Research Data Management – Introduction

UMONS Université de Mons

UMONS Université de Mons — DIRECTION AVRE Administration et Valorisation de la Recherche

Sébastien Hoyas

Data Project Manager & Business Analyst

Place du Parc, 22
Office 1.12
7000 Mons
+32 65 37 34 54
sebastien.hoyas@umons.ac.be

Image created with Dall-E by OpenAI

Outline

Scope of this training session

What are (research) data?

What is data management?

FAIR principles

Data management plan (DMP)

Data management advices

Closing remarks

Q&A

Typical journeys in research… ☺

Start of a project

Typical journeys in research… ☺

Start of a project

Name your exciting results like ‘results.txt’ in your Downloads folder

Months later, your supervisor asks you to make a presentation for tomorrow 8 AM

You cannot find the results to make your graph

Your supervisor gets angry

Typical journeys in research… ☺

Start of a project

Name your exciting results like ‘results.txt’ in your Downloads folder

Store the results of an experiment on your laptop

Months later, your supervisor asks you to make a presentation for tomorrow 8 AM

A colleague spills coffee on your laptop that dies

You cannot find the results to make your graph

You’ve lost hours of work

Your supervisor gets angry

Typical journeys in research… ☺

Start of a project

Name your exciting results like ‘results.txt’ in your Downloads folder

Store the results of an experiment on your laptop

You analyze your **data** and notice something wrong

You redo the experiment and tweak a few parameters and obtain nice results

Months later, your supervisor asks you to make a presentation for tomorrow 8 AM

A colleague spills coffee on your laptop that dies

You want to publish the results and a reviewer asks for details about the settings of the experiment

You cannot find the results to make your graph

You’ve lost hours of work

You did not keep track of them

Your supervisor gets angry

Scope of this training

All these problems are related to ***data*** and their ***management***

Scope of this training:

- To teach you current regulations in terms of data (FAIR, open).
- To try help you manage your data.
- To make you aware about resources to help you.

What I will *not* be able to do:

- Provide a unique answer to manage your data.

What are (research) data?

What are actually *data*?

- **You need to think about *what is needed to validate or reproduce your research?***
  - **Factual elements**: figures, texts, images, sounds, measurements, results of recordings, computer programmes, etc.
  - **Raw** (not processed, manipulated or transformed in any way) *or* **derived from raw data** (obtained after transformation of raw data).
  - **Quantitative** (figures, measurements, statistics) *or* **qualitative** (interview, speeches, recordings of speakers, videos).
  - Can be digital, paper-based, etc.

What is data management?

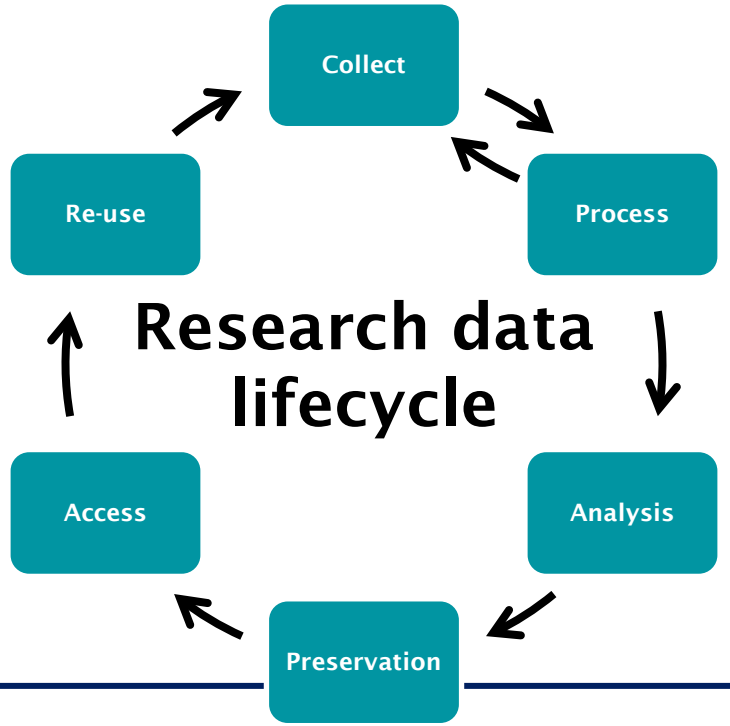
Research Data Management = all actions related to **collecting**, **organizing**, **storing**, **preserving**, and **sharing *data*** during a research project and on the long-run.

When are data involved in a research project?

What is data management?

Research Data Management = all actions related to **collecting**, **organizing**, **storing**, **preserving**, and **sharing *data*** during a research project and on the long-run.

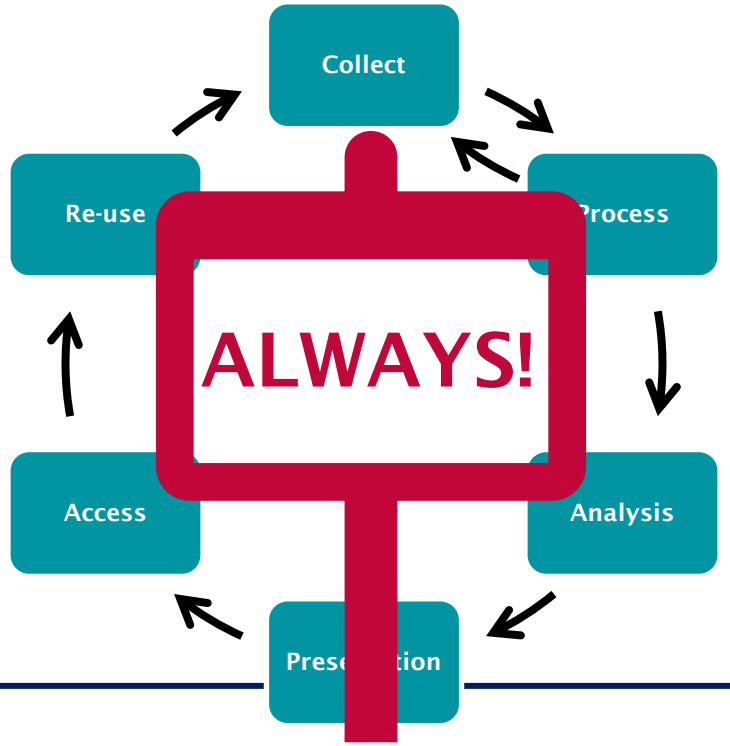
When are data involved in a research project?



What is data management?

Research Data Management = all actions related to **collecting**, **organizing**, **storing**, **preserving**, and **sharing *data*** during a research project and on the long-run.

When are data involved in a research project?



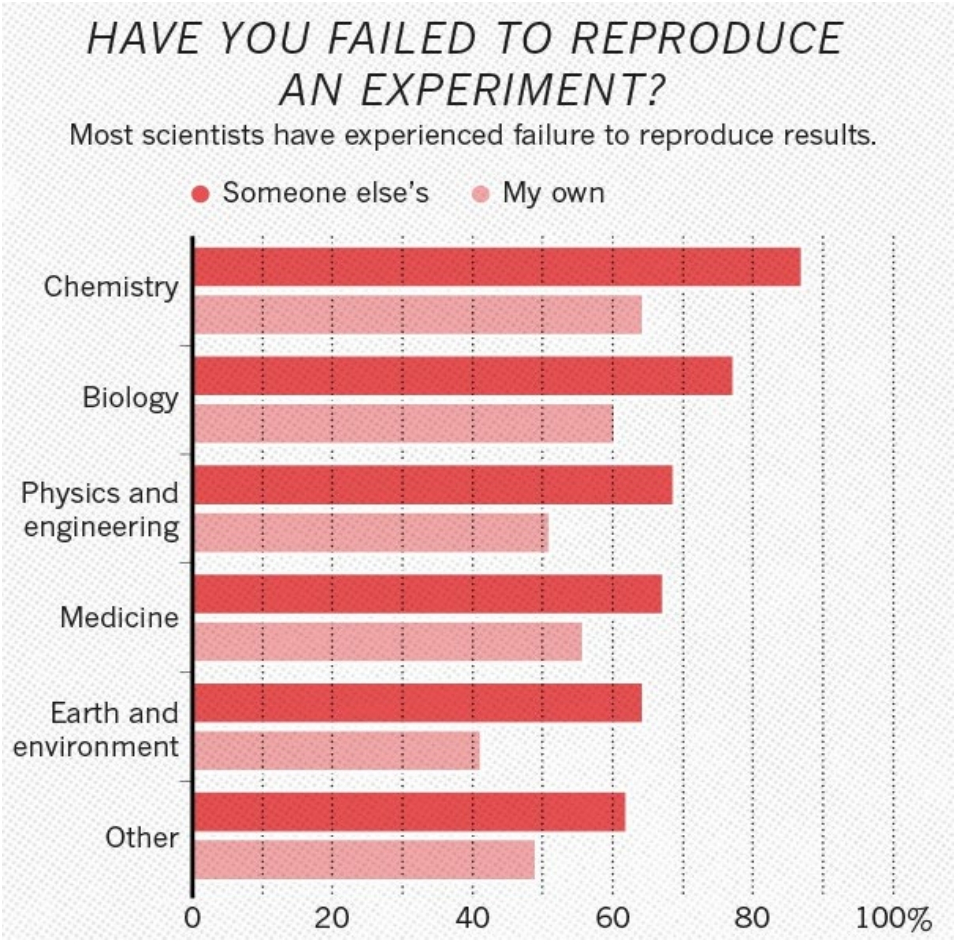
What is data management?

Research Data Management = all actions related to **collecting**, **organizing**, **storing**, **preserving**, and **sharing** ***data*** during a research project and on the long-run.

Why is RDM so important?

Why is it so important?

