

“Introduction to Research Data Management – Why? When? How?” 1st Workshop 15/06/21

DATA MANAGEMENT



cc-by https://hochschulresearch.wordpress.com/

Dr. Birte Pfeiffer
Leibniz Centre of Tropical Marine Science (ZMT)
ZMT Research Data Team

Who we are The Research Data Team at ZMT

Joscha Schmiedt,
Head and Data
Technology

Miroslav Shaltev,
Data Technology



Birte Pfeiffer
Data Curation

Team email address: research-data@leibniz-zmt.de
My email address: birte.pfeiffer@leibniz-zmt.de

HELLO 🖐️

where are you located?

Drop a pin on the map and add a comment (or hit C on your keyboard), listing your:

- name
- title + organisation
- location
- fun fact about yourself



miro

02

Workshop Agenda

Topic
Introduction
> Your experience with research data management (Mentimeter Survey)
> What is research data?
> Your turn: How would you classify your research data?
> What is research data management?
Why manage research data?
When to manage research data?
How to manage research data?
> Data Management Plan
> Data Management Plan Tools
> Short Film
> Data Organisation: Formats - Folder Structure - Naming
> Your turn - Breakout Session: Order and Structure
> Data Organisation: Versioning
Summary and Outlook
Your turn: Meeting Reflection

Aim of the Workshop: Transfer foundational knowledge and skills in research data management

03

Your Experience with Research Data Management (RDM)

Go to www.menti.com and use the code 5588 5275

How experienced are you in research data management? 



Go to www.menti.com and use the code 5588 5275

What topics are you interested in? 



04

What is Research Data?

"Research data means all data that was created in the course of research or results from it."*

"Research data can be placed into two broad categories: quantitative or qualitative."**



How would you classify your research data?

"Research data takes many different formats."**

"Data may be grouped into four main types based on methods for collection: observational, experimental, simulation, and derived."**

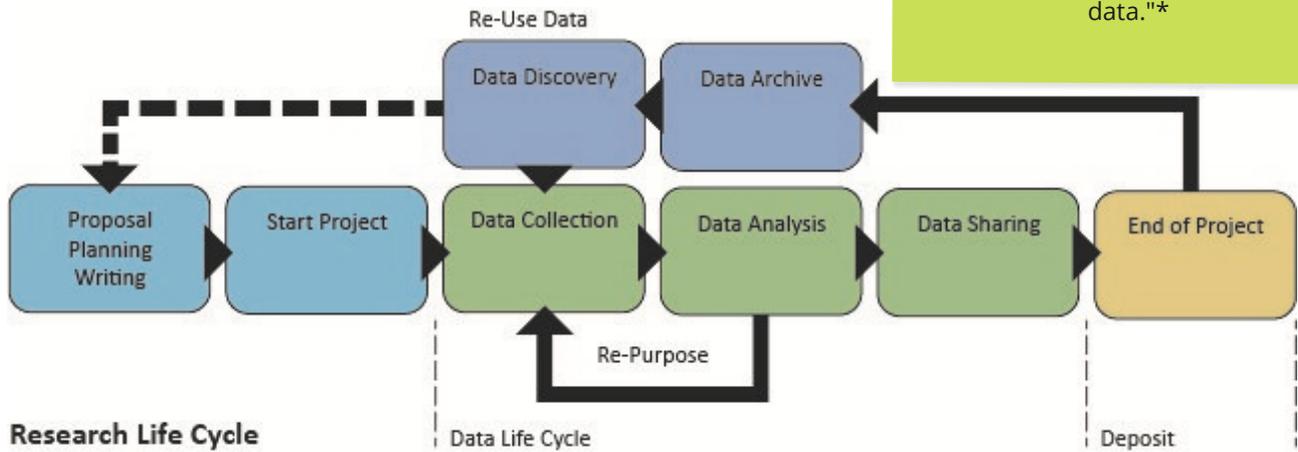
05

*Source: Biernacka et al. (2020)

** Source: <https://libguides.maclester.edu/c.php?g=527786&p=3608583>

What is Research Data Management?

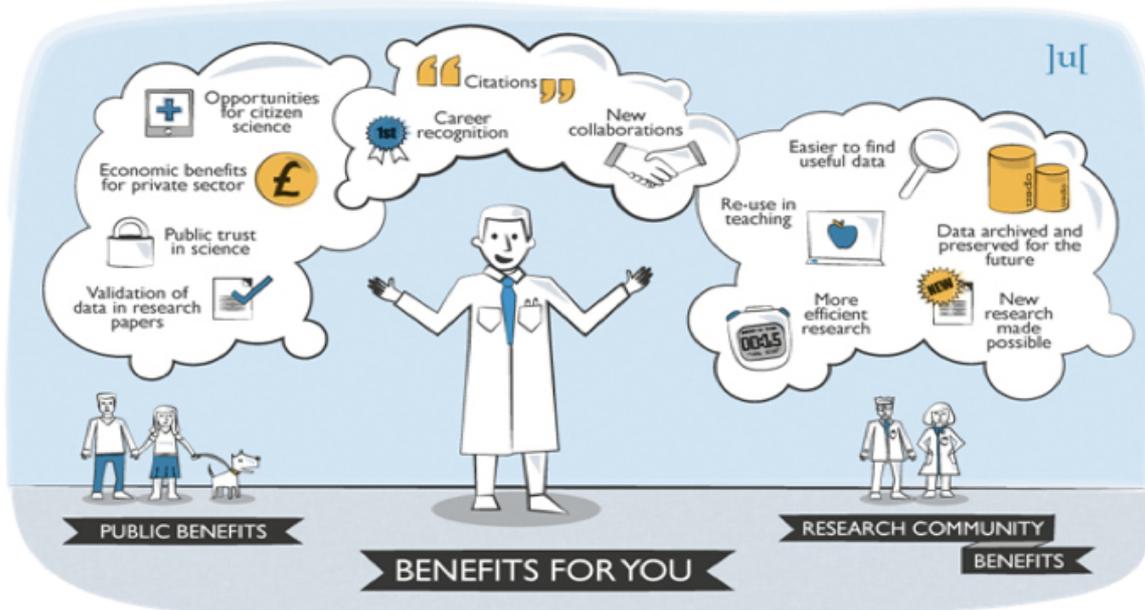
"Research data management [...] complements research from the initial planning stage all the way to archiving, reuse or deletion of the data."*



06

Source of Figure: <https://library.maastrichtuniversity.nl/research-support/rdm/guide/>
 *Source: *Source: Biernacka et al. (2020)

Why manage research data?

