Data Management Policy

SFB 1552

Defects and Defect Engineering in Soft Matter

JOHANNES GUTENBERG UNIVERSITÄT MAINZ

JG|U

MAX PLANCK INSTITUTE FOR POLYMER RESEARCH

Funded by

DFG Deutsche Forschungsgemeinschaft

German Research Foundation

I. Preamble

In the awareness of the importance of scientific integrity, ethics, and responsibility towards the scientific community and the Deutsche Forschungsgesellschaft (DFG), we, the members of the Collaborative Research Centre SFB 1552, have obligated ourselves to jointly issue a declaration of self-commitment.

As researchers working together in the Collaborative Research Centre 1552, we recognize that research data are the results of scientific work and the basis for future scientific knowledge. We recognize that sustainable and responsible management of research data is essential/crucial to ensure the integrity, reproducibility and quality of our research.

We are aware of the responsibility we owe not only towards the current scientific community, but also towards future generations of researchers. We are therefore committed to establish, develop and implement responsible research data management over the entire duration of the SFB 1552 (of a maximum of 3 funding periods) and thereby also enabling subsequent use of the data.

We make this joint declaration of commitment to strengthen and promote the principles of good scientific practice as the cornerstone of our scientific cooperation in the Collaborative Research Centre 1552.

We acknowledge the importance of the guidelines and regulations for good scientific practice issued by various institutions such as the German Research Foundation (DFG), the Johannes Gutenberg University (JGU), the Max Planck Institute for Polymer Research (MPIP), the Leibniz Institute for Polymer Research (IPF), and the Fraunhofer Institute for Microstructure of Materials and Systems (IMM).

Through our joint declaration of self-commitment, we reaffirm our commitment to scientific integrity, ethics, and responsibility. We are determined to maintain the highest standards in our research and to make a significant contribution to scientific progress within the framework of the Collaborative Research Centre 1552. We are also committed to the common goal of maintaining a sustainable research data lifecycle to achieve FAIR data. The acronym FAIR stands for findable, accessible, interoperable and reusable. [1]

We sign this declaration of self-commitment in the confidence that it forms the fundament for successful and ethically responsible scientific collaboration within the Collaborative Research Centre 1552.

II. Participants

Area A – Doping Defects		
A01	Real-Time Nanoscale Investigation of Soft Matter Defects	Dmitry Budker
		Anna Ermakova
A02	Tuning Hybrid Molecule-Magnetic Interfaces by Defect Engineering	Angela Wittmann
A04	Energy Transfer towards Engineered Organic Dyes that Prevent Charge Carrier Trapping	Paul Blom
A05	Visible Light Driven Hydrogen Generation with Engineered Photocatalyst-Doped Micelles in Water	Christoph Kerzig
A06	Multicomponent Supramolecular Copolymerization for Controlled Defect Engineering	Shikha Diman

Area B – Connectivity Defects		
B01	Enhanced Mobility in Supramolecular Polymer Networks by Connectivity Defects	Arash Nikoubashman
		Sebastian Seiffert
B02	Molecular-Defect Regulated 1D Supramolecular Polymer and Hydrogel Formation using Multidomain Peptides	Pol Besenius
B03	Homeostatic System Behaviour to Engineer Resilience against Defects	Pol Besenius
		Andreas Walther
B04	Engineering water connectivity defects on nano-scale protein interfaces	Paul Czodrowski
		Lukas Stelzl

Area C – Topological Defects		
C01	Impact of Colloids as Defects on the Structure Formation of Lamellae Forming Amphiphiles and Amphiphilic Block Copolymers	Regina Bleul
		Michael Maskos
C02	Manipulating Defects with Defects in Block Copolymer based Materials	Friederike Schmid
C03	Dynamic Defect Annealing within Enzymatic Reaction Networks to Design Autonomous Reconfigurable ATP-Fueled Non-Equilibrium Multicomponent Systems	Susanne Gerber
		Andreas Walther
C04	In-Situ Characterization of Membrane Defects by Mechanical and Electrical Atomic Force Microscopy	Dirk Schneider
		Johannes Wahl

Methodological Cross-Section Projects		
Q01	Interdisciplinary Sensing and Spectroscopy	Dmitry Budker
		Ronald Ulbricht
Q02	Optical Super-Resolution Imaging in Soft Matter Systems	Xiaomin Liu
Q03	Dynamical Information through X-Rays	Katrin Amann-Winkel

Central Service Projects		
Z01	Administration and Coordination	Sebastian Seiffert
Z02	Integrated Research Training Group “Defects to Effects Engineering in Materials Sciences”	Pol Besenius

III. Content

IV. Abbreviations

BMBF ..... Bundesministerium für Bildung und Forschung
CC-BY 4.0 ..... Creative Commons License Attribution
DFG ..... Deutsche Forschungsgemeinschaft
DMP ..... Data Management Plan
FDG ..... Forschungsdatengesetz
IMM ..... Fraunhofer-Institut für Mikrotechnik und Mikrosysteme Mainz
IPF ..... Leibniz-Institut für Polymerforschung Dresden
JGU ..... Johannes Gutenberg-Universität Mainz
M.Sc. ..... Master of Science
MPIP ..... Max-Planck-Institut für Polymerforschung Mainz
PhD ..... Doctoral Students
PI. ..... Principal Investigator
PID ..... Persistent Identifiers
PostDoc ..... Postgraduate Students
RDM ..... Research Data Management
RDS ..... Research Data Steward
RS ..... Researcher
SFB ..... Sonderforschungsbereich

V. Validity

Our guidelines have been reviewed and verified by the community of project leaders and the Steering Committee and jointly published in their first version on 20.08.2024. Consequently, these framework conditions apply to each subproject for the entire duration of the Collaborative Research Centre's existence (of a maximum of 3 funding periods).