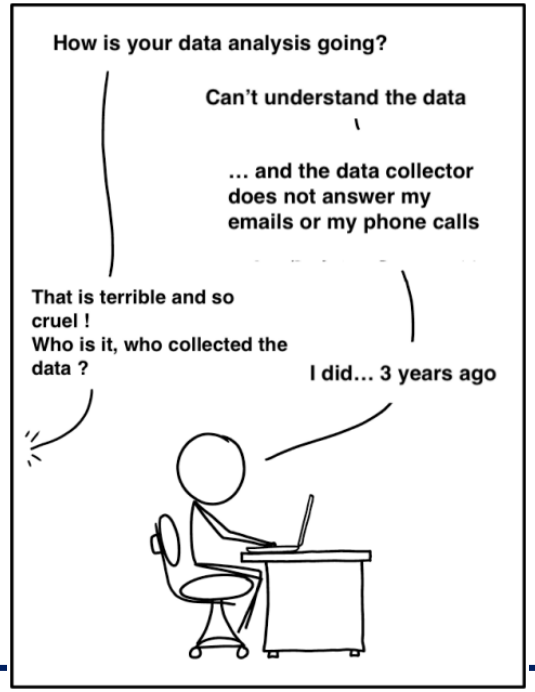
- Most scientific results are difficult, even **impossible**, to reproduce and/or replicate because **research integrity** is not favoured (*publish or perish*)
- Avoid reiventing the wheel



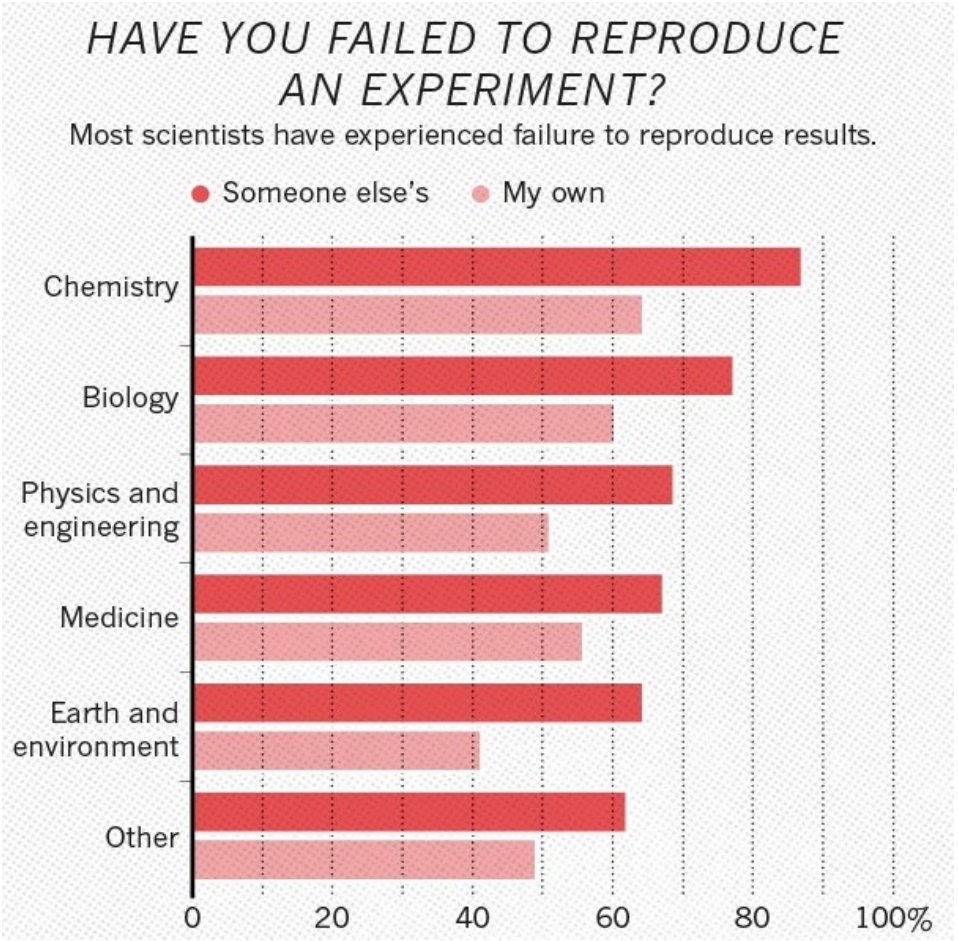
Nature 533, 452–454 (26 May 2016) doi:10.1038/533452a

Why is it so important?

- Most scientific results are difficult, even **impossible**, to reproduce and/or replicate because **research integrity** is not favoured (*publish or perish*)
- Avoid reiventing the wheel



your future self, by Julien Colomb, CC-BY-NC, derived from .NORM Normal File Format, CC-BY-NC, by Randall Munroe



Nature 533, 452–454 (26 May 2016) doi:10.1038/533452a

Benefits and requirements for data management

1. **Efficiency**: Proper data management can make your research process more efficient, saving you time and effort in the long run.
2. **Reproducibility**: It ensures that the future you and others can understand and replicate your work in the future.
3. **Collaboration**: It makes it easier to share and collaborate with others, both within and outside your research team by setting proper licences.
4. **Compliance**: Many funding agencies require a **DMP** and FAIR data management (*see later*).
5. **Preservation**: It ensures your data is preserved for future use and **reuse**.
6. **Return on investment**: Making data discoverable, accessible, and reusable maximizes the research potential of the data and provides greater returns on public investments in research.

Benefits and requirements for data management

1. **Efficiency**: Proper data management can make your research process more efficient, saving you time and effort in the long run.
2. **Reproducibility**: It ensures that the future you and others can understand and replicate your work in the future.
3. **Collaboration**: [illegible] within and outside your res[illegible]
4. **Compliance**: [illegible]ment (*see later*).
5. **Preservation**: It ensures your data is preserved for future use and **reuse**.
6. **Return on investment**: Making data discoverable, accessible, and reusable maximizes the research potential of the data and provides greater returns on public investments in research.

These benefits align with the **FAIR principles**.

FAIR principles

The FAIR principles are a set of instructions formulated to **maximize the (re)use of data** and other digital objects such as code and software.

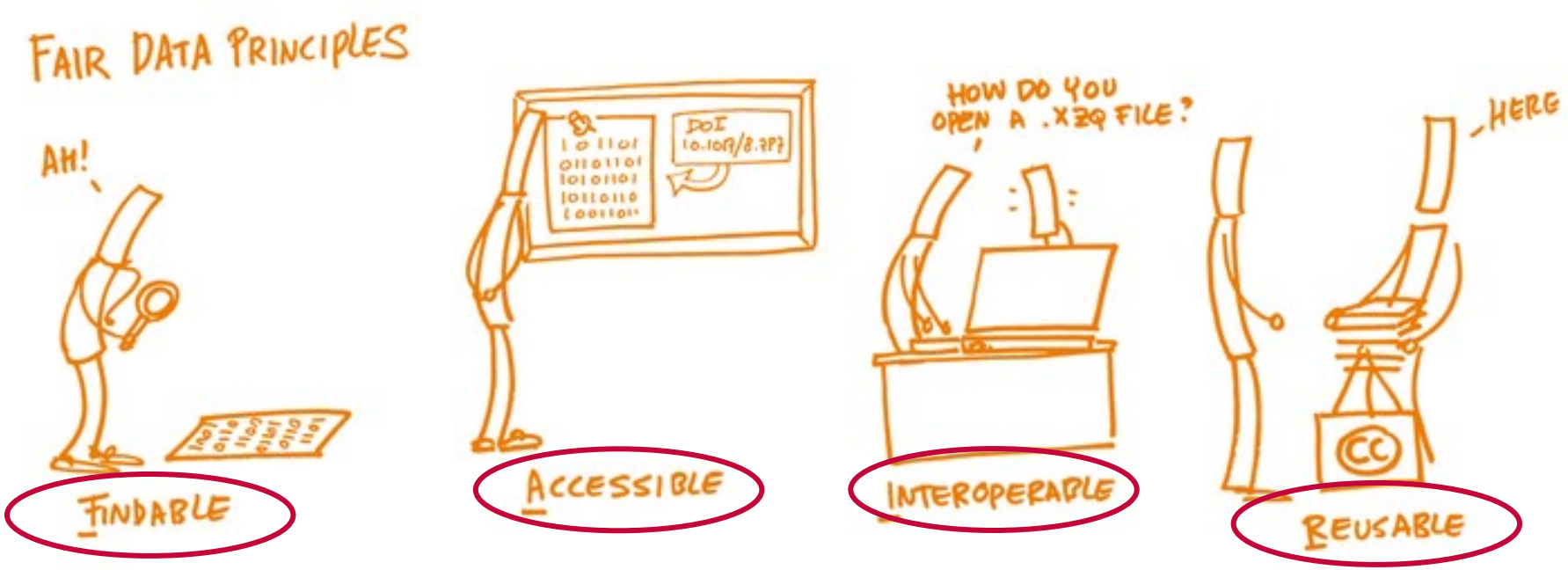


Image by Patrick Hochstenbach, CC0 1.0
The FAIR principles - Research Data Management Support - Utrecht University

FAIR principles

Findable: your data can be discovered by others.

👉 **Where to find data ❓**

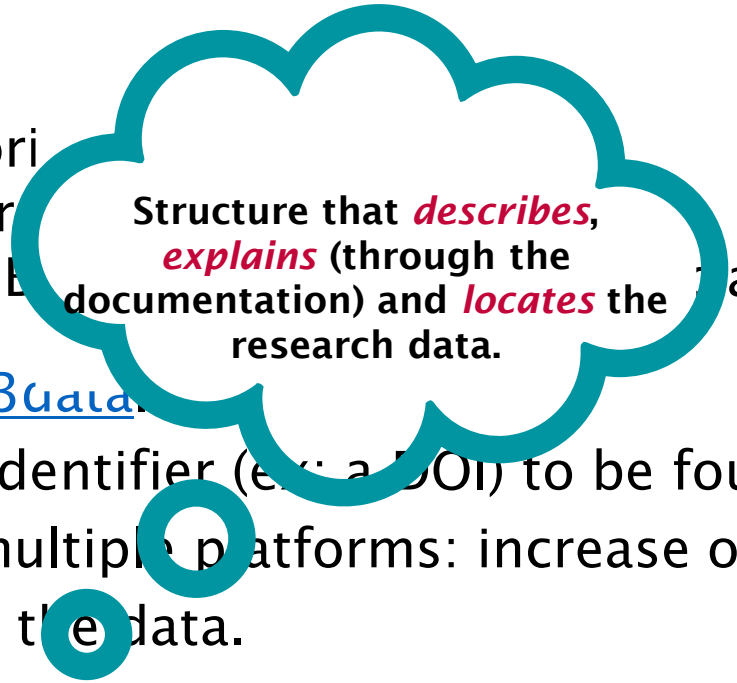
- Online, in data repositories!
  - Zenodo: general purpose repository.
  - SODHA: the federal Belgian data archive for social sciences and the digital humanities.
  - And many more: re3data.
- Unique and persistent identifier (ex: a DOI) to be found and be cited.
- Content harvested by multiple platforms: increase of visibility and impact.
- ***Meta***data associated to the data.

