FAIRmat Guide

Legal Aspects in Research Data Management

This guide was created by FAIRmat. Cite it as "FAIRmat Guide to Legal Aspects in Research Data Management", version 1.1, March 22, 2024.

Contents

Preamble

Welcome to this practical guide to legal aspects of research data management (RDM). As a researcher, it is essential to be aware of the legal considerations relevant to your research, particularly when it comes to disseminating your findings or using data and results from other studies. This guide focuses on practical advice tailored to the fields of condensed-matter physics and materials science in Germany. It offers an overview of the key legal considerations relevant to the various stages of the data lifecycle, namely planning, collection, analysis, preservation, sharing, and reuse of the data. The guide is particularly intended for graduate students, postdoctoral researchers, and principal investigators who wish to understand and navigate the legal aspects of RDM in their research.

The rise of digitalization has significantly reshaped scientific practices. The traditional approach of merely publishing the conclusions derived from analyzed research data is no longer sufficient. Instead, we have seen the emergence of new research paradigms that use data beyond their original research aims. In parallel with Open Science, a movement that promotes transparency, collaboration, and accessibility across all scientific research facets,[1] an increasing need for comprehensive data management plans has emerged.[2] As a result of these transformations in scientific practices, effective RDM, with consideration of its legal aspects, has become significantly important.

Research data are the foundation of scientific work. The term is usually used correctly by its self-explanatory sense, referring to the collected materials as primary sources to support, develop, or validate scientific hypotheses, models, or theories. Research data may include


Photo by [ThisisEngineering RAEng on Unsplash]

measurement data, laboratory values, audiovisual information, various forms of sensor data in the form of images or N-dimensional numerical arrays, software, simulations and computations, or source codes, among others.

When navigating the legal aspects of RDM, understanding the rights associated with data is essential. It is important to distinguish between the concept of *holding the rights related to the data* and the 'ownership'* of data themselves. In this guide, the main attention is given to holding the rights related to data

*Ownership of data is a topic of ongoing discussions in Germany. For further details, see the references provided at the end of this guide.[3–5]

Contract Law

Contract law refers to the set of rules and principles that govern the agreements (contracts) between stakeholders. With respect to research data, the stakeholders include researchers, funding agencies, universities, research institutes, grant holders, companies, publishers, software providers, and those who own or run repositories. Contracts may specify details about data management such as access rights, publication conditions, ownership of rights, dissemination, and intellectual property rights (IPR). These contracts are crucial in shaping one's RDM strategy and data management plan, ensuring that all planned uses of the data align with the contractual obligations and terms specific to each party. Figure 1 shows a selection of key contracts relevant to RDM along with legal information that they typically include.

Contract law serves as the primary legal framework in RDM. The terms agreed upon in a contract generally take precedence, for example, in the case of a legal dispute. Therefore, these contracts should be the first point of reference for researchers when considering legal aspects of research data management. For instance, a signed data confidentiality agreement; i.e., a non-disclosure agreement (NDA), overrides other rights that may allow wider use of the data. It is important to note that contracts themselves must align with overarching laws and regulations of the relevant jurisdiction, such as the EU General Data Protection Regulation (GDPR)[6]* or intellectual property laws. Researchers must be aware of and adhere to these overarching legal frameworks. For instance, when signing additional

*See section 'Privacy and Data Protection' in 'Cybersecurity Measures.'

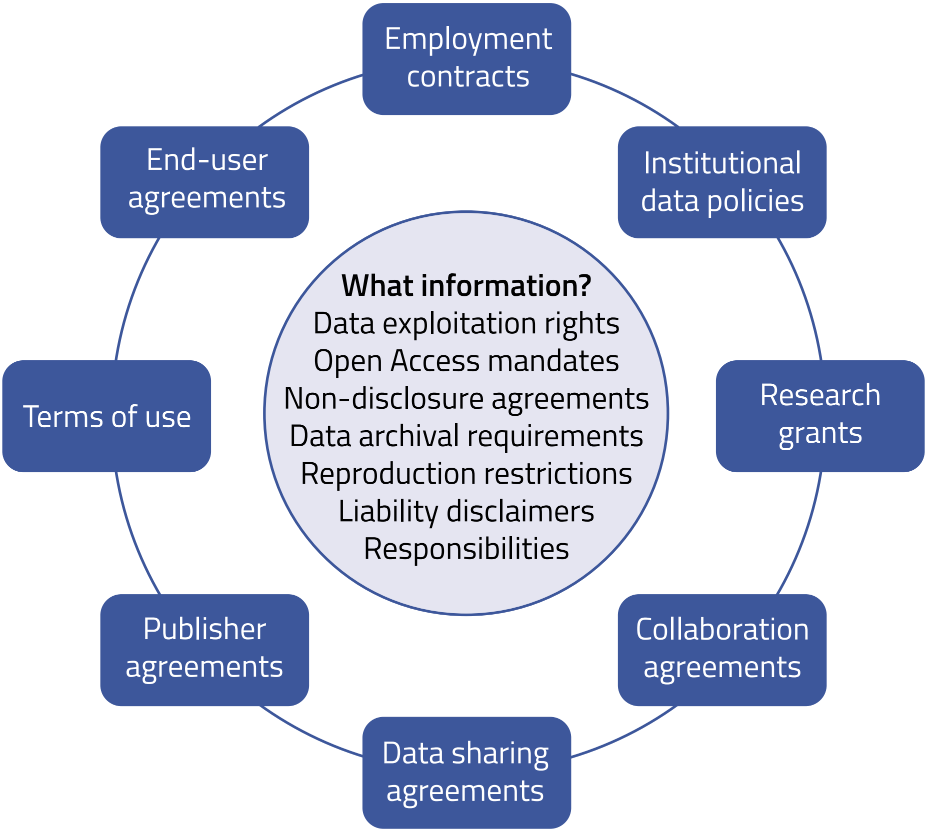


Figure 1: Overview of essential contracts that researchers encounter in their RDM. The core of the schematic highlights the legal provisions provided by these contracts.

agreements, such as data-sharing agreements among collaborators or accepting the terms and conditions of a data repository, researchers should ensure that their terms do not violate broader laws or any existing contracts with their institutions or funders. Furthermore, it is advisable to seek institutional legal advice when drafting additional contracts.