This and all further versions can be found under the following Digital Object Identifier (DOI): 10.5281/zenodo.13347606

VI. Versioning

Date	Version	Created by	Description of changes
31.03.2024	0.1	Johannes Berg	Data Policy Draft
24.04.2024	0.2	Johannes Berg	Data Organisation added
26.06.2024	0.3	Johannes Berg	RDMO and Standards added
10.07.2024	0.4	Johannes Berg	Publication Form added
20.08.2024	1.0	Johannes Berg	First Version published

1. General Information

The purpose of this document "Data management policy SFB 1552 (Defects and Defect Engineering in Soft Matter)" is intended to standardize and improve the research guidelines and research data management within the Collaborative Research Centre 1552 and to raise it to a sustainable and future-oriented level in accordance with the FAIR guidelines.

The Collaborative Research Centre 1552 consists of a large number of scientists at the following institutions. Johannes Gutenberg-University Mainz (JGU), Max Planck Institute for Polymer Research (MPIP), Leibniz Institute for Polymer Research (IPF), and Fraunhofer Institute for Microstructure of Materials and Systems (IMM).

The steering committee of the Collaborative Research Centre consists of the spokesperson, Prof. Dr. Sebastian Seiffert, and the vice-spokesperson, Prof. Dr. Michael Maskos, as well as the heads of the associated Integrated Research Training Group, Prof. Dr. Katrin Amann-Winkel and Prof. Dr. Pol Besenius.

Spokespersons:

Prof. Dr. Sebastian Seiffert
Johannes Gutenberg University Mainz, Department of Chemistry
Duesbergweg 10-14, D-55128 Mainz, Germany
+49 (0) 6131 39 23887
sebastian.seiffert@uni-mainz.de

Prof. Dr. Michael Maskos
Johannes Gutenberg University Mainz, Department of Chemistry
Jakob Welder Weg 11, D-55128 Mainz, Germany
+49 (0)6131-990 100
michael.maskos@imm.fraunhofer.de

Further Governing Board and Spokespersons of the IRTG:

Jun.-Prof. Dr. Katrin Amann-Winkel
Max Planck Institute for Polymer Research Mainz
Ackermannweg 10, D-55128 Mainz, Germany
+49 (0)6131 379 196
katrin.amannwinkel@mpip-mainz.mpg.de

Prof. Dr. Pol Besenius
Johannes Gutenberg University Mainz, Department of Chemistry
Duesbergweg 10-14, D-55128 Mainz, Germany
+49 (0)6131 39 22355
besenius@uni-mainz.de

Furthermore, responsible for the organization, coordination and research data management are the office assistant Julia Bender, the scientific coordinator Nora Fribiczer and the research data steward Johannes Berg.

Organizational & Coordinational Board:

Julia Bender (Organizational assistance)
Johannes Gutenberg University Mainz, Department of Chemistry
Duesbergweg 10-14, D-55128 Mainz, Germany
+49 (0) 6131 39 26142
j.bender@uni-mainz.de

Dr. Nora Fribiczer (Scientific Coordination)
Johannes Gutenberg University Mainz, Department of Chemistry
Duesbergweg 10-14, D-55128 Mainz, Germany
+49 (0) 6131 39 31247
nofribic@uni-mainz.de

Johannes Berg, M.Sc. (Research Data Steward)
Johannes Gutenberg University Mainz, Department of Chemistry
Duesbergweg 10-14, D-55128 Mainz, Germany
+49 (0) 6131 39 28815
joberg@uni-mainz.de

This Data Management Policy is based on the "Model Policy for Research Data Management (RDM) at Research Institutions/Institutes", published in 2017: ”LEARN Toolkit of Best Practice for Research Data Management” (pages 132-140) under the terms of the Creative Commons Attribution License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.[2]

In order to ensure that this data policy is also in line with the generally recognised FAIR data principles, this data management policy has been reviewed against the aspects of the FAIR-enabling data policy checklist. The aim is to ensure that sustainable and future-proof research data management can be implemented in the SFB 1552.[3]

2. Legal & ethical compliances

In order to ensure that this data policy is also in line with the generally recognised FAIR data principles, this data management policy has been reviewed against the aspects of the FAIR-enabling data policy checklist.

In addition to this Data Management Policy, the guidelines of the affiliated institutions must be followed, which include the following:

- § Research Data Law (Forschungsdatengesetz, FDG) of the Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung, BMBF)
- § Guidelines for ensuring good scientific practice in the Deutsche Forschungsgemeinschaft (DFG)
- § Regulations for safeguarding good scientific practice at Johannes Gutenberg-University Mainz (JGU)
- § Rules for safeguarding good scientific practice at the Max Planck Institute for Polymer Research Mainz (MPIP)
- § Code of Good Scientific Practice at Leibniz Institute of Polymer Research Dresden (IPF)
- § General principles of scientific work at Fraunhofer Institute for Microtechnology and Microsystems in Mainz (IMM)

These policies for the safe, sustainable and effective handling of research data within the Collaborative Research Centre 1552 apply to both researchers and organizational staff such as undergraduate students, graduate students, research trainees and module students, doctoral students (PhD), postgraduate (PostDoc) researchers, group and project leaders, principal investigators (PI) technical employees, coordinators and operational staff.

Intellectual property rights (IPR) are defined in an employment contract between a researcher and the corresponding employer. Other agreements (e.g. grant or consortium agreements) may also define the IPR. In cases where the intellectual property rights are held by the employing organization, the organization has the authority to decide how the data should be published and shared.