FAIR principles

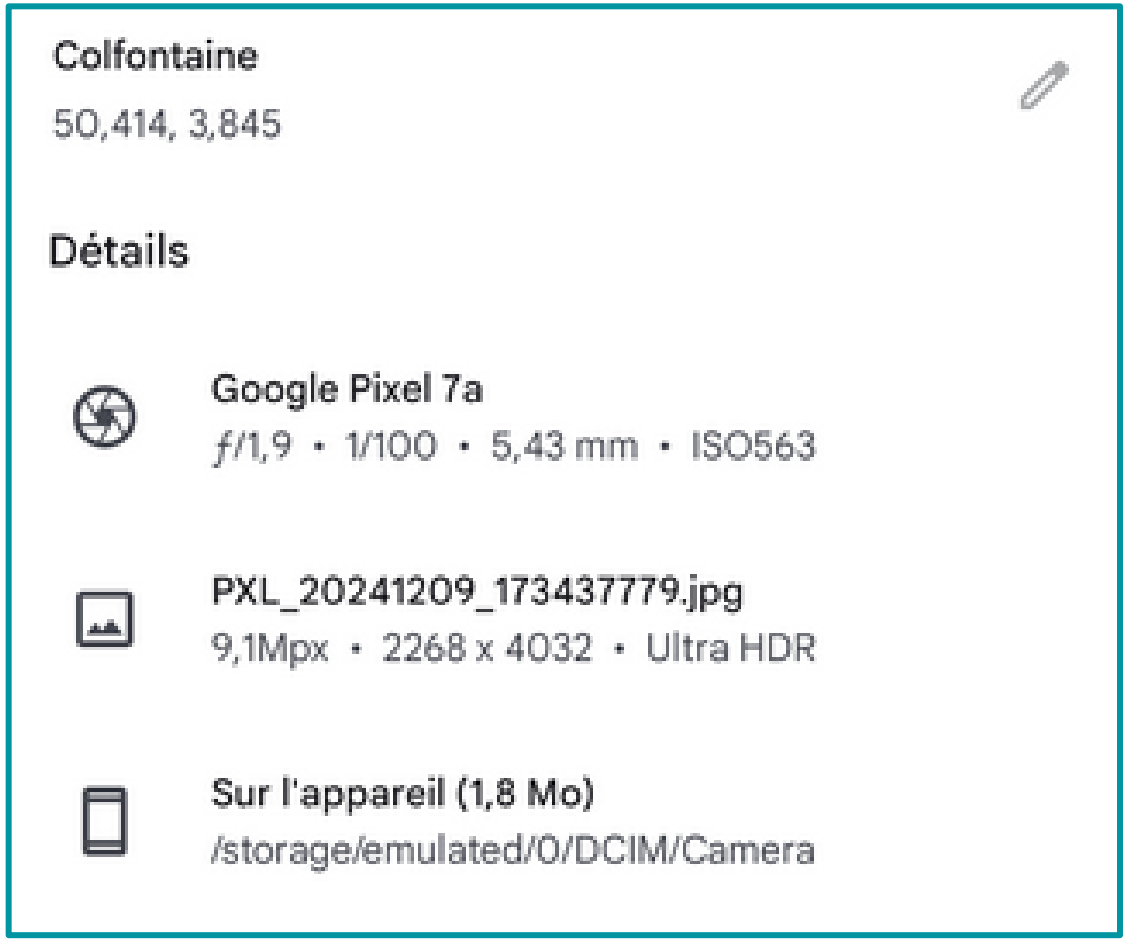
Findable: your data can be discovered by others.

👉 **Where to find data ❓**

- Online, in data repositori
  - Zenodo: general pur
  - SODHA: the federal E al sciences and the digital humanities.
  - And many more: re3data.
- Unique and persistent identifier (ex: a DOI) to be found and be cited.
- Content harvested by multiple platforms: increase of visibility and impact.
- ***Meta***data associated to the data.

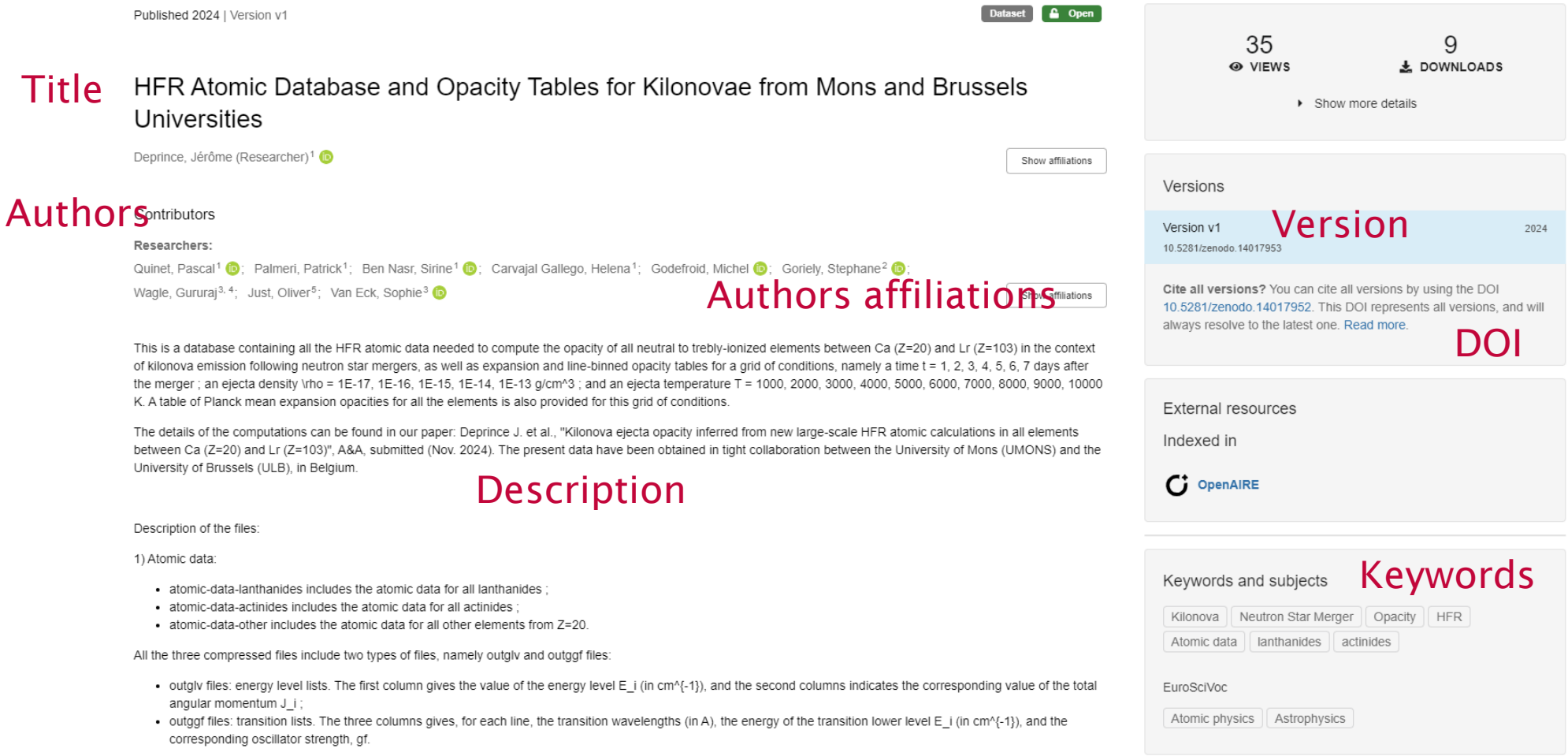


FAIR principles



The FAIR Data Principles – FORCE11

FAIR principles



Published 2024 | Version v1

Dataset Open

Title

HFR Atomic Database and Opacity Tables for Kilonovae from Mons and Brussels Universities

Deprince, Jérôme (Researcher)[1]

Show affiliations

Authors

Contributors

Researchers:

Quinet, Pascal[1]; Palmeri, Patrick[1]; Ben Nasr, Sirine[1]; Carvajal Gallego, Helena[1]; Godefroid, Michel; Goriely, Stephane[2]; Wagle, Gururaj[3, 4]; Just, Oliver[5]; Van Eck, Sophie[3]

Show affiliations

Authors affiliations

This is a database containing all the HFR atomic data needed to compute the opacity of all neutral to trebly-ionized elements between Ca (Z=20) and Lr (Z=103) in the context of kilonova emission following neutron star mergers, as well as expansion and line-binned opacity tables for a grid of conditions, namely a time t = 1, 2, 3, 4, 5, 6, 7 days after the merger ; an ejecta density \rho = 1E-17, 1E-16, 1E-15, 1E-14, 1E-13 g/cm^3 ; and an ejecta temperature T = 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000 K. A table of Planck mean expansion opacities for all the elements is also provided for this grid of conditions.

The details of the computations can be found in our paper: Deprince J. et al., "Kilonova ejecta opacity inferred from new large-scale HFR atomic calculations in all elements between Ca (Z=20) and Lr (Z=103)", A&A, submitted (Nov. 2024). The present data have been obtained in tight collaboration between the University of Mons (UMONS) and the University of Brussels (ULB), in Belgium.

Description

Description of the files:

1) Atomic data:

- atomic-data-lanthanides includes the atomic data for all lanthanides ;
- atomic-data-actinides includes the atomic data for all actinides ;
- atomic-data-other includes the atomic data for all other elements from Z=20.

All the three compressed files include two types of files, namely outglv and outggf files:

- outglv files: energy level lists. The first column gives the value of the energy level E_i (in cm^{-1}), and the second columns indicates the corresponding value of the total angular momentum J_i ;
- outggf files: transition lists. The three columns gives, for each line, the transition wavelengths (in A), the energy of the transition lower level E_i (in cm^{-1}), and the corresponding oscillator strength, gf.

35 VIEWS

9 DOWNLOADS

Show more details

Versions

Version v1 2024

10.5281/zenodo.14017953

Version

Cite all versions? You can cite all versions by using the DOI 10.5281/zenodo.14017952. This DOI represents all versions, and will always resolve to the latest one. Read more.

DOI

External resources

Indexed in

OpenAIRE

Keywords and subjects

Keywords

Kilonova | Neutron Star Merger | Opacity | HFR | Atomic data | lanthanides | actinides

EuroSciVoc

Atomic physics | Astrophysics

[The FAIR Data Principles - FORCE11]

FAIR principles

Findable: your data can be discovered by others.

