(Collaborative) RDM

FibReLoop Training

Wouter Grouve

Faculty of Engineering Technology
Mechanics of Solids, Surfaces & Systems (MS^3)
University of Twente

July 1, 2025
Enschede

Quick questionnaire

Raise hands

- Did you follow courses on research data management?
- Who has already written a research data management plan?
- Who has data archived on the web?
- Did you read the sections on Open Science and RDM in the proposal?

Quick questionnaire

Raise hands

- Did you follow courses on research data management?
- Who has already written a research data management plan?
- Who has data archived on the web?
- Did you read the sections on Open Science and RDM in the proposal?

Relevant deadlines in the FibReLoop proposal (1):

- Individual Data Management Plan (DMP) in M9
- Common consortium-wide DMP in M12

What is data?

And what is data management?

Research data

Research data is any information collected or generated for the purpose of analysis, in order to generate or validate scientific claims.

What is data?

And what is data management?

Research data

Research data is any information collected or generated for the purpose of analysis, in order to generate or validate scientific claims.

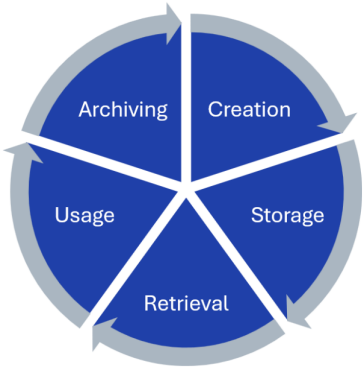
- Research data management refers to the activities that you perform to create, store, maintain, disclose and archive these data sustainably
- In the data management plan (DMP) you describe these activities. A DMP encourages you to think beyond your current stage and plan for later.

Information life cycle

Typical data life cycle:

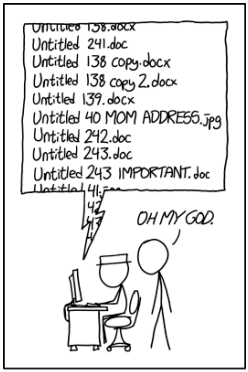
1. Creation (experiments, simulation, . . .)
2. Storage (often closed)
3. Usage (leading to output)
4. Archival (at least closed, preferably open)
5. Re-use

It makes sense to consider use and re-use in an early stage of the process.



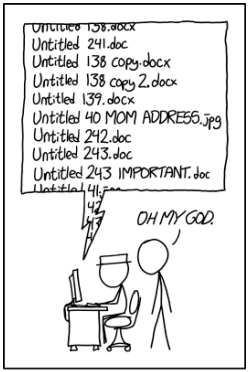
See yourself as your future collaborator

Or the colleague you would like to have



See yourself as your future collaborator

Or the colleague you would like to have



- Organize your data – I can find it again
- Backing up your data – I won't loose my data
- Use open or widely used file formats – I will always be able to open my files
- Document your data – I will understand what I have done
- Consider legal issues – I am allowed to reuse the data without problems

Prevent any misfortune

Don't be a fool

https://data.blogs.bristol.ac.uk/bootcamp/storing/

Or see others as collaborators

Why you should care? Because (2, 4):

- Funders or publishers may say that you should
- Your data is unique and difficult to collect
- Data can be reused - perhaps in unpredictable ways
- It can create opportunities for collaboration
- Sharing boosts innovation
- It promotes research integrity
- Data ages more slowly than publications

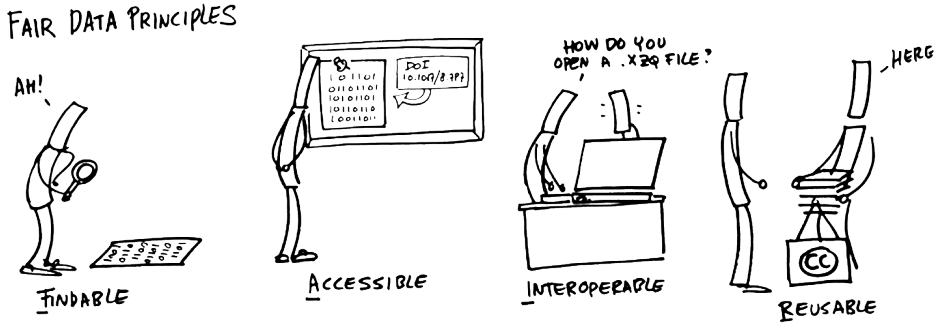
FAIR principles

Findable, Accessible, Interoperable, Reusable

- The FAIR principles are the standard for responsible data management and practicing open science
- FAIR stands for Findable, Accessible, Interoperable and Reusable (3)
- Each letter represents a list of principles with a total of 15 principles altogether
- Encourages researchers to think about the bigger picture of where their data sits in the context of their research domain (5)
- The principles are a resource for optimal choices and so are not set in stone