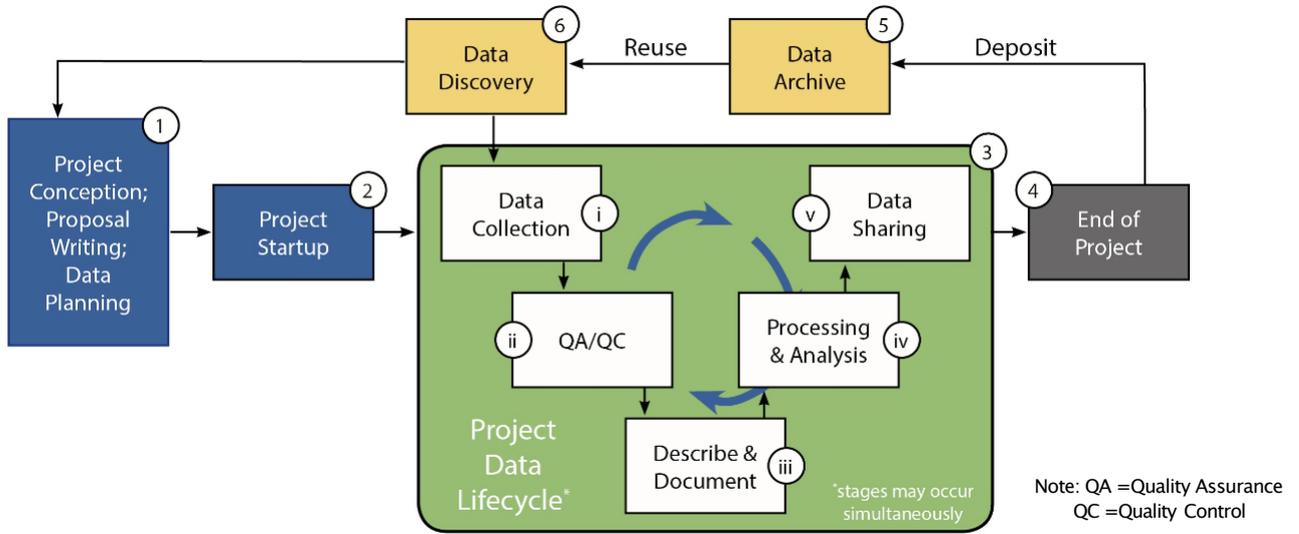


07

Source: <https://www.forschungsdaten.info/themen/bewahren-und-nachnutzen/open-data-open-access-und-nachnutzung/>

When to Manage Research Data?

Data Management During the Research Lifecycle



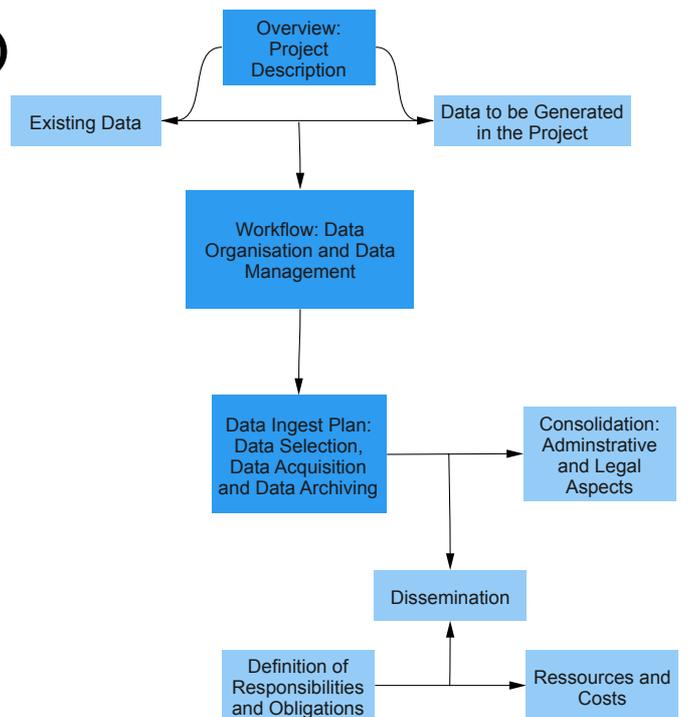
08

Source: <https://www.ohsu.edu/library/plan-and-organize>

Data Management Plan (DMP)

"A data management plan (DMP) is a document that describes the intended management of research data."*

*Source: Biernacka et al. (2020)



Source: Based on <https://www.forschungsdaten.info/themen/informieren-und-planen/datenmanagementplan/>

09

Data Management Plan (DMP) Tools



[DMPonline](#)
(Digital Curation
Centre, DCC)



University of California
Curation Center der California
Digital Library (CDL)



Research Data
Management Organizer
(RDMO), LeibnizInstitut
für Astrophysik Potsdam



[ARGOS](#) (OpenAIRE)

Source:
https://www.openaire.eu/images/easyblog_articles/1077/b2ao3_large_argos-what-it-is.png