👉 **Where to fin**

- Online, in dat
  - Zenodo: g
  - SODHA: t
    humanitie
  - And many
- Unique and p
- Content harv
- ***Meta***data ass

e digital

pact.

Even when not using a data repository, you should always **attach metadata** to your data (within their folder):

- Title
- Creator(s)
- Collection date
- Contact
- Version
- Format
- Etc.

(You future you or next colleague may thank you for that)

TIPS!

Dublin Core standard generator: The Simple Dublin Core Generator!

FAIR principles

Findable: your data can be dis

👉 **Where to fin**

- Online, in dat
  - Zenodo: g
  - SODHA: t
    humanitie
  - And many
- Unique and p
- Content harv
- ***Meta***data ass

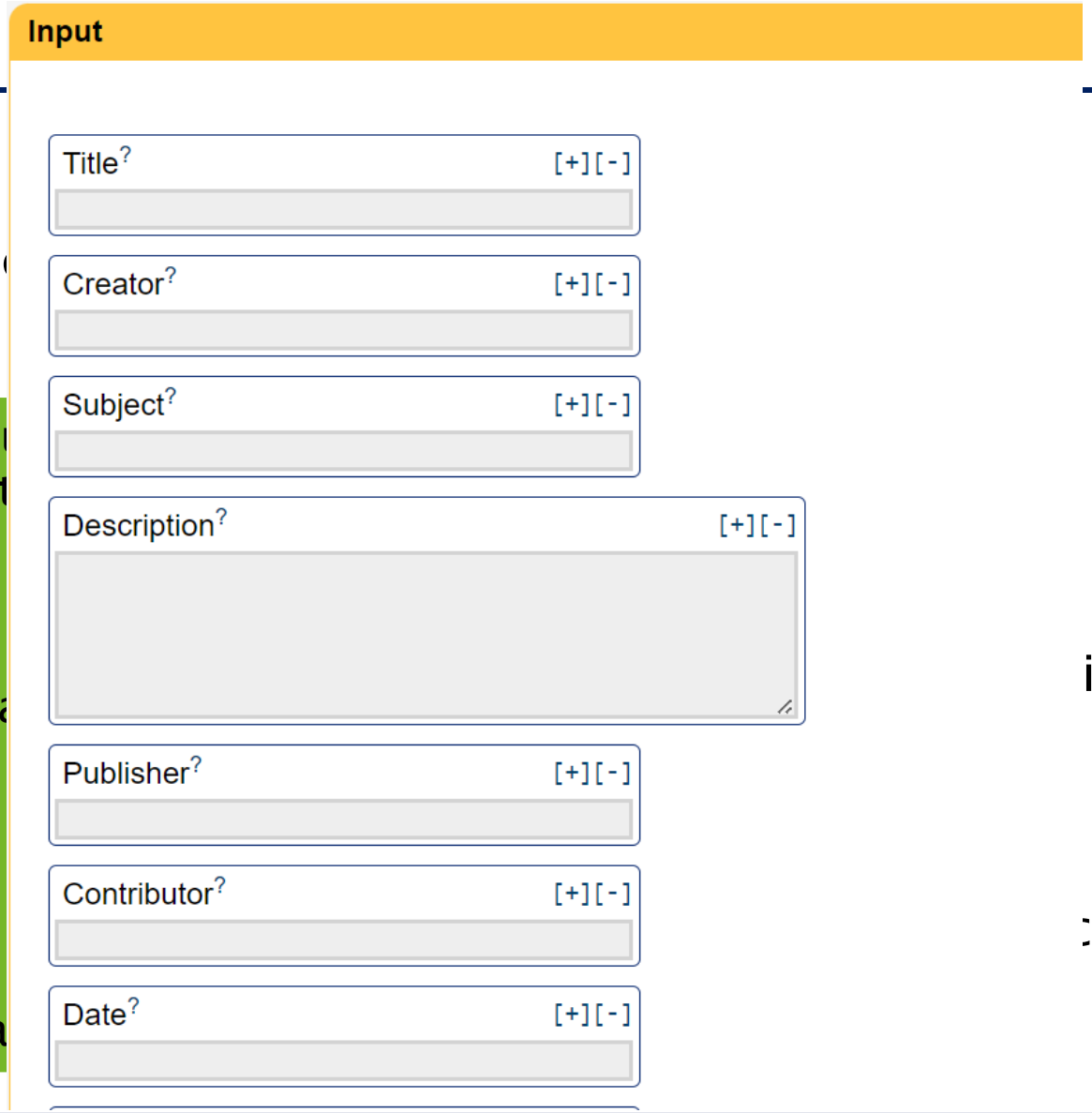
Even when not
attach metadat

- Title
- Creator(s)
- Collection da
- Contact
- Version
- Format
- Etc.

Dublin Core sta
Generator!

igital

:t.



FAIR principles

Accessible: your data can be made available to others.

Data can be accessed in a data repository.

- **A in FAIR means Accessible, not necessarily publicly available!**
- Metadata must be publicly available and they specify if and how the data can be accessed (open or with restrictions).
- Metadata should remain accessible**, even if the data are no longer available** (archive policy, etc.).

FAIR principles

Interoperable: your data can understood, exchanged, and reused across different systems, tools, and disciplines without losing meaning or functionality.

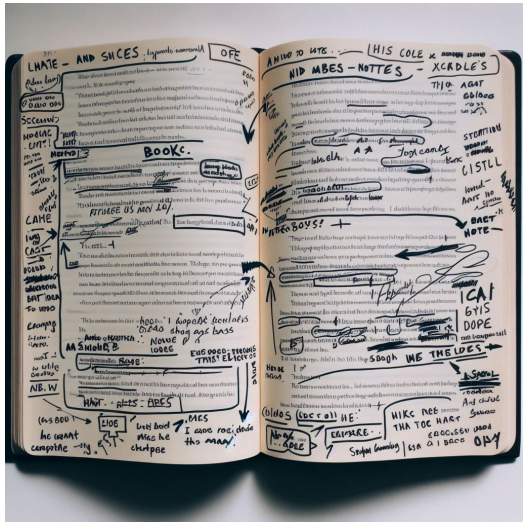
Humans and machines should be able to exchange and interpret each other's data.

- Use standards from your discipline in (meta)data for dates (ISO), coordinates, etc.
- If possible, save the data in a widely used file format compatible with different operating systems and softwares (.txt, .md, .pdf, .csv, etc.).

FAIR principles

Reusable: your data can be reused by others (and you ☺).

Your data should be accompanied by thorough **documentation** to enhance the reproducibility and a clear description about **who can use them and under which conditions** (see later licenses).



- Documentation should be included in the data repository, for example in a **README** file.
- Metadata should inform other researchers on how to reuse, where the data came from, where to find additional documentatio, etc.

FAIR principles

FAIR data are more and more important, **even for publishing results in scientific journals**:

> Retraction of an article because access to the underlying data was not granted.

*"[...] Because all the authors were not granted access to the raw data and the **raw data could not be made available to a third-party auditor, we are unable to validate the primary data** sources underlying our article, "Cardiovascular Disease, Drug Therapy, and Mortality in Covid-19." We therefore **request that the article be retracted**. We apologize to the editors and to readers of the journal for the difficulties that this has caused.""*

How to ensure that your data are well-managed?

You need to ***think and plan*** at the earliest stage of the project(s) about the data that will be involved in your project, having the FAIR principles in mind.

"What happens if data is lost, inaccessible, or poorly documented?"

It will allow you **organize** your work more efficiently and **comply** with the funding requirements as well as **strengthen** your own scientific **integrity** and the **impact** of your research.

How to ensure that your data are well-managed?

You need to ***think and plan*** at the earliest stage of the project(s) about the data that will be involved in your project, having the FAIR principles in mind.

"What happens if data is lost, inaccessible, or poorly documented?"

It will allow you **organize** your work more efficiently and **comply** with the funding requirements as well as **strengthen** your own scientific **integrity** and the **impact** of your research.

How?

👉 **By filling in a Data Management Plan (DMP)!**

What is a Data Management Plan – DMP?

DMPs are structured documents that outline how research data will be handled **throughout a research project and *beyond***.

DMPs are required by more and more funders to assess how **future-proof** is the research by evaluating multiple criteria.

They consist of a set of questions to make you think of how to apply the FAIR principles to your research and comply with the funding requirements.

You should not see the DMP as a simple document to apply for or receive funding but rather as a **new tool to improve your scientific method**.

Data Management Plan – DMP

Content of a DMP, common to nearly all DMPs:

What data you will collect/reuse?

Where you will store them, who will be able to access them?

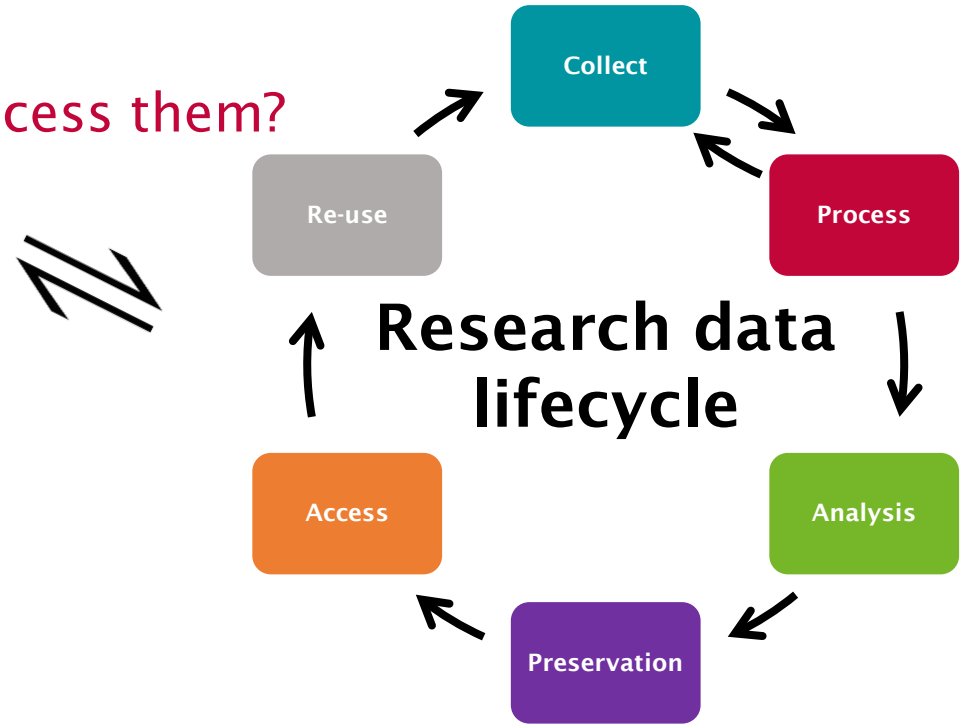
How you will document the data?

