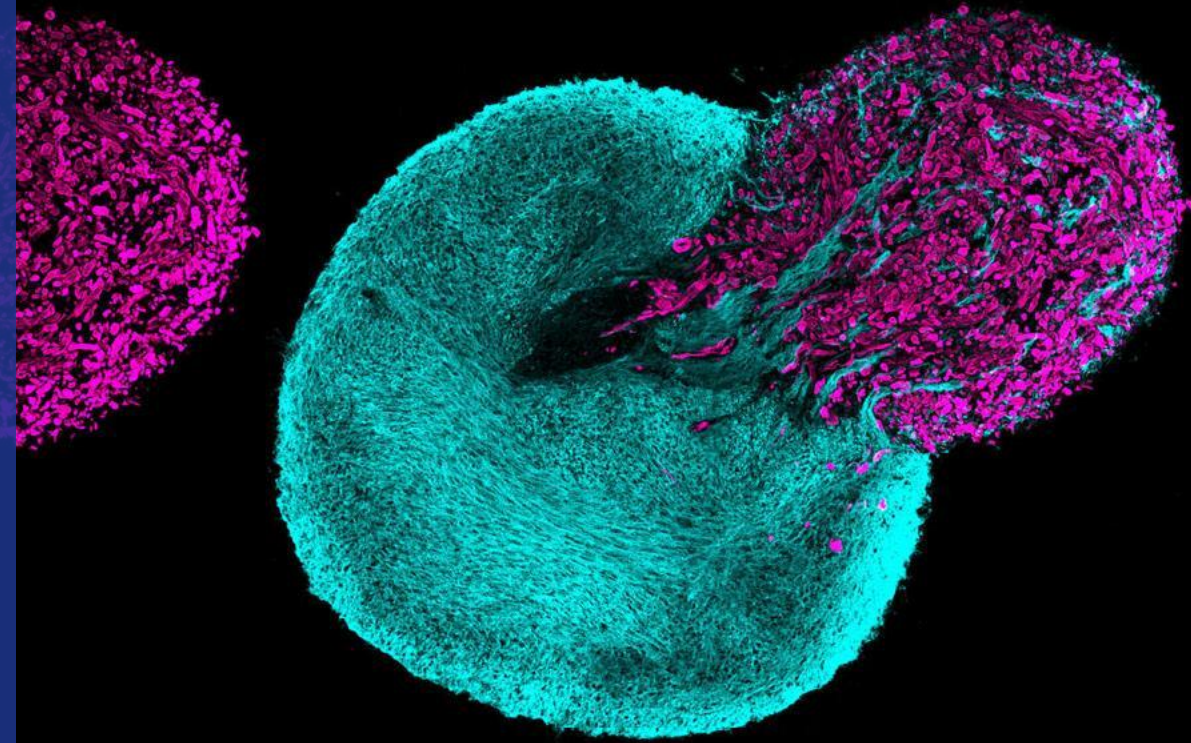


Intro to RDM & DMPs

- Open Science Café -
UNIRI 27.02.2024.

Inga Patarčić, PhD
<https://orcid.org/0000-0003-0527-6252>



Agenda

RDM introduction and 10 reasons to do it

RDM as a career path

Overview of requirements from funders, publishers & institutes for RDM

DMPs & FAIR Wizard

A Data Management Cautionary Tale

https://www.youtube.com/watch?v=N2zK3sAtr-4&ab_channel=NYUHealthSciencesLibrary



Search



DEFINITION OF RDM

Research Data Management is the process of providing the appropriate labeling, storage, and access for data at all stages of a research project.



Research Data Lifecycle

VISION AND MISSION OF RDM

To support researchers in all their data related activities in order to achieve a more **effective use of resources, transparency, re-use, participation, accountability and reproducibility.**



Research Data Lifecycle

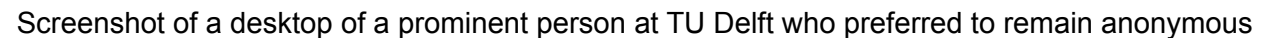
FOLLOWS:

- Why Manage Data?
- RDM in practice

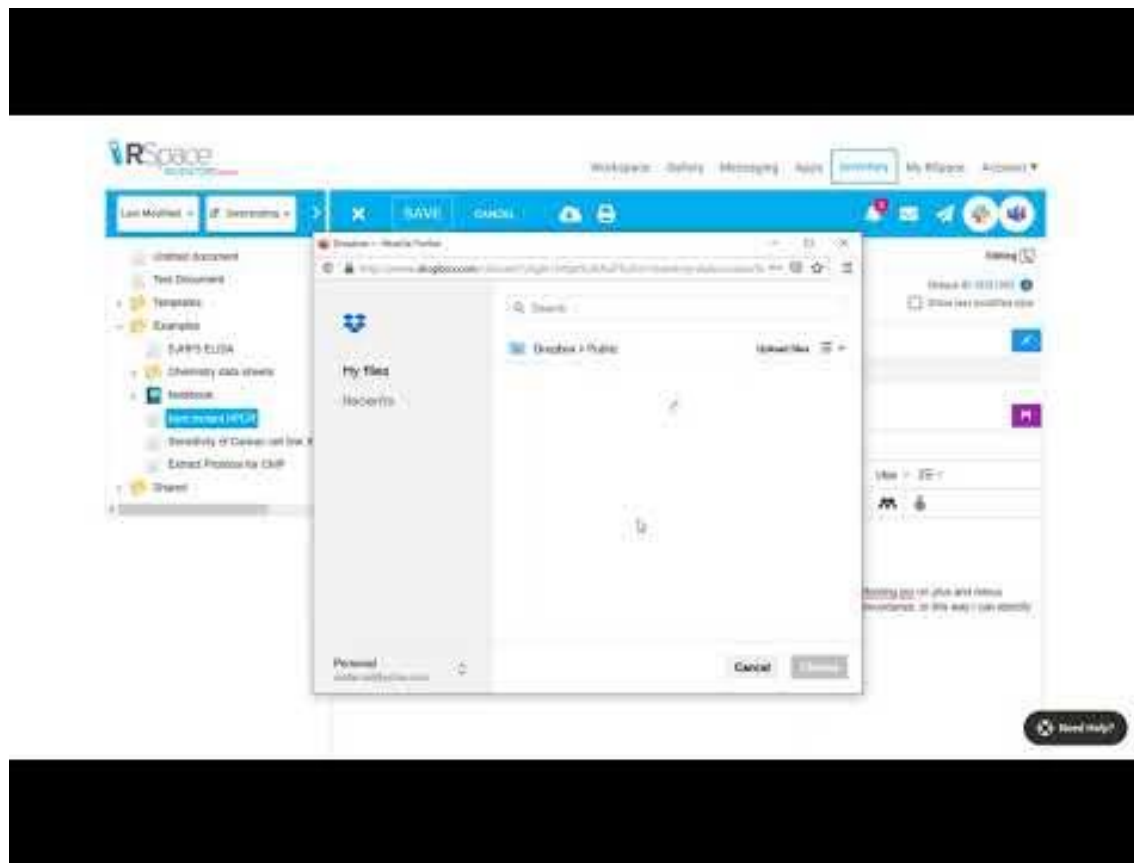
10 reasons to do data management



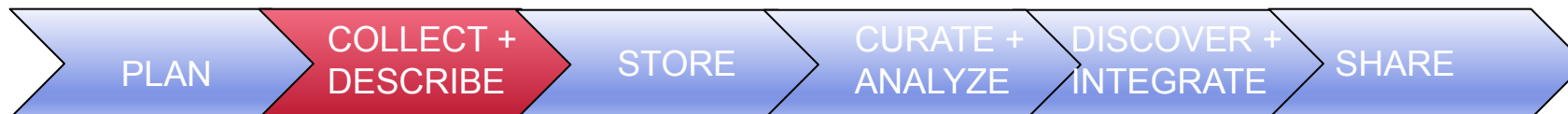
MAX
DELBRÜCK
CENTER



2. Ensures accountability



https://www.youtube.com/watch?v=q2CEnZRC1_o&t=1s&ab_channel=researchspace



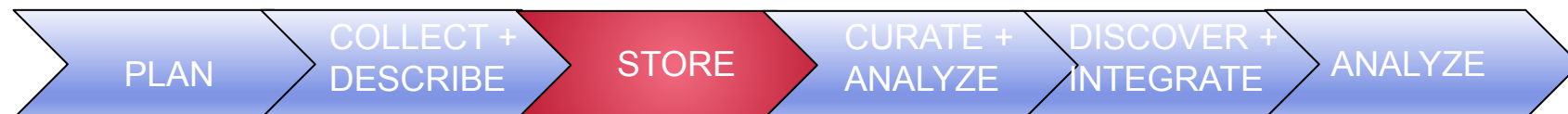
3. Data is easily lost.

Most Scientific Research Data From the 1990s Is Lost Forever

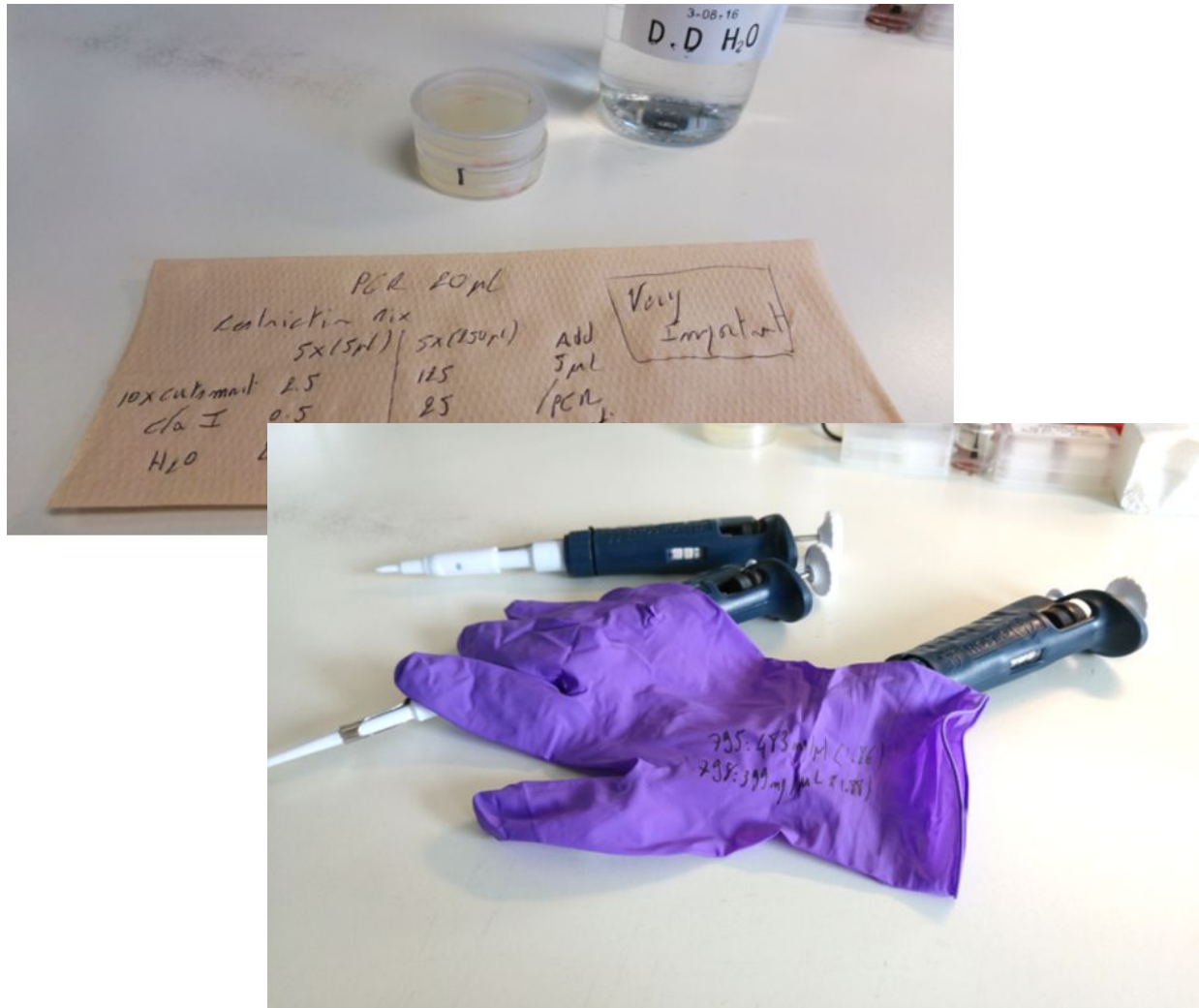
A new study has found that as much as 80 percent of the raw scientific data collected by researchers in the early 1990s is gone forever, mostly because no one knows where to find it.