It is intended that these guidelines will be reviewed at regular intervals, at the latest annually, to ensure that they are up to date, adapted and subsequently verified by the bodies of the Collaborative Research Centre. It is the responsibility of the research

data management and the scientific coordination to make any necessary changes to this form and subsequently submit them to the consortium for approval.

In the case of partial or co-funding of research by third parties, or in the case of cooperation with external projects outside the Collaborative Research Centre, these cooperating parties commit themselves to accept and apply the contents and provisions of this document. Agreements with third parties concerning intellectual property rights, access rights, retention periods or storage of research data must not conflict with this Research Data Policy. The requirements of this Research Data Policy shall be taken into account when concluding new contracts. This policy can be specified and expanded for the handling of research data within the research fields (chemistry, biology, physics, etc.), units or projects by means of guidelines, a corresponding handbook and research data management plans.

3. Terminology

a. Research Data

Research data refers to all data generated by and resulting from the research process. Depending on the research question and topic, the experiments carried out (methodology, analysis devices, measuring equipment etc.) and the subsequent processing of the data, research data can be very diverse. In general, research data can be divided into the four categories: raw data, processed data, published data, and metadata.

Raw Data

Raw data refers to all unprocessed data that is generated during experiments, investigations, modulation or simulation and has not been processed or evaluated at this point in time.

Processed Data

Processed data refers to research data that is generated from raw data through cleaning, processing and evaluation.

Published Data

Published data are raw and/or processed data that are made available to scientists who are not involved in the experimental data collection and management of the

research area. In line with data management principles, published data should also be citable and have a persistent identifier (e.g. DOI).

Metadata

Metadata are any additional data that further describe the experimental data generated in the research data cycle. Metadata provides information relevant to the publication and reuse of the experimental data, i.e. how the experimental data were obtained, so that the research dataset can be understood by external parties. Metadata includes, but is not limited to, persistent object identifier, creator, title, publisher, date, time of publication, subject, language, type of resource, research and measurement equipment, method descriptions, solvent, wavelength and file format.

b. Fair Data Principles

The FAIR Data Principles are a set of fundamental data management principles defined by a consortium of scientists and scientific organisations in 2016, and have since been adopted as good scientific practice. It is the best practice to ensure that data and software are findable, accessible, interoperable and reusable for both humans and machines.[1]

c. Research Data Management (RDM)

Research data management is concerned with the handling and organization of data throughout the lifetime of a research project. Efficient and sustainable research data management begins with the planning of methodologies, implementation and standards of experiments as well as the planning of data collection, processing and management. Finally, research data management also includes backup, deletion or long-term archiving in accordance with the guidelines for good scientific practice.

d. Data Management Plans (DMPs)

Data management plans constitute the underlying basis for the management of research data in a research area. A research data management plan is also a living document and a kind of checklist for the organisation and management of research data. It is created early in a research data process and is continually updated to reflect scientific progress.

A DMP should therefore include at least the following information: Project description; administrative information and organisational structures; legal and ethical compliance;

timelines and resources; experimental description; data collection and generation; data types and formats; metadata standards; data organisation with folder structure and file naming conventions; secure data storage and backup; long-term archiving; data management responsibilities for data and data access, sharing and curation, and preservation for data re-use.

e. Smart Laboratory Components

As interlinked building blocks, intelligent/smart laboratory infrastructures such as electronic lab notebooks and repositories support meaningful, effective and sustainable research data management. [4]

Electronic lab notebooks (ELNs) are a global approach to implement a seamless data flow within a smart laboratory and at the same time enable a workflow according to the FAIR principles. This allows the necessary metadata to be directly assigned to a data set and proprietary data formats to be converted into open data formats. At the same time, it is possible to link experiment descriptions and framework conditions with the collected, processed and selected data.

Repositories are digital platforms that are used for secure long-term archiving and fair publication of research data and research results. Access to these research data can be restricted, making them available to the public or to a limited group of users.

f. Persistent Identifiers (PIDs)

Persistent identifiers (PIDs) are permanent, external references to (research) data, metadata files, documents, digital objects or personal profiles.

The Digital Object Identifier (DOI) has become the standard for data and digital objects. A DOI refers to the elements mentioned in a standardised form and to cite them reliably.[5]

The Open Researcher Contributor Identification Initiative (ORCID) was developed to identify researchers. Profiles can be continuously maintained in an ORCID profile, allowing for unambiguous referencing even if the name or research institution changes.[6]

g. Principal Investigator (PI)

The Principal Investigators are the organizational and scientific leaders of the individual research projects associated with the Collaborative Research Centre.

h. Researchers (RS)

Researchers are all persons actively involved in research within the Collaborative Research Centre SFB 1552, e.g. undergraduate students, graduate students, research assistants, doctoral students and postdoctoral researchers.

i. Research Data Steward (RDS)

The Research Data Steward is the contact person for all issues related to research data management. The Research Data Steward is the liaison between the funding and administrative bodies, the Collaborative Research Centre's Management Committee and all group leaders and researchers, and is responsible for all matters relating to research data.

