeling of complex systems, Systems Biology Markup Language (SBML) for biological processes, SPICE for modeling electronic circuits and AIMMS or LINGO as modeling language for integer optimization
10. Scientific illustrations, visualizations or animations.

In terms of good scientific practice, research software and research data should be viewed together. Research software is an important tool for obtaining research data. The use of research software at the WIAS is based on the Guidelines for handling research software from the Priority Initiative Digital Information of the Alliance of German Science Organizations ⁶ and the software engineering recommendations of the German Aerospace Center Aerospace Association (DLR)⁷.

2 Scope

This WIAS RD-policy is aimed at all (scientific) members of the WIAS. The RDM must be carried out exclusively in accordance with the applicable legal provisions. The RDM will be conducted in accordance with applicable laws protecting individuals and intellectual property and subject to specific research funding agreements with third parties. In the context of research collaborations, these guidelines apply as long as the other parties involved do not have equivalent ones or meet stricter requirements. If the research is funded by a third party and the underlying funding agreement contains specific provisions on research data management, these provisions take precedence.

⁶<https://doi.org/10.5281/zenodo.1172970>

⁷https://rse.dlr.de/01_guidelines.html

3 Legal and ethical aspects

3.1 Ownership and Use Rights

Data must be marked personally and filed under the name of the person responsible. Data authors and other persons involved (such as data collectors, data curators, project managers, others) must be mentioned.

3.2 Privacy

Personal information is provided in accordance with applicable legislation process and use.

3.3 Transfer of rights

When transferring subsequent use or publication rights, care should be taken to ensure that the data remains freely available for scientific purposes. All rights to data, in particular the right to use or publish the data further, should be reserved for the data authors and should not be assigned to third parties.

4 Handling research data

4.1 Research Data Lifecycle

The life cycle of research data includes the creation, storage, archiving and deletion of sorted data. The computing technology (IT) department and the scientific library of the WIAS support the employees in dealing with research data throughout the entire life cycle. In particular, IT provides the technical infrastructure for data security (backup, see 4.5) and long-term storage (data repository). In order to be able to find and access research data at any time, it must be structured and provided with metadata. The scientific library of the WIAS supports and advises on the selection of suitable metadata schemas. In addition, it ensures the provision of persistent identifiers (e.g. DOI, URI etc) for research data objects. Persistent identifiers are used for the unique and permanent addressing of digital data. The WIAS recommends creating a data management plan as early as possible, which should describe the handling of research data within the framework of a research project (see 5.3.1). The WIAS follows the recommendations of the DFG for handling research data in research projects and the guidelines of the Leibniz Association⁸.

4.2 Basic principles

The research data must be correctly and completely managed, unadulterated in a reliable manner, and the integrity of the research data must be guaranteed. According to

⁸https://www.dfg.de/download/pdf/foerderung/basisn_dfg_foerderung/forschungsdaten/forschungsdaten_checkliste_de.pdf; see 5.3.1); https://www.leibniz-gemeinschaft.de/fileadmin/user_upload/Bilder_und_Downloads/Forschung/Open_Science/Leitlinie_Forschungsdaten_2018_EN.pdf

the FAIR principles⁹, the research data should be kept identifiable, traceable, available and, if possible, reusable and interoperable.

4.2.1 Scope of application

The WIAS distinguishes between:

- sustainable data in relation to the publication performance of the institute (monographs, publications in scientific journals, preprints, WIAS reports and technical reports, contributions to collective works and conference proceedings),
- project-specific data that arise in the processing of projects with non-disclosure agreements or special ethics guidelines, and
- further, unpublished data that arise in the research work.

4.3 Criteria for data selection

In consideration of the interests of exploitation, contractual restrictions and ethical issues, the WIAS supports the publication of research data both alongside publications and in dedicated data journals and promotes their public access in accordance with the principles of Open Access. For selected research data, the WIAS strives for the implementation of an institute's own data repository and the assignment of a DOI for the purpose of data registration and citability. This underscores the importance of research data as research result in its own right.

4.3.1 Sustainable data related to publications

Sustainable research data that are created in connection with a scientific publication are subject to special requirements. It is recommended that the research data and research software on which a scientific publication is based be made publicly available. When publishing, the software and Data Citation Principles¹⁰ should be observed. Research data referenced in this way can provide comprehensive evidence of the research work. In order to implement the rules of good scientific practice, the research data used in each case is stored in a structured manner and provided with metadata, as well as the associated data registration and long-term storage. In order to enable the traceability of the research results by third parties, the associated research data and, if necessary, the scientific software used for the data generation can also be stored on the WIAS-own repository together with a publication. For research data stored in a data repository, research data and publications reference each other.

4.4 Access and Licensing

The WIAS supports and promotes free access to research data according to the principles of Open Access as described in the "Berlin Declaration on Open Access to Scientific

⁹Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>

¹⁰<https://doi.org/10.25490/a97f-egykh>

Knowledge”¹¹ from 2003, unless intellectual property rights, data protection rights, legal requirements, ethical aspects or the rights of third parties are opposed to this. In this way, the WIAS supports the Open Access strategy of the State of Berlin¹² and has adopted an institute’s own Open Access Policy. Following the adoption of the Open Access Policy, WIAS recommends making research data and scientific publications publicly accessible. Important data for which no publication is intended should nevertheless be processed and stored in a structured manner according to the FAIR principles so that they can be made available on request. Free standard formats should preferably be selected for the format of the data in order to facilitate interoperability and ensure long-term readability.