By Danielle Wiener-Bronner

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



3. Data is easily lost. (But, it's not going to happen to me!)

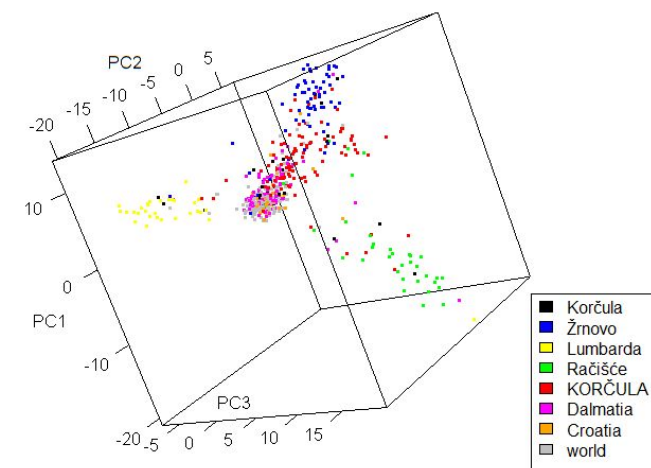


Credit for the photo: Peter Murray-Rust
<http://blogs.ch.cam.ac.uk/pmr/2011/08/01/why-you-need-a-data-management-plan>

4. High quality data is crucial for developing any (AI) model

Garbage-in garbage out, ChatGPT

	A	B	C	D	E	F	G	H	I
1	,,,Research Integrity,Research Integrity,Citizen Science ,Citizen Science ,Citizen Science ,Scholarly Publishing ,Scholarly Pu								
2	,,,INFORMATION & DATA LITERACY ,DIGITAL CONTENT CREATION ,COMMUNICATION & COLLABORATION ,DIGITAL CON								
3	Institute,Title of course,Description,"Research Integrity & Ethics, GDPR","Copyright, Intellectual Property (in the Digital E								
4	VIB,Biomedical Research Ethics,"During this training, different aspects of ,Äöresponsible research,Äö will be discussed.B								
5	VIB,Image Ethics & Poster Design,"In this training course, we want to teach the basic ethic aspects of image editing and								
6	VIB,Respons IP ZI,,,,,,1								
7	VIB,Scientific to gain insig to make a st IP ZI,,,LB, If IP, ZI" ,,,1								
8	VIB,Scientific Writing & Publishing with impact,"Learn how to make an impact with your research! You are cordially invit								
9	IGC,INFORM ,,b) Students ,,b) Students ,,b) Students ,,c) Students ,,d) Students ,,e) Students ,,b) Students ,,c) Students I								
10	IGC,RESEAR IP,ZI" ,LB IP,ZI",1								
11	IGC,Social Media for Scientists Course,Members of the public are increasingly using online environments and specially sc								
12	IGC,SCIENCE COMMUNICATION COURSE,"n an increasingly complex world, communicating science to non-specialists is								
13	IC,Intellectua ZI IP,,,,,,1								
14	IC,"Raising a ZI IP,,, "IP, ZI" ,,,1								
15	IC,Science Co ZI IP,,,,,,1								
16	IC,Scientific i ZI IP,,, "lb, ZI" ,,,1								
17	MDC,Good S ZI IP,,,,,,Z IP,1								
18	MDC,Open S IP,,,LB IP,,,LB IP,,,LB IP,,,LB IP,,,LB IP,1								
19	MDC,Orienta IP,,,LB IP,,,LB IP,,,LB IP,,,LB IP,,,LB IP,1								
20	MDC,Resear IP,,,LB ZI IP,,,LB ZI IP,,,LB ZI IP,,,LB ZI IP,1								
21	MDC,Science ZI IP,,,1								
22	CEITEC,Align ZI IP,,,LB ZI IP,,,IP ZI,L ZI IP,1								
23	CEITEC,HOW ZI IP,,,1								
24	CEITEC,"Ope ZI IP,,,LB ZI IP,,,LB ZI IP,,,1								
25	CEITEC,Open Science in the Context of Data Management Tools,"Proper and robust research data management can,Äöt t								
26	CEITEC,Open ZI IP,,,LB ZI IP,,,IP ZI,L ZI IP,1								
27	CEITEC,OPEN ZI IP,,,LB ZI IP,,, "IP, lb, ZI" ,LB IP,1								
28	CEITEC,Open Science Workshop 2020,"Translation: Open access to scientific information is one of the basic conditions fo								



Master thesis: Genetička struktura
humane populacije otoka Korčule



5. To support additional discoveries.

The New York Times

Sharing of Data Leads to Progress on Alzheimer's



By Gina Kolata
Aug. 12, 2010

[See how this article appeared when it was originally published on NYTimes.com.](#)

In 2003, a group of scientists and executives from the National Institutes of Health, the Food and Drug Administration, the drug

Data sharing and the future of science

[Nature Communications](#) 9, Article number: 2817 (2018) | [Cite this article](#)

15k Accesses | 30 Citations | 417 Altmetric | [Metrics](#)

Who benefits from sharing data? The scientists of future do, as data sharing today enables new science tomorrow. Far from being mere rehashes of old datasets, evidence shows that studies based on analyses of previously published data can achieve just as much impact as original projects.

COVID-19 Open Data Repository



6. Data is crucial to validate and replicate findings.

Review > PLoS One. 2009 May 29;4(5):e5738. doi: 10.1371/journal.pone.0005738.

How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data

Daniele Fanelli ¹

Affiliations + expand

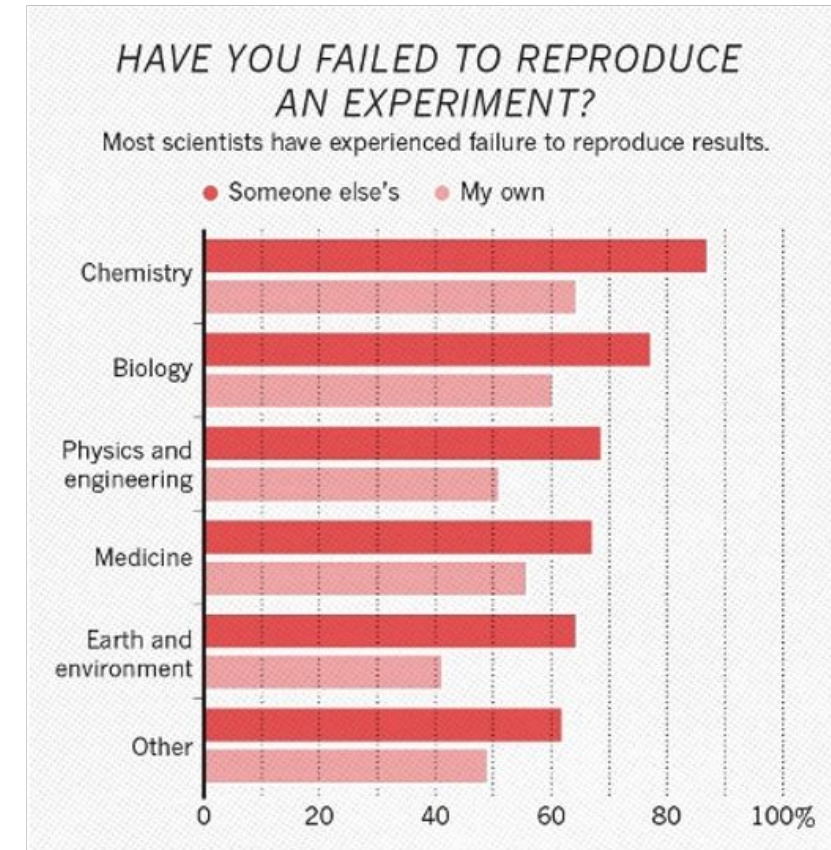
PMID: 19478950 PMCID: PMC2685008 DOI: 10.1371/journal.pone.0005738

[Free PMC article](#)

Abstract

The frequency with which scientists fabricate and falsify data, or commit other forms of scientific misconduct is a matter of controversy. Many surveys have asked scientists directly whether they have committed or know of a colleague who committed research misconduct, but their results appeared difficult to compare and synthesize. This is the first meta-analysis of these surveys. To

[DOI: 10.1371/journal.pone.0005738](https://doi.org/10.1371/journal.pone.0005738)



Reproducibility crisis Nature 533, 452
454 (26 May 2016) <https://doi.org/10.1038/533452a>

13



7. Data can transform the way we understand and engage with the world.

- allow a fuller understanding of the global problems and universal issues.
- help **fight** global problems such as **disease** or crime or famine.
- empower citizens and hence can **strengthen democracy**.
- give a big boost to businesses.
- **streamline** the **processes** and systems that the society and governments have built. It can help



8. Proper data management helps to comply with legislative requirements

According to [GDPR enforcement tracker](#), up to now **13 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 10 entries

ETid	Country	Date of Decision	Fine [€]	Controller/Processor	Quoted Art.	Type	Source
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
		2021-08-20	1,500	MOVE Ireland	Art. 5 (1) f) GDPR, Art.	Insufficient technical and	link

Home License Privacy Imprint



9. Proper data management helps to comply with publisher's requirements

nature [View all journals](#) [Search](#)

[Explore content](#) [About the journal](#) [Publish with us](#) [Subscribe](#) [Sign up for alerts](#)

[nature](#) > for authors

For Authors

- [Editorial criteria and processes](#)
- [Formatting guide](#)
- [Presubmission enquiries](#)
- [Initial submission](#)
- [Final submission](#)
- [Supplementary information](#)
- [Forms and declarations](#)
- [Matters Arising](#)
- [Other types of submissions](#)
- [Publishing options](#)

For Authors

The sections below provide essential information for authors and we recommend that you take the time to read them before submitting a contribution to *Nature*. These instructions refer to Articles, Reviews and Perspectives. Separate guidelines are available for [Matters Arising](#) and for [other types of submission](#).

Editorial criteria and processes

Please read this section before submitting anything to *Nature*. This section explains *Nature*'s editorial criteria, and how manuscripts are handled by our editors between submission and acceptance for publication.

Formatting guide

This section provides a description of all types of contribution and detailed instructions for preparing, formatting and writing.

PLOS ONE

- Qualifying for Authorship
- Group Authorship
- Author Contributions
- Acknowledgments
- Corresponding Author Responsibilities
- Professional Medical Writers
- Authorship Changes
- Author Identification
- Editor and Reviewer Requirements

Authorship

The following policies apply to all PLOS Journals, unless otherwise noted.

Everyone listed as an author should meet our criteria for authorship. Everyone who is not listed as an author.

We expect that all authors will take public responsibility for the content of the manuscript. Authors must be described.

All authors will be contacted by email at submission to ensure that they are aware of its content, and its authorship. Some PLOS journals require that all co-authors confirm.

Qualifying for Authorship

Authorship criteria is based on the International Committee of Medical Journal Editors' [Manuscripts Submitted to Biomedical Journals](#).

Science

[Current Issue](#) [First release papers](#) [Archive](#) [About](#) [Submit manuscript](#)

[HOME](#) > [SCIENCE](#) > [INFORMATION FOR AUTHORS](#)

Information for authors

Science is a weekly, peer-reviewed journal that publishes significant original scientific research, plus reviews and analyses of current research and science policy. We seek to publish papers that are influential in their fields or across fields and that will substantially advance scientific understanding. Selected papers should present novel and broadly important data, syntheses, or concepts. We welcome submissions from all fields of science and from any source.

Manuscripts should be submitted at our manuscript submission and information portal <https://cts.sciencemag.org>. The status of submitted manuscripts can also be tracked at this portal.

Authors should familiarize themselves with the overall editorial policies for the *Science* Journals before submitting their paper. These policies spell out the rights and responsibilities of authors and agree to when submitting and publishing their papers. [Access this information](#).

F1000Research

[Search](#)

[BROWSE](#) [GATEWAYS & COLLECTIONS](#) [HOW TO PUBLISH](#) [ABOUT](#) [BLOG](#)

[Home](#) > [Article Guidelines](#)

How to Publish

[Submit your Research](#)

[My Submissions](#)

[Article Guidelines](#)

[Article Guidelines \(New Versions\)](#)

[Data Guidelines](#)

[Posters and Slides Guidelines](#)

[Document Guidelines](#)

[Article Processing Charges](#)

[Peer Review](#)

Article Guidelines

F1000Research publishes a number of different articles types. We aim to make it easy for authors and, where possible, offer some flexibility in terms of formats and structure. Specific requirements do apply to some article types, however; please choose from the article type-specific instructions listed below.

Please review the details of F1000Research's post-publication [peer-review model](#) and our [policies](#) before you submit.

Manuscripts can be submitted as Word (DOC or DOCX) or rich text format (RTF) files. For LaTeX, please use the [F1000Research article template](#) and submit a ZIP file of the project, which must also include the PDF. LaTeX users can also submit via [Overleaf](#), using an F1000Research [journal article template](#) or a [software tool article template](#). If you have any questions about suitable file formats, please [email us](#).

Research Articles

Research Articles should present originality in findings and insights and offer theoretical, empirical, experimental and/or methodological advances to their respective fields of research. Null and negative findings and reanalyses of previous studies leading to new results, as well as confirmatory results, are also encouraged.