4. Handling Research Data

In order to ensure the sustainable and efficient generation and use of research data, the handling of research data within the SFB must be carried out in accordance with the following rules. A coherent and consistent life cycle of research data can only be ensured if these rules are followed. Compliance with these rules and guidelines is particularly important because inadequate planning, execution and recording of experiments, as well as inadequate documentation or storage and archiving of data, will lead to a failure of the FAIR data life cycle. The principles of research data management within SFB 1552 defined in this Research Data Policy are further described and specified in the accompanying guideline paragraphs for handling Research Data Management.

a. FAIR Principles

As described above, the FAIR Principles are generally accepted guidelines on how research data and research results should be handled in order to make them (re)usable and thus sustainable for the researchers themselves and the scientific community. In the course of research data management within this special research area, all research data must be processed and organised in such a way that it complies with the FAIR principles and is therefore findable, accessible, interoperable and reusable.

b. Research Data Life Cycle

If research data is handled FAIRly, it can be used as a starting point for new research. The aim is to design all research in such a way that the experiments carried out and the associated data processing and data organization result in a continuous research data cycle. The aim should not be to generate unorganised data paths. It is therefore necessary to plan the research project in a structured way based on the current state of research. At the same time, the participants undertake to plan the associated data management at the beginning of this process. When the research results are then collected and provided with descriptive data, the metadata, the results can be analyzed and interpreted after the experimental data have been checked, validated and cleaned. When the resulting findings are published and the data generated are archived, the research results can be cited and the data re-used, forming the basis for a new 'research data life cycle'.

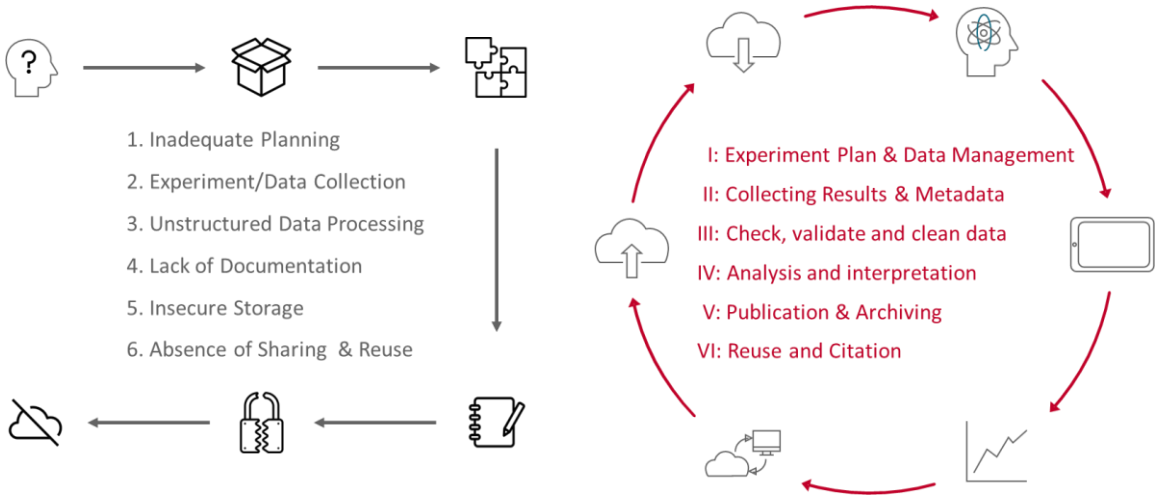

Figure 1: Disorgnised Data Trail and Research Data Life Cycle (Image remixed: NFDI4Chem, KnowledgeBase, ResearchDataLifeCycle CC BY-SA 4.0)

c. Data Organisation

Different types of research data are generated during the course of a research project. Research data can be laboratory values, measurement data in raw, analyzed, processed or published form, visual representations such as graphics, illustrations and plots, texts, software or simulation data. In addition to these direct research data in raw, processed and published form, research data also includes so-called metadata. It is important that all data generated as part of a research project are organised in a consistent manner so that they remain accessible and traceable long after the scientific research has been planned and carried out.

File naming conventions

It is necessary to introduce conventions for naming files. When naming files, it is important that they can be easily sorted and that their contents can be recognised at a glance. Short and meaningful file names make it easier to identify the data records. The date and version should also be included. The naming of research data files should follow a standard scheme within the Collaborative Research Centre as set out in this Policy. There are some general basic conditions for this file naming convention. Use standardised file names with meaningful abbreviations if necessary. There should be no special characters such as ?!& , % # ; ( ) @$. The date format (YYYYMMDD) and the time format (HHMMSS) should be used in the same way, as well as leading zeros to ensure correct sorting of the version numbering.

Date/Time_Project Identifier_Experiment_Datatype_Version_Fileformat

Some examples are listed in the following for analytical data:

20240331-011600_JoBe001_UVVis_rawdata_01.csv
20240331-011800_JoBe001_UVVis_metadata_01.txt
20240331-120405_JoBe001_UVVis_processeddata_01.csv
20240331-143103_JoBe001_UVVis_plotted_01.opju
20240331-150500_JoBe001_UVVis_graph_01.jpeg

Metadata

Metadata must be created for each type of research data. Without knowledge of the metadata, datasets cannot be correctly evaluated or reproduced. Therefore, the provision of metadata is essential to enable the correct interpretation and

contextualization, reproducibility and reuse of data. The metadata generated in SFB 1552 must be created according to a uniform standard.

```
{
  "project_identifier": "JoBe000",
  "title":"RDM Policy",
  "creator":
  {
    "last_name":"Berg",
    "first_name":"Johannes"
  },
  "date":"20240331",
  "time":"011600",
    "contributor":
  {
    "last_name":"Musterfrau",
    "first_name":"Martina"
  },
  "publisher":"SFB 1552 Z01",
  "publication_date":"20260626",
  "publication_time":"120000",
  "subject":"Research Data Management",
  "language":"english",
  "resource_type":"personal computer",
  "application":"MS Office Word",
  "description":"further description",
  "settings":"further settings",
  "solvent":"no solvent",
  "wavelenght":"no wavelenght",
  "file_format":".pdf/a",
  "rights":"CC BY 4.0",
  "persistent_identifier":"DOI-Text"
}
```

Figure 2: Sample form as a .json file for the machine-readable and open storage of metadata using the example of this Research Data Management Policy.