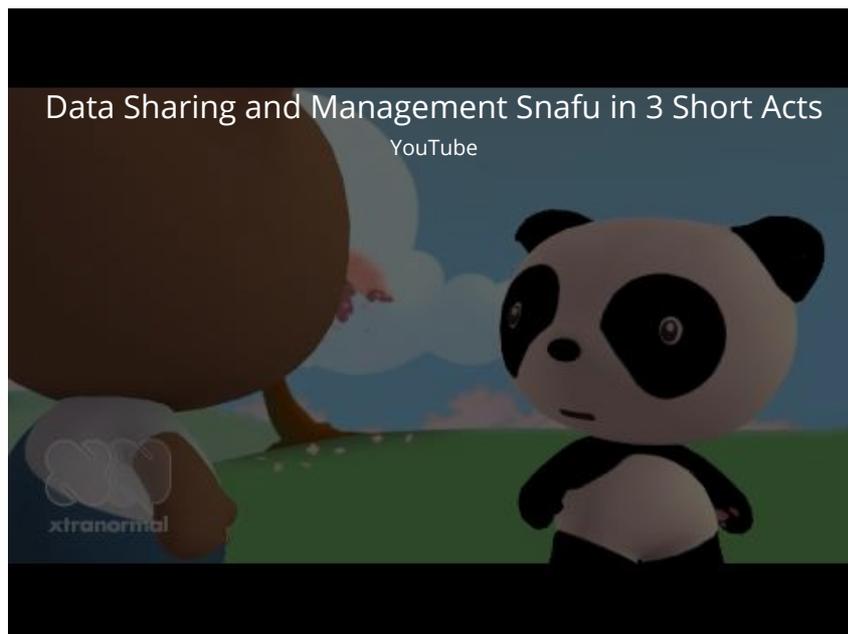


[GFBio Data Management Plan Tool](#)

10

and others... (overview, in German: <https://www.forschungsdaten.info/themen/informieren-und-planen/datenmanagementplan/>)

Data Sharing and Management Snafu in 3 Short Acts



11

Data Organisation File Formats

Best Practices*

- Non-proprietary / patent-encumbered
- Unencrypted
- Uncompressed
- Open, documented standard
- Standard representation (ASCII, Unicode)

Type of data	Recommended formats	Acceptable formats
Geospatial data vector and raster data	ESRI Shapefile (.shp, .shx, .dbf, .prj, .sbx, .sbn optional) geo-referenced TIFF (.tif, .tfw) CAD data (.dwg) tabular GIS attribute data Geography Markup Language (.gml)	ESRI Geodatabase format (.mdb) MapInfo Interchange Format (.mif) for vector data Keyhole Mark-up Language (.kml) Adobe Illustrator (.ai), CAD data (.dxf or .svg) binary formats of GIS and CAD packages
Textual data	Rich Text Format (.rtf) plain text, ASCII (.txt) eXtensible Mark-up Language (.xml) text according to an appropriate Document Type Definition (DTD) or schema	Hypertext Mark-up Language (.html) widely-used formats: MS Word (.doc/.docx) some software-specific formats: NUD*IST, NVivo and ATLAS.ti

3

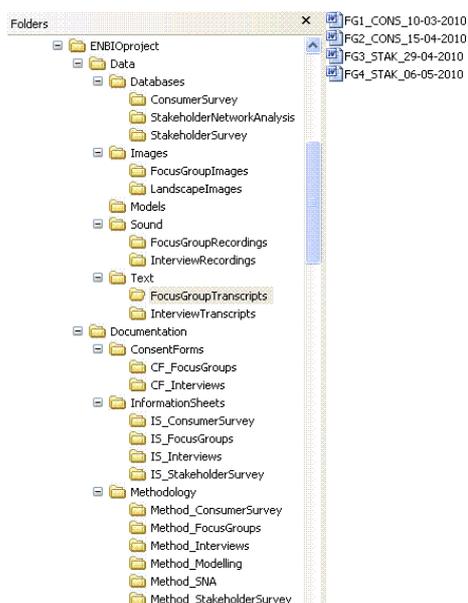
12

*Source: *Source: Biernacka et al. (2020)



Source: Screenshot of https://zenodo.org/record/4041512/files/OpenAIRE_Guides_DataFormat_sept-2020.pdf?download=1

Data Organisation Folder Structure



Best Practices*

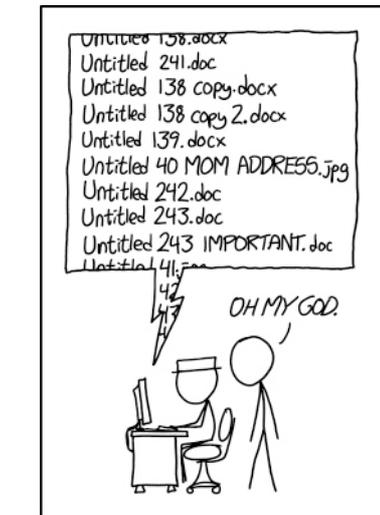
- Plan the folder hierarchy in the beginning of a project
- Establish a folder hierarchy that aligns with the project
- Be consistent
- Separate ongoing and completed work
- Review records
- Backup your files

*Sources: Biernacka et al. (2020), <https://libraries.mit.edu/data-management/store/organize/>

13

Source of Figure: <https://www.ukdataservice.ac.uk/managing-data/format/organising>

Data Organisation Naming



PROTIP: NEVER LOOK IN SOMEONE ELSE'S DOCUMENTS FOLDER.

14

Source of Comic: xkcd.
"Documents." <https://xkcd.com/1459>.
Shared under [CC-BY-NC](https://creativecommons.org/licenses/by-nc/4.0/) license

Best Practices*

- Comprehensive names
- Uniform scheme
- Define naming conventions in a README file
- Logical structure
- Date for chronological sorting in the following form: YYYYMMDD
- Avoidance of spaces and special characters

Tools/Mac:*
[Renamer 5 \(for Mac\)](#), [Name Changer](#),
[ExifRenamer](#)

Tools/Windows:* [A nt Renamer](#),
[RenameIT](#), [Bulk Rename Utility](#),
[Total Commander](#)

Documented naming conventions or abbreviations:

- [sediment]_[sample]_[instrument]_[YYYYMMDD].csv
- [project]_[interview]_[place]_[personID]_[YYYYMMDD].mp4

*Source: *Source: Biernacka et al. (2020)

Breakout Session: Order and Structure

Task 1: Please discuss with your partners: Which of the following examples follow a good naming convention?