Researchers should also be aware of their institution's data policy, if available. These internal guidelines typically describe the expected institutional standards for data management, including storage, sharing, and usage. Such policies are likely to be found on the institution's official website under sections relevant to RDM, or can be requested from the institutional body responsible for RDM, such as the library or data stewards. The 'Data Policies' section on Forschungsdaten.org offers a comprehensive overview of institutional data policies.[7]

Funding agencies often impose specific conditions on how research data should be managed or published. For instance, they may require research outcomes to be made available through Open Access or specify a minimum archival period (see sections Open Access and Data Preservation).

Intellectual Property Rights

Intellectual property is an intangible asset resulting from creations of the mind, i.e., products of human intellect and creativity, such as artistic, literary, technical, or scientific creations.[8] Intellectual property rights (IPR) are the rights granted to creators of intellectual property, providing legal protection for their creations and ideas. These rights allow creators or any other rights holders to decide how, when, and where their creations are used and/or exploited.[8] In the context of research data, various forms of intellectual property may arise, making it important to understand and manage IPR carefully. For instance, when researchers use, share, or adapt the results of other studies, they must always consider the IPR associated with the original work.

Photo by [Elf-Moondance on Pixabay]

When considering the legal aspects of RDM in condensed-matter physics and materials science, the following forms of IPR, among others, are particularly relevant:

- **Copyright** grants authors exclusive rights to their literary, scientific, and artistic works. These rights include the use, reproduction, distribution, and public presentation of their works.[9]

- **Database rights** grant the database producer exclusive rights to reproduce and distribute the database as a whole or a qualitatively or quantitatively substantial part of the database and to make it available to the public.

- **Industrial property rights** include patents, trademarks, and utility models. Patents are granted for all inventions, provided that they are new, involve an inventive step, and are susceptible of industrial application.[10] Utility models offer faster and shorter protection for technical innovations than patents.[11,12] Trademarks protect brand names, logos, and distinctive products or services.[13]

Next, we explore the different forms of IPR in more detail and provide tips on what researchers need to consider when handling research data in the respective context.

Copyright

According to the copyright law, the creators of original works of literature, science, and art enjoy protection for their works.[9] In general, the eligibility of a work for copyright protection depends on two main factors:[14]

- **Individuality:** The work should be a personal creation or be derived from personal choices of its maker.

- **Originality:** The work should be an intellectual creation showing a certain level of originality; this means that the work must exhibit a degree of creativity, which exceeds standard or routine effort and reflects the author's personal intellectual creation.

Copyright is granted automatically to the creator without the need for registration, meaning that the exclusive rights of the copyright holder are effective from the moment the work is created, regardless of its publication.[15] Under German law, the copyright protection lasts for the entire life of the authors, and continues for 70 years after author's death.[16] Furthermore, copyright cannot be transferred; however, it may be inherited upon the author's death.[17]

> Content created using generative artificial intelligence (AI), such as the ChatGPT or the DALL-E, does not qualify for copyright protection under German law, because it lacks individuality, i.e., personal intellectual creation. As a result, there are no copyright or exploitation rights in AI-generated content.

Eligibility of research data for copyright protection

Whether research data can be protected by copyright laws, necessitates their case-by-case examination. For example, research data when considered as 'facts', such as machine-generated raw data*, typically do not fulfill the conditions of an intellectual creation, and therefore may not be protected by copyright. On the other hand, when data created by a researcher meet a certain degree of originality and individuality, such as a computer program for simulations, copyright is granted. Other examples that may be covered by copyright protection include: representations of a scientific or technical description such as drawings, sketches, plans, figures, micrographs, tables, etc.[18]

With respect to computer programs special provisions are in place. Computer programs are protected, if they represent individual works in the sense that they are the result of the author's own intellectual creation. The copyright protection applies to all forms of expression of a computer program including their drafts and preparatory design material. Therefore, distribution and reproduction of (parts of) a computer program, by any means, require the consent of the rights holder. It is worth noting that ideas and principles underlying elements of a computer program related to interfaces are not protected.[21]

*Raw data refer to original data that have been collected from a source (e.g., a scientific instrument) and not yet processed or analyzed. They are the primary information collected to perform an analysis and are defined by the absence of processing.[19,20]

Shared copyright

Scientific research in physics and materials science often involves collaboration among several researchers. When researchers contribute jointly to a project, their combined efforts may result in the creation of research results that are eligible for copyright protection. According to German law, if several persons have jointly created a work without being able to separately exploit their individual contributions to the work, they are joint authors (co-authors) of the work.[22] As co-authors, they collectively hold the rights to the work, meaning that the consent of each individual co-author is necessary for any action involving the work, such as modification, exploitation, publication, or reproduction.[22]

Photo by Scott Graham on Unsplash

In line with this law, the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) further defines a co-author in the research context in its 'Guidelines for Safeguarding Good Research Practice – Code of Conduct,' as someone who has made a genuine and identifiable contribution to a research project. This includes participation in the development and conceptual design of the project, the acquisition or provision of data, software, or sources, their analysis or interpretation, or the drafting of a manuscript.[23] It is worth noting that contributions that do not qualify for co-authorship should be appropriately acknowledged, e.g., in footnotes. Honorary authorship is not permitted, and holding a supervisory or leadership role does not in itself justify co-authorship.[23]

A co-author has the option to waive or give up their share of exploitation rights associated with the work, such as reproduction, distribution, and communication to the public,[24] but not the copyright.[17] This can be done by a clear statement expressing the decision to waive certain rights. For instance, scientists at research institutions often transfer exploitation rights to their institutions through employment contracts, or to journal publishers upon submission of their works, while retaining the copyright (see Contract Law).

Usage of copyrighted works in research

Using copyright-protected works in research activities is a common practice, yet it requires careful consideration of both legal and ethical dimensions. It is important to understand and to follow these aspects in order to maintain high standards of professionalism and to uphold the principles of responsible research. Below are key steps to guide researchers to adhere to these standards when using copyrighted works in research:

Identify copyrighted works: Be sure to recognize and acknowledge all materials you intend to use, such as text, images, graphics, and data sets. Look for copyright notices in metadata, watermarks, captions, README files, or the terms and conditions of hosting platforms or data repositories. Please note that the absence of a copyright notice does not imply the absence of copyright protection.

