

Research Data Policies: Best Practices and Essential Elements

Dr. Özlem ÖZKAN

Training Officer / Helmholtz Metadata Collaboration, Hub Matter Berlin

Available via: <https://doi.org/10.5281/zenodo.10451379>

www.helmholtz-metadaten.de



Key Points

Background Research Data Policies

Purpose of Research Data Policies

Best Practices & Key Elements of Data Policies



ÖZKAN, Ö. (2023) <https://doi.org/10.5281/zenodo.10198909>

Why Research Data Management?

Data is **fragile** and **easily lost**.

Most Scientific Research Data From the 1990s Is Lost Forever

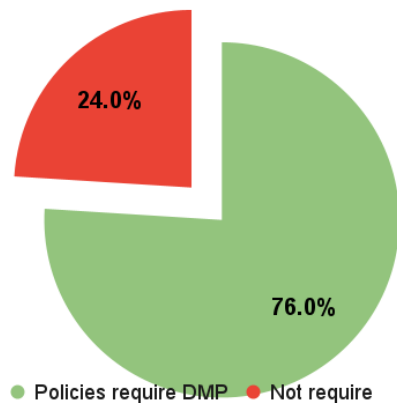
A new study has found that as much as 80 percent of the raw scientific data collected by researchers in the early 1990s is gone forever, mostly because no one knows where to find it.

By Danielle Wiener-Bronner

<https://www.theatlantic.com/national/archive/2013/12/scientific-data-lost-forever/356422/>

Why Research Data Management?

Growing number of **funders** and **publishers mandate research data requirements**



Analysis of international funder data policies

Joris van Rossum

In October 2021, STM commissioned a research on funder data policies. The top 100 funders based on number of Crossref records were selected, and analyzed for the availability of data policies. These data policies were analyzed according to the elements of the journal data policy framework as developed by Hrynaszkiewicz et al. (<https://datascience.codata.org/article/10.5334/dsj-2020-005/>). This research will be used as input for more alignment between funder and data policies.

Files (2.1 MB)		
Name	Size	
STM analysis funder data policies (October2021).xlsx	2.1 MB	Download
md5:e517d7e79ab4f1035849ee131d436aa4		

Joris van Rossum. (2021). Analysis of international funder data policies (1.0) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.5643352>

Why Research Data Management?

Good management helps to **prevent errors** and **increases the quality of the analyses**, **saves time** and **resources** in the long run.



In May 2018, the EU Commission published a report: not having FAIR research data costs the European economy at least **€10.2bn/year**

Cost of not having FAIR research data

Cost-Benefit analysis for FAIR research data

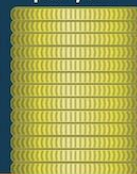
[doi: 10.2777/02999](https://doi.org/10.2777/02999)

here's one example of the gains arising from open research data

Bioinformatics Institute

€1.3 billion
per year

Benefits identified by the European Bioinformatics Institute to users and their funders just by making scientific information freely available to the global life science community...



equivalent to **more than 20 times** the direct operational cost of the Institute

Source: Charles Beagrie Ltd. for EMBL-EBI

<https://www.embl.org/documents/wp-content/uploads/2021/10/EMBL-EBI-impact-report-summary-2021.pdf>

Why Research Data Management?

Helps to **comply with legislative requirements** such as GDPR.

According to [GDPR enforcement tracker](#), up to now **23 universities** have been imposed fines and penalties.

GDPR Enforcement Tracker








tracked by **CMS**
law-tax-future

The CMS Law GDPR Enforcement Tracker is an overview of fines and penalties which data protection authorities within the EU have imposed under the EU General Data Protection Regulation (GDPR, DSGVO). Our aim is to keep this list as up-to-date as possible. Since not all fines are made public, this list can of course never be complete, which is why we appreciate any [indication of further GDPR fines and penalties](#). Please note that we do not list any fines imposed under national / non-European laws, under non-data protection laws (e.g. competition laws / electronic communication laws) and under "old" pre-GDPR laws.

New features: "ETId" and "Direct URL"
We have assigned a unique and permanent ID to each fine in our database, which makes it possible to precisely address fines, e.g. in publications. Once an "ETId" has been assigned to a fine, it remains the same, even if the fine is overturned or amended by courts at a later date, or if we add fines that were issued chronologically before. The "Direct URL" (click "+" or on a specific ETId to view details of a fine) can be used to share fines online, e.g. on Twitter or other media.