How you will preserve the data?

How you will apply the FAIR principles?

How you will share your data?

…



Data Management Plan – DMP

DMPs can be overwhelming at first, especially Horizon Europe requirements.

Between 2024 and 2025, the Fédération Wallonie-Bruxelles funded a project, CoARA FWB.

One goal was to develop a quick and easy tool, Compass to RDM, to show attention points about RDM and FAIR data to researchers before diving into DMPs.

Live demo

Data Management Plan – DMP

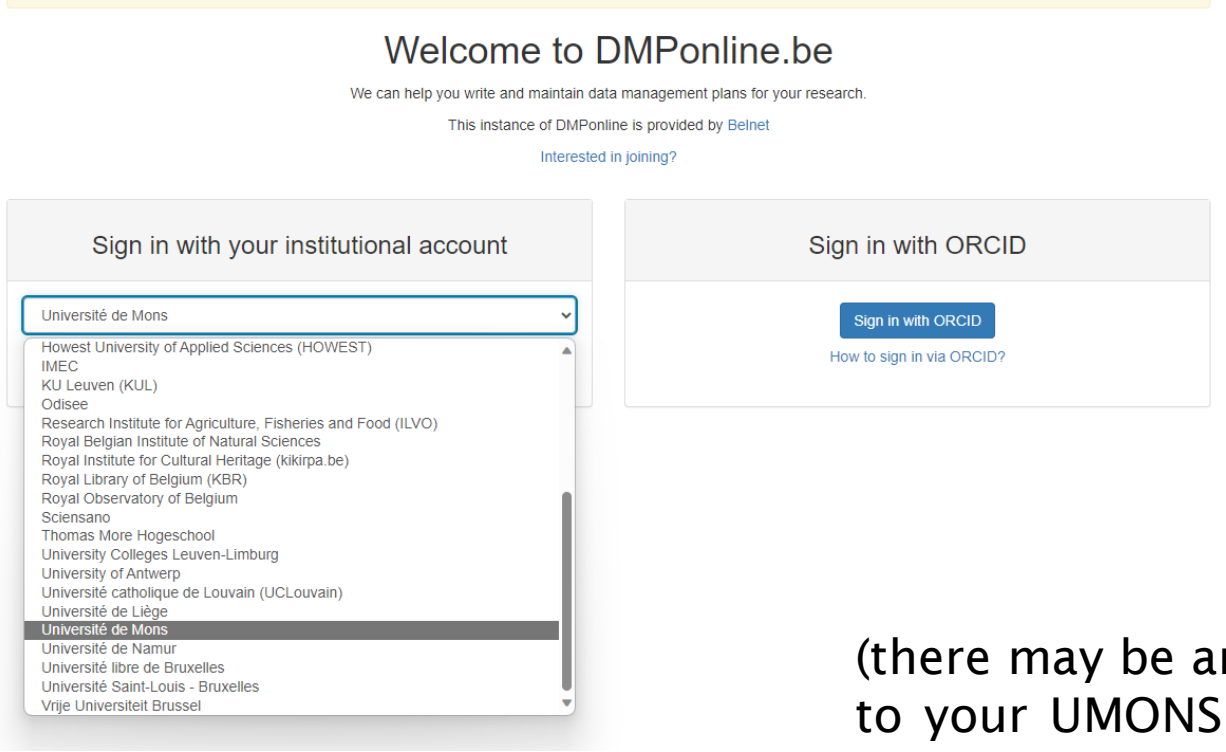
Where to fill in a DMP?

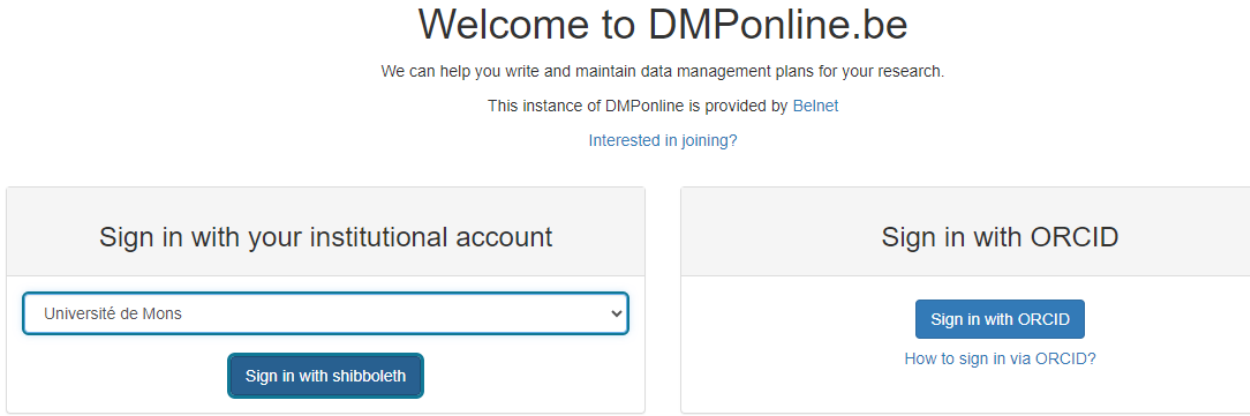
👉 On https://dmponline.be/

Research Data Management - How to use DMP Online? - Université de Mons Training on February 2nd

DMP Online

Log in using your institutional credentials





(there may be an extra step if you are not connected to your UMONS account on your web browser or if you are outside of the university network)

DMP Online

Welcome to DMPonline.be

We can help you write and maintain data management plans for your research.

This instance of DMPonline is provided by the DMPbelgium Consortium, which was founded in 2017 by:

- Instituut voor Natuur- en Bosonderzoek
- Université Libre de Bruxelles
- Universiteit Antwerpen
- Universiteit Gent
- Universiteit Hasselt
- Vrije Universiteit Brussel
- Wetenschappelijk Instituut Volksgezondheid – Institut Scientifique de Santé Publique (Sciensano)

In 2018 they were joined by:

- Université Catholique de Louvain
- Université de Liège
- Université de Mons
- Université de Namur
- Vlaamse Instelling voor Technologisch Onderzoek

Since then, the Consortium has been joined by:

- Arteveldehogeschool
- Instituut voor Landbouw-, Visserij- en Voedingsonderzoek
- Universitair Ziekenhuis Gent
- Vlaams Instituut voor de Zee
- Vlerick Business School
- Hogeschool Gent

Interested in joining the Consortium?

Sign in with your institutional account

Flanders Make (Belgium) (flandersmake.be)

Flanders Marine Institute

Flemish Institute for Technological Research

Ghent University

Ghent University (UZGent)

Hasselt University

Hogeschool Gent (HOGENT)

Hogeschool VIVES (VIVES)

IMEC

KU Leuven (KUL)

Research Institute for Agriculture, Fisheries and Food (ILVO)

Royal Institute for Cultural Heritage (kikirpa.be)

Royal Library of Belgium (KBR)

Royal Observatory of Belgium

Sciensano

Thomas More Hogeschool

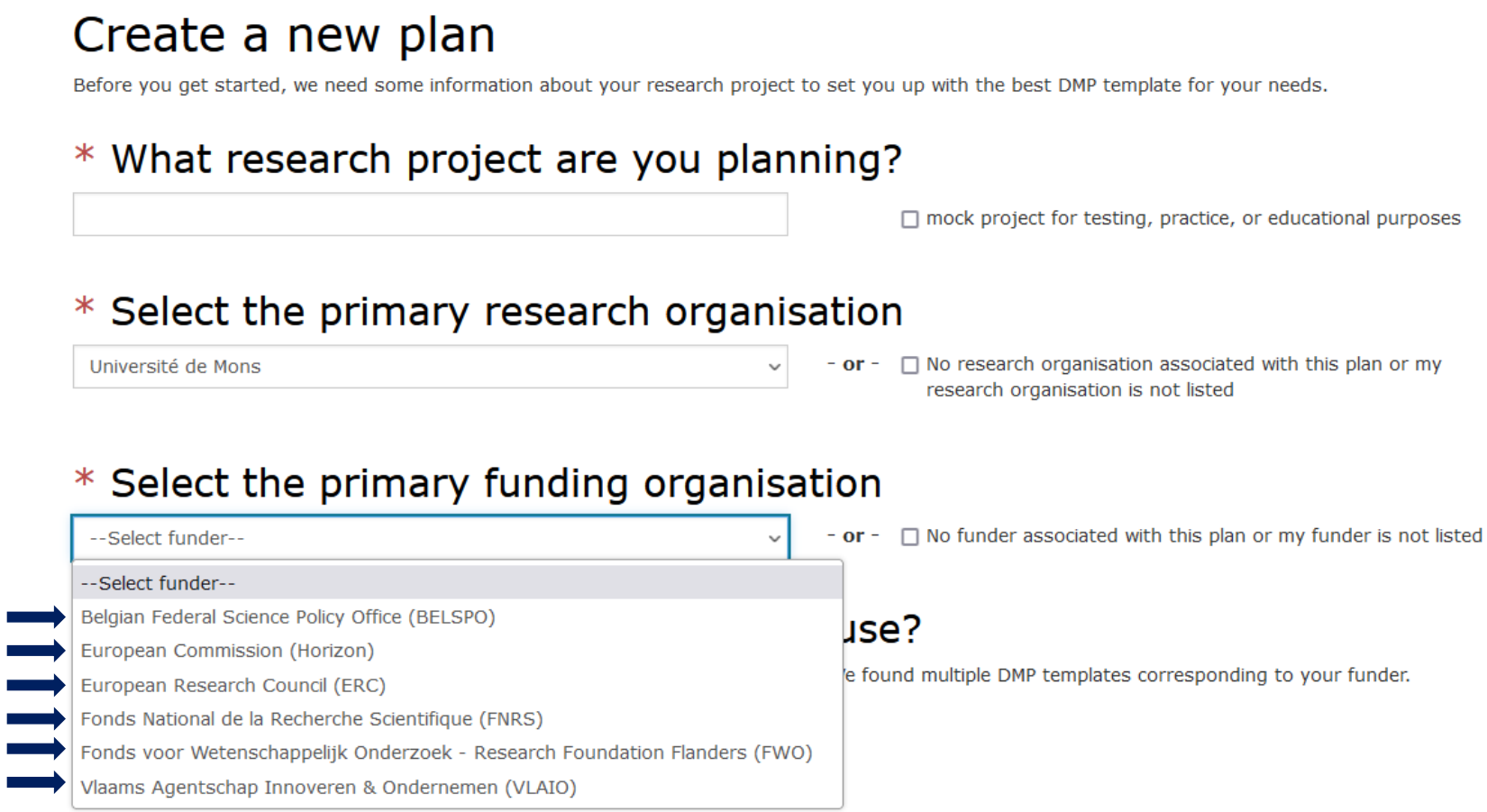
Université catholique de Louvain (UCLouvain)