Identify the rights holder: Determine who holds the rights to the work you want to use. For instance, the rights to a figure in a poster which is published in Zenodo are held by the author(s) of the poster. However, when reusing content from publishers, such as an article in a scientific journal, the exploitation rights (i.e., the rights of use) are typically transferred from the author(s) to the publisher.

Check for licenses and permissions: Review the licenses or permissions associated with copyrighted works. Make sure that you understand and comply with their terms.

Obtain and document permission: If necessary, obtain written consent from the rights holder for your intended use. Keep a record of all permissions to ensure legal compliance.

Give proper attribution: When you are granted the right to use copyrighted works, always ensure proper attribution and acknowledgment of the original author(s). This practice is important, not only to comply with legal requirements, such as a license, but also to uphold standards of academic integrity and research ethics.

Usage without permission

In general, a copyrighted work can be used without the author's permission in the following ways:

- Copying or rephrasing the mere facts from a protected work in one's own style or words (e.g., paraphrasing the methodology or results from a research paper in your own words for a literature review);

- Quoting from the work (e.g., directly citing a specific equation from another publication);

- Using a work produced by government agencies (e.g., data from a government-funded experiment that is publicly available, unless specifically protected);

- Making a copy of the work for personal use (e.g., downloading a copyrighted research article to your personal device for reference or study, but not for redistribution).

In addition to these cases, there are specific exemptions in the German copyright law that allow the reuse of copyrighted works for educational, noncommercial, and scientific research purposes without author's permission.[25] For example, up to 15% of a published work may be reproduced, distributed, and made available to the public on a noncommercial basis for the illustrative purposes in educational institutions. However, the law prohibits reproduction of a work by means of recording on video or audio media or communication to the public of a work, while it is being publicly recited, performed, or presented.[26] As another example, the German copyright law addresses the use of works for non-commercial scientific research. It allows for up to 15% of a work to be reproduced, distributed, and made available to the public for personal scientific research or quality monitoring of scientific research.25 For personal scientific research, up to 75% of a work may be reproduced.[25] The law has special provisions for such exemptions, such as restrictions on recording on a video or audio medium for distribution. Given these complexities, it is crucial to ensure legality in every instance or to seek professional legal advice.

The free flow of knowledge is essential to scientific research. Licenses serve as practical tools to navigate the complexities of copyright law, allowing authors a way to manage how their work is accessed, used, and shared.

Licenses

A license is a permission granted by the copyright holder (licensor) to another party (the licensee) to use the copyrighted work in specific ways, often defining terms such as distribution, reproduction, modification, or commercial use.

The Creative Commons (CC) suite of six licenses is the most widely used framework for licensing academic outputs, offering different levels of openness.[27] CC licenses are available in simple language, legal, and machine-readable formats, ensuring that they are comprehensible for researchers, legal professionals, and digital systems used in data management, respectively. By answering a few simple questions regarding the permission for commercial use, creating derivative works, and restrictions on licensing of derivative works using tools like the CC License Chooser,[28,29] researchers are guided to the most appropriate CC license for their work. Issuing a CC

Table I: Creative Commons licenses.[31]

CC license	Description, permissions and restrictions
CC BY	Use, adapt, share. Author must be credited.
CC BY SA	Like CC BY. Derivatives must retain the original license.
CC BY ND	Like CC BY. No adaptations are allowed.
CC BY NC	Like CC BY. No commercial use is allowed.
CC BY NC ND	Like CC BY-NC. No adaptations are allowed.
CC BY NC SA	Like CC BY-NC. Derivatives must retain the original license.

license does not affect the copyright, but simply standardizes the way permission is granted to use research outcomes. All CC licenses allow assigning other licenses for the same work (dual or multi-licensing), and once granted, they are not revocable.[30] The core of six CC licenses are presented in Table I.

For researchers interested in maximum dissemination and visibility, the CC0 tool — also known as 'No Rights Reserved' — lets authors dedicate their protected work to the public domain and waive all related rights in their works to the fullest extent allowed by law. Additionally, the 'Public Domain Mark' serves as an indicator for works that have been identified as free of copyright, thereby facilitating their unrestricted use and further contributing to an open research environment. Table II shows the two CC public domain tools.

Table II: Creative Commons tools for freely available contents on public domain.[31]

Symbol	Description	Permissions
PUBLIC DOMAIN	Identified as copyright-free.	Use, adapt, share for any purpose.
PUBLIC DOMAIN	Dedicated to public domain.	

Both 'CC BY' and 'CC BY-SA' meet the open access,[32] with 'CC BY' becoming the standard for open access journal publishing recently.

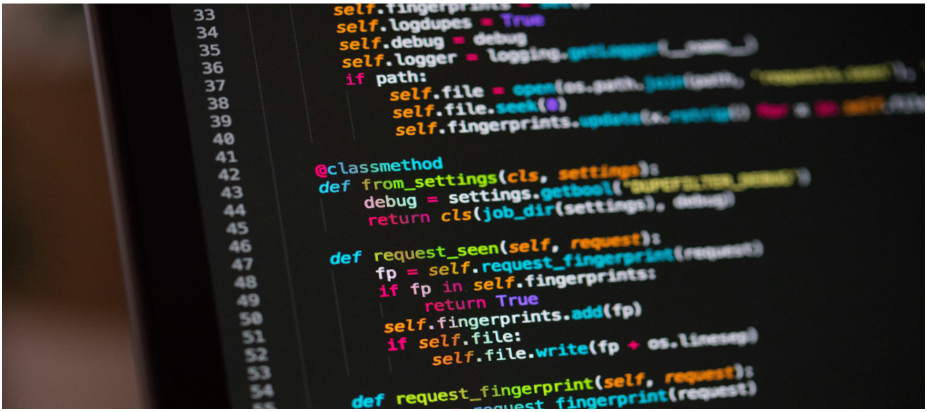
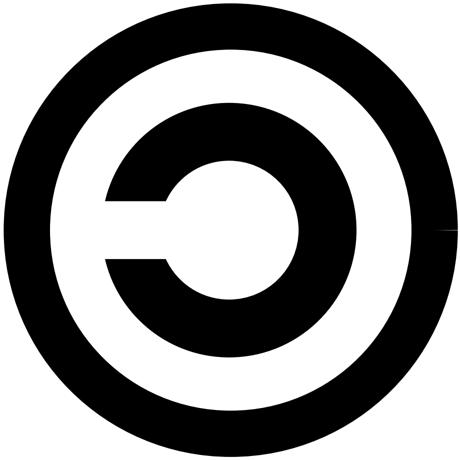


Photo by [Chris Ried on Unsplash]

Creative Commons licenses cannot be used for computer programs.[30] Placing software with no license in the public domain can lead to vulnerabilities, such as proprietary claims after minor modifications. On the users' side, when encountering unlicensed software or code, the terms of use on the corresponding website or hosting platform are decisive. If no such information is provided, users should treat the software as if it were under copyright protection.