10. To save time and resources in the long run.

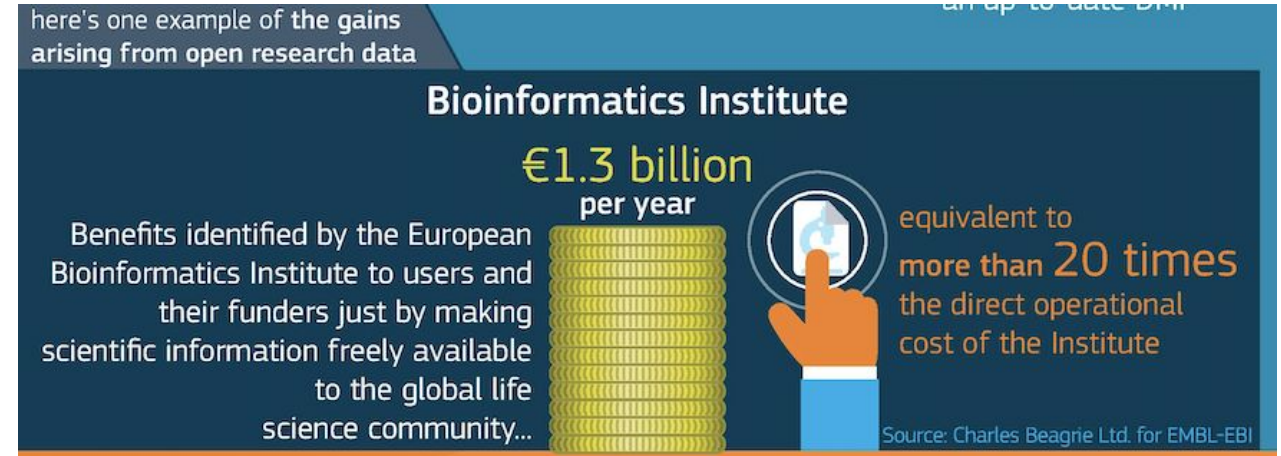


Cost of not having FAIR research data

Cost-Benefit analysis for FAIR research data

€10.2bn/year

doi: 10.2777/02999



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



RDM in practice (for researchers)

- writing data management plans (DMPs)
- data organization, adding metadata, using electronic lab notebooks
- 3-backup data storage and archiving
- cleaning data, file and folder naming schemas
- adding persistent identifiers, ontologies, metadata schemas
- depends on the previous steps
- FAIR data, adding licence, uploading data into repository secure work with sensitive and personal data and complying to the GDPR, submitting DTA (data transfer agreements),



RDM in practice (research data managers and curators)

Building tools, workflows and computational infrastructure

IDEA: Helping you to appropriately provide labeling, storage, and access for data at all stages of a research project.

GOAL: Complete, clean, curated, organized (clearly labelled, separated & documented) data + metadata



MDC'S RESEARCH DATA MANAGEMENT UNIT

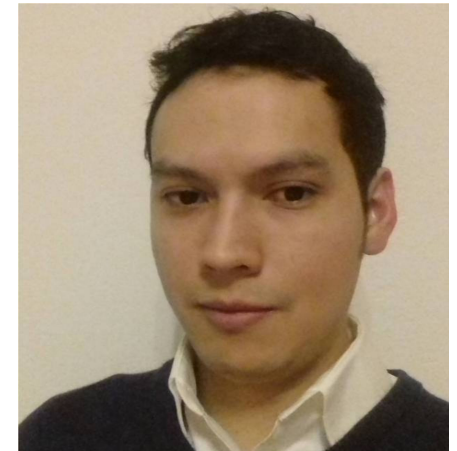
The Max Delbrück Center for Molecular Medicine is one of the world's leading biomedical research centers (1992).

The MDC has **88 labs** and **833** researchers who analyze how the human body works in both health and disease.

The Research Data Management Unit was established in **2020** as a part of the MDC's Scientific Infrastructure Department.

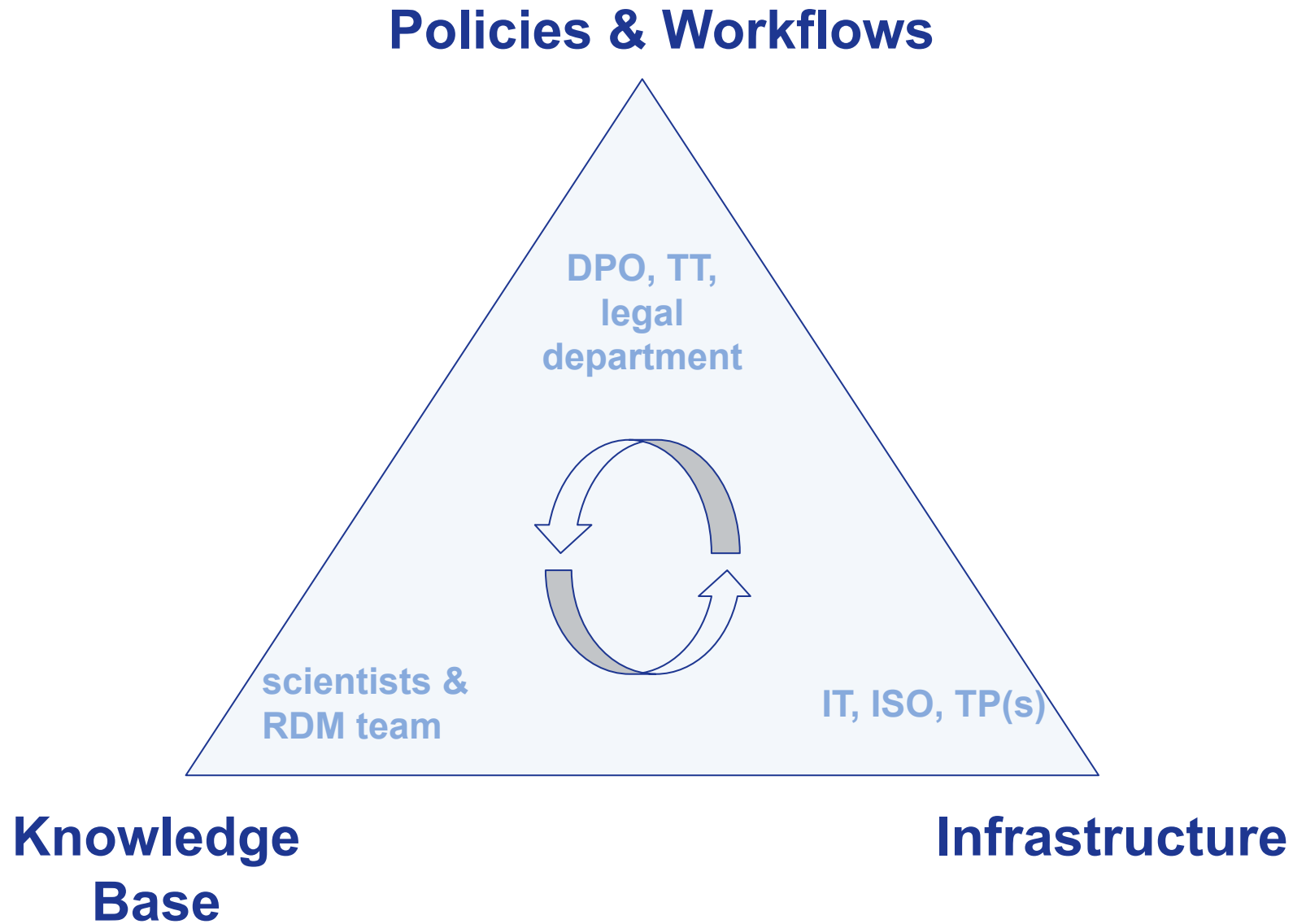


Dr. I. P.



O. M.

WE WORK WITH:



OVERVIEW: From the RDM Roadmap

1. Developing RDM infrastructure

➤ ELN: 

➤ DMP: 

➤ Image data management:


➤ digTools database:



2. Writing policies & workflows

➤ Policy framework for research data management

➤ Data organization guidelines

➤ Data storage, access and backup guidelines

➤ ...



3. Building knowledge base

- Training
- Support
- RDM website
- Outreach



Typical workflow for an infrastructural project: establishing an ELN



STEPS:

1. Assessments of available tools
2. Purchasing Tool
3. Resolve Documentation
4. Installation at the MDC
5. Onboarding & Training
6. Tool customisation

A	B	C	D	E	F	G
Features/ELN solution	OneNote	RSpace	ELN solutions Benchling	LabArchives	Labfolder	SciNote
Features and Design						S
Intuitive Interface Design	1	2	2	2	2	2
Able to search for documents within the platform	1	2	2	2	2	2
Able to search beyond file formats	2	2	2	2	2	2
Able to search beyond typographical errors	2	2	2	2	2	2
Transparency of file structure	2	2	2	2	2	2
Can manipulate text documents	2	2	2	2	2	2
Can manipulate spread sheets	2	2	2	2	2	2
Can manipulate images	2	2	2	2	2	1
Can manipulate datasets	2	2	2	2	2	0
Can manipulate other elements	2	2	1	2	0	0
Support for multiple open windows	2	2	2	2	2	2
Ability to link out	2	2	2	2	2	2
Training	0	2	2	2	2	2
Documentation	2	2	2	2	2	2
Links to open repositories	0	2	0	1	2	0
Links to DropBox/ OneDrive/GoogleDrive	1	2	1	1	1	0
Recovery Options	2	2	1	2	2	2
Single Sign-on (institutional ID)	2	2	2	2	2	2
Windows Compatible	2	2	2	2	2	2
Macintosh Compatible	0	2	2	2	2	2
Linux Compatible	0	2	2	2	2	2
Android Compatible	2	2	2	2	2	2
iOS Compatible	0	2	2	2	2	2
Web interface	2	2	2	2	2	2
Open Science/Open Data Efforts	0	2	0	1	2	1
Academically Oriented	0	2	2	2	2	2
Can be expanded to LIMs/Integrated with another	0	2	2	-	0	2
Content creation tools						
Integrated Apps	1	2	1	2	2	1
Connect with instruments	0	2	2	0	1	2
Import/ Data Entry/ Data Migration from Other S	2	2	2	2	2	2
Export/ Exit Strategies (outgoing migration)	2	2	1	1	2	2
Metadata Creation Prompts	0	2	2	2	2	2
Rights Management (licensing)	0	2	2	2	2	2
Protocol Integration	0	2	1	1	2	2
Adaptable to a Variety of Workflows	0	2	2	1	2	2
Hosting						
Cloud	2	2	2	2	2	2
Local	0	2	0	2	2	2
Hybrid	0	2	0	2	2	2
Free cloud storage	-	Unlimited	10GB	100 GB	3 gb	10 GB per team

Typical workflow for an policy-making project: writing RDM policy

Incorporate legal obligations, ethical responsibilities and the rules of institutional, governmental and funding bodies



HELMHOLTZ

MDC
MAX-DELBRÜCK-CENTRUM
FÜR MOLEKULARE MEDIZIN
IN DER HELMHOLTZ-GEMEINSCHAFT

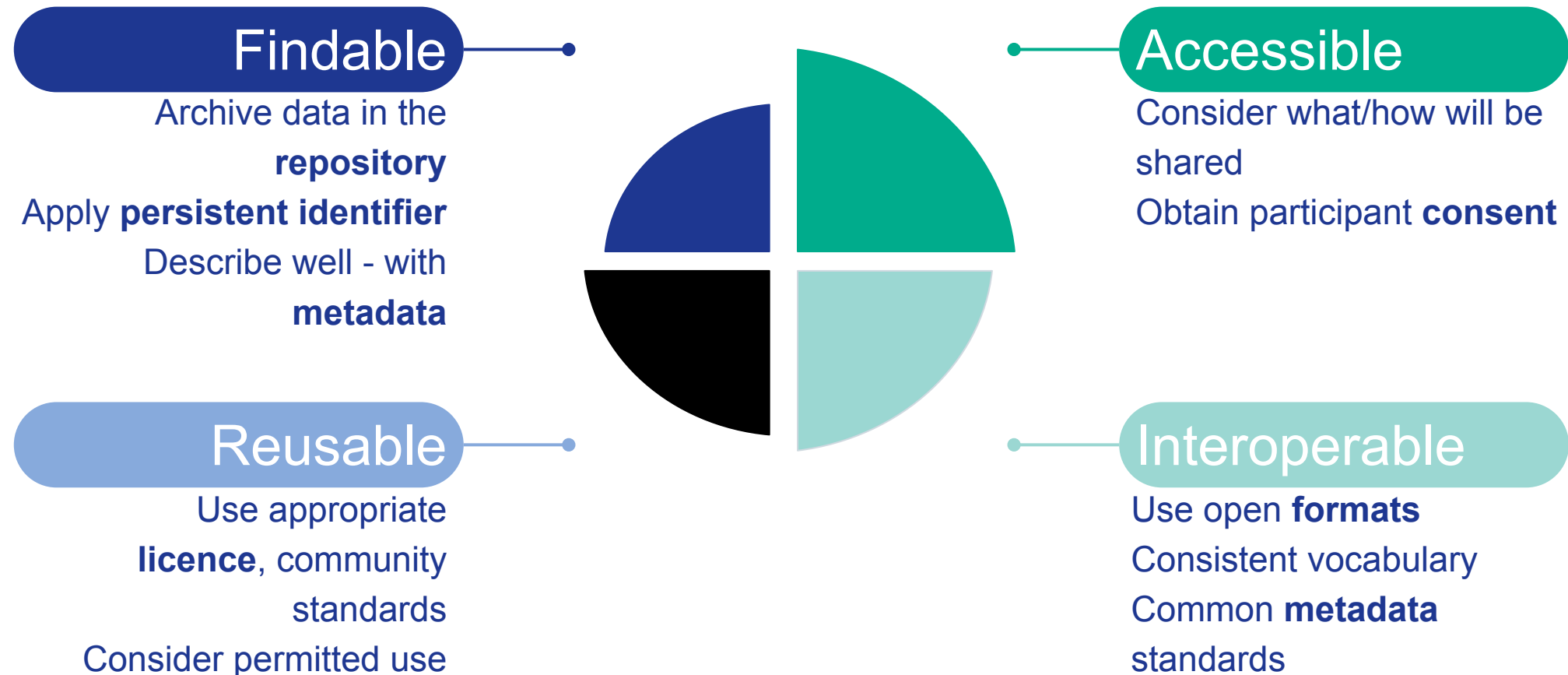
Policy Framework for Research Data Management (RDM)

Max-Delbrück-Centrum für Molekulare Medizin in der
Helmholtz-Gemeinschaft

The MDC is committed to a goal of making data created as part of the research process compliant with the **FAIR** principles [1] Data should be: **Findable, Accessible, Interoperable** and **Reusable**.