Show entries

Search:

ETId	Country	Date of Decision	Fine [€]	Controller/Processor	Quoted Art.	Type	Source
<input type="text" value="Filter Column"/>	<input type="text" value="Filter Column"/>	<input type="text" value="Filter Column"/>	<input type="text" value="Filter Column"/>	<input type="text" value="Filter Column"/>	<input type="text" value="Filter Column"/>	<input type="text" value="Filter Column"/>	<input type="text" value="Filter Column"/>
 ETId-896	 LUXEMBOURG	2021-10-13	18,000	Unknown	Art. 37 (7) GDPR, Art. 38 (1), (2) GDPR, Art. 39 (1) b) GDPR	Insufficient involvement of data protection officer	link
 ETId-895	 LUXEMBOURG	2021-10-13	13,200	Unknown	Art. 38 (1) GDPR, Art. 39 (1) b) GDPR	Insufficient involvement of data protection officer	link
 ETId-894	 LUXEMBOURG	2021-10-06	5,300	Unknown	Art. 5 (1) c) GDPR, Art. 13 GDPR	Non-compliance with general data processing principles	link
	 IRELAND	2021-08-20	1,500	MOVE Ireland	Art. 5 (1) f) GDPR, Art.	Insufficient technical and	link

Home License Privacy Imprint



These should be regulated in a written document!

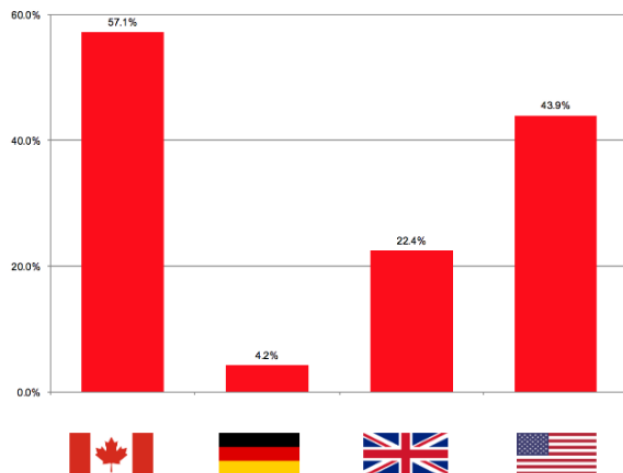
Research Data Policy



What is a Research Data Policy?

A data policy is a set of rules/guidelines that govern the **collection**, **management**, and **sharing** of research data.

Percentage of universities with a Research Data Policy: Canada, Germany, UK, US



Laurence Horton. (2016, June 30). 7.1 'The Road to Data Sharing is Paved with Good Intentions': Looking at UK Research Data Policies. Zenodo.

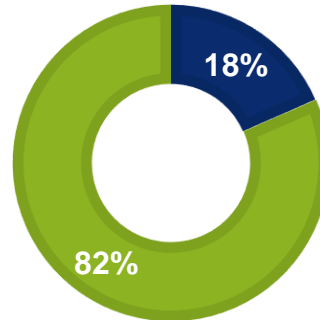
<https://doi.org/10.5281/zenodo.3607367>

Now



According to [forschungsdaten.org](https://www.forschungsdaten.org/index.php/Hauptseite)¹, **71 German universities** have a Research Data Policy and as of 2023, there are **385 universities**² in Germany.

■ Data Policy ■ No Data Policy



[1] Forschungsdaten.org <https://www.forschungsdaten.org/index.php/Hauptseite>

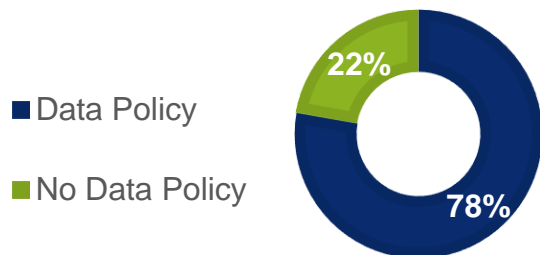
[2] All 385 Universities in Germany | Rankings & Reviews 2024 (universityguru.com) <https://www.universityguru.com/universities--germany>

The Helmholtz Centers are implementing high level requirements

Helmholtz Open Science Policy: all centers should have a data policy in place

“All Helmholtz Centers will establish detailed procedures for managing research data in publicly available policies...”³

14 out of 18 Helmholtz institutes have data policy.