- Olga_170413_probe17k
- Naturepaper karl britta james finished!
- Vm4520132Schmidt.pdf
- 647749157.pdf
- 170413_sample17k_olga
- Naturepaper+karl+britta+james &nal
- Olga170413sample17k
- Krst_765_spct_1203
- Naturepaper+karl+britta+james finished! reworked
- Cristal_765_spectr_20161203
- Nature_karlbrittajames_endendversion
- 28q8QGIHKwrRw.pdf
- Conference_Digital_Science.pdf

BREAKOUT SESSION

Task 2: Discuss the naming conventions in your disciplines!

Data Organisation Version Control

Version control software: Git or Subversion

Best Practices*

- Use sequential numbering
- Include date and version number in the name (e.g. v1.0.0)
- Use a version control table
- Define responsibilities for completion of files
- Save milestone versions
- Store obsolete file versions separately after a backup
- For large amounts of data, use version control software

For collaborative documents and storage locations such as Wiki, Google Docs or cloud, versioning and change tracking is available.

Examples of file labelling with version control:

- NumberingSystem 1: Lipid_analysis_rate_V2.1
- NumberingSystem 2: Doe_interview_July2010_V1
- NumberingSystem 3: Level 0, Level 1, ...

Australian National Data Service: <https://www.ands.org.au/working-with-data/data-management/data-versioning>

16

*Source: Biernacka et al. (2020)

Summary and Outlook

Topic
Summary of Last Workshop
Documentation and Metadata
Ethical and Legal Norms
Storage, Backup and Archiving
Data Sharing, Access and Re-Use

Group Work

Exercises

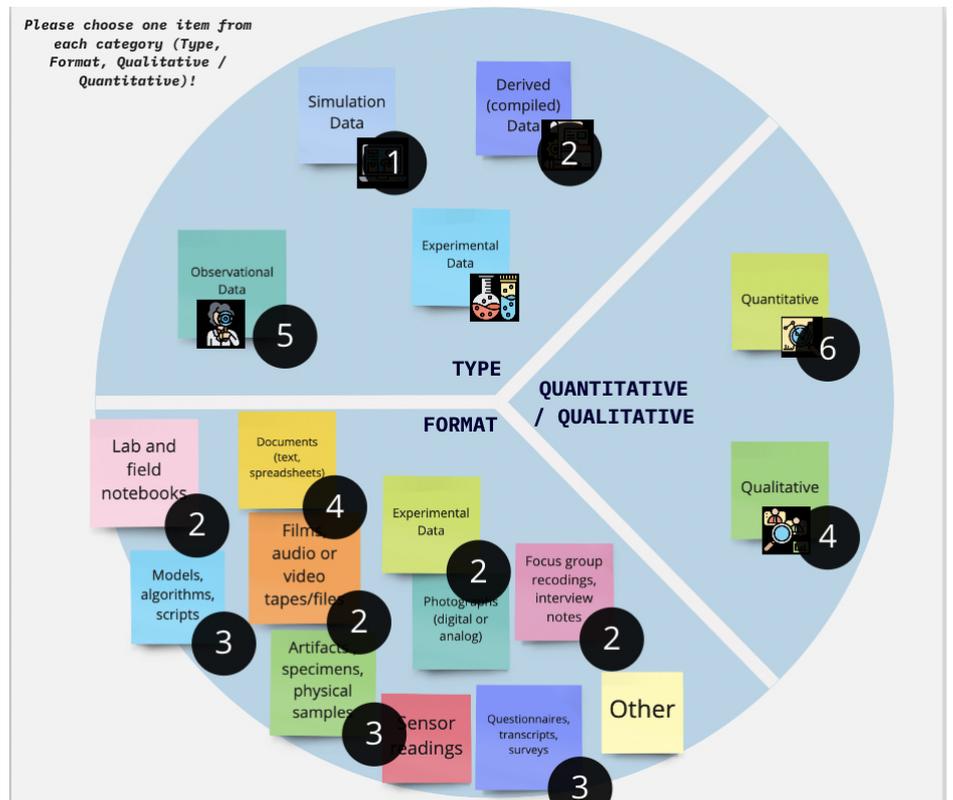
17

Meeting Reflection

<p>What have you learned?</p>	<p>What was surprising?</p>
<p>What user problems or opportunities we should fix before the next workshop?</p>	<p>What have you missed? What would you like to see more of?</p>

18

Supplement: How Would You Classify your Research Data?



19

Meeting Reflection

What have you learned?

- Some things were quite similar to what we have learned before, but nicely condensed in the introduction to data management. I missed it a bit in the studies before.
- Property naming our files, different places where we can save our data.
- I learned about different online DMS platforms, I would like to try some of them.
- I liked the links such as <https://dmsonline.dst.ac.uk/> for getting assistance with data management plans.

What was surprising?

- Someone can ask for our data anytime and we should be prepared for moving.
- Learned that data management starts as early as project conceptualization.
- The short video was quite informative.

What user problems or opportunities we should fix before the next workshop?

- Maybe to give some actual examples on data storage.

What have you missed? What would you like to see more of?

- How to store data on the different platforms, please.
- I would like to see more practical.
- I would like to get more information about ZMT Digiz and with examples for practical support in data management.