FAIR principles

Findable, Accessible, Interoperable, Reusable



https://www.fosteropenscience.eu/content/open-science-training-handbook

Key RDM Practices

Data documentation

contents can = data

Key RDM Practices

Data documentation

+

contents can = data

can label = metadata

Key RDM Practices

Data documentation

+

=

contents can = data

can label = metadata

can with label =

findable and understandable data without opening the can

Key RDM Practices

Data documentation

Document your data set:

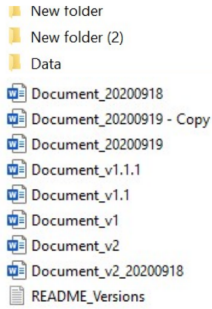
- Ensure that other researchers from the same field can understand, verify, redo and re-use your data, without having to ask you.
- Include description of context, content, and the generation process.
- What information is relevant for re-use depends strongly on the field of research.
- Examples of required information: process history, used test standards, software versions, device settings/configurations, etc.

Key RDM Practices

Data documentation

File and folder structure:

- Be consistent in your (logical and hierarchical) folder structure and file naming to prevent your data becoming unfindable.
- Where possible, be consistent in the data file structure too! E.g., use the same column names and order for your tabular data.
- How will you handle version control? Consider systems such as git.



"Copy", "New folder", "final" "old", ... Bad examples!

Key RDM Practices

Data storage

Recommended storage practices during the project:

- Network storage of your organization, there may be several options
- Encrypted laptop
- Make sure that your data is accessible by your supervisor

Key RDM Practices

Data storage

Recommended storage practices during the project:

- Network storage of your organization, there may be several options
- Encrypted laptop
- Make sure that your data is accessible by your supervisor

- Avoid external hard drives, USB drives and personal cloud storage
- Data sharing possibilities may differ per organization (we'll discuss later)

Key RDM Practices

Data preservation

What to preserve?

- Closed vs. open? Who has read the FibReLoop research proposal?

Key RDM Practices

Data preservation

What to preserve?

- Closed vs. open? Who has read the FibReLoop research proposal?
- Research output other than publications will be shared if deemed necessary (1)...

Key RDM Practices

Data preservation

What to preserve?

- Closed vs. open? Who has read the FibReLoop research proposal?
- Research output other than publications will be shared if deemed necessary (1)...

If you choose to store the data in a public repository:

- Use a persistent identifier (DOI)
- Choose a proper license for data (e.g., CC0, CC-BY) or software (e.g., MIT)

In any case, make sure your data is FAIR and store everything for a significant amount of time.

Exercise

Dataset evaluation

Go to: `https://github.com/wjbg/fibreloop`

Assignment (15 minutes):

- Subdivide in smaller groups of three to four persons
- Select a dataset from the list on Github and inspect it carefully
- Discuss the questions on Github in small group
- Discuss findings in big group

Collaborative RDM

Sharing *during* the project

- Multiple partners means different practices, tools, and expectations
- Data must remain usable across institutions, disciplines, and time zones
- Poor data management can lead to duplication, errors, or even loss of results
- RDM is not just about compliance – it's about enabling effective collaboration
- Good RDM helps ensure the project meets open science and FAIR data goals
- Data managemetn needs constant attention

Data sharing in a project

Infrastructure and organization

Infrastructure:

- MS Teams via one of the institutes
- Shared drives, such as OneDrive or Nextcloud
- Collaborative platforms, such as OSF or perhaps GitHub

How to manage:

- Access control and permissions
- Documentation
- Versioning and backups
- Data security

Data standardizations

Consistency and clarity

Standardization

What happens if each partner uses a different format? This not only holds for file types but also contents.

Data standardizations

Consistency and clarity

Standardization

What happens if each partner uses a different format? This not only holds for file types but also contents.

As a group, you would ideally:

- Agree on common data formats
- Use controlled terminology or vocabulary (e.g., for metadata)
- Include clear README files (a FibReLoop template may be sensible)
- Harmonize data collection protocols
- Agree on the licenses your using

Data governance and quality control

Define who is responsible for curating and validating shared data

Things that need to be addressed:

- Agreements on data ownership and stewardship
- Policies for data access, use and publication rights
- Version control
- Validation of data before sharing?
- Centralized wiki for information sharing?

Exercise

Collaborative data planning

Go to: `https://github.com/wjbg/fibreloop`

Assignment (30 minutes):

- Subdivide in smaller groups of three to four persons
- Discuss the questions on Github in small group
- Discuss findings in big group

Wrap-up

References

[1] FibReLoop Research Proposal, *Closing the fibre-reinforced composites loop: recycling materials for recycled components*, Horizon MSCA-2024-DN-01-01, 2023.
[2] Giglia, E., *FAIR data basics / IMIBAS / ISPAS project*, Zenodo, 2022.
[3] Go FAIR, *FAIR Principles*, https://www.go-fair.org/fair-principles/, .
[4] Research Data, Northumbria, *Benefits of data sharing and basic data management*, Northumbria University, 2022.
[5] Wilkinson, M.D. and others, *The FAIR Guiding Principles for scientific data management and stewardship*, Nature Publishing Group, 2016.

Thank you!

Enschede, July 1, 2025

w.j.b.grouve@utwente.nl