For licensing research software and analysis codes, researchers often choose between two broad categories of open-source licenses, each with its unique characteristics and philosophy: copyleft licenses or permissive licenses.



> *"Copyleft is a general method for making a computer program free software and requiring all modified and extended versions of the program to be free software as well."*[33]

Copyleft licenses: These licenses require that any modifications to the licensed work must be distributed under the same license terms. They promote a collaborative and open environment, ensuring that derived works also remain open-source and accessible to the community. The most popular copyleft open-source licenses, in order of increasing permissiveness, are the **GNU General Public License (GPL)** followed by the **GNU Lesser General Public License (LGPL)**.[33]

The **GPL** maintains notifications of license and copyright terms and is appropriate for use in commercial, patent, and private contexts. Any software licensed with GPL is required to distribute all of its source code under the same license. Also when incorporating a module licensed under GPL into a software, the entire software must be licensed under the GPL, if distributed. This restriction makes the GPL a strong copyleft license. The LGPL, a more permissive variant of the GPL, allows linking to non-GPL code. The main difference to the GPL is that smaller projects accessed through larger licensed works do not require distribution of the larger project. Moreover, any modifications to the source do not have to be distributed under the same terms as the larger code project. This can, for example, be suitable for libraries and frameworks, making it easier to combine open-source and proprietary licensed components.[33]

The GPL permits private derivatives of software even for organizational use. Only if derivatives are publicly released, they must be shared under the GPL.

Permissive licenses: A permissive license is an open-source license that ensures the freedom to use, modify, and redistribute, while also allowing proprietary derivative. They typically impose fewer restrictions, making them appealing for interdisciplinary collaborations. The most common permissive licenses used in licensing scientific software are the **MIT license** and the **Apache 2.0 license**.

The **MIT license** is the most widely used open-source license and known for its simplicity and flexibility.[34] It permits anyone to do whatever they wish with the original source code as long as the license terms are included with the code or software. It does not impose copyleft restrictions, making it an attractive option for developers who want to allow their code to be used in a wide variety of contexts, including commercial projects. This license encourages collaboration by enabling a broader audience to utilize the work, while still recognizing the original authorship. Similar to the MIT license in its permissiveness, the **Apache 2.0 license** goes a step further by including explicit provisions regarding patents and contributions.[35] It provides legal clarity and protection for both contributors and users, ensuring that the IPR are clearly defined. The Apache 2.0 license also requires any modifications to be documented, facilitating collaboration and transparency in the development process. This makes Apache 2.0 suitable for projects that involve multiple contributors or that require clear legal protections.

The choice of license often depends on various factors such as the goals of the project, compatibility with other licenses and projects, and the desired level of control over the work. Table III offers a comparative overview of the main features of the four predominant open-source software licenses in research. For a broader selection of licenses and more detailed information, please refer to the suggested resources provided at the end of this document.[36–38]

Checking license compatibility becomes important when merging code under different licenses. Permissive licenses like MIT are generally more flexible, allowing their integration into projects with other licensing terms. However, licenses might have strict terms

Table III: Key features of commonly used open-source licenses.

	License	Integration rules	Compatibility	IPR provisions	Distribution obligations
Copyleft	**GPL**	Strict	Limited	No	Share entire source code
	LGPL	Moderate	Moderate	No	Share LGPLed changes
Permissive	**MIT**	Flexible	High	No	None
	Apache 2.0	Flexible	Moderate	Yes	None

that can conflict. Copyleft provisions, which require similar licensing for derivative work, along with patent clauses and distribution requirements, are the main reasons for license incompatibility. For example, the Apache license 2.0 has patent retaliation clauses that conflict with GPLv2 terms. However, these issues are resolved in GPLv3, making it compatible with the Apache license 2.0.[39] Official websites of licenses usually provide useful information regarding license compatibilities; for example, GNU maintains a list of GPL-compatible licenses.[40] Moreover, there are platforms such as the Joinup Licensing Assistant (JLA), which offer tools for compatibility check[41] as well as license analysis by categorizing licenses based on compatibility, legal aspects, and the level of support.[42]

Database Rights

A database is defined as a collection of works, data, or other materials arranged in a systematic or methodical way and individually accessible by electronic or other means.[43] In scientific contexts, databases rarely qualify for copyright protection, since they don't meet the threshold of originality; the arrangement of data is determined by practical needs which follow established protocols, rather than creative choices of database producer. Therefore, scientific databases generally fall outside the scope of copyright protection. However, the so-called 'related rights' might still apply to databases as intellectual achievements in collecting datasets.

Database-related rights protect the producer, usually a research institute or funding agency, which has made a substantial investment to produce the database. This investment may include financial resources, labor, time, or effort devoted to obtain and verify the data or to present the database.[44,45] The producer has the exclusive right of communication to the public, reproduction and distribution of the database contents.[44] As a result, using a database requires the consent of its producer, e.g., either through a license or as outlined in the terms of use of the database.

Database rights regulate the usage of data in specific ways. For example, consent from the database producer is necessary for the repeated and systematic reproduction, distribution, or public presentation of even minor parts of the database.[44] However, reproduction of a substantial portion a publicly available database is allowed without the database producer's consent for the purpose of:[18]

- Personal use (e.g. to satisfy personal curiosity);

- Personal scientific research (e.g. a copy is necessary for non-commercial purposes, with no intention for disclosure to others);

- Text and data mining;

- Illustration for teaching.

For the purposes of personal scientific research and illustration for teaching, e.g., illustrations in meetings and lectures at university, the law requires that the original works, from which the data have been derived, must be clearly cited.[44]

Table IV provides a general overview of actions protected by copyright or database rights that require author's consent (e.g., a license) as well as those which can be done without consent.

Table IV: Consent requirement for actions on copyright protected data or obtained from public databases.