Resolve

Guest 15 Jun, 13:43

I am very delighted about data organization. It is a first step to data management that will surely help me in my work

2

Documentation and Metadata

Health Survey for England, 2010

Project-level

Title	File name	Size (MB)
"Interviewer, Nurse, Coding and Editing Instructions"	6986supcorintodocs.pdf	4.26
"Questionnaires, Showcards, Coding Frames and Consent Booklets"	6986interviewingdocs.pdf	5.87
Additional data for renal analysis	6986userguide_renal_analysis_2009_2016.pdf	0.55
Lists of Variables and Derived Variables	6986datadocs.pdf	3.38
Study information and citation	UKDA_Study_6986_Information.htm	0.01

Data-level

Name	Type	Width	Decimals	Label	Values	Missing
175	numeric	2	0	Which of the qualifications do you hold? Do you have? 10	0-9	No data
176	numeric	2	0	Actively class for last week	0-9	No data
177	numeric	2	0	Manager/foreman	0-9	No data
178	numeric	2	0	Ever had paid employment or self-employed	0-9	No data
179	numeric	2	0	Full-time or part-time	0-9	No data
180	numeric	2	0	How long have you been looking	0-9	No data
181	numeric	2	0	Allow to call week with 2 weeks (Government training scheme)	0-9	No data
182	numeric	2	0	Looking paid and unpaid combined	0-9	No data
184	numeric	5	1	HS (SIC - long version (three decimal))	0-9	No data
185	numeric	2	0	Ever had other employment (week) to start week	0-9	No data
186	numeric	2	0	Age when last had a paid job	0-9	No data
187	numeric	4	0	Year left last paid job	0-9	No data
188	numeric	2	0	Month last left paid job	0-9	No data
189	numeric	2	0	Social class	0-9	No data
190	numeric	2	0	Sex in Forensic Group	0-9	No data
191	numeric	2	0	Self-employed, how many employees	0-9	No data
192	numeric	2	0	Age last birthday	0-9	No data

Source: <https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=6986&type=Data%20catalogue#!#documentation>, Accessed 2/7/21

Source: <https://ukdataservice.ac.uk/media/604455/spssdocu.gif>, ssed 2/7/21

Catalogue Metadata
"Data about data"



Metadata Standards: In order to be useful, metadata needs to be standardized. This includes agreeing on language, spelling, date format, etc.

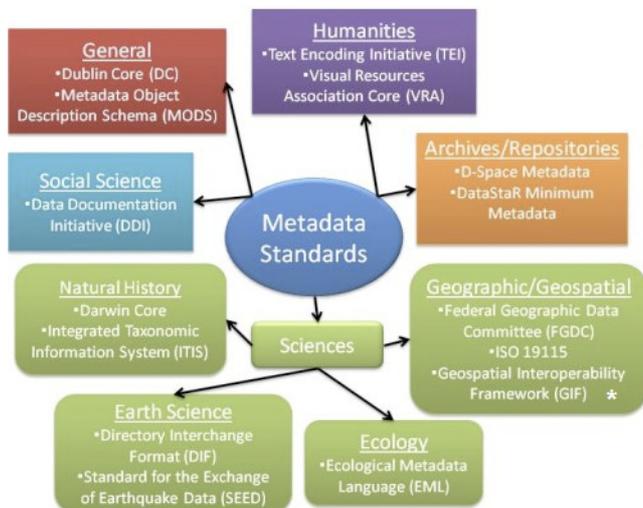
A key component of metadata is the schema. Metadata schemas are the overall structure for the metadata. It describes how the metadata is set up, and usually addresses standards for common components of metadata like dates, names, and places.

3

Source: <https://tdan.com/toward-a-better-understanding-of-metadata-metadata-defined/19154>, Accessed 2/7/21

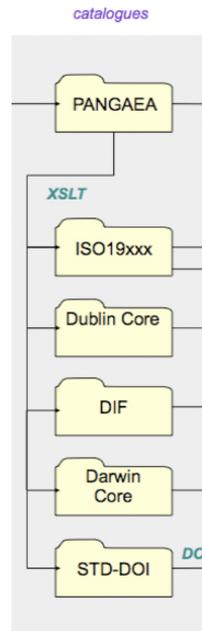
Source: <https://guides.lib.uc.edu/metadata/standards>, Accessed 2/7/21

Documentation and Metadata: Metadata Standards



Source: Metadata Concept Map by [Amanda Tarbet](#) is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License](#), <https://libguides.bc.edu/dataplan/metadastandards>, Accessed 2/7/21

4



Source: https://wiki.pangaea.de/wiki/File:Pangaea_metadata.png, Accessed 2/7/21

Discipline-specific Metadata Standards
<https://www.dcc.ac.uk/guidance/standards/metadata>

Example

Documentation and Metadata: Electronic Lab Notebooks (ELN)

Electronic Lab Notebooks are designed to document the conception, execution and evaluation of scientific experiments, observations or other studies and the research data generated in this context.*