The MDC supports access to research data following the European Commission for Open Data principle [12], "**as open as possible, as closed as necessary**". This includes metadata, methods, protocols and software/code needed to support effective data use.

*FAIR Data



Working as a Research Data Manager

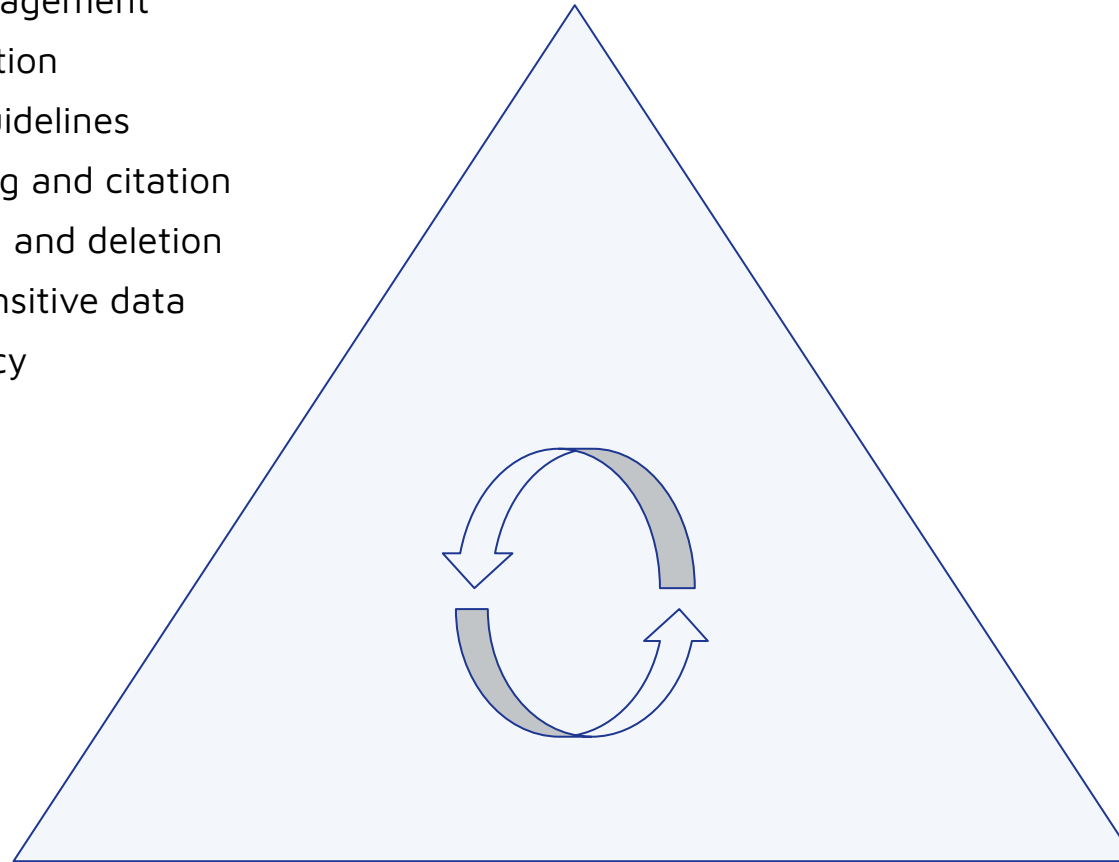
- According to forschungsdaten.org page, there are currently **67 RDM** units in Germany.
- There are plenty of positions in **Netherlands and the UK**
- There is **no** specific university education to be a Research Data Manager so you can take the trainings.
 - [Research Data Management and Sharing \(Coursera\)](#)
 - [ORION MOOC for Open Science in the Life Sciences](#)
 - [Mantra- Free online Research Data Management Training](#)
 - [FOSTER- Open Science Training Courses](#)
 - [Datatree Free online course on research data management](#)
 - [Open Science MOOC](#)



Elements of the RDM Roadmap

Policies & Workflows

1. Policy on research data management
2. Guidelines on data organization
3. Storage and Data Access Guidelines
4. Guidelines on data publishing and citation
5. Guidelines on data archiving and deletion
6. Guidelines on processing sensitive data
7. Research Data Security Policy



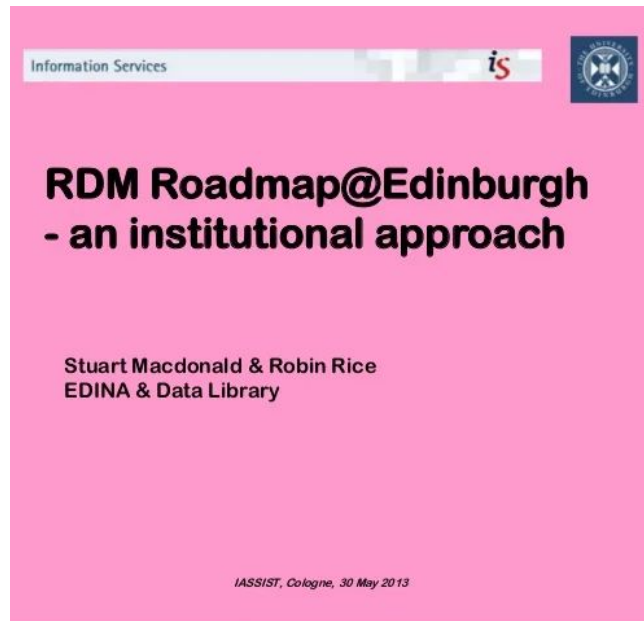
Data organization
 External training on GDPR
 Digital Resources at the MDC
 Metadata and metadata schemas
 General training on RDM
 (overview)
 FAIR and open data

**Knowledge
Base**

- I.1. Provide the ELN solution
- I.2. Provide a DMP solution
- I.3. Provide an archiving solution for the MDC that fulfills requirements for data publishing
- I.4. Provide solution for management of imaging data

Infrastructure

How to write a RDM Roadmap?



HOW - TO GUIDE

January 2017

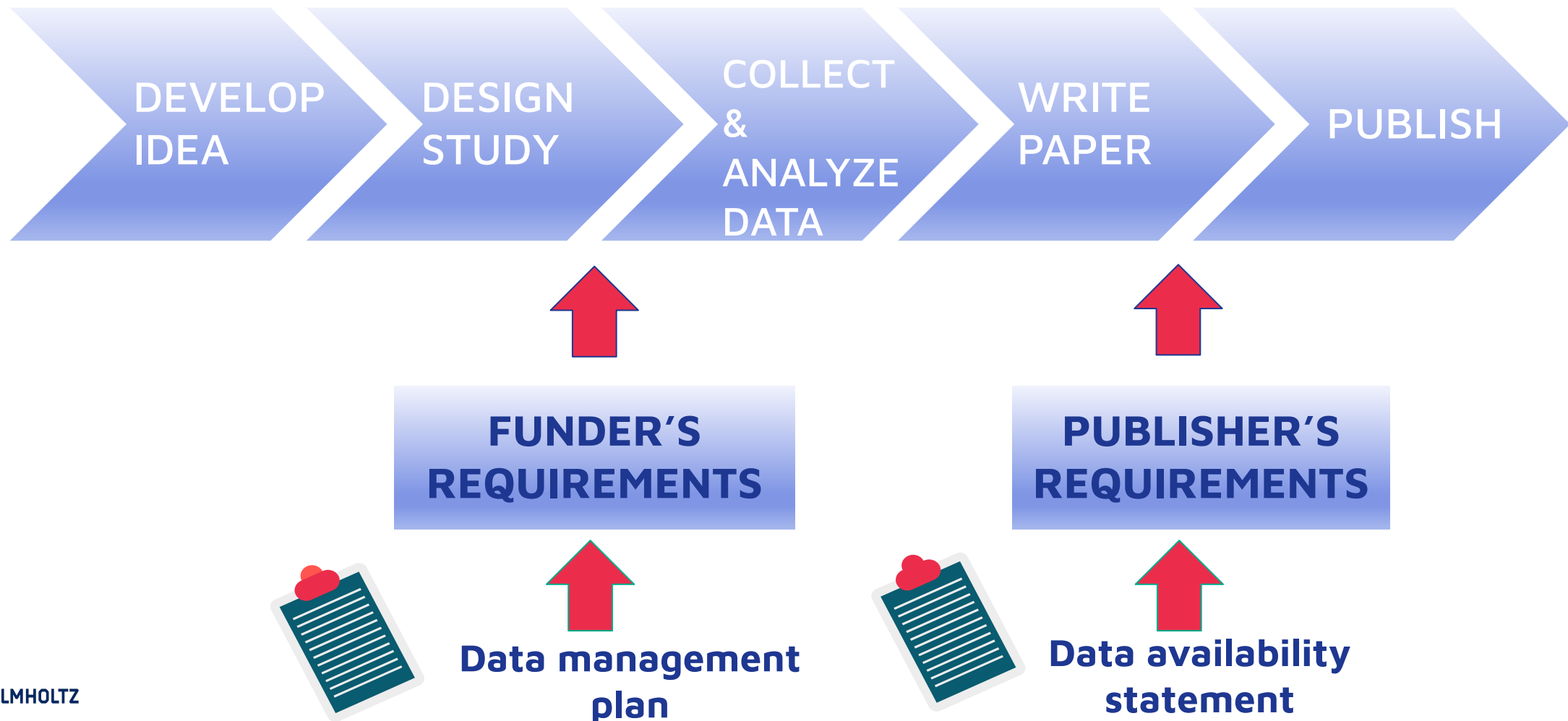
Version 1.1

Using RISE the Research Infrastructure Self-Evaluation Framework

https://www.dcc.ac.uk/sites/default/files/documents/publications/UsingRISE_v1_1.pdf