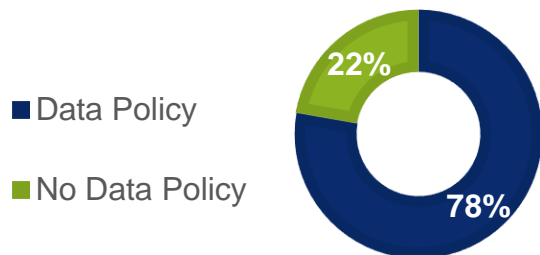


The Helmholtz Centers are implementing high level requirements

Helmholtz Open Science Policy: all centers should have a data policy in place

“All Helmholtz Centers will establish detailed procedures for managing research data in publicly available policies...”³

14 out of 18 Helmholtz institutes have data policy.



Other reasons to have a Research Data Policy

Need to respond national research funder's and other funders' guidelines, and organizations' principles:

- **DFG - German Research Foundation:** Guidelines for Safeguarding Good Research Practice¹
*Institutions had a transitional period **until July 31, 2023**, to fully implement the guidelines **in a legally binding manner**.
The implementation of this Code is essential for institutions **to be eligible for DFG funding**.*
- **Science Europe** strongly recommends research organizations to have a data policy for sustainability of research data management.²
- **European Open Science Cloud (EOSC), European Research Council (ERC), International Science Council (ISC)**, etc. underline importance of open science and FAIR (Findable Accessible Interoperable Reusable) principles.

[1] DFG Guidelines for Safeguarding Good Research Practice (2022) <https://doi.org/10.5281/zenodo.6472827>

[2] Practical Guide to Sustainable Research Data (2021). <https://doi.org/10.5281/zenodo.4769703>

Best Practices & Key Elements of Data Policies



FAIRsFAIR: FAIR-enabling data policy checklist



Metadata sharing	The policy should make clear any expectations around metadata sharing in particular when the data cannot be shared openly or if data are no longer accessible. An emphasis should be placed on making clear whether metadata sharing is required or is suggested.		• The policy clearly states that sharing metadata for selected data outputs is required.
			• The policy encourages metadata sharing but does not require it.
			• The policy does not address metadata sharing or lacks clarity over what is expected of researchers when it comes to sharing metadata.
Data Management Plan (DMP)	Policies should provide clarity over whether there is an expectation for researchers to develop a DMP as part of their research.		• The policy makes clear whether a data management plan should be developed.
			• The policy does not clearly state whether a data management plan should be developed.
Timing of DMP	Where DMPs are required, policies should provide clarity over the timing of their preparation and delivery (pre award, in award, post award). If multiple versions are required at different stages, this should be made clear.		• The policy makes clear at what stage the DMP should be prepared.
			• The policy lacks clarity about when the DMP should be prepared.
			• The policy does not include an expectation for a DMP.

This checklist is designed to assist policy-makers at all levels to ensure that their data policies are in alignment with the FAIR (Findable, Accessible, Interoperable, Reusable) Principles.

→ The checklist consists of 28 policy elements to be addressed in order to align with the FAIR principles.

Davidson, J., et al. (2022). FAIR-enabling Data Policy Checklist (2.0). Zenodo. <https://doi.org/10.5281/zenodo.6225775>

Clear Titling: Policies should have a clear title indicating the scope and applicability.

*Ensure your policy has **a title** that clearly identifies its owner and subject matter.*

... University Research Data Policy

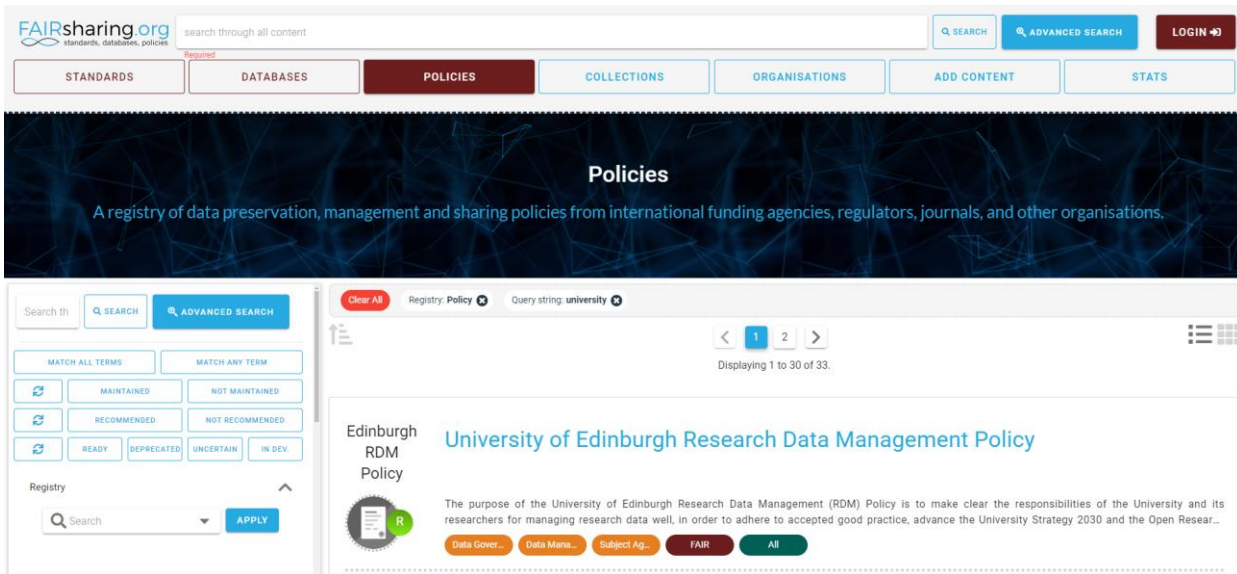
Validity and Versioning: Differentiating between the creation and implementation dates, and using PIDs for versioning.