Université de Liège

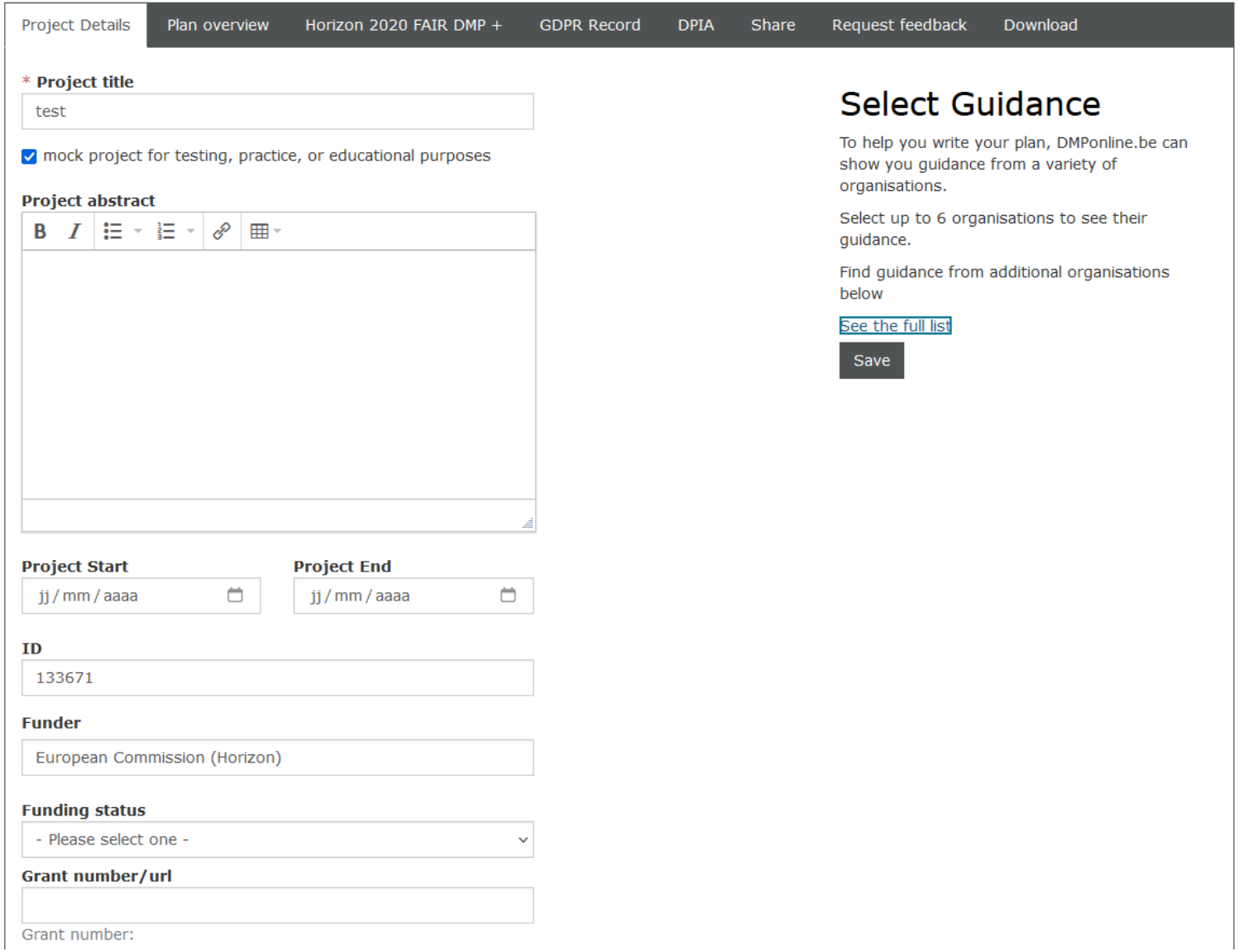
Université de Mons

Université de Namur

DMP Online


Create a new plan
Before you get started, we need some information about your research project to set you up with the best DMP template for your needs.
* What research project are you planning?
mock project for testing, practice, or educational purposes
* Select the primary research organisation
Université de Mons
- or -
No research organisation associated with this plan or my research organisation is not listed
* Select the primary funding organisation
--Select funder--
- or -
No funder associated with this plan or my funder is not listed
--Select funder--
Belgian Federal Science Policy Office (BELSPO)
European Commission (Horizon)
European Research Council (ERC)
Fonds National de la Recherche Scientifique (FNRS)
Fonds voor Wetenschappelijk Onderzoek - Research Foundation Flanders (FWO)
Vlaams Agentschap Innoveren & Ondernemen (VLAIO)
use?
e found multiple DMP templates corresponding to your funder.

DMP Online


Project Details
Plan overview
Horizon 2020 FAIR DMP +
GDPR Record
DPIA
Share
Request feedback
Download
* Project title
test
mock project for testing, practice, or educational purposes
Project abstract
Project Start
jj / mm / aaaa
Project End
jj / mm / aaaa
ID
133671
Funder
European Commission (Horizon)
Funding status
- Please select one -
Grant number/url
Grant number:
Select Guidance
To help you write your plan, DMPonline.be can show you guidance from a variety of organisations.
Select up to 6 organisations to see their guidance.
Find guidance from additional organisations below
See the full list
Save

DMP Online

Funder Templates

Templates are provided by a funder.

Templates for data management plans are based on the specific requirements listed in funder policy documents. DMPonline.be maintains these templates, however, researchers should always consult the funder guidelines directly for authoritative information.

Template Name	Download	Organisation Name	Last Updated	Funder Links	Sample Plans (if available)
BELSPO DMP +		Belgian Federal Science Policy Office (BELSPO)	27-09-2021		
ERC DMP +		European Research Council (ERC)	27-09-2021		
DCC Template		Digital Curation Centre	27-09-2021		
Horizon 2020 FAIR DMP +		European Commission (Horizon)	27-09-2021		
BRAIN 2.0		Belgian Federal Science Policy Office (BELSPO)	13-05-2022	www.belspo.be	
FNRS DMP		Fonds National de la Recherche Scientifique (FNRS)	19-05-2022		
VLAIO cSBO DMP (Flemish Standard DMP)		Vlaams Agentschap Innoveren & Ondernemen (VLAIO)	02-09-2022		
Horizon Europe DMP +		European Commission (Horizon)	12-10-2022		
FWO DMP (Flemish Standard DMP)		Fonds voor Wetenschappelijk Onderzoek - Research Foundation Flanders (FWO)	24-10-2022		

DMP: Examples

DMP publics (opidor.fr)

Data Management Advices

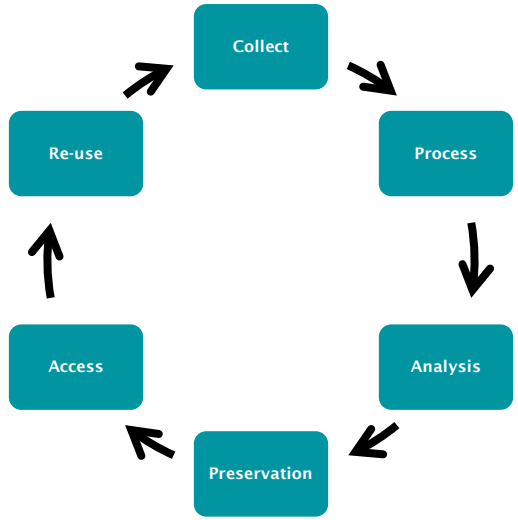
Data management advices

Data management is not easy (who said research is easy?) and heavily **depends on *you*, *your project* and *your peers***

To help you, general advices in:

- **Data organization**
- **Data description & documentation**
- **Data storage**
- **Data sharing**

All these « categories » are linked together and reminiscent of the data lifecycle



Data organization

Data organization

It is better to think in advance about ***how*** you will organize your data, but you can also change your mind during your research as long as you keep track of those changes.

Take into consideration that you may not be the only one working on those data, so **make it clear for *anyone***!

In general, most research data are digitalized, but it also applies to physical data (samples, etc.).

Tips & tricks

Existing procedures: check if there are already established ways to organize your data in your team (and check if it suits your needs!).

File organization: there is no universal answer ☹ You need to apply an organization that is compatible with you, your project and your team.

Above all, you must be ***CONSISTENT*** accross your project.