Each measurement file or raw data file must be linked to a metadata record, for example as .json file like shown in Figure 2. Metadata is essential to describe the data set in more detail and contributes immensely to reuse, reproduction and reorganisation.

SFB 1552 Metadata Standard:

Identifier	JoBe000	Resource type	Personal Computer
Title	RDM Policy	Application	MS Office Word
Creator(s)	Johannes Berg	Description	further description
Date/Time	20240331-011600	Settings	further settings
Contributor	Martina Musterfrau	Solvent	no solution
Publisher	SFB 1552 Z01	Wavelength	no wavelength
Publication	20240630-120000	Fileformat	.pdf/a
Subject	Research Data Management	Rights	CC BY 4.0
Language	English	Persistent ID	DOI-Text

Folder structure

A basic folder structure is essential for a successful organization of data storage. The filing and storage of research data must therefore follow a structure of folders and subfolders that corresponds to the project structure and workflow, using only self-explanatory and unique names. An exemplary description of a possible folder structure can be found in

Figure 3. A total of eight projects, divided into four main projects and four secondary projects, are shown. In addition, eight experiments are divided into the first two projects and the various folders for literature, planning, implementation, observation and evaluation are shown for one experiment.

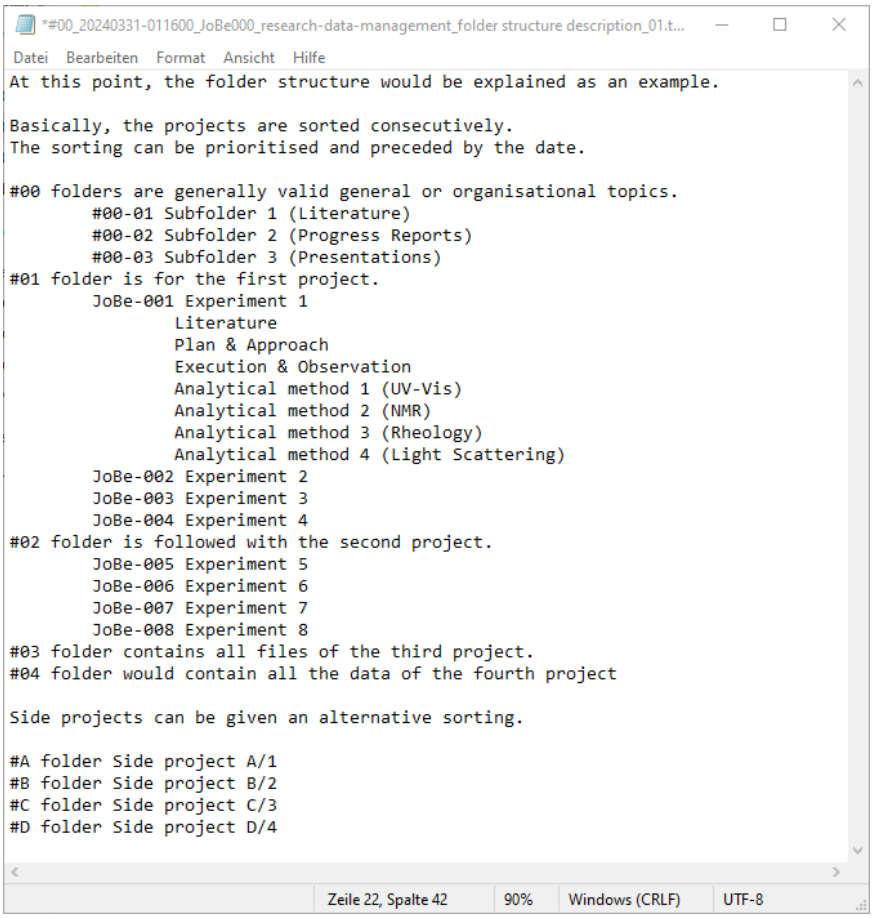

Figure 3: Description of the folder structure in order to understand it.

Folders with the same name should be avoided. A text file describing the top-level folder structure is essential, and may include the file naming convention on which the project is based. shows an exemplary presentation of the above described folder structure.

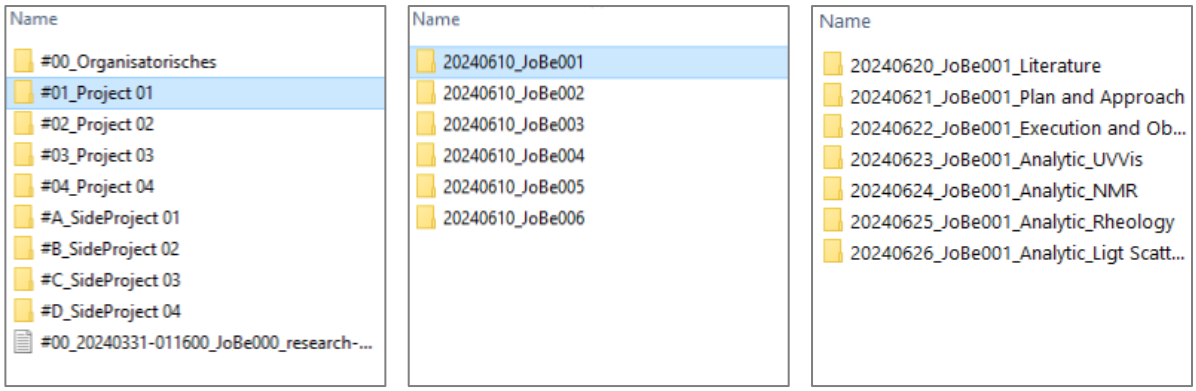

Figure 4: Exemplary folder structure of four main projects and four secondary projects with description of the folder structure, six experiments in the first project and some experimental and analytic data folders in experiment JoBe001.