*Clearly state the policy's effective **date and establish a scheduled review date.***

*Assign **Persistent Identifiers** (PIDs) to versioned policies for seamless tracking.*



Registration and Availability: Registering policies with services like FAIRsharing and making them openly available online in a structured format.

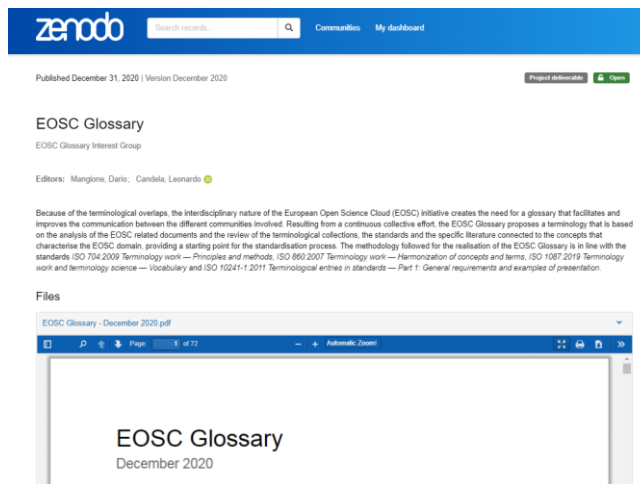


The screenshot shows the FAIRsharing.org website interface. At the top, there is a search bar with the text "search through all content" and buttons for "SEARCH", "ADVANCED SEARCH", and "LOGIN". Below the search bar are navigation tabs: "STANDARDS", "DATABASES", "POLICIES" (highlighted), "COLLECTIONS", "ORGANISATIONS", "ADD CONTENT", and "STATS". The main content area has a dark blue header with the word "Policies" and a subtitle: "A registry of data preservation, management and sharing policies from international funding agencies, regulators, journals, and other organisations." Below this, there is a search bar with "Search th" and "ADVANCED SEARCH" buttons. A sidebar on the left contains filters for "MATCH ALL TERMS", "MATCH ANY TERM", "MAINTAINED", "NOT MAINTAINED", "RECOMMENDED", "NOT RECOMMENDED", "READY", "DEPRECATED", "UNCERTAIN", and "IN DEV.". The main content area displays a list of policies, with the first one being the "University of Edinburgh Research Data Management Policy". The policy description states: "The purpose of the University of Edinburgh Research Data Management (RDM) Policy is to make clear the responsibilities of the University and its researchers for managing research data well, in order to adhere to accepted good practice, advance the University Strategy 2030 and the Open Resear...". Below the description are buttons for "Data Govern...", "Data Mana...", "Subject Ag...", "FAIR", and "All".

<https://fairsharing.org/search?fairsharingRegistry=Policy&q=university>

Definition of Scope: Providing clear definitions for what the policy covers, such as research data and software.

EOSC Glossary:



EOSC Glossary Interest Group. (2020). EOSC Glossary.

Zenodo. <https://doi.org/10.5281/zenodo.4472643>

Data Management Planning (DMP): Requiring the development of a DMP as part of research practice.

Four options, the policy :

- requires the development of a data management plan.
- encourages the development of a data management plan.
- refers to data management plans.
- does not address data management plans.

* But **requirement of a DMP** requires many resources to consider:

- Data Stewards/Managers
- Infrastructure to support short and long term preservation; data sharing; sensitive data handling etc.
- A DMP tool

Data Sharing Requirements: Stating requirements for data sharing, including valid reasons for not sharing and exceptions.