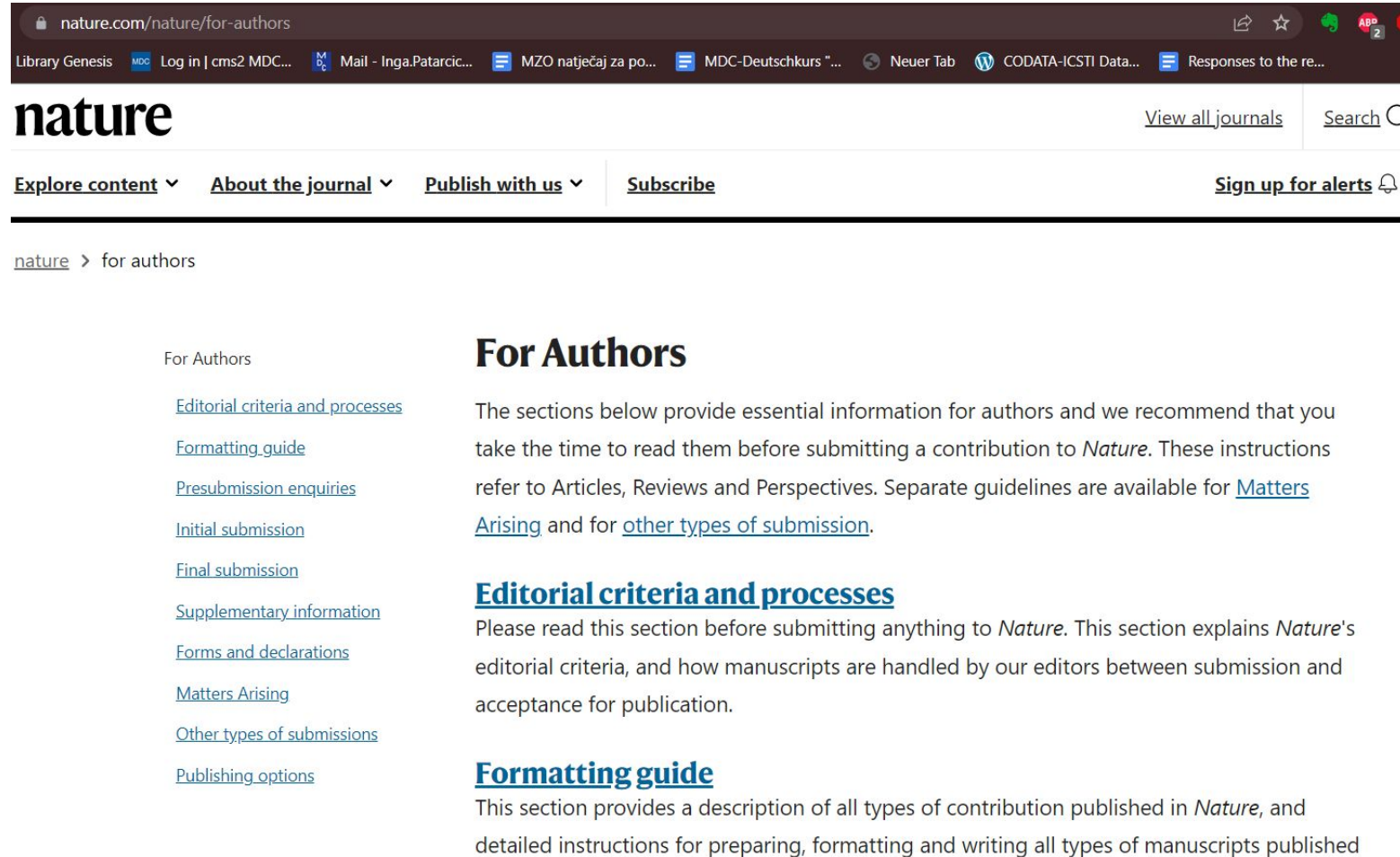
Basic RDM training

Requirements for data management



01

Requirements of the publishers



The screenshot shows the 'nature.com/nature/for-authors' page. The browser's address bar and tabs are visible at the top. The page features the 'nature' logo, navigation links like 'View all journals' and 'Search', and a menu with 'Explore content', 'About the journal', 'Publish with us', and 'Subscribe'. A 'Sign up for alerts' button is also present. The main content area is titled 'For Authors' and includes a list of links on the left: 'Editorial criteria and processes', 'Formatting guide', 'Presubmission enquiries', 'Initial submission', 'Final submission', 'Supplementary information', 'Forms and declarations', 'Matters Arising', 'Other types of submissions', and 'Publishing options'. The main text explains that the sections provide essential information for authors and recommends reading them before submission. It also highlights the 'Editorial criteria and processes' and 'Formatting guide' sections.

nature.com/nature/for-authors

Library Genesis MDC Log in | cms2 MDC... Mail - Inga.Patarcic... MZO natječaj za po... MDC-Deutschkurs "... Neuer Tab CODATA-ICSTI Data... Responses to the re...

nature [View all journals](#) [Search](#)

[Explore content](#) [About the journal](#) [Publish with us](#) [Subscribe](#) [Sign up for alerts](#)

[nature](#) > for authors

For Authors

- [Editorial criteria and processes](#)
- [Formatting guide](#)
- [Presubmission enquiries](#)
- [Initial submission](#)
- [Final submission](#)
- [Supplementary information](#)
- [Forms and declarations](#)
- [Matters Arising](#)
- [Other types of submissions](#)
- [Publishing options](#)

For Authors

The sections below provide essential information for authors and we recommend that you take the time to read them before submitting a contribution to *Nature*. These instructions refer to Articles, Reviews and Perspectives. Separate guidelines are available for [Matters Arising](#) and for [other types of submission](#).

[Editorial criteria and processes](#)

Please read this section before submitting anything to *Nature*. This section explains *Nature*'s editorial criteria, and how manuscripts are handled by our editors between submission and acceptance for publication.

[Formatting guide](#)

This section provides a description of all types of contribution published in *Nature*, and detailed instructions for preparing, formatting and writing all types of manuscripts published

<https://www.nature.com/nature/for-authors>

Do >1,000 journals have >1,000 different policies?

nature

[View all journals](#)[Search](#)

[Explore content](#)[About the journal](#)[Publish with us](#)[Subscribe](#)

[Sign up for alerts](#)

[nature](#) > for authors

For Authors

- [Editorial criteria and processes](#)
- [Formatting guide](#)
- [Presubmission enquiries](#)
- [Initial submission](#)
- [Final submission](#)
- [Supplementary information](#)
- [Forms and declarations](#)
- [Matters Arising](#)
- [Other types of submissions](#)
- [Publishing options](#)

For Authors

The sections below provide essential information for authors and we recommend that you take the time to read them before submitting a contribution to *Nature*. These instructions refer to Articles, Reviews and Perspectives. Separate guidelines are available for [Matters Arising](#) and for [other types of submission](#).

Editorial criteria and processes

Please read this section before submitting anything to *Nature*. This section explains *Nature's* editorial criteria, and how manuscripts are handled by our editors between submission and acceptance for publication.

Formatting guide

This section provides a description of all types of contribution published in *Nature*, and detailed instructions for preparing, formatting and writing all types of manuscripts published

PLOS ONE

- Qualifying for Authorship
- Group Authorship
- Author Contributions
- Acknowledgments
- Corresponding Author Responsibilities
- Professional Medical Writers
- Authorship Changes
- Author Identification
- Editor and Reviewer Requirements

Authorship

The following policies apply to all PLOS journals, unless otherwise noted.

Everyone listed as an author should meet our criteria for authorship. Everyone who is as an author.

We expect that all authors will take public responsibility for the content of the manuscript and its authorship. Some PLOS journals require that all co-authors confirm

All authors will be contacted by email at submission to ensure that they are aware of its content, and its authorship. Some PLOS journals require that all co-authors confirm

Qualifying for Authorship

Authorship criteria is based on the International Committee of Medical Journal Editors: [Manuscripts Submitted to Biomedical Journals](#).

Science

[Current Issue](#)[First release papers](#)[Archive](#)[About](#)[Submit manuscript](#)

[HOME](#) > [SCIENCE](#) > [INFORMATION FOR AUTHORS](#)

Information for authors

Science is a weekly, peer-reviewed journal that publishes significant original scientific research, plus reviews and analyses of current research and science policy. We seek to publish papers that are influential in their fields or across fields and that will substantially advance scientific understanding. Selected papers should present novel and broadly important data, syntheses, or concepts. We welcome submissions from all fields of science and from any source.

Manuscripts should be submitted at our manuscript submission and information portal <https://cts.sciencemag.org>. The status of submitted manuscripts can also be tracked at this portal.

Authors should familiarize themselves with the overall editorial policies for the *Science* Journals before submitting their paper. These policies spell out the rights and responsibilities that authors agree to when submitting and publishing their papers. [Access this information here](#).

F1000Research

[Search](#)

[BROWSE](#)[GATEWAYS & COLLECTIONS](#)[HOW TO PUBLISH](#)[ABOUT](#)[BLOG](#)

[Home](#) » [Article Guidelines](#)

How to Publish

- Submit your Research
- My Submissions
- Article Guidelines (New Versions)
- Data Guidelines
- Posters and Slides Guidelines
- Document Guidelines
- Article Processing Charges
- Peer Review

Article Guidelines

F1000Research publishes a number of different articles types. We aim to make it easy for authors and, where possible, offer some flexibility in terms of formats and structure. Specific requirements do apply to some article types, however, please choose from the article type-specific instructions listed below.

Please review the details of F1000Research's post-publication [peer-review model](#) and our [policies](#) before you submit.

Manuscripts can be submitted as Word (DOC or DOCX) or rich text format (RTF) files. For LaTeX, please use the [F1000Research article template](#) and submit a ZIP file of the project, which must also include the PDF. LaTeX users can also submit via [Overleaf](#), using an F1000Research [journal article template](#) or a [software tool article template](#). If you have any questions about suitable file formats, please [email us](#).

Research Articles

Research Articles should present originality in findings and insights and offer theoretical, empirical, experimental and/or methodological advances to their respective fields of research. Null and negative findings and reanalyses of previous studies leading to new results, as well as confirmatory results, are also encouraged.

Unifying journal's TOP (transparency and openness) guidelines



- A set of best practices that journals across scientific disciplines can adopt.
- They cover various topics in scientific publishing from data sharing to preregistration.
- The goal is to translate scientific norms and values into concrete actions and change the current incentive structures to drive researchers' behavior toward more open research practices.

Journals implementing TOP guidelines (COS)

	Not Implemented	Level I	Level II	Level III
Citation Standards	No mention of data citation.	Journal describes citation of data in guidelines to authors with clear rules and examples.	Article provides appropriate citation for data and materials used consistent with journal's author guidelines.	Article is not published until providing appropriate citation for data and materials following journal's author guidelines.
Data Transparency	Journal encourages data sharing, or says nothing.	Article states whether data are available, and, if so, where to access them.	Data must be posted to a trusted repository. Exceptions must be identified at article submission.	Data must be posted to a trusted repository, and reported analyses will be reproduced independently prior to publication.
Analytic Methods (Code) Transparency	Journal encourages code sharing, or says nothing. No mention Encourage	Article states whether code is available, and, if so, where to access them. Availability statement	Code must be posted to a trusted repository. Exceptions must be identified at article submission. Storage in repository	Code must be posted to a trusted repository, and reported analyses will be reproduced independently prior to publication. Is reproducible?
Research Materials Transparency	Journal encourages materials sharing, or says nothing.	Article states whether materials are available, and, if so, where to access them.	Materials must be posted to a trusted repository. Exceptions must be identified at article submission.	Materials must be posted to a trusted repository, and reported analyses will be reproduced independently prior to publication.
Design and Analysis Transparency	Journal encourages design and analysis transparency, or says nothing.	Journal articulates design transparency standards.	Journal requires adherence to design transparency standards for review and publication.	Journal requires and enforces adherence to design transparency standards for review and publication.
Study Preregistration	Journal says nothing.	Article states whether preregistration of study exists, and, if so, where to access it.	Article states whether preregistration of study exists, and, if so, allows journal access during peer review for verification.	Journal requires preregistration of studies and provides link and badge in article to meeting requirements.
Analysis Plan Preregistration	Journal says nothing.	Article states whether preregistration of study exists, and, if so, where to access it.	Article states whether preregistration with analysis plan exists, and, if so, allows journal access during peer review for verification.	Journal requires preregistration of studies with analysis plans and provides link and badge in article to meeting requirements.

An example of TOP guidelines for individual journals

	Level 1	Level 2	Level 3
Data Citation	✓		
Data Transparency	✓		
Materials Transparency	✓		
Code Transparency		✓	
Design & Analysis		✓	
Study Preregistration	✓		
Analysis Preregistration	✓		
Replication	✓		



An example of TOP guidelines for individual journals

	Level 1	Level 2	Level 3
Data Citation		✓	
Data Transparency		✓	
Materials Transparency	✓		
Code Transparency		✓	
Design & Analysis	✓		
Study Preregistration	✓		
Analysis Preregistration	✓		
Replication	✓		

Science
AAAS

Data Transparency



Level 1: Disclosure

Article states whether or not data are available. If so, give URL.



Level 2: Mandate

Data MUST be posted in a trusted repo (exceptions permitted for legal or ethical constraints).



Level 3: Verified Mandate

Level 2 + Can results be replicated? Reported analysis will be reproduced independently prior to publication





Data Transparency: Level 1

- Data available at submission
- Must include a **data availability statement**

Data availability template:

- 1 The datasets generated during and/or analysed during the current study are available in the [NAME] repository, [PERSISTENT WEB LINK TO DATASETS]
- 2 The datasets generated during and/or analysed during the current study are not publicly available due [REASON WHY DATA ARE NOT PUBLIC] but are available [STATE CONDITIONS FOR ACCESS].
- 3 Data sharing not applicable to this article as no datasets were generated or analysed during the current study.
- 4 All data generated or analysed during this study are included in this published article [and its supplementary information files].



Data availability statement:

Data availability

DNA and RNA sequencing data have been deposited in the European Genome-phenome Archive under the accession code [EGAS00001004521](#) and will be made available upon reasonable request for academic use and within the limitations of the provided informed consent by the corresponding author upon acceptance. Every request will be reviewed by the institutional review board of the Netherlands Cancer Institute; the researcher will need to sign a data access agreement with the Netherlands Cancer Institute after approval. Sequencing data correspond with Figs. [3](#) and [4](#). Multiplex immunofluorescence raw quantification data corresponding to Figs. [3d](#) and [4](#) will be made available upon reasonable academic request within the limitations of informed consent by the corresponding author upon acceptance.

Mandates for the specific datasets (Springer Nature)

Mandatory deposition	Suitable repositories
Protein sequences	Uniprot
DNA and RNA sequences	Genbank
	DNA DataBank of Japan (DDBJ)
	EMBL Nucleotide Sequence Database (ENA)
DNA and RNA sequencing data	NCBI Trace Archive
	NCBI Sequence Read Archive (SRA)
Genetic polymorphisms	dbSNP
	dbVar
	European Variation Archive (EVA)
Linked genotype and phenotype data	dbGAP
	The European Genome-phenome Archive (EGA)
Macromolecular structure	Worldwide Protein Data Bank (wwPDB)
	Biological Magnetic Resonance Data Bank (BMRB)
	Electron Microscopy Data Bank (EMDB)
Gene expression data (must be MIAME compliant)	Gene Expression Omnibus (GEO)
	ArrayExpress
Crystallographic data for small molecules	Cambridge Structural Database
Proteomics data	PRIDE
*Earth, space & environmental sciences	Recommended Repositories

SOURCE:
<https://www.springernature.com/gp/authors/research-data-policy/recommended-repositories>

Generalist repositories

Recommended by Springer Nature (below) and F1000 (right):

- [Dryad Digital Repository](#)
- [figshare](#)
- [Harvard Dataverse](#)
- [Open Science Framework](#)
- [Science Data Bank](#)
- [Zenodo](#)



DATA TYPE	WHERE TO SUBMIT*	WHAT TO INCLUDE IN THE DATA AVAILABILITY SECTION OF YOUR ARTICLE
Any	Dryad	Title, DOI
Any, but especially data in SAV and POR formats	Dataverse	Title, DOI
Any	Figshare ^s	Title, DOI
Any, but especially deposits with mixed data, materials and documents	Open Science Framework [†]	Title, DOI
Any, but especially deposits with mixed data and code	Zenodo	Title, DOI
Deposits of mixed data and code	Code Ocean	Title, DOI, embed code for interactive reanalysis tool
Any biological data, but especially data linked to studies in other databases	BioStudies	Title, accession number

Source: <https://f1000research.com/>

Where to search for repositories/data?