	Protected by copyright	Protected by database rights
Without consent	▪ Copying for personal use (e.g. analysis, learning) ▪ Copying merely bare facts, presenting in own words ▪ Quoting	▪ Copying substantial portions for - personal use (e.g., analysis) - personal scientific research* - teaching* *under conditions[18]
With consent	▪ Adapting ▪ Sharing ▪ Publishing	▪ Repeated and systematic reproduction, distribution, or public presentation of even minor portions ▪ Copying of substantial portions

Industrial Property Rights

In the context of intellectual property law, patents, trademarks, and utility models are typically categorized as industrial property rights. A patent is a form of intellectual property that protects inventions and grants the patent holder a legal monopoly to prevent others from making, using, selling, or importing the patented invention. Trademarks protect brands and logos, which distinguish the products and services of one trader from those of others.[46] Utility models protect confidential business information that provides a competitive edge, such as formulas, practices, designs, patterns, or compilations of information. Finally, another example of industrial property rights, which might become relevant to research in the field of condensed-matter physics is the topography protection. This protection, which bears similarities to utility models, covers the rights associated with the three-dimensional arrangements of semiconductors and circuits.[47]

While the industrial property rights are less likely to intersect directly with RDM, considering them becomes important in specific situations. For instance, if research data support an invention that is novel, involves an inventive step, and has the potential for seeking a patent. Managing the sharing and utilization of data linked to potential patentable inventions is a critical aspect of RDM. Another scenario is when, for example, the physical properties of a research product are compared to those of an industrial product protected by industrial property rights.

International Data Transfer

Certain types of data are highly regulated, such as those related to the investigation, handling, storage, and disposal of specific substances such as radioactive materials or toxins. Adherence to these regulations in managing such data may become particularly important when storing data in publicly accessible repositories is intended, or international collaborations are involved.

In this context, the EU provides a guide for research on dual-use items – goods (including data), software, and technology typically used for civil purposes but also applicable to military uses.[48] The guide helps researchers and organizations in identifying, managing, and mitigating risks related to dual-use export controls. In Germany, the Federal Office for Economic Affairs and Export Control (Bundesamt für Wirtschaft und Ausfuhrkontrolle - BAFA), offers resources specifically designed for academics to enhance understanding and legal compliance on this topic. The presented materials are intended to raise awareness for the aims of export control, as outlined in the leaflet 'Export Control in Science & Research',[49] and to further support academics in implementing foreign trade law, as detailed in the handbook 'Export Control and Academia Manual (2nd Edition)'.[50] For instance, the manual describes a use-case regarding sharing the data in globally accessible databases; storing data in such databases requires a license if the data are listed in one of the relevant control lists and are neither already in the public domain nor part of basic scientific research. The fact that the origin of a database is in the EU does not affect the licensing requirement.[50]

Open Access

Open access refers to the online availability of scientific information, including peer-reviewed articles and research data, free of charge to the reader. Open access enhances research quality and efficiency and accelerates the progress of knowledge and innovation by sharing results, making them easier to reuse and verify.

Beyond its inherent advantages, open access transforms into a legal consideration in RDM, if it is given in grant agreements as a requirement (see section Contract Law). For example, Horizon Europe requires that all peer-reviewed scientific publications arising from funded projects be made available via open access. For this, two options are provided: publishing in a fully open-access journal, book, or platform, or self-archiving the publication in a trusted repository for immediate open access under a permissive license (e.g., CC BY, CC BY-NC, CC BY-ND or equivalents), especially when publishing in a subscription or hybrid journal. In addition to publications, research data must also be openly accessible under the principle 'as open as possible, as closed as necessary,' licensed under CC BY or dedicated to the public domain using CC0 or equivalent. Metadata should also be open under CC0 or equivalent.

Public research funding agencies in Germany also encourage and support open-access publication.[51,52] For instance, publications from DFG-funded projects are not required to be open access, yet the DFG encourages grant recipients to make their research outcomes publicly available This can be done by either directly publishing in recognized open-access journals or platforms (gold open-access route), or by uploading electronic archive copies of articles to the internet, such as to subject-specific or institutional repositories, in addition to the publication in traditional (paid) journals (green open-access route).[53,54]

In open-access publishing, researchers can decide to publish their findings either directly or after an embargo, to maintain competitive advantage. Funding agencies, while promoting open access, also acknowledge the potential commercial value of research results. Generally, open access does not imply an obligation to publish nor restricts commercial exploitation such as patenting.[52,56] The decision to publish research outcomes (regardless of being open access) should come after deciding whether to immediately publish the results or first protect them, e.g., by applying an embargo period, to pursue commercial opportunities or maintain competitive advantage. This is an important part of effective RDM.

‘Sherpa Romeo’ is a web-based service that gathers and analyzes global publisher and journal open access policies. It offers simplified summaries, helping authors to understand their rights.[55]

Data Preservation

Drawing by [GGbot on Pixabay]

Data preservation can generally be categorized into short-term preservation, i.e., during a project for the purpose of ongoing analyses and collaboration, and long-term preservation for archival purposes. Researchers are responsible for managing access to the preserved data, particularly when data contain third-party IPR (see Intellectual Property Rights) and sensitive data (see 'Privacy and Data Protection' in 'Cybersecurity measures').

Furthermore, preserved research data are valuable, since they enable building upon earlier knowledge and facilitate comprehensive meta-analyses. Researchers are often committed to the secure storage, processing, and preservation of their research data on the long-term for the purpose of validation and replication. Institutional data policies or funding agreements usually outline the specifics concerning data preservation, such as the medium or duration of data preservation. For example, the DFG recommends research data (generally raw data) to be archived in an accessible and identifiable way for a period of ten years at the institution where the data were produced or in cross-location repositories.[57] Non-compliance with these requirements can affect the evaluation of future research proposals, thereby influencing funding decisions.

Besides institutional repositories, several platforms and repositories provide data preservation solutions. 're3data.org' is a global registry of research data repositories from all academic disciplines offering an overview of existing repositories to help researchers find suitable platforms for preserving their data. Zenodo is a general-purpose open repository developed under the European OpenAIRE program and operated by CERN. Notable features include the ability to assign digital object identifiers (DOI) to datasets, DOI versioning for updated files, integration with GitHub for software archiving, and robust access control options (closed, open, or embargoed).