The license is selected accordingly to enable the use of the data according to the FAIR principles. CC0 or CC BY licenses are recommended, an example of source code licenses is the General Public License. Research data that cannot be made accessible – due to legal, contractual provisions or for data and other protection reasons or ethical or legal aspects – are stored in a confidential storage service or in the institutional repository of the WIAS.

4.5 Backups

The storage and archiving of digital research data takes place in the IT infrastructure of the WIAS, in the institute’s own repository or publication server or in recognized national or international (subject-specific) archives. Data storage with external providers or based on specifications by third party providers remains unaffected.

4.6 Storage time/storage period

Within the legal framework, the researchers determine when research data is to be made publicly accessible as early as possible. Research data on which a publication is based are kept for at least ten years from the date of publication.

4.7 Deletion of data

In the event that research data is to be deleted, these measures will be carried out after the required archiving period has expired and taking into account all legal and ethical aspects. The interests and contractual obligations of third parties and other stakeholders and the aspects of confidentiality and security must also be taken into account.

5 Responsibilities of Researchers

The handling of research data at WIAS is regulated in the projects. The respective project or research group leader is responsible for the implementation of the research

¹¹https://openaccess.mpg.de/67605/berlin_declaration_engl.pdf

¹²See also printed paper 17/2512, October 21, 2015 of the Berlin House of Representatives - Open Access Strategy for Berlin: ”Make scientific publications accessible and usable for everyone”.

data guidelines. The responsibility for the life cycle¹³ of research data, in particular for securing and providing the research data for long-term archiving, lies primarily with the principal investigator (PI).

5.1 Requirements

With RDM, the researchers must first meet, and once met, observe all organizational, statutory, legal and contractual requirements.

5.2 Project Regulations

Those responsible for the project make regulations for RDM in their projects, for example specifying when and under what conditions the research data are made accessible. The heads of research projects shape the RDM in their working groups and make regulations when their members change. These include keeping the original data at the place of origin, precautions for the appropriate transfer of primary data and for clarifying access rights, as well as the option of creating a duplicate of the data when changing jobs, provided that data protection regulations do not conflict with this.

5.3 Good scientific practice and data management plan

In the spirit of good scientific practice, the documentation of the scientific work should be stored. The WIAS is oriented towards the guidelines for safeguarding good scientific practice - Code of the German Research Foundation (2019)¹⁴. According to the requirements of the funding institutions, a data management plan adapted to the requirements of the project is created in third-party funded projects. This should ensure the use of the data for scientific work and the reusability of the research data in later research. The data management plan regulates, among other things, the structure of the metadata schema and, if necessary, access to research data, such as that of external partners.

5.3.1 Data management plan

It is recommended that as part of any research project, a data management plan be maintained that explicitly addresses how the data should be managed in terms of quality assurance¹⁵. Subject-specific features are taken into account. Research projects with research data require (even before the start of the research project) a data management plan that applies to all emerging relevant research data:

1. where the research data is stored,
2. in which form the data is stored,

¹³From the generation to the final deletion of research data.

¹⁴https://www.dfg.de/download/pdf/foerderung/srechten_rahmenbedingungen/gute_scientific_practice/kodex_gwp.pdf, in particular for research data, see guideline 13 on creating public access to research results.

¹⁵See: https://www.dfg.de/download/pdf/foerderung/basisn_dfg_foerderung/forschungsdaten/forschungsdaten_checkliste_de.pdf

3. time and duration of storage and reasons for restrictions,
4. how the access is designed.

6 Responsibilities of the institute

6.1 Basic equipment

The WIAS implements and maintains a basic set of research data infrastructure and thus ensures appropriate storage and the technical availability of digital research data.

6.2 Infrastructure

An institutional research data repository for long-term storage and publication, for verification and for reuse according to the access rights defined by the scientists is made available.

6.3 Advice and support measures

The WIAS supports researchers in the planning, collection and storage of research data, in the formulation and maintenance of standards for the handling of research data and with appropriate consulting services. A successful implementation of research data management and the writing of data management plans are based on concrete support measures. The WIAS offers data curation and data registration.

6.4 Cooperation on common standards and structures

As the applicant and coordinating institution of the NFDI consortium MaRDI (Mathematical Research Data Initiative), the WIAS promotes the development of standards and infrastructures for mathematical research data. In particular, it contributes to the work area *Statistics and Machine Learning* as well as the interdisciplinary use of mathematics at the interface to other disciplines.

7 Validity

7.1 Validity and Review

This policy was approved on July 8th, 2022 at the Director's Meeting of the WIAS in Berlin. The policy came into effect on July 11th, 2022. This policy is checked regularly, at least every 3 years, by the research data and research software committee of the WIAS to ensure that it is up to date. As part of the review, monitoring is carried out to check compliance with legal and regulatory standards and to adapt the research data policy to infrastructural developments and the current needs of researchers.

7.2 Contact

If you have any questions about the content of the research data policy, please contact the WIAS library.