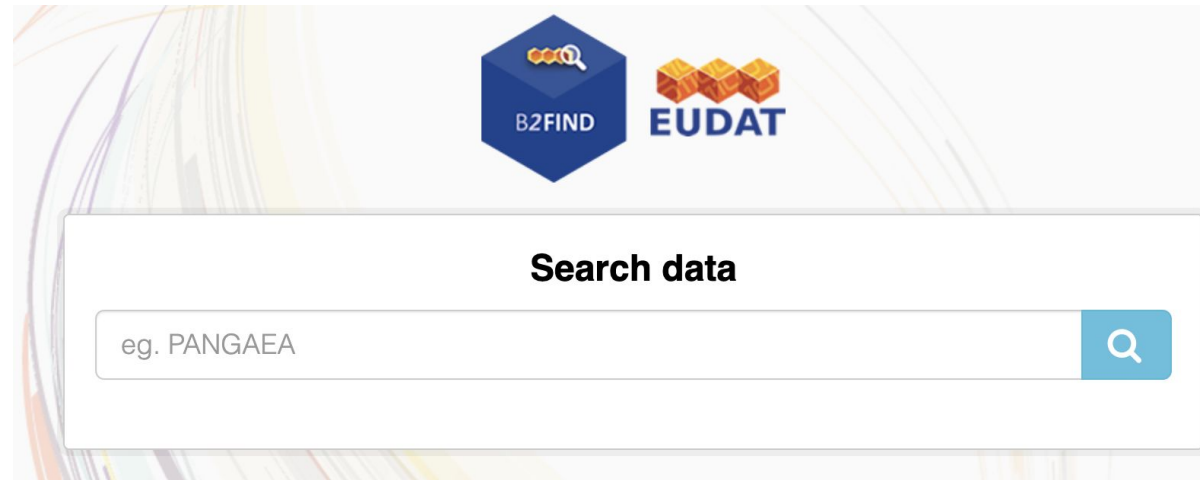
<https://www.re3data.org/>

· CROSDA

CROSSDA (Croatian Social Science Data Archive)

<https://data.crossda.hr/>

Where to search for repositories/data?

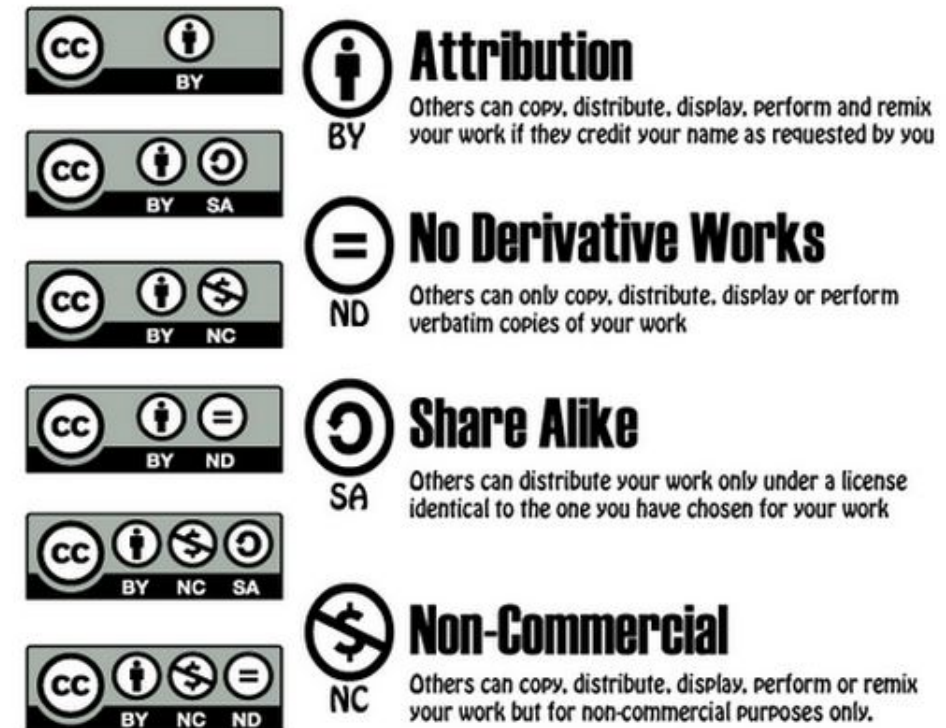


<https://b2find.eudat.eu/>

How do I open data?

- CC licenses offer several levels of openness:
- **CC BY** license (requiring only attribution) is a good option for works such as articles, books, working papers, and reports while a dedication to the public domain using
- **CC0** is recommended for datasets and databases

TO READ: [Creative Commons](#)



VIDEO: [Creative Commons licences explained - YouTube](#)

How do I open software?

Licences that allow software to be freely used, modified, and shared:

- [Apache License 2.0](#)
- [BSD 2-Clause "Simplified" or "FreeBSD" license](#)
- [GNU General Public License \(GPL\)](#)
- [GNU Library or "Lesser" General Public License \(LGPL\)](#)
- [MIT license](#)
- [Mozilla Public License 2.0](#)
- [Common Development and Distribution License](#)

Begin license text.

Copyright <YEAR> <COPYRIGHT HOLDER>

Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.



End license text.

SOURCE: [MIT license](#)

02

Requirements of the funders

The requirements for the retention and preservation of research of the top funders

FUNDER	WHAT?	HOW LONG?	STARTING WHEN?	WHERE?
 Horizon 2020	Research data, unpublished data, code ¹	-	Immediately	Any repository ¹
 European Research Council (ERC)	Research data, unpublished data, code ¹	-	Within six months after the publication ²	Any relevant repository (Suggestions: GenBank and PDB) ^{1,2}
 NIH	"Financial and programmatic records, supporting documents, statistical records, and all other records that are required by the terms of a grant, or may reasonably be considered pertinent to a grant" ³	Period of three years ³	The date the annual FFR* is submitted ³	-

Sherpa Juliet

[About](#)[Search](#)[Statistics](#)[Contact](#)[Admin](#)

Research Funders' Open Access Policies

Sherpa Juliet is a searchable database and single focal point of up-to-date information concerning funders' policies and their requirements on open access, publication and data archiving.

Search for a funder policy

Bundesministerium für Bildung und
Forschung (**BMBF**)

[Search](#)

<https://v2.sherpa.ac.uk/juliet/>

Horizon Europe

Requirements

- Provide outline the measures planned in the project that tend to increase reproducibility
- **DMP** template (usually not required at submission stage)
- **OA without embargo** under open licenses (such as CC),
- Information in the **repository**
- Open peer review encouraged
- OA to **software, models, algorithms, workflows, protocols, simulations, electronic notebooks** and others is not required but strongly recommended.

- Access to 'physical' results like cell lines, biospecimens, compounds, materials, etc. is also strongly encouraged.
- **Citizen**, civil society and end-user engagement should be implemented in project, if appropriate
- Early and open sharing of research (preregistration, registered reports, pre-prints, or crowd-sourcing) encouraged

03

DMP

DMP

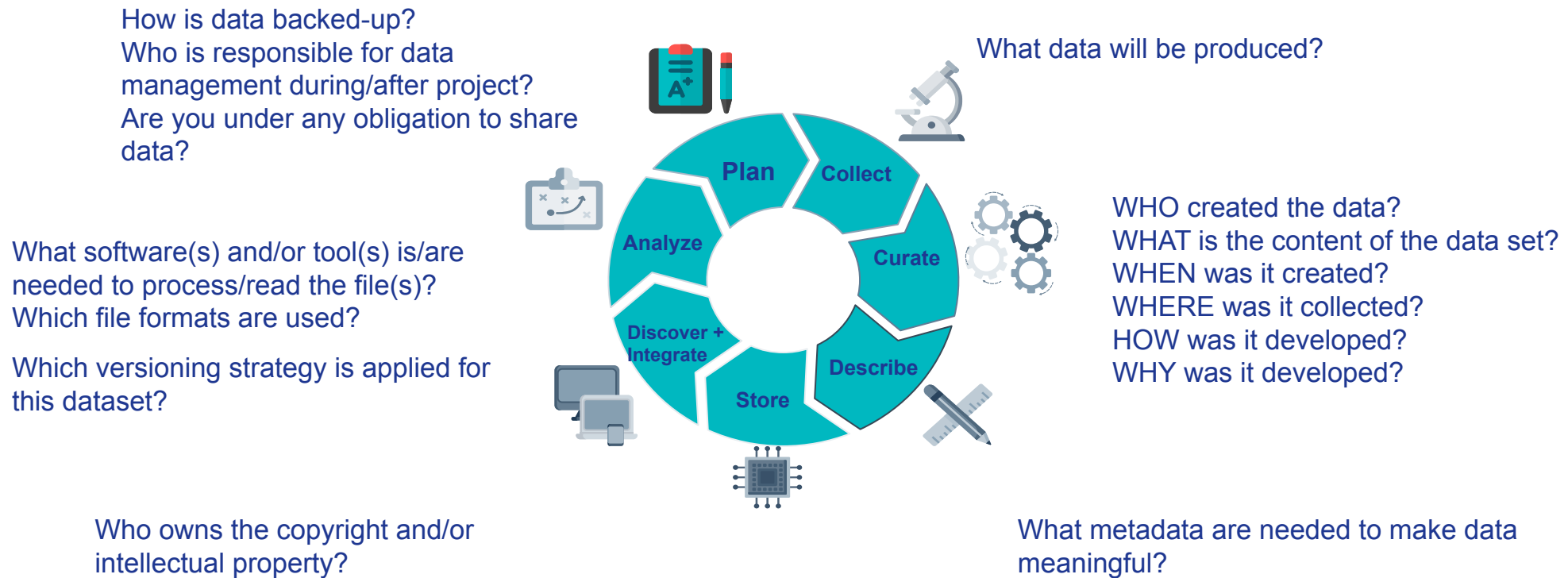
What is a data management plan?

A document that describes what will scientists do with their **data** during the research and after completion of project.

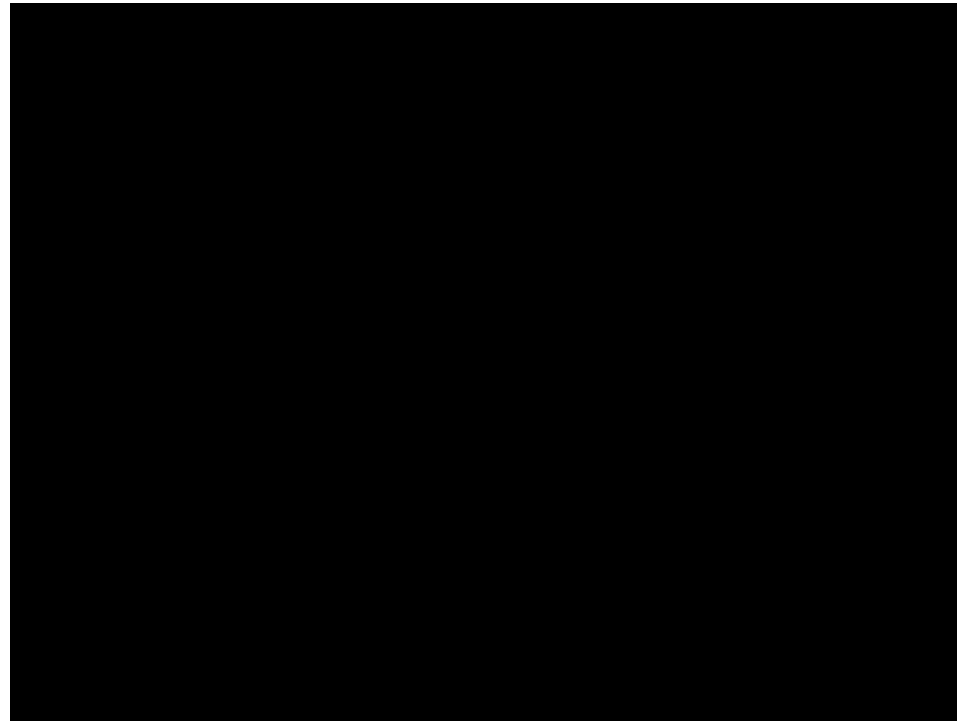


Research Data Lifecycle

2. What are components of Data Management Plan?



How does a Data Management Plan look like?



Which tools can be used to write a DMP?



DMPOnline

Based on the DMPRoadmap
(co- developed by the DCC
and the University of
California Curation Center
(UC3).