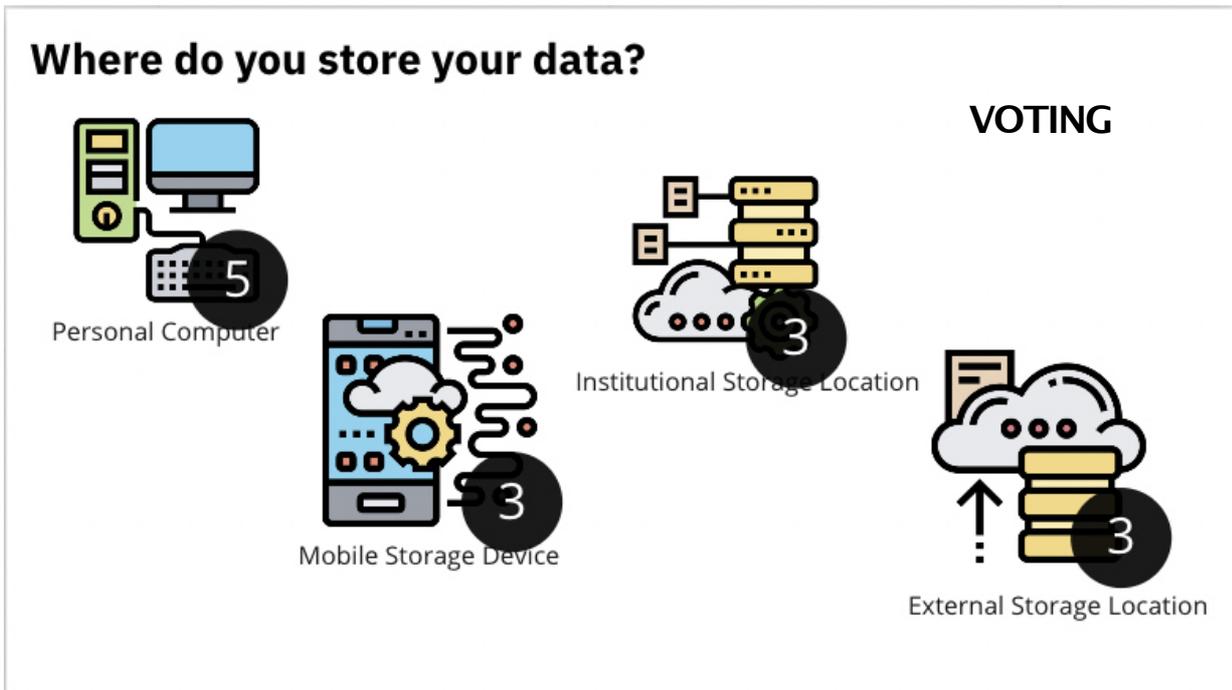
Examples of ELN Software Products*

- eLabFTW (Open Source) <https://www.elabftw.net>
- Labfolder (commercial) <https://www.labfolder.com>
- openBIS (Open Source) <https://openbis.ch>
- Rspace ELN (commercial) <https://www.researchspace.com>

5

*Source: Biernacka et al. (2020)

Storage, Backup and Archiving: Storage Locations



6

Storage, Backup and Archiving: Storage Locations

Advantages	Disadvantages
Own PC Easy accessible, private Easy accessible all time available easy access, personal control	Can't get access if the device is damaged or lost Can get lost no back up can be broken
Mobile Storage (USB stick, external hard drive) easy transfer eg from lab to PC Portable good as back up	Can get lost vulnerable to physical damage can get lost, more vulnerable to physical damage
Institutional storage (cloud, virtual drives) Accessible via internet from many devices save to physical storage, easy access through institutional account backup and shared can be accessed from many devices	Needs internet connection storage limitations can't be changed personally, restricted sharing with external, need for internet connection to access after personal change no access
External Storage (cloud of a free or commercial provider) Accessible, sharing, backups easy access to large capacities, easy sharing availability from everywhere	Limited storage possible lack of security, need of internet access Space

GROUP WORK

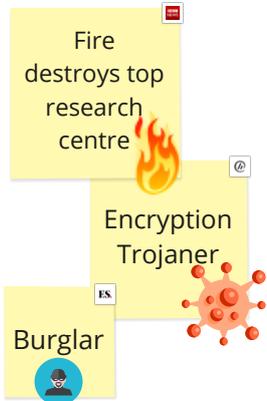


Advantages and disadvantages of different storage locations

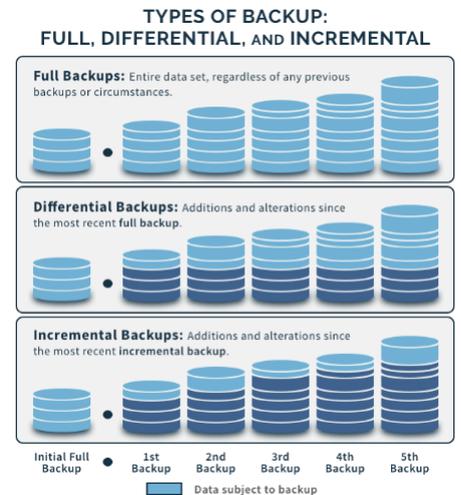
7

Storage, Backup and Archiving: Backup Options

Incremental / Differential Backup Options:*



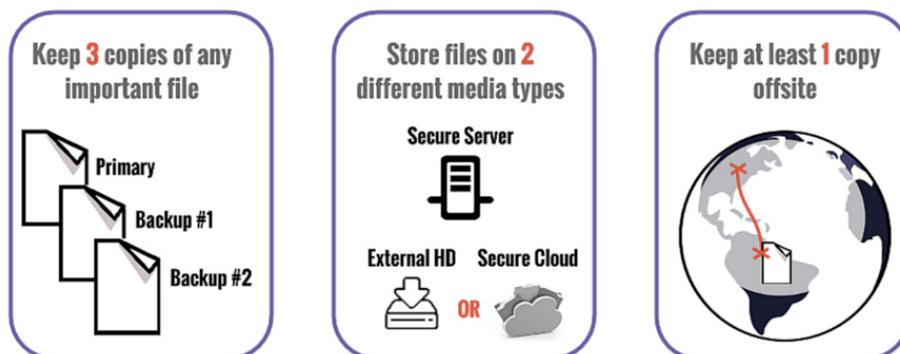
*Source: Biernacka et al. (2020)



Wallen (2020), <https://spanning.com/blog/types-of-backup-understanding-full-differential-incremental-backup/>, Accessed 2/7/21

8

Storage, Backup and Archiving: Strategies for Secure Backup



Source: https://guides.nyu.edu/data_management/storage-backup

9

Storage, Backup and Archiving: Long-Term Archiving Requirements*

- Technical requirements
- Cost of services
- Making the data accessible
- Longevity of the service provider
- Seal for trustworthy long-term archives

Archiving research data means submitting it to a data centre, archive or repository where it will be protected in the long term against loss, deterioration, unauthorised or inappropriate access, and future incompatibility.**

A trusted digital repository is one whose mission is to provide reliable, long-term access to managed digital resources to its designated community, now and in the future.***

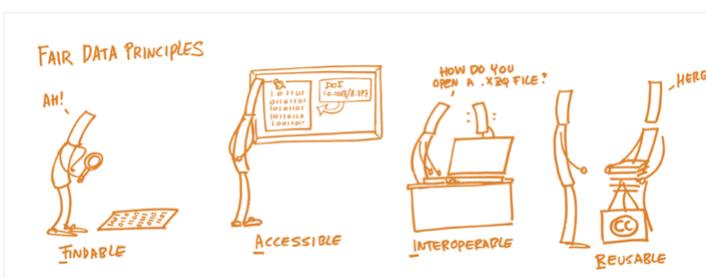


*Source: Biernacka et al. (2020)