NOMAD, developed and operated by FAIRmat, is an open data infrastructure for managing, sharing, and publishing research data from materials science and condensed-matter physics and chemistry. In addition to DOI assignment for datasets, it enables users to incrementally create datasets. NOMAD processes uploaded files to extract structured data and rich metadata, providing a unified way to find, access, interoperate with, and reuse data from different codes, sources, and workflows.

Cybersecurity Measures

Cybersecurity is an important aspect of modern RDM, since research data are often sensitive and valuable, making them targets for cyber threats.[58] Improper RDM practices may lead to security risks, including breaches of confidentiality, integrity, privacy, and thus risk legal violations. In Germany, the importance of IT security is increasingly recognized due to the widespread reliance on safe and reliable information technology in different sectors, including scientific research.[59] Effective management of data security contains various organizational, contractual, and technical measures. A comprehensive recent study based on literature reviews and expert interviews has identified eight key requirements for the application of data sovereignty in the context of data exchange. These include access control, usage control, location of data servers, technical and technological aspects, legal considerations, organizational compliance, monetization, and data quality.[60] This section briefly reviews aspects of cybersecurity measures essential for researchers in condensed-matter physics. For a comprehensive understanding of data security measures in RDM, researchers are referred to the resources provided at the end of this guide.[59,61,62]

Data Security

In our digital world, cyber threats have become increasingly frequent and more sophisticated.[59] As a result, implementing security measures to protect data and information from unauthorized access has become more important. Depending on the need or the extent of implementation, there are several recommended practices that help researchers ensure data security:

▪ **Secure passwords:** Implementing complex and regularly updated passwords helps to protect data and systems against unauthorized access.

▪ **Authorization and access control:** Thoughtful setting of access control rules and mechanisms ensures that data are accessed only by authorized individuals, such as members within a research team and collaborators. This involves setting appropriate permissions and access levels for different users.

▪ **Data encryption:** Encrypting data is essential for keeping information secure and accessible only to authorized individuals. This practice is particularly important for research findings with patenting potential or in collaborative projects involving industrial external partners, e.g., to comply with provisions of non-disclosure agreements (see Contract Law).

▪ **Contracts for collaborative data usage and sharing:** When working in collaborative projects, drafting detailed contracts that clearly define the terms of data-usage rights, sharing, and confidentiality provisions is important. These contracts should also comply with legal obligations of each party (see Contract Law).

Data Integrity

Data integrity ensures the accuracy, completeness, and consistency of research data throughout its lifecycle. It involves implementing discipline-specific quality control measures, such as validation and verification, to maintain the correctness of data. Furthermore, regular backups and long-term archival of research data, as highlighted in the DFG Guidelines for Safeguarding Good Research Practice (guideline number 17),[23] enhance the traceability and verifiability of research outcomes. These practices are essential in protecting research data against loss, corruption, or unauthorized alterations, thereby preserving their integrity and supporting the future verification of research findings.

Privacy and Data Protection

In research, privacy concerns emerge when personal data are collected and stored. The primary challenge lies in using, preserving, and sharing such data while protecting identifiable information to ensure personal privacy. Personal data may cover various information such as health, financial, or behavioral details. The EU General Data Protection Regulation (GDPR) along with other regional laws enforce strict handling requirements for personal data.[6,63] While privacy and data protection are not the primary concerns in condensed-matter physics and materials science when compared to fields like medical or social sciences, they must be considered in certain scenarios. For example, when testing biomaterials on volunteers, the collected research data may include personal data subject to privacy and data protection laws.[62]

Research Ethics

Research ethics extend beyond legal considerations. The DFG guidelines highlight the researchers' responsibility to maintain a continual awareness of the risks associated with the misuse of research results such that risks can be recognized, assessed, and evaluated.[23] The potential consequences of a research project should be evaluated in detail, and the ethical aspects should be assessed. The book 'Ethics for Researchers' outlines twelve golden rules,[64] among which the following become particularly important when performing research in natural sciences: (i) respecting individual's privacy and data protection rights, (ii) trying to prevent research from being openly available for misuse or dual-use, (iii) respecting biodiversity and not imposing irreversible change that threatens the environment or ecological balance, and finally (iv) ensuring that research should serve the well-being of human society.

Although very rare in research in materials science, if human-related data arise, researchers must align their data practices with ethical standards, e.g., obtaining informed consent from participants. Informed consent has three key elements: clear research information, voluntariness, and participants' comprehension and awareness of consent implications.[64]

In the context of research ethics, the DFG acknowledges the role of generative artificial intelligence (AI) in scientific research and emphasizes the need for ethical responsibility. In a recent statement, the DFG also highlights the importance of transparency in AI usage, adherence to intellectual property laws, and the commitment of researchers to comply with research integrity principles.[65] They recommend clearly disclosing the use of AI in research, detailing the specific models employed, their intended purposes, and the scope of their application, in order to ensure that such usage enhances scientific quality and aligns with ethical research standards.[65]

Summary

Research data management involves controlling the data lifecycle - comprising planning, collection, analysis, preservation, sharing, and reuse - each stage associated with legal facets. Observing research ethics and ensuring cybersecurity, in particular privacy and data protection, where applicable, are integral across all stages of the data lifecycle. However, certain legal aspects become especially important when handling research data in the last three stages, namely preservation, sharing, and reuse. For data preservation, the main legal considerations are often outlined in institutional and contractual data policies. These policies typically provide details about the duration and medium for data archiving, along with guidelines for open-access data preservation. In the case of dual-use data, compliance with regulations on international data transfer is crucial. When it comes to sharing research data and publishing research outcomes, the core legal considerations are respecting third-party's IPR such as copyright and database rights and licenses, adhering to open-access publishing requirements. For reusing data from other research, the primary consideration should be respecting IPR associated with the original work.

Finally, it should be noted that while this guide tries to offer a comprehensive overview of the legal landscape relevant to the handling of research data, the complexities of individual research activities, institutional policies, and diverse laws at local, national, and international levels may necessitate specific legal advice.

References

1. European Commission, *Monitoring the Open Access Policy of Horizon 2020 – Final Report* (2021), https://data.europa.eu/doi/10.2777/268348.

2. A. E. Mansour, L. Rotheray, K. Helbig, S. Botti, H. B. Weber, M. Aeschlimann, and C. Draxl, FAIRmat, *Guide to Writing a Research Data Management Plan* (2023).