DMPTool

DMPTool is a service
of the California Digital
Library



RDMO

DMP tool developed by
AIP, FHP und
KIT-Bibliothek



Argos

online machine-actionable
tool developed by OpenAIRE





DS Wizard

machine-actionable
tool based on FAIR
principles by ELIXIR CZ
and ELIXIR NL

...

Other tools like:

Selecting DMP tool of choice

					
MA DMPs (system connections e.g. ethics review, RIM, repository)	✓	✓	✓	✓	✓
Versioning of DMPs	✓		✗	✓	✓
Notification of changes					✓
Archiving and publishing plans				✓*	✓
Assessing the FAIRness of data plans	✗	✗	✗	✗	✓
Usage statistics	✗	✗	✗		✓

BUT!

DMP ≠ extensive static document with a long text created to satisfy grant agency

- Machine-actionable DMPs
 - System connections e.g. ethics review, RIM, repository...
- Plan lifecycle
 - Versioning of DMPs**
 - Notification of changes**
 - Archiving and publishing plans
- **Assessing the FAIRness of data plans**
- **Usage statistics**

 **FAIRWizard** **DSW**

FAIR WIZARD GUIDES SCIENTISTS TOWARDS BETTER RESEARCH PRACTICES

Question

Title

Description

Answers

Who answered

Advice

1.a.4.b.1.a.1 What repository will this data be stored in?

Domain repositories often have the best functionality to make the data findable and reusable: even though it may look like a database that could be reused in a completely different field would be better findable in a generic repository, the limited availability of domain-specific metadata make that less valuable.

Many repositories are listed in <https://fairsharing.org/>

If a repository offers to give your data set a DOI or alternative persistent identifier it is a good idea to use that option.

External links: [FAIRSharing](#), [Registry of Research data Repositories](#)

☐ a. A domain-specific repository **Findability**

☒ b. Our national repository **Findability**

☐ c. Our institutional repository **Findability**

☐ d. A special-purpose repository for the project **Findability**

Clear answer

Answered in less than 5 seconds by Albert Einstein.

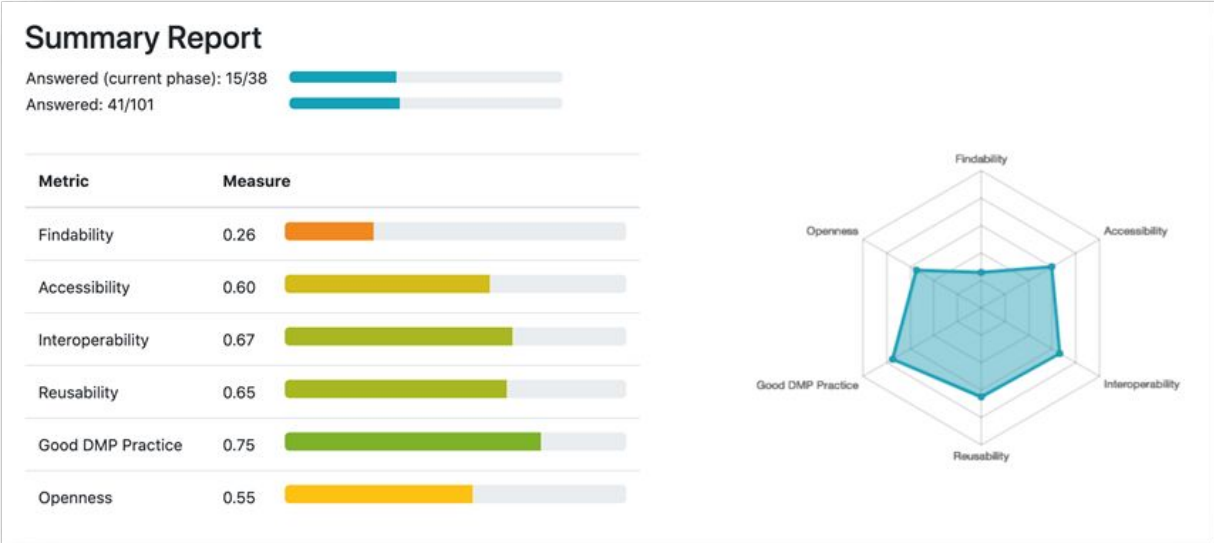
Disadvantage of a general purpose repository is the lack of data-specific features (e.g. 'play' instead of 'download' for an audio file) and limited findability

References

FAIR metrics

FAIR WIZARD PROVIDES SUMMARY REPORT

Assessing FAIRness of the data



The MDC “is committed to a goal of making data created as part of the research process compliant with the FAIR principle” (Policy Framework for Research Data Management, 2021).

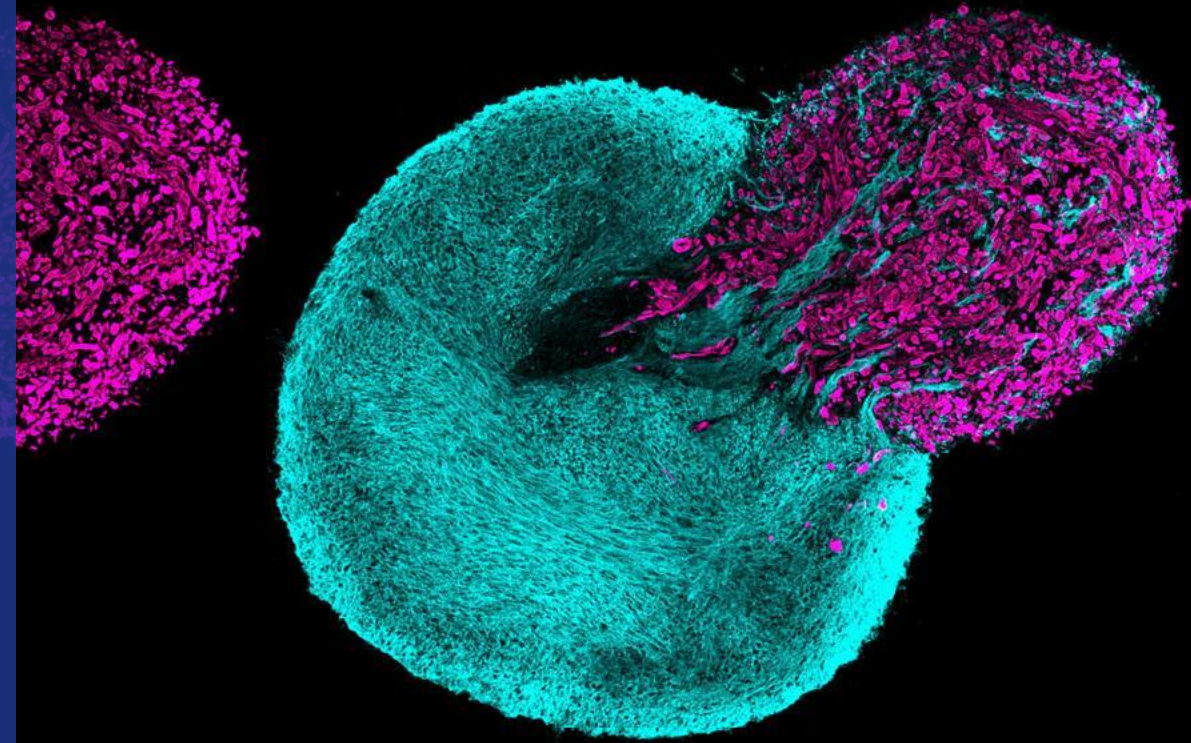
RDM in practice (for researchers)

- writing data management plans (DMPs)
- data organization, adding metadata, using electronic lab notebooks
- 3-backup data storage and archiving
- cleaning data, file and folder naming schemas
- adding persistent identifiers, ontologies, metadata schemas
- depends on the previous steps
- FAIR data, adding licence, uploading data into repository secure work with sensitive and personal data and complying to the GDPR, submitting DTA (data transfer agreements),



Open for discussion

Thank you for your attention!



Comparison of different data policies across journals

Journal	Data Citation	Data Transp.	Material Transp.	Code Transp.	Design & Analysis	Study Prereg	Analysis Prereg.	Replica tion	Sum
Ncomms, Nature, Nature Machine Intelligence, Genome Biology	1	1	3	1	2	1	1	1	11
Bioinformatics	2	2	0	2	2*	0	0	0	8
NAR	0	2	2	2	0	0	0	0	6
F1000	1 or 2	1 or 2**	1 or 2	1 or 2***	2	3	3	NA	16
Gigascience	2	2	2	3***	3*	0	0	0	12
Science	2	2	2	1	1	1	1	1	11
PNAS	1	1	1	1	1	1	1	1	8
Cell, Cell Reports	2	1	2	1	2	1	1	1	11
Biophysical Journal	0	2	2	2	2	0	0	0	8
PLOS Computational Biology	1	2	2	2-3****	2	0	0	0	13
The EMBO Journal	1	1	1	1	1	0	0	0	5
PLOS Biology	1	2	2	2	2	0	0	NA	9
Nature Methods	1	2	1	1	2	1	1	1	10
eLife	2	2	2	2	2	1	1	1	13
Genome Research	2	2	2	2	0	0	0	0	8

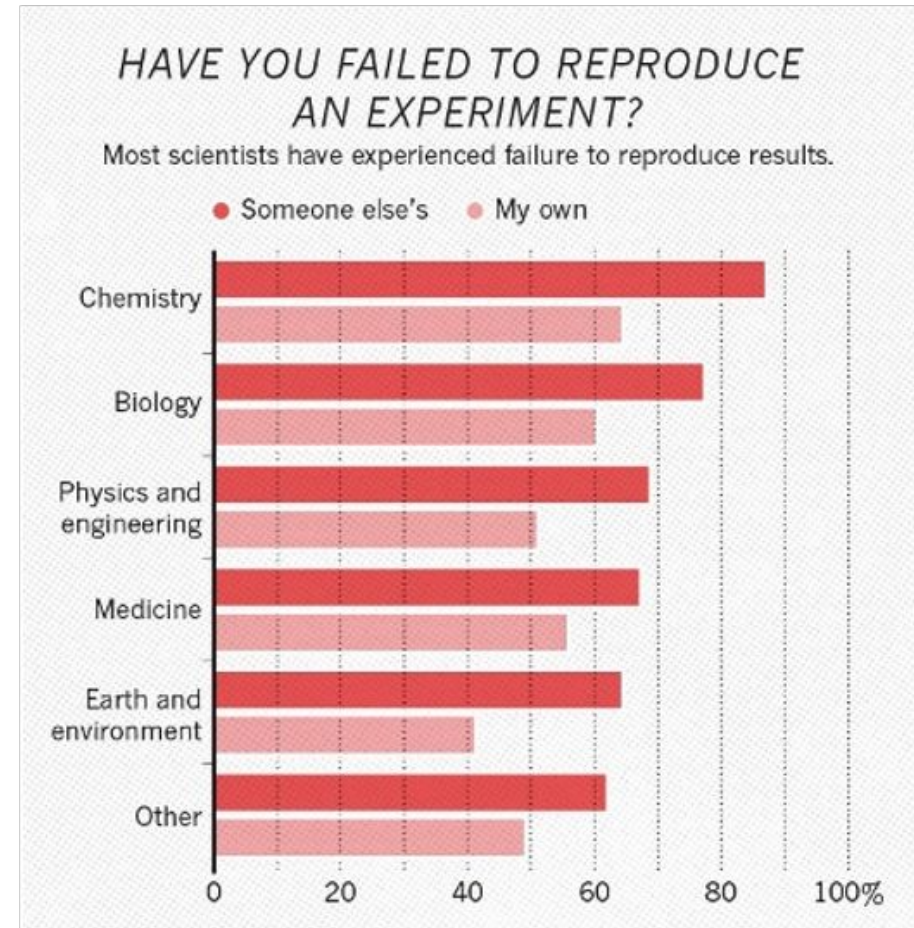
Repository	Costs	Size	Retention period
Dryad Digital Repository	DPC per data submission is \$120 + charges excess storage fees for data totaling over 50GB (\$50 for each additional 10GB)	300GB per data publication individual files should not exceed 10GB.	permanently archived and available through the California Digital Library's Merritt Repository .
figshare	250GB/\$745 to 5TB/\$11,860	250GB - 5TB	AWS, lifetime of the repository
Harvard Dataverse	free	up to 1TB and files of up to 2.5GB	NA
Open Science Framework	free	individual files must be 5GB, public projects and components to 50 GB	Always welcome to deactivate or delete your account. Uses Google Cloud (optional storage in Frankfurt).
Science Data Bank	Do not charge individual users at present. All fees incurred throughout the publishing process will be waived. However, we reserve the rights to charge users for data storage, review and publishing,	Any	Do not guarantee life-long free services. China
Zenodo	free	<50GB per dataset. If bigger, contact	Lifetime of the repository = lifetime of the host laboratory CERN (exp. programme defined for the next 20 years at least.)
Code Ocean HELMHOLTZ	Free version: 20Gb, 10h/month	NA	NA

Definitions

Reuse: using the same data to answer a different question than originally intended.

Reproduce: being able to follow the same footsteps with same data to trace back and recreate the same conditions to *hopefully* arrive at the same conclusion!