Data organization

File organization: examples

Hierarchical structure

```
ProjectName/
│
├── LiteratureReview/
│   ├── Paper1.pdf
│   ├── Paper2.pdf
│   └── ...
│
├── Data/
│   ├── RawData/
│   │   ├── Experiment1/
│   │   ├── Experiment2/
│   │   └── ...
│   │
│   ├── ProcessedData/
│   │   ├── Experiment1/
│   │   ├── Experiment2/
│   │   └── ...
│   │
│   └── Metadata/
│
├── Analysis/
│   ├── Script1.R
│   ├── Script2.py
│   └── ...
│
└── Outputs/
    ├── Figures/
    ├── Tables/
    └── ...
```

Data organization

File organization: examples

Hierarchical structure

```
ProjectName/
│
├── LiteratureReview/
│   ├── Paper1.pdf
│   ├── Paper2.pdf
│   └── ...
│
├── Data/
│   ├── RawData/
│   │   ├── Experiment1/
│   │   ├── Experiment2/
│   │   └── ...
│   │
│   ├── ProcessedData/
│   │   ├── Experiment1/
│   │   ├── Experiment2/
│   │   └── ...
│   │
│   └── Metadata/
│
├── Analysis/
│   ├── Script1.R
│   ├── Script2.py
│   └── ...
│
└── Outputs/
    ├── Figures/
    ├── Tables/
    └── ...
```

Chronological structure

```
ProjectName/
│
├── 2022/
│   ├── Q1/
│   │   ├── January/
│   │   ├── February/
│   │   └── March/
│   │
│   ├── Q2/
│   │   ├── April/
│   │   ├── May/
│   │   └── June/
│   │
│   ├── Q3/
│   │   ├── July/
│   │   ├── August/
│   │   └── September/
│   │
│   └── Q4/
│       ├── October/
│       ├── November/
│       └── December/
│
└── 2023/
    └── ...
```

Data organization

File organization: examples

Hierarchical & chronological structure

→ basically, any combination that allows you to be organized AND consistent AND that suits your needs

```
ProjectName/
│
├── LiteratureReview/
│   ├── Paper1.pdf
│   ├── Paper2.pdf
│   └── ...
│
├── Data/
│   ├── RawData/
│   │   ├── 2022/
│   │   │   ├── January/
│   │   │   │   ├── Experiment1/
│   │   │   │   ├── Experiment2/
│   │   │   │   └── ...
│   │   └── 2023/
│   │       ├── January/
│   │       └── ...
│   │
│   ├── ProcessedData/
│   │   ├── Experiment1/
│   │   ├── Experiment2/
│   │   └── ...
│   │
│   └── Metadata/
│
├── Analysis/
│   ├── Script1.R
│   ├── Script2.py
│   └── ...
│
└── Outputs/
    ├── Figures/
    ├── Tables/
    └── ...
```

Data organization

Folder names: keep folder names short (max 15-20 characters) and make them descriptive of what is inside, without being redundant with the folder structure.

- Bad example

```
ProjectName/
│
├── 2022/
│   ├── ProjectName Q1 Results of Experiments/
│   │   ├── Results of January/
│   │   ├── Results of February/
│   │   └── …
```

- Good example

```
ProjectName/
│
├── 2022/
│   ├── Experiments/
│   │   ├── Spectrometry/
│   │   ├── Spectroscopy/
│   │   └── …
```

NOTES

- Avoid using spaces, dots and special characters (&, ?, etc.)
- Use hyphens (-) or underscores (_) to separate elements
- Use a minimum of two leading zeros for padding (001, 002, etc.) to properly sort folders by names

Mode	LastWriteTime	Length	Name
-a---	13-01-26 10:26	0	1_a.txt
-a---	13-01-26 10:26	0	10_a.txt
-a---	13-01-26 10:27	0	100.txt
-a---	13-01-26 10:26	0	11_a.txt
-a---	13-01-26 10:26	0	2.txt
-a---	13-01-26 10:26	0	3.txt
-a---	13-01-26 10:26	0	4.txt
-a---	13-01-26 10:26	0	5.txt
-a---	13-01-26 10:26	0	6.txt
-a---	13-01-26 10:26	0	7.txt
-a---	13-01-26 10:26	0	8.txt
-a---	13-01-26 10:26	0	9.txt

Mode	LastWriteTime	Length	Name
-a---	13-01-26 10:26	0	001_a.txt
-a---	13-01-26 10:26	0	002.txt
-a---	13-01-26 10:26	0	003.txt
-a---	13-01-26 10:26	0	004.txt
-a---	13-01-26 10:26	0	005.txt
-a---	13-01-26 10:26	0	006.txt
-a---	13-01-26 10:26	0	007.txt
-a---	13-01-26 10:26	0	008.txt
-a---	13-01-26 10:26	0	009.txt
-a---	13-01-26 10:26	0	010_a.txt
-a---	13-01-26 10:26	0	011_a.txt
-a---	13-01-26 10:27	0	100.txt

Data organization

File names: keep file names brief and explicit, without being redundant with the folder structure.

- Avoid using spaces, dots and special characters (&, ?, ;, etc.)
- Use **hyphens** (-) or **underscores** (_) to separate elements (easier to recover in the OS)
- Use a minimum of two **leading zeros** for padding (001, 002 ,etc.) to properly sort files by names
- Use an extension that matches the file format
- If your files cannot be integrated in a versioning tool like Git, include a **version number** at the end. Keep a logfile where you briefly state changes in each of the new versions
- Include elements such as the date (**YYYYMMDD** format, best to sort) at the beginning of the file name
- Avoid starting to name files with « draft », « final » or the version number

Example: 20230130_RDMTraining_**V001.pptx**

https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide/2.-Organise-Document/File-naming-and-folder-structure
https://library.si.edu/sites/default/files/tutorial/pdf/filenamingorganizing20180227.pdf

Data organization

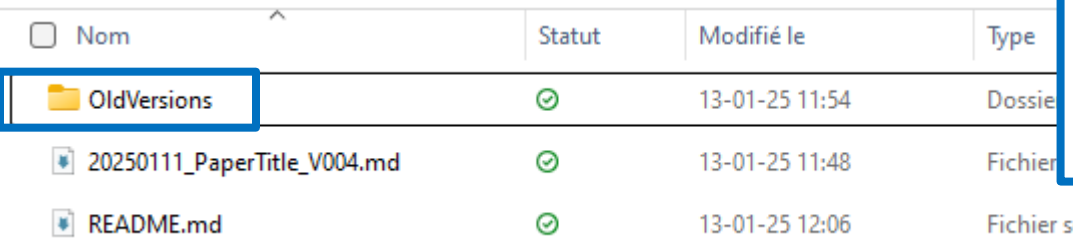
Versioning: keep older versions in a separate folder, and do not delete them unless you are absolutely sure you can. Keep a logfile that briefly explains the changes in each version.

Nom	Statut	Modifié le	Type	Taille
OldVersions	✓	13-01-25 11:54	Dossier de fichiers	
20250111_PaperTitle_V004.md	✓	13-01-25 11:48	Fichier source Mar...	0 Ko
README.md	✓	13-01-25 12:06	Fichier source Mar...	1 Ko

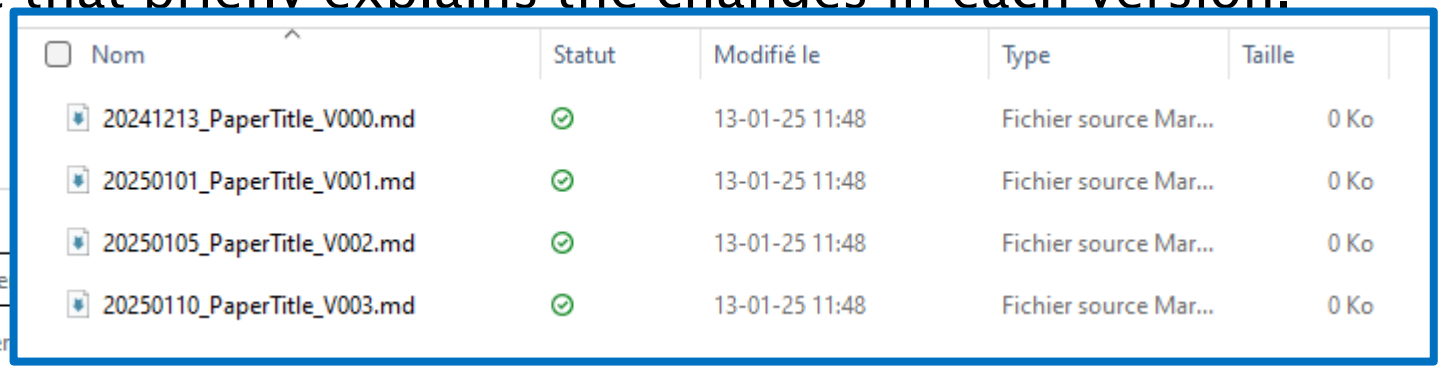
https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide/2.-Organise-Document/File-naming-and-folder-structure
https://library.si.edu/sites/default/files/tutorial/pdf/filenamingorganizing20180227.pdf

Data organization

Versioning: keep older versions in a separate folder, and do not delete them unless you are absolutely sure you can. Keep a logfile that briefly explains the changes in each version.



Nom	Statut	Modifié le	Type
OldVersions	✓	13-01-25 11:54	Dossie
20250111_PaperTitle_V004.md	✓	13-01-25 11:48	Fichier
README.md	✓	13-01-25 12:06	Fichier source Mar... 1 Ko

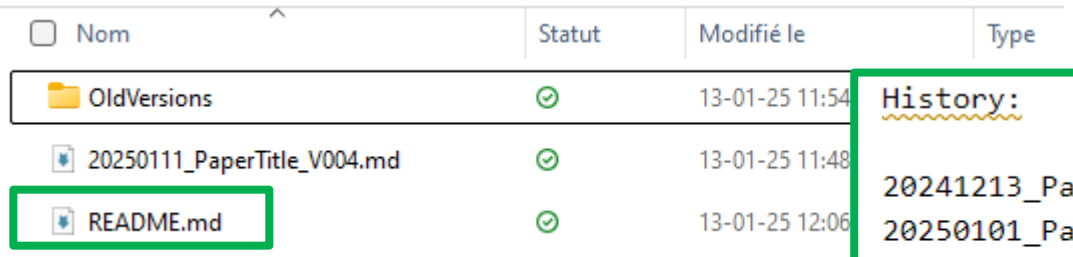


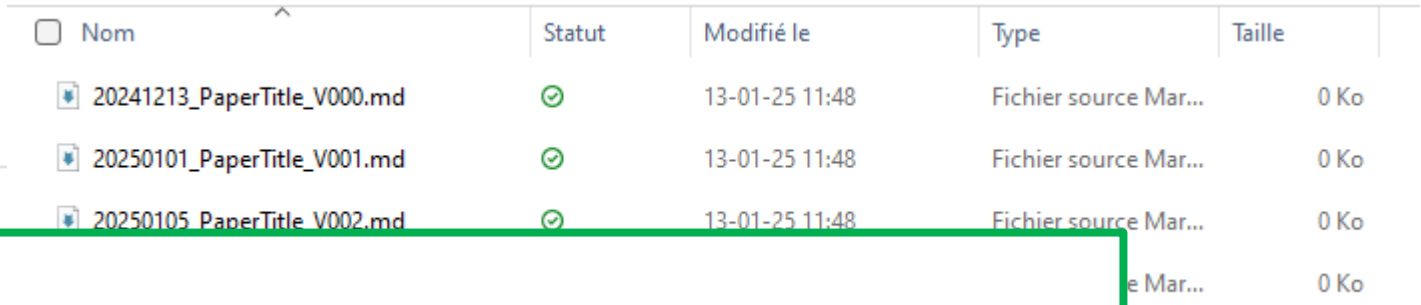
Nom	Statut	Modifié le	Type	Taille
20241213_PaperTitle_V000.md	✓	13-01-25 11:48	Fichier source Mar...	0 Ko
20250101_PaperTitle_V001.md	✓	13-01-25 11:48	Fichier source Mar...	0 Ko
20250105_PaperTitle_V002.md	✓	13-01-25 11:48	Fichier source Mar...	0 Ko
20250110_PaperTitle_V003.md	✓	13-01-25 11:48	Fichier source Mar...	0 Ko

https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide/2.-Organise-Document/File-naming-and-folder-structure
https://library.si.edu/sites/default/files/tutorial/pdf/filenamingorganizing20180227.pdf

Data organization

Versioning: keep older versions in a separate folder, and do not delete them unless you are absolutely sure you can. Keep a logfile that briefly explains the changes in each version.