3. L. Kuschel, *Wem "gehören" Forschungsdaten?*, accessed on 22.03.2024, https://www.forschung-und-lehre.de/forschung/wem-gehoeren-forschungsdaten-1013.

4. J. Drexl, *The (Lack of) Coherence of Data Ownership with the Intellectual Property System*, in *Transition and Coherence in Intellectual Property Law* (Cambridge University Press, 2021), pp. 213–223.

5. R. M. Hilty, J. Drexl, and D. Harhoff, *Arguments Against Data Ownership - Ten Questions and Answers* (2017).

6. *General Data Protection Regulation (GDPR) – Official Legal Text*, accessed on 22.03.2024, https://gdpr-info.eu/.

7. Forschungsdaten.org, *Data Policies*, accessed on 22.03.2024, https://www.forschungsdaten.org/index.php/Data_Policies.

8. EUR-Lex, *Intellectual Property*, accessed on 22.03.2024, https://eur-lex.europa.eu/EN/legal-content/glossary/intellectual-property.html.

9. Based on FAIRmat Interpretation from UrhG §§ 1-2, https://www.gesetze-im-internet.de/englisch_urhg/englisch_urhg.html#p0013.

10. Based on FAIRmat Interpretation from PatG § 1, https://www.gesetze-im-internet.de/englisch_patg/englisch_patg.html.

11. Based on FAIRmat Interpretation from GebrMG § 1, https://www.gesetze-im-internet.de/gebrmg/__1.html.

12. Deutsches Patent- und Markenamt (DPMA), *Utility Models*, accessed on 22.03.2024, https://www.dpma.de/english/utility_models/index.html.

13. Based on FAIRmat Interpretation from MarkenG §§ 1-3, https://www.gesetze-im-internet.de/englisch_markeng/englisch_markeng.html#p0014.

14. Urheberrecht.de, N. P., *Schöpfungshöhe Im Urheberrecht: Wann Ist Diese Erreicht?*, accessed on 22.03.2024, https://www.urheberrecht.de/schoepfungshoehe/.

15. Based on FAIRmat Interpretation from UrhG § 12, https://www.gesetze-im-internet.de/englisch_urhg/englisch_urhg.html.

16. Based on FAIRmat Interpretation from UrhG §§ 64-65, https://www.gesetze-im-internet.de/englisch_urhg/englisch_urhg.html.

17. Based on FAIRmat Interpretation from UrhG §§ 28-29, https://www.gesetze-im-internet.de/englisch_urhg/englisch_urhg.html.

18. M. de Cock Bunin, B. van Dinther, C. G. Jeppersen de Boer, and A. Ringnalda, *The Legal Status of Research Data in the Knowledge Exchange Partner Countries*, 2011.

19. K. Briney, *Data Management for Researchers - Organize, Maintain and Share Your Data for Research Success* (Pelagic Publishing, 2015).

20. S. Venkataraman and P. Moura, *Raw Data, Backup and Versioning: What You Need to Know to Preserve Your Research Data.*, 2020.

21. Based on FAIRmat Interpretation from UrhG § 69a, https://www.gesetze-im-internet.de/englisch_urhg/englisch_urhg.html#p0602.

22. Based on FAIRmat Interpretation from UrhG § 8, https://www.gesetze-im-internet.de/englisch_urhg/englisch_urhg.html.

23. Deutsche Forschungsgemeinschaft (DFG), *Guidelines for Safeguarding Good Research Practice - Code of Conduct*, 2021.

24. Based on FAIRmat Interpretation from UrhG §§ 15-22, https://www.gesetze-im-internet.de/englisch_urhg/englisch_urhg.html.

25. Based on FAIRmat Interpretation from UrhG § 60c, https://www.gesetze-im-internet.de/englisch_urhg/englisch_urhg.html.

26. Based on FAIRmat Interpretation from UrhG § 60a (3) No. 1 and UrhG § 60c (4), https://www.gesetze-im-internet.de/englisch_urhg/englisch_urhg.html.

27. Creative Commons, *About CC Licenses*, accessed on 22.03.2024, https://creativecommons.org/share-your-work/cclicenses/.

28. Creative Commons, *Choose a License Beta*, https://chooser-beta.creativecommons.org/.

29. Creative Commons, *Choose a License*, accessed on 22.03.2024, https://creativecommons.org/choose/.

30. Creative Commons, *Frequently Asked Questions*, accessed on 22.03.2024, https://creativecommons.org/faq/.

31. Creative Commons, *Downloads*, accessed on 22.03.2024, https://creativecommons.org/mission/downloads/.

32. Max Planck Open Access, *Berliner Erklärung*, accessed on 22.03.2024, https://openaccess.mpg.de/Berliner-Erklaerung.

33. GNU Project - Free Software Foundation, *Licenses*, accessed on 22.03.2024, https://www.gnu.org/licenses/.

34. Open Source Initiative, *The MIT License*, accessed on 22.03.2024, https://opensource.org/license/mit/.

35. The Apache Software Foundation, *Apache License Version 2.0*, accessed on 22.03.2024, https://www.apache.org/licenses/LICENSE-2.0.

36. Open Source Initiative, *Licenses*, accessed on 22.03.2024, https://opensource.org/licenses/.

37. Choose a License, *Choose an Open Source License*, accessed on 22.03.2024, https://choosealicense.com/.

38. forschungslizenzen.de, *Lizenzierung von Forschungsdaten Open Science & Open Access*, accessed on 22.03.2024, https://forschungslizenzen.de/.

39. The Apache Software Foundation, *Apache License v2.0 and GPL Compatibility*, accessed on 22.03.2024, https://www.apache.org/licenses/GPL-compatibility.html.

40. GNU Project - Free Software Foundation, *Various Licenses and Comments about Them,* accessed on 22.03.2024, https://www.gnu.org/licenses/license-list.en.html#GPLIncompatibleLicenses.

41. Joinup Licensing Assistant (JLA), *Compatibility Checker*, accessed on 22.03.2024, https://joinup.ec.europa.eu/collection/eupl/solution/joinup-licensing-assistant/jla-compatibility-checker.

42. Joinup Licensing Assistant (JLA), *Find and Compare Software Licenses*, accessed on 22.03.2024, https://joinup.ec.europa.eu/collection/eupl/solution/joinup-licensing-assistant/jla-find-and-compare-software-licenses.