Replicate: collecting new data to arrive at the same scientific findings as a previous study.





“Data is open if it can be freely accessed, used, modified and shared by anyone for any purpose - subject only, at most, to requirements to provide attribution and/or share-alike.”

Open Knowledge Foundation

A. Legally open: open (data) license

B. Technically open: data available in machine-readable and bulk form.

Code Transparency

Level 1: Disclosure

Article state whether or not code is available. If so, give URL.



Level 2: Mandate

Code MUST be posted in a trusted repo (exceptions permitted for legal or ethical constraints).



Level 3: Verified Mandate

Level 2 + Can code be replicated? Reported analysis will be reproduced independently prior to publication



Code Transparency: Level 2

- All computer **code** central to the findings being reported should be **available** to readers to ensure **reproducibility**
- Commercially available **software** or publicly archived **source code** should be **appropriately referenced** (with the version included).
- Author-written **source code** should be **archived** in a **permanent** public repository **prior** to publication and cited (see [Data and Code Deposition](#)).
- In exceptional cases where for example, security concerns, legal restrictions, or proprietary hardware preclude sharing of custom code, an alternate means of ensuring reproducibility must be **arranged with the editor**. Preferred option: **include pseudocode that fully clarifies the underlying algorithms**; this pseudocode will be subject to peer review and may require further elaboration in accord with reviewers' feedback.
- The reason(s) for the **code restrictions** in these special cases should be **explained clearly** in the acknowledgments.

Journals with written guidelines on ML & statistics:

- **Bioinformatics**

Provide support / **maintenance** for min 3yrs post publication

- **Gigascience**

Use of **literate programming outputs**, such as **Jupyter notebooks**, and/or publishing environments, such as **Code Ocean** or **Gigantum**

The manuscript **should be available for testing by reviewers**, and also include test data

- **Science**

Data pre-processing steps such as transformations, re-coding, re-scaling, normalization, truncation, and handling of below detectable level readings and outliers should be **fully described**; any **removal** or **modification** of data values must be fully acknowledged and **justified**.

- **PLOS Computational Biology**

Papers using **leave-one-out** will be **editorially rejected** unless there is a special circumstance

- **EMBO Journal**

Data pre-processing steps such as transformations, re-coding, re-scaling, normalization, truncation, and handling of below detectable level readings and outliers should be fully described; any removal or modification of data values must be fully acknowledged and justified.

The number of sampled units, n , upon which each reported statistic is based must be stated.

For continuous variables, distributions should be described using graphical displays such as scatterplots, boxplots, or histograms or by reporting measures of central tendency (e.g., mean or median) and dispersion (e.g., SD, interquartile range).

For continuous variables that are approximately normally distributed, mean and SD are suitable measures for center and dispersion, respectively. For continuous variables with asymmetrical distributions, median and range (or interquartile range) are preferred to mean and SD. All measures of central tendency or dispersion that are used should be identified.

For very small samples sizes (e.g., $n < 20$), presentation of all data values in tabular format is desirable unless presentation would violate restrictions for privacy or confidentiality for human subjects.

Methods used for conducting statistical tests (e.g., t-test, Wilcoxon signed rank test, Wald test of regression coefficient) and for constructing confidence intervals (e.g., normal-based 95% CI: mean \pm 2SD, likelihood ratio-based interval) should be clearly stated.

The testing level (α) and whether one-sided or two-sided testing was used should be reported for each statistical test; typically, two-sided testing is appropriate, but if one-sided testing is used its use should be justified.

Adjustments made to alpha levels (e.g., Bonferroni correction) or other procedure used to account for multiple testing (e.g., false discovery rate control) should be reported.

When Bayesian analyses are conducted, any assumptions made for prior distributions must be fully described. Sufficient information should be supplied to allow readers to judge whether any assumptions necessary for the validity of statistical approaches (e.g., data are normally distributed, survival data are consistent with proportional hazards in a Cox regression model) have been verified.

An accounting of missing data values should be provided; if imputed data values are used in statistical analyses, the methods used for imputation should be fully described.

Authors should present results in a complete and transparent fashion so that stated conclusions are backed by appropriate statistical evaluation and limitations of the study are frankly discussed

Point estimates of population parameters (e.g., mean, correlation coefficient, slope) or comparative measures (e.g., mean difference, odds ratio, hazard ratio) should be accompanied by a measure of uncertainty such as a standard error or a confidence interval.

How do I open protocols/software?

In repositories for experimental workflows and protocols:

- [Protocol Exchange](#) open repository for sharing scientific research protocols
- [Protocols.io](#) platform for data management and protocol sharing



MDC repository:

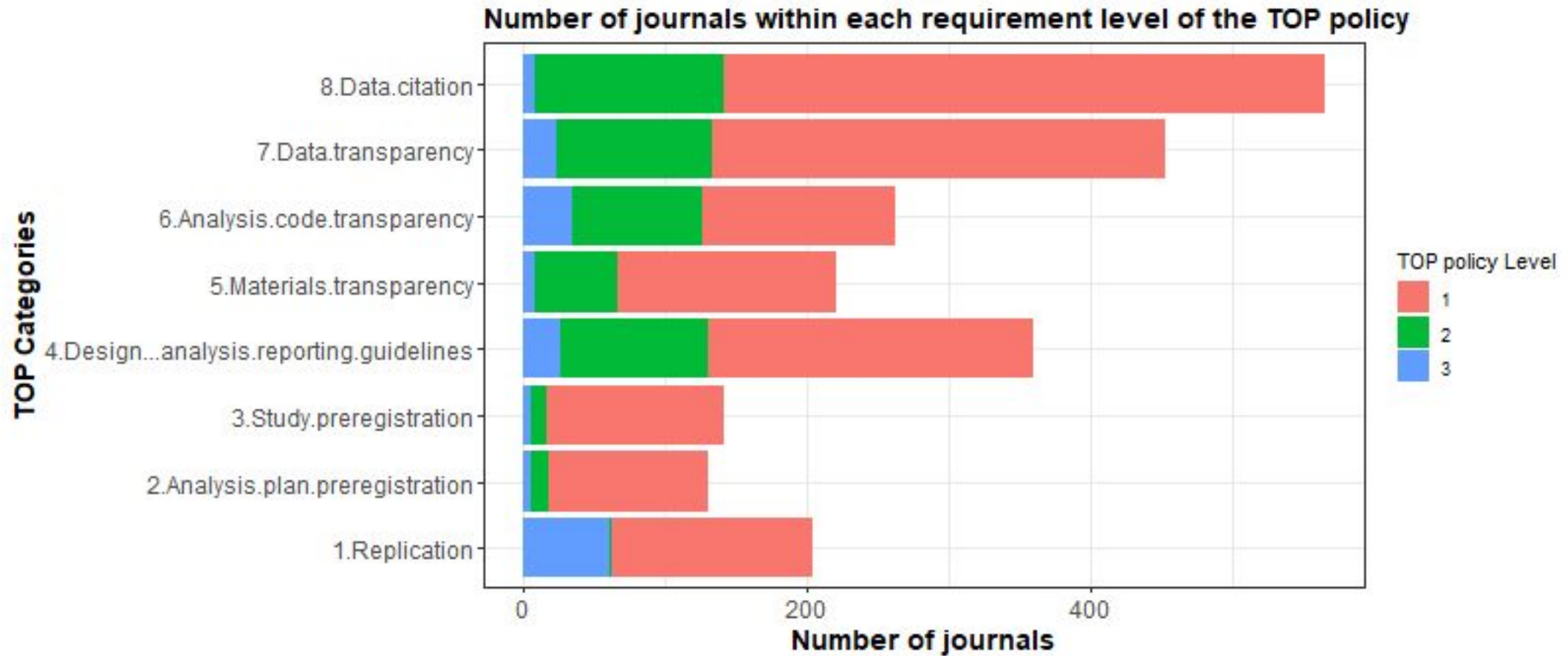
- <https://edoc.mdc-berlin.de/>

In software repositories:

- [Github](#) - development platform to host and review code, manage projects, and build software
- [Savannah](#) - hosts free projects that run on free operating systems, with a focus on GNU software
- [SourceForge](#) - Open Source software community and hosting platform



The current state of journal requirement trends



Is Open Data equal to FAIR Data?

Data that is subject to **IPR** issues and data that is **sensitive** in nature.



While established for data-sharing methods in health research are:

- **anonymisation** and/or
- **access control**
- making sure that there are also **consent forms** in use where necessary

Information about general purpose repositories

Recommended by Springer Nature (below) and F1000 (right):

[Dryad Digital Repository](#)

[figshare](#)

[Harvard Dataverse](#)

[Open Science Framework](#)

[Science Data Bank](#)

[Zenodo](#)

DATA TYPE	WHERE TO SUBMIT*	WHAT TO INCLUDE IN THE DATA AVAILABILITY SECTION OF YOUR ARTICLE
Any	Dryad	Title, DOI
Any, but especially data in SAV and POR formats	Dataverse	Title, DOI
Any	Figshare ^s	Title, DOI
Any, but especially deposits with mixed data, materials and documents	Open Science Framework [†]	Title, DOI
Any, but especially deposits with mixed data and code	Zenodo	Title, DOI
Deposits of mixed data and code	Code Ocean	Title, DOI, embed code for interactive reanalysis tool
Any biological data, but especially data linked to studies in other databases	BioStudies	Title, accession number



Source: <https://f1000research.com/>

Repositories for unstructured and/or Large Data*

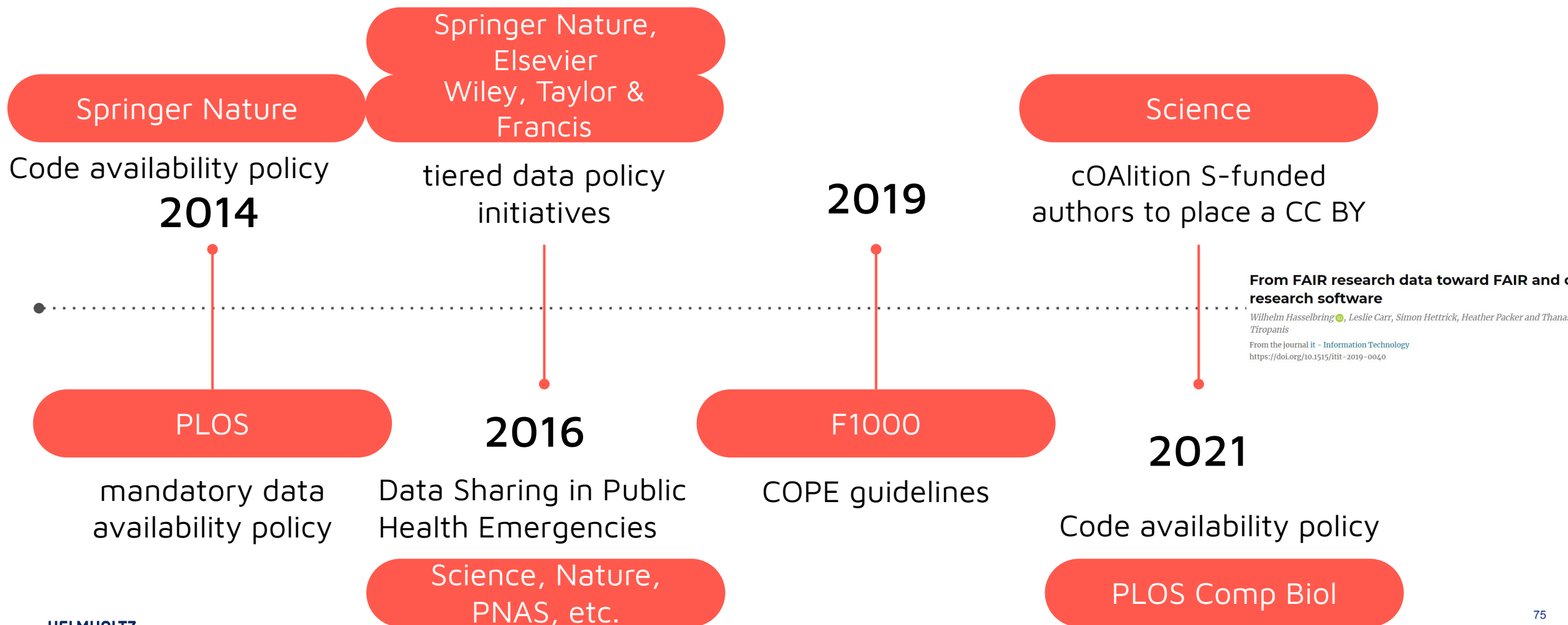
Recommended by PLOS Comp Biol:

- [BioStudies](#)
- [CSIRO Data Access Portal](#)
- [Data Archiving and Networking Services \(DANS\)](#)
- [GigaDB](#)
- [SimTK](#)
- [Swedish National Data Service](#)



*PLOS thanks the Open Access Nature Publishing Group journal, *Scientific Data*, for their own [list of recommended repositories](#)
HELMHOLTZ

Increasing levels of demands



Data Transparency: Level 1



- Providing large datasets in supplementary information is strongly discouraged
- Min dataset may be provided through preferred deposition in:
 1. public community/ discipline-specific repositories,
 2. custom proprietary repositories for certain types of datasets,
 3. or general repositories
- Mandatory deposition of data is required for certain data types