Data formats

Only open, documentable and non-proprietary data formats should be used. Data compression must be strictly prohibited and the migration of research data into current data formats is essential for the success of fair research.

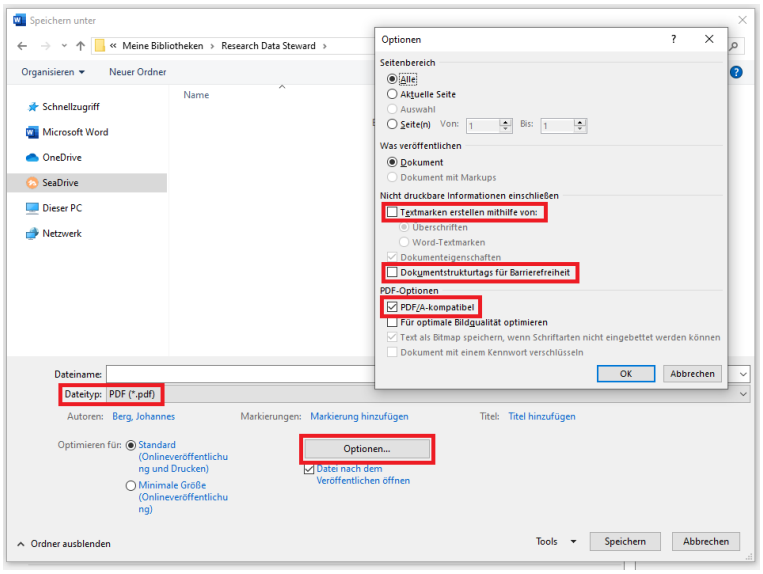

Figure 5: Description of the procedure for saving a .pdf/a compatible file.

Common and recommended formats in the SFB 1552 are different for tables, text, images and other multimedia files. For tables beneath .csv there is also .tsv and .spss. Text should be stored, backuped and archived as .txt, .html, .rdf or as .pdf/a formats. Images should be used as .tiff, .jpeg (JPEG2000) or as .png. The formats for multimedia files such as videos should be handeled as .avi, .wav or .mp3 / .mp4. Figure 5 describes the procedure for saving a .pdf/a compatible file for making the pdf format more FAIR.

d. Research Data Management Plans

In order to document and plan the data management tasks within a research project, a research data management plan must be established at the start of each research project. The data management plan must clearly state how a sustainable data life cycle will be ensured. As research projects evolve over time and the current state of good scientific practice adapts to new circumstances, it is essential that the research data management plan is updated accordingly. Therefore, the DMP is a living document and needs to be continuously updated, reviewed and adapted to the situation of the research project.

It is highly recommended to use the Research Data Management Organiser (RDMO) tool. The checklist published by the German Research Foundation (DFG), which has been enriched by the NFDI4Chem for a better understanding with science-specific content and supplemented with necessary questions, should be used. This DFG question catalogue configured by the NFDI4Chem can be found at https://rdmo.nfdi4chem.de/ and is shown in Figure 6.

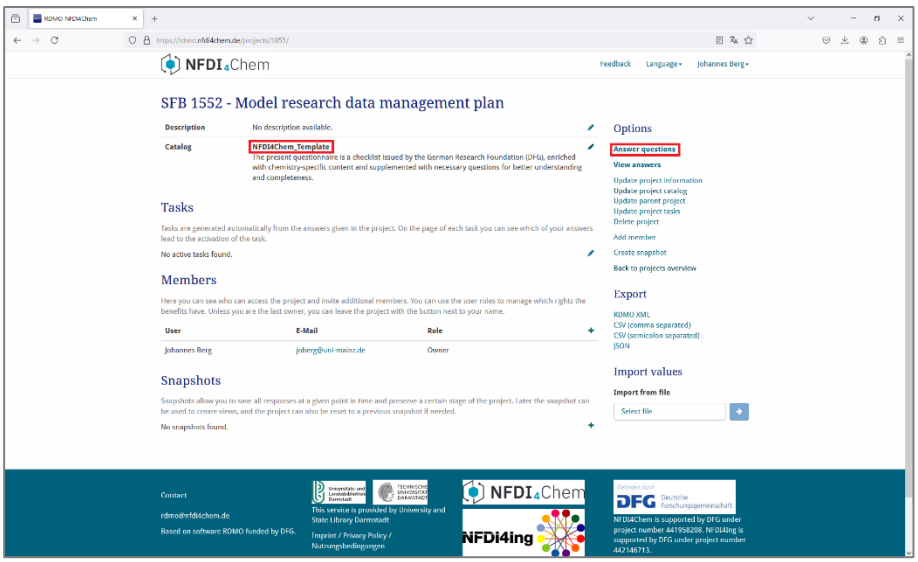

Figure 6: Example of a research data management plan in the RDMO tool. To create your DMP, you simply need to answer the questions provided. Your entries will be used to create a standardised research data management plan in accordance with DFG guidelines.