43. Based on FAIRmat Interpretation from UrhG § 4, https://www.gesetze-im-internet.de/englisch_urhg/englisch_urhg.html#p0030.

44. Based on FAIRmat Interpretation from UrhG §§ 87a-c, Division 6, https://www.gesetze-im-internet.de/englisch_urhg/englisch_urhg.html#p0737.

45. EUR-Lex, *Legal Protection: Databases*, accessed on 22.03.2024, https://eur-lex.europa.eu/EN/legal-content/summary/legal-protection-databases.html.

46. Deutsches Patent- und Markenamt (DPMA), *Trade Marks*, https://www.dpma.de/english/trade_marks/index.html.

47. Deutsches Patent- und Markenamt (DPMA), *Topographies*, accessed on 22.03.2024, https://www.dpma.de/english/utility_models/topographies/index.html.

48. European Comission, *EU Compliance Guidance for Research Involving Dual-Use Items*, 2019.

49. Bundesamt für Wirtschaft und Ausfuhrkontrolle (BAFA), *Export Control in Science & Research*, 2019.

50. Bundesamt für Wirtschaft und Ausfuhrkontrolle (BAFA), *Export Control and Academia Manual (2nd Edition)*, 2023.

51. Deutsche Forschungsgemeinschaft (DFG), *DFG Statement on the Establishment of "COAlition S" to Support Open Access*, Information Für Die Wissenschaft Nr. 26, accessed on 22.03.2024, https://www.dfg.de/de/aktuelles/neuigkeiten-themen/info-wissenschaft/2022/info-wissenschaft-22-26.

52. Bundesministerium für Bildung und Forschung (BMBF), *Open Access in Deutschland - Die Strategie Des Bundesministeriums Für Bildung Und Forschung* (2018).

53. Deutsche Forschungsgemeinschaft (DFG), *Guidelines Publication Grants DFG Form 51.10 - 09/22*.

54. J. Fournier, *Roads to Knowledge: Activities for Promoting Open Access by the DFG*.

55. Securing a Hybrid Environment for Research Preservation and Access (SHERPA), accessed on 22.03.2024, https://www.sherpa.ac.uk/romeo/.

57. Deutsche Forschungsgemeinschaft (DFG), *DFG Guidelines on the Handling of Research Data* (2015).

58. Forschung & Lehre, *Cyberangriff: Helmholtz-Zentrum in Berlin Gehackt*, accessed on 22.03.2024, https://www.forschung-und-lehre.de/management/helmholtz-zentrum-berlin-gehackt-5700.

59. Bundesministerium für Bildung und Forschung (BMBF), *IT-Security*, accessed on 22.03.2024, https://www.bmbf.de/bmbf/en/research/digital-world/it-security/it-security_node.html.

60. S. Biehs and J. Stilling, *Identification of Key Requirements for the Application of Data Sovereignty in the Context of Data Exchange*, in *Proceedings of the 57th Hawaii International Conference on System Sciences* (2024), pp. 4248–4257.

61. P. A. Grassi et al., *Digital Identity Guidelines: Authentication and Lifecycle Management* (2017).

62. European Commission, *Ethics and Data Protection* (2021).

63. Datenschutz-Grundverordnung (DSGVO) – Finaler Text der DSGVO Inklusive Erwägungsgründe, accessed on 22.03.2024, https://dsgvo-gesetz.de/.

64. European Comission, *Ethics for Researchers - Facilitating Research Excellence in FP7* (2013).

65. Deutsche Forschungsgemeinschaft (DFG), *Statement by the Executive Committee of the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) on the Influence of Generative Models of Text and Image Creation on Science and the Humanities and on the DFG's Funding Activities* (2023).

About FAIRmat

FAIRmat is a consortium of the German Research-Data Infrastructure, NFDI. FAIRmat is creating a federated data infrastructure for materials data and supporting scientists in the field of condensed-matter physics and the chemical physics of solids in making their research data FAIR (Findable, Accessible, Interoperable, and Repurposable).

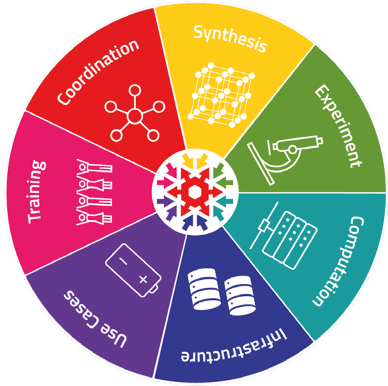
Our mission

1. Develop and maintain a federated FAIR data infrastructure for materials data with built-in tools and standards to support scientific collaboration and proper research data management (RDM) practices.

2. Support the scientific community to introduce and maintain high standards of reproducibility, research integrity, and compliance with ethical and legal requirements by adopting proper RDM practices based on the FAIR data principles.

How we work

FAIRmat is governed by a complementary team of internationally renowned researchers, who are actively embedded into their (sub) communities. Its infrastructure is developed by a dedicated team of coworkers with complimentary expertise. FAIRmat integrates data obtained from synthesis, experimental characterization, theory and computations, and various applications. The organization is divided into seven Areas, each addressing specific key aspects of the project.



Our services

NOMAD is a free web-service that lets you share your data or use comprehensive data that others provide. You can use NOMAD to organize, analyze, share, and publish your materials-science data, as well as explore, download, and analyze your colleagues' data.

NOMAD Oasis lets you create your own NOMAD. Get an instant overview of all your group's data, increase productivity, and implement research data management plans with ease.

The NOMAD Encyclopedia allows users to see, compare, explore, and understand materials data.

FAIRmat offers training and educational materials on RDM, writing RDM plans, and using NOMAD tools.

Disclaimer

This guide is designed to aid researchers and is offered for informational purposes only. It is not meant to serve as a substitute for legal consultation or interpretation of applicable laws. Neither FAIRmat nor the authors assume liability for any complications arising from the use of information within this document.

Notes

www.fairmat-nfdi.eu
fairmat@physik.hu-berlin.de
@fairmat.bsky.social
company/fairmat-nfdi
@TheNOMADLaboratory

This guide provides an overview of the key legal considerations relevant to the various stages of the data lifecycle. It is specifically designed for graduate students, postdoctoral researchers, and principal investigators who wish to understand and navigate the legal aspects of research data management in their research projects.

FAIRmat