**Source: <https://library.bath.ac.uk/research-data/archiving-and-sharing/choosing-an-archive>

***Source: RLG/OCLC Working Group (2002)

Data Sharing, Access and Re-Use: FAIR & CARE



Source: <https://book.fosteropenscience.eu/en/>, Accessed 2/7/21

Expert Information:
 Tanhua (2019) et al. "Ocean FAIR Data Services": Article outlines how the FAIR principles apply to ocean data
<https://www.frontiersin.org/articles/10.3389/fmars.2019.00440/full>



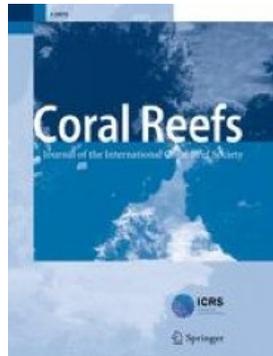
Source: <https://www.gida-global.org/care>, Accessed 2/7/21

Data Sharing, Access and Re-Use: Publication Strategies



Source: Auke Herrema, https://libguides.unisa.ac.za/rdm_platform_trial

Data for an "Enriched Publication"



Source: <https://data.globalchange.gov/journal/coral-reefs.html>



Source: <https://www.earth-system-science-data.net>

Data Journals

Repositories



PANGAEA.

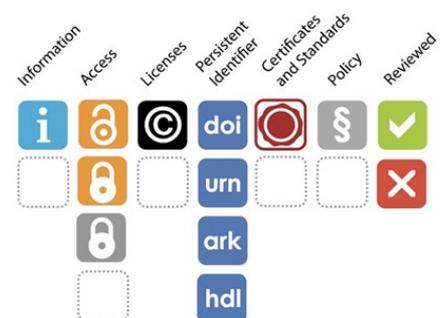
Source: <https://www.pangaea.de/>

12

Data Sharing, Access and Re-Use: Repository Selection I

General Criteria

1. Use an external data archive or repository already established for your research domain
2. If available, use an institutional research data repository
3. Use a cost-free data repository such as [Zenodo](https://zenodo.org/)
4. Search for other data repositories here: re3data.org



Source: <https://www.openaire.eu/opendatapilot-repository-select-data-repository>

13

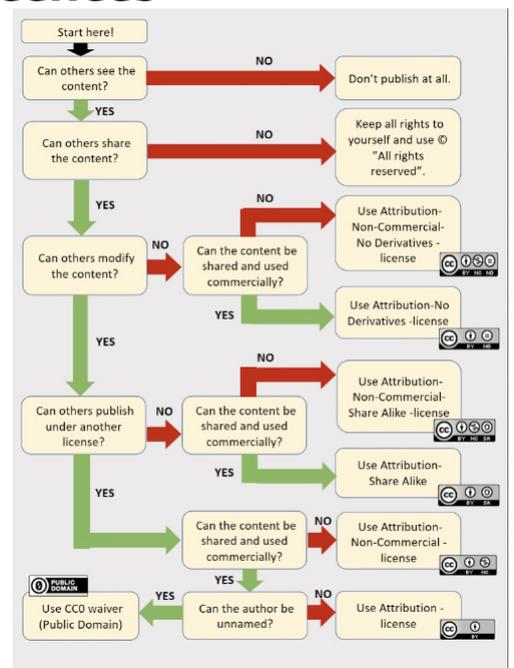
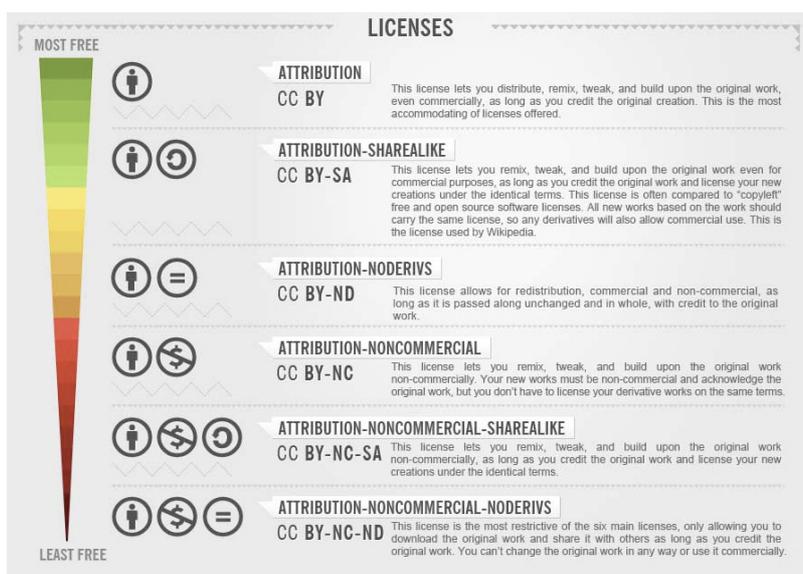
Data Sharing, Access and Re-Use: Repository Selection II

Specific Criteria



Source: <https://doi.org/10.1371/journal.pone.0078080.g001>, Accessed 2/7/21

Data Sharing, Access and Re-Use: Licences



Source: https://libapps-eu.s3.amazonaws.com/accounts/139959/images/License_selector.jpg, Accessed 2/7/21

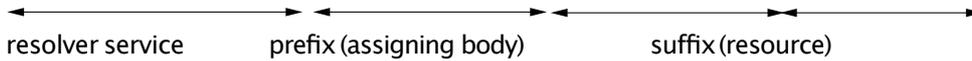
Source: Creative Commons License Spectrum™ by Shaddim (CC BY), <https://creativecommons.org/share-your-work/public-domain/freeworks/>, Accessed 2/7/21

Data Sharing, Access and Re-Use: Persistent Identifier

Anatomy of a Digital Object Identifier (DOI)



<https://doi.org/10.1594/PANGAEA.897645>



- Resolver Service: Ensures the DOI resolves to an online metadata record about the dataset or collection
- Prefix: Assigned by a DOI Registration Agency (i.e. DataCite for research datasets and collections) and always starts with '10.' This distinguishes it as a DOI as opposed to other types of Handle
- Institution Suffix
- Unique Suffix

Source: Australian National Data Service, https://www.ands.org.au/_data/assets/pdf_file/0006/715155/Digital-Object-Identifiers.pdf

Citation: Likendey, Julian; Pisternick, Timo; Neumann, Sarah Isabel; Dumur Neelayya, Danishta; Bröhl, Stefanie; Neehaul, Yashvin; Moosdorf, Nils (2019): Age at length data for juvenile Grey demoiselle Chrysiptera glauca (Pomacentridae) in two contrasting tide pools in Mauritius. PANGAEA, <https://doi.org/10.1594/PANGAEA.897645>.