Appropriate tools such as RDMO should be used to create a meaningful data management plan. A research data management plan created by this tool will include at least the following information:

- Project Description
- Administrative Information and Organizational Structures
- Legal and Ethical compliance
- Schedule and Resources
- Experimental Description (Methods und Standards)
- Data Collection and Generation
- Data Types and Formats
- Metadata Standards
- Secure Data Storage & Backup
- Long-term Archiving
- Data Management Responsibilities
- Data Access, Sharing & Re-use

e. Lab Notebook and SmartLab

Safeguarding the methods and procedures of any research project requires detailed and comprehensible documentation. This is the only way to meet the high-quality standards of reusable research. The use of electronic lab notebooks is mandatory in the Collaborative Research Centre. In working groups where only analogue lab notebooks are used, these must be replaced immediately by digital equivalents. The previous documentation must then be transferred to electronic laboratory journals. Care must be taken to ensure that the electronic laboratory notebook used is adapted to the expected data formats, data volumes and methods and that the data is managed in accordance with the FAIR principles and the principles of good scientific practice. The use of a digital notebook (digitally written “paper notebook”, e.g. use of the program “Microsoft One Note”) is not a suitable electronic laboratory notebook, but is already a development step from an analogue paper laboratory notebook to an electronic one. In the context of the first funding period and the migration from analogue to digital laboratory notebooks, the use of such notebooks is permitted. However, projects are encouraged to look for suitable ELNs, the use of which will be mandatory in the second funding period.

Any electronic lab notebook is better than no lab notebook or only an outdated, non-FAIRly paper lab notebook. The number of diverse needs of the numerous projects within the Collaborative Research Centre are different, so it is up to the individual projects within the CRC to choose the ELN that best suits their needs. Nevertheless, we would like to recommend the two ELNs Chemotion and eLabFTW to all those who are still looking for an adequate platform.

f. Secure Data Storage and Backup

In our research initiative, data are mainly obtained by spectroscopic, spectrometric and other instrumental characterization methods, e.g. in the form of light, X-ray and neutron scattering profiles, electrical measurements or microscopic images and films. In some cases, however, the data are also obtained using particle-based computer simulations.

It is important to store the research data securely and for the long term. If the computers at the local monitoring stations are the primary storage location for the raw and processed data, this is not a sufficient primary storage location for the research data, so there must be a backup location. Storing research data only on the institute's own measurement computers, private data infrastructures and external removable storage media does not constitute adequate data backup. The PIs are responsible for ensuring that data are adequately backed up frequently. Researchers shall ensure that data are regularly and promptly transferred to a server-based storage system at the respective research institution if a measurement device is not automatically backed up, at least once a week.

The university's data centre (ZDV) regularly backs up the data stored on this internal university server (e.g. SeaFile). At MPI-P, the institute's computers are also backed up centrally (IBM Spectrum Protect and Personal Backup) and the central group drive is backed up to tape daily. The Max Planck Digital Library also offers all staff and project partners the possibility to synchronise project data. A similar central data infrastructure is also available at Fraunhofer ICT-IMM and Leibniz Institute for Polymer Research Dresden IPF. This type of automatic data backup is used for all measurement computers and user accounts connected to the network.

To ensure the integrity of research data, research data must be kept accurate, complete, unaltered and reliable. Adherence to the FAIR principles is essential, as is adherence to the principles of good scientific practice in the handling of research

data. In addition to these principles, the requirements of third-party funding bodies (Deutsche Forschungsgemeinschaft) and all applicable legal and contractual provisions of employment relationships (Forschungsdatengesetz) must also be observed.

The storage systems intended for the storage and backup of active research data must be named in the DMPs of the subprojects in such a way that it can be traced at any time where all research data generated as part of the laboratory documentation and the description of the research work in electronic laboratory journals have been stored. The intended naming convention for the storage/backup of active research data must be specified in the DMP so that the storage location of the corresponding research data can be unambiguously assigned.

The long-term storage of observational data and their origin, i.e. the applied experimental methodology used, is the primary objective. If there are comprehensible reasons for not retaining certain data (e.g. data sets from simulations that are easily and reliably reproducible but are likely to be relevant only for a short period of time), this must be documented in the DMP in a comprehensible and justified manner. If research data and documents have to be deleted or destroyed prematurely at the end of the prescribed archiving period has expired or for other reasons, this can only be done after all aspects have been considered. In particular, the interests and contractual requirements of third-party funding bodies and project partners in the Collaborative Research Centre as well as confidentiality and security aspects must be taken into account when deciding on storage or destruction. All of the above measures must be documented and made available for possible review.

g. Publication of Data and Results

The fundamental aim of the scientific research carried out in the SFB is the publication of the resulting research results and research data. Research results will continue to be published in the relevant literature according to current standards and citation norms. To ensure that all experimental data that are required for the verification of scientific results can be used at a later date, they must also be made publicly available in suitable repositories, unless there are legal, contractual or ethical reasons not to do so. Both the results and the data publication must be signed with a persistent identifier. Original and used data sources therefore must be traceable in every publication and any other presentations.

The choice of an appropriate repository will depend on the data format, the expected volume of data and the methodology used to generate the data. Users can orientate themselves better with the help of a register for research data repositories like https://www.re3data.org/. Examples of (data) repositories for chemical research data are the Chemotion repository, RADAR4Chem and the Gutenberg Open Science repositories. The Johannes Gutenberg University also offers a data archive, named iRODS where research data could be archived on long time use. All research data must be actively made publicly available at the earliest possible stage in order to ensure the fairest possible interaction. The publication of results and data is the responsibility of the project leader, but always in consultation with the researchers involved. Research data and records must be archived for an appropriate period of time. According to the current state of research, a period of 10 years is the usual standard.

When publishing results in the form of a paper or comparable media, the SFB 1552 publication form must be completed. This form can be found as an appendix to this SFB 1552 Research Data Policy. In addition to the actual publication of results, the place of data publication and all other important information are also recorded in order to demonstrate FAIR research data management. The CRC ID is assigned in consultation with the central service project Z01 in order to centrally record the SFB's publication performance.