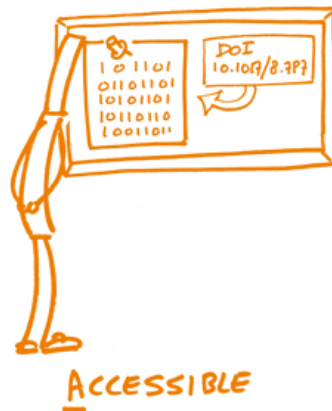
The policy

- can **require** data sharing.
- can **encourage** data sharing.
- **does not** address data sharing.

Metadata Sharing Expectations: Clear expectations around metadata sharing, **especially** when data can't be shared openly.

- Ensure metadata is stored in accessible databases (*use standard metadata schemas relevant to your field*)

Data Protection Legislation: Outlining expectations for compliance with laws like **GDPR**.



Open Science Training Handbook
<https://doi.org/10.5281/zenodo.1212496>

Encouragement of Reuse and Licensing: Requiring appropriate licenses and providing guidance to help researchers select these.

"Choose licenses that promote data sharing while respecting data ownership."

Helmholtz Open Science Policy:

"... the provision of open access to the deposited **data** as early as possible and within the time limits specified in the DMP under the latest available version of the **Creative Commons Attribution International Public License (CC BY)** or the **Creative Commons Public Domain Dedication License (CC0)**, or a license that grants equivalent rights"

"... the **metadata** of the deposited publications are accessible under a **Creative Commons Public Domain Dedication (CC0)** license or another equivalent license."¹

[1] Helmholtz Open Science Policy. Version 1.0. (2022). <https://doi.org/10.48440/os.helmholtz.056>

Retention and Preservation: Clear guidance on the retention period for outputs and linking to relevant preservation policies.

According to DFG Guidelines for Safeguarding Good Research Practice “*Retention of research data should be guaranteed for **10 years**.*”¹

However, in **funder policies** this number vary or there **can legal or contractual reasons** so it is better to formulate the statement considering them.

Example: MDC Research Data Policy

*“Research data and related material should be retained for a **minimum** of ten years after acquisition or generation based on the recommendation of the DFG. **Longer or shorter retention periods** prevail in accordance to legal regulations, funders' and other contractual requirements (e.g. clinical trials, patents).”²*

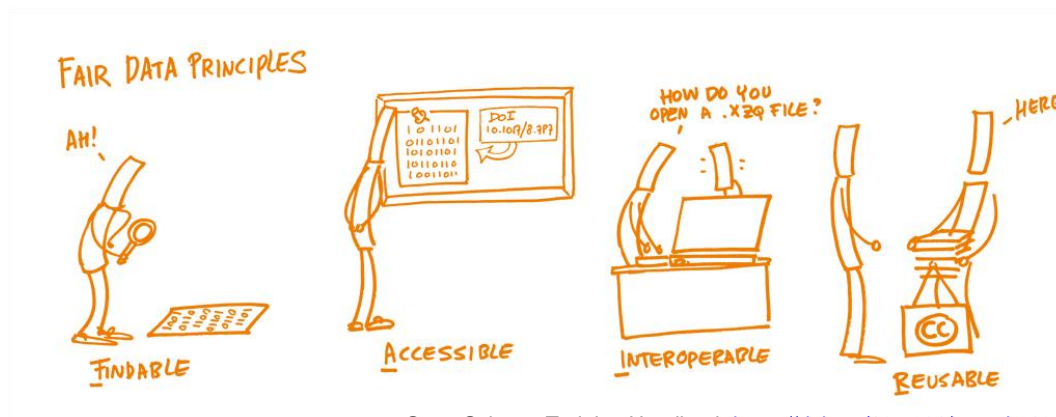
[1] DFG Guidelines for Safeguarding Good Research Practice (2022) <https://doi.org/10.5281/zenodo.6472827>

[2] <https://www.mdc-berlin.de/research-data-management/research-data-management-policy>

Alignment with FAIR Principles: Ensuring that the policy aligns with FAIR Principles.

There can be three options:

- The policy makes **explicit reference** to the FAIR Principles and **aligns with** FAIR.
- The policy **does not specifically refer** to the FAIR Principles **but aligns with** FAIR.
- The policy **does not address** the FAIR Principles explicitly or implicitly.



Open Science Training Handbook <https://doi.org/10.5281/zenodo.1212496>

Thank you

Contact: oezlem.oezkan@helmholtz-berlin.de