The DOI name directly links to the storage location of the object.



Data Sharing, Access and Re-Use: Access



OPEN ACCESS

- Metadata is fully discoverable
- Data are accessible and immediately downloadable
- Preferred option for non-sensitive data from completed projects



MEDIATED OPEN ACCESS

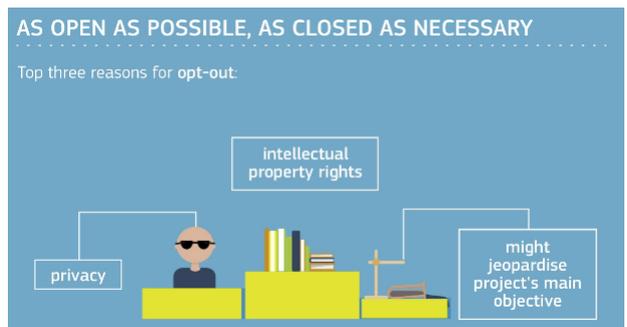
- Metadata is fully discoverable
- Mediated access to data via data custodian
- Good option for sensitive or confidential data



CLOSED ACCESS

- Metadata is not publicly available
- Data not discoverable or available to third parties
- Safest option for highly sensitive data

Source: Own compilation.



Burgelman (2018)

Data Sharing, Access and Re-Use: Re-Use

Searching for existing data

- Direct browsing of discipline-specific and multidisciplinary repositories
- Search by means of meta search engines, e.g.



- Search in data journals



18

*Source: Biernacka et al. (2020)

Data Sharing, Access and Re-Use: Re-Use

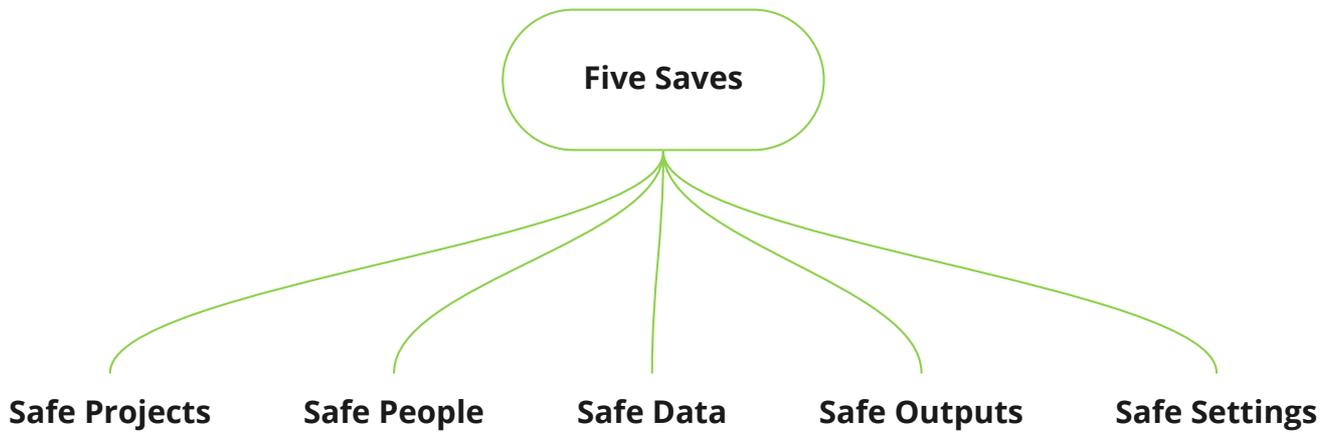
Things to consider when re-using data:

- Access and terms of use
- Evaluation of reusability
- Proper citation

19

*Source: Biernacka et al. (2020)

Ethical and Legal Norms: Five Saves



20

Source: <https://www.ukdataservice.ac.uk/manage-data/legal-ethical/access-control/five-saves.aspx>, Accessed 2/7/21

Feedback

<p>What have you learned?</p> <p>where to look for possible/suitable repositories</p> <p>links and labbook software, CARE principle</p> <p>how to decide on legal terms</p>	<p>What was surprising?</p> <p>the amount of storage possibilities</p> <p>the https://www.re3d-ata.org/ site is awesome very useful.</p>
<p>What user problems or opportunities should be fixed before the next workshop?</p>	<p>What have you missed? What would you like to see in another workshop?</p>

21

References (except Websites)

Biernacka, Katarzyna, Bierwirth, Maik, Buchholz, Petra, Dolzycka, Dominika, Helbig, Kerstin, Neumann, Janna, ... Wuttke, Ulrike. (2020). Train-the-TrainerConcepton Research Data Management (Version 3.0). Zenodo. <http://doi.org/10.5281/zenodo.4071471>.

Burgelman, Jc. (2018), The future of science is open Rationale, goals and milestones of the EU policies COIMBRA High Level seminar on research policy Venice 7–12–2018, HoU Open Science, DG RTD, <https://www.coimbra-group.eu/wp-content/uploads/Burgelman2018-OS-COIMBRA-december.pdf>.

RLG/OCLC Working Group, Research Libraries Group (2002), Trusted Digital Repositories: Attributes and Responsibilities, An RLG–OCLC Report, <https://wiki.lib.sun.ac.za/images/e/e1/Tdr-oclc.pdf>.