The project leaders must draw up a project-internal data usage agreement with all members of the consortium. This must consider points relating to the authorship of data and result publications and the rights of use of the individual project members during the project term and after the completion of individual sub-projects.

5. Responsibilities, rights and obligations

Within the Collaborative Research Centre, responsibility for research data management extends to three levels. A distinction must be made between the responsibilities of the Principal Investigator, the Researcher and Research Data Management. Their responsibilities are described in more detail below:

a. Principal Investigator

Principal Investigators are responsible for the scientific work carried out by the researchers. This requires the provision of appropriate infrastructure and resources to support research, staff training and Research Data Management. The project leaders of the corresponding projects are responsible for the proper implementation of research data management in accordance with this policy. Furthermore, they are responsible for the research data management that is carried out in their working group and for providing the necessary research data infrastructure to maintaining Research Data Management.

b. Researchers

Researchers are responsible for the generation, processing and the analysis of research data and creating metadata. The researchers in the sub-projects are generally responsible for the execution and implementation of research data management and the management of the research data generated. In addition, RDM for publication, storage and long-term archiving for these topics will also be carried out by the researchers in accordance with the established general agreements and regulations of the SFB.

c. Research Data Steward

The Research Data Management Service Unit within the central service project, represented by the Scientific Coordinator (SC) and the Research Data Steward (RDS), is responsible for establishing common rules, standard procedures and (meta-)data formats for the entire SFB in order to implement, execute and maintain the research data management described in our initiative. The Research Data Steward is responsible for drafting and updating the general agreements on research data management, such as this policy. The RDS is also responsible for supporting the creation of data management plans. Courses on research data management will be organised and delivered by the RDS and the SC. In addition, the Research Data

Steward is the link between the central facilities of the affiliated institutes, the central facilities of the SFB, the Principal Investigators and the researchers within the subgroups.

6. Documentation data management modules

11.03.2024	Introduction about good Scientific Practice and Research Data Management
12.03.2024	Grundlagendiskussion zum Forschungsdatenmanagement im SFB
31.03.2024	Start of the creation of a data management guideline @ SFB 1552
04.04.2024	Launching of an internal knowledge database to standardize and optimize research and data collection methods within the Collaborative Research Centre 1552
20.08.2024	Publication of ResearchDataManagementPolicy@SFB1552
28.08.2024	General RDM Workshop SFB 1552
29.08.2024	Introducing Research Data Management Policy @SFB1552 and Creation of Research Data Management Plans @SFB1552.

7. Disclaimer and Severability Clause

This research data policy has been prepared in accordance with the current state of research data management and to the best of our knowledge and belief. This policy will be reviewed and updated at least annually to ensure that it is as accurate, complete, up-to-date and correct as possible. However, it is not possible to guarantee that this policy is complete and up to date. The authorship is not liable for ensuring that research data is free from third party rights, errors, viruses, malware or other malfunctions through the aforementioned institutions and research data management procedures. Liability claims can only be made in cases of malice, intent or gross negligence.

Should any provision of this policy be invalid, this shall not affect the validity of the remaining regulations. The participating members undertake to replace the invalid

provision with a valid provision that comes as close as possible to the purpose pursued by the authors with the invalid provision.

8. References

[1] M. D. Wilkinson, M. Dumontier, I. J. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. Da Silva Santos, P. E. Bourne et al., *Scientific data* **2016**, *3*, 160018.

[2] a) P. Budroni, B. Sánchez Solís, I. D. Traub, *LEARN Toolkit of Best Practice for Research Data Management*. *Development of a Model Policy for Research Data Management (RDM) at Austrian Research Institutions*, LEARN, **2017**; b) P. Budroni, B. Sánchez Solís, I. D. Traub, *LEARN Toolkit of Best Practice for Research Data Management*, LEARN, **2017**.

[3] J. Davidson, M. Grootveld, M. Verburg, R. van Horik, R. O'Connor, C. Engelhardt, F. Garbuglia, A. Vieira, E. Newbold, V. Proudman et al., *FAIR-enabling Data Policy Checklist*, Zenodo, **2022**.

[4] "Knowledge Base. SmartLab", can be found under https://knowledgebase.nfdi4chem.de/knowledge_base/docs/smartlab/, **2024**.

[5] O. Wannenwetsch, T. A. Majchrzak in *2016 Federated Conference on Computer Science and Information Systems (FedCSIS)*, **2016**, pp. 1031–1040.

[6] L. L. HAAK, M. FENNER, L. PAGLIONE, E. PENTZ, H. RATNER, *Learned Publishing* **2012**, *25*, 259.

Publication Form SFB 1552 SFB-ID: 000-Z01

This form contains all important data related to a publication within the SFB1552. Furthermore, it serves as proof both within the SFB 1552 and to the DFG as the funding organisation that the publication process has been in line with the guidelines of the research data management policy. This includes the publication of the research data and metadata associated with the publication in order to ensure verifiability, reproducibility and transparency of results along with long-term availability and reusability of data.

Information on research publication:

Title	
Article Type	Research article, communication, review, …
Journal	
Volume	
Issue	
Pages	
Persistent identifier	DOI, ISBN, …
Date of publication	YYMMDD (online or printed)
First Author	Name, Institution, contact details, ORCiD
Correspondent Author	Name, Institution, contact details, ORCiD
Project Number	A0X, B0Y, C0Z, Z01, …
Further Author	Names
Supporting information	Identifier, Description

Information on data publication:

Infrastructure	Zenodo, Gutenberg Open Science, Chemotion Repository
Date of publication	YYYYMMDD
Persistent identifier	DOI, ISBN, …