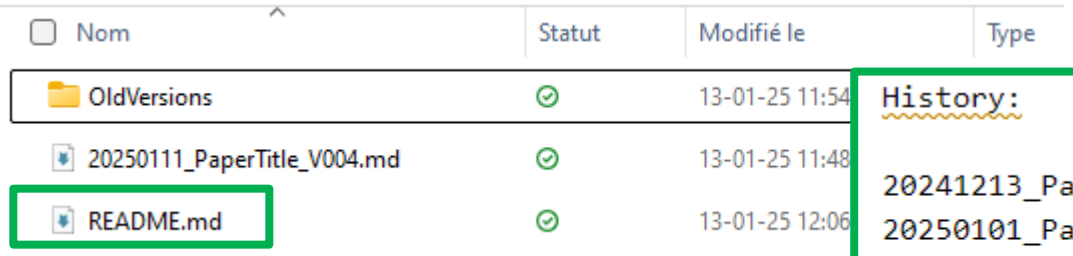


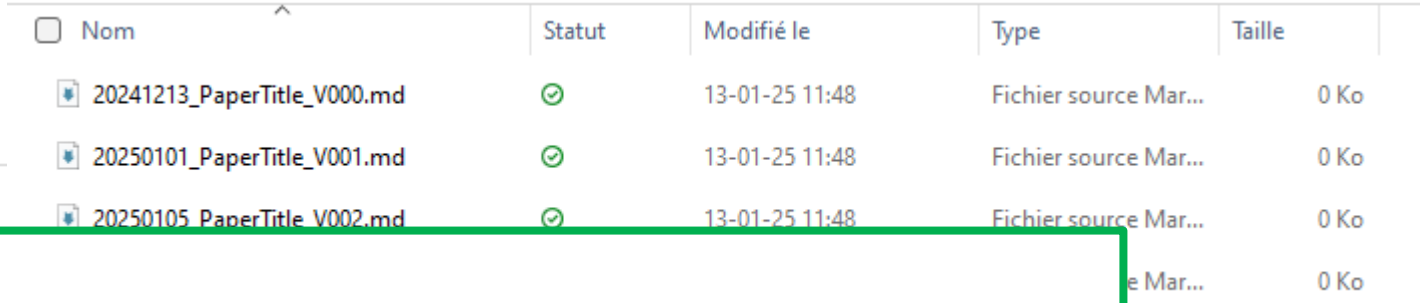

```
History:

20241213_PaperTitle_V000: first draft, introduction and state of the art
20250101_PaperTitle_V001: revision of the references
20250105_PaperTitle_V002: methodology
20250110_PaperTitle_V003: update methodology
20250111_PaperTitle_V004: draft of results
```

https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide/2.-Organise-Document/File-naming-and-folder-structure
https://library.si.edu/sites/default/files/tutorial/pdf/filenamingorganizing20180227.pdf

Data organization

Versioning: keep older versions in a separate folder, and do not delete them unless you are absolutely sure you can. Keep a logfile that briefly explains the changes in each version.






```
History:

20241213_PaperTitle_V000: first draft, introduction and state of the art
20250101_PaperTitle_V001: revision of the references
20250105_PaperTitle_V002: methodology
20250110_PaperTitle_V003: update methodology
20250111_PaperTitle_V004: draft of results
```

Key considerations: there are bad practices, but there are ***no unique answer*** or best method to organize data. You must be ***consistent*** throughout your project so that you and your team can work on.

https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide/2.-Organise-Document/File-naming-and-folder-structure
https://library.si.edu/sites/default/files/tutorial/pdf/filenamingorganizing20180227.pdf

Data description & documentation

Data description and documentation

Once you know how you will organize your data, you can start collecting them.

You should also start describing them using **metadata** and **documentation**.

- **Description through metadata** to make data findable.
- **Documentation** of your data should include everything needed to reproduce what you did (the method(s) you used to obtain them, how they were analyzed, processed, where you can find them, etc.).

Those information should be stored in distinct files in the relevant folders.

Data description

Data description and documentation

Standard metadata → depends on your discipline/data type/purpose

Metadata are important to **find** the data, ensure **reproducibility** and **reuse**

Common elements

- **Title**: The name given to the dataset.
- **Author**: The main researchers involved in producing the data.
- **Date of creation**
- **Identifier**: A unique code assigned to the dataset. May be added later, when *sharing* the dataset.

Slides 19-23

Standards commonly used:

- Dublin Core: 15 properties for describing a wide range of resources (general purpose)
  Dublin Core standard generator: The Simple Dublin Core to generate your metadata file!
- Digital Curation Center: social sciences
- Biology, earth sciences, physical sciences

Data documentation

Data description and documentation

Data documentation: document that explains your data. It will help others that would use your data, but also yourself to remember how you obtained and processed those data (*try to do it on the fly*). It also helps you and others to reproduce your results.

- **Folder organization:** you should explain how data are organized so that anyone starting to collaborate with you can understand what is going on.
- **Data collection:** explain how you obtained the data (from known datasets or from an experiment, survey, simulations, etc.).
- **Data cleaning**: explain your investigation and why you removed part of them (errors, inaccuracies, etc.) .
- **Data analysis:** how you analyzed the cleaned data (which software, parameters, results of the analysis, etc.).
- **Plan for change**: keep in mind that your data may change over the course of your project. Plan for how you will document and manage these changes.
- **Consider your audience**: remember that the description of your data may be read by people who are not experts in your field. Try to write in a way that is accessible to non-experts.

Data storage & sharing

Data storage and sharing

Data Storage

- **Reliability**: Use reliable storage solutions that ensure data integrity and availability **NOT YOUR LOCAL LAPTOP/DESKTOP!**
- **Accessibility**: Ensure that data is easily accessible to authorized users.
- **Scalability**: Choose storage solutions that can grow with the size of your data.

NOTE: Backup is different than preservation

- **Backup** = periodic snapshots in case current version is lost or destroyed (***cloud, NAS, etc.***)
- **Preservation** = archival, usually the final version of a dataset, stored for long-term and further use (***data repositories***)

Data storage and sharing

Backup: *when?*

- **Frequency**: Regularly back up data according to the importance and frequency of change.
- **3-2-1 Rule**: Keep at least three copies of your data, on two different media, with one backup offsite[1].
- **Verification**: Regularly verify the integrity of backup copies.
- **Disaster Recovery**: Have a disaster recovery plan in place to restore data if needed.

Nice **free open source** software for Linux, Windows and MacOS: FreeFileSync: Open Source File Synchronization & Backup Software

https://library.si.edu/research/best-practices-storing-archiving-and-preserving-data

Data storage and sharing

Backup: *where?*

- **Cloud**: UMONS provides 1To/user or team (see sharepoint) for free
- **Cloud**: if you need extra space or you want to backup somewhere else for collaborative work outside the university, you must pay for another service (not offered by UMONS)
- **Local server**: you can setup a Network Attached Server (NAS) for you and your team with backup services (not offered by UMONS)

https://library.si.edu/research/best-practices-storing-archiving-and-preserving-data

Data storage and sharing

Long-term preservation: ***data repositories***

A trusted digital repository provides reliable long-term access (10+ years) to managed digital resources to its designated community, now and in the future!

Only completed datasets with the purpose to ***publish***, ***share*** and/or ***preserve*** them should be uploaded (**not all research data**).

Typical files included in a dataset repository: inputs, outputs, method/documentation and metadata, not all intermediary data! With only those information, anyone should be able to reach your conclusions.

Examples:

- Zenodo: general purpose repository.
- SODHA: the federal Belgian data archive for social sciences and the digital humanities.
- Re3data.org: find other trusted data repositories.

Data storage and sharing

Data repositories also allow you to ***share*** your data.

Your data should meet the **FAIR data principles**:

Findable → **Metadata**

- Data should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets and services.

Accessible → **Data repository and unique identifier (DOI)**

- Once the user finds the required data, they need to know how they can obtain them. The data might be publicly accessible, or access may be restricted but metadata should remain accessible.

Interoperable → **Metadata, documentation**

- The data usually need to be integrated with other data regardless of the systems or tools being used → non-proprietary data formats (.txt, .csv, .md, .pdf, etc.)

Reusable → **Documentation**

- The ultimate goal of FAIR is to optimize the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.

Data storage and sharing

UMONS does not currently have an institutional data repository.

Therefore, we recommend using reliable and sustainable data repositories. **This approach allows researchers to retain freedom in choosing the repository that best suits their needs.**

We recommend using Zenodo for its versatility and ease of use.

Zenodo is free of charge.

A dataset is limited to 100 files and a total of 50 GB.

Before publishing, make sure:

- you hold the necessary rights to the data;
- no personal or confidential data is included without anonymization or explicit consent;
- all contractual obligations (research agreements, NDAs, etc.) have been respected.

Data storage and sharing

If you deal with personal data, *i.e.*, data that may directly or indirectly allow to identify someone, there are conditions and restrictions to share them : **GDPR**.

Only fully anonymized data can be shared.

In some cases, the approval of an ethics committee is mandatory before proceeding to data collection.

Please make sure to check with your **Data Protection Officer** (DPO) and with your local ethics committee contact to ensure compliance.

Contact: dpo@umons.ac.be

Microsoft Word - Fiche 11 Recherches scientifiquesV3.docx

Modèles de textes "RGPD" à utiliser

10 steps towards privacy compliance in research

/ ! \ Data Privacy Handbook

Data storage and sharing

Datasets are valorized in activities reports since this edition!

If you have datasets hosted on data repositories, you can reference them in ORBi and they will appear in the activities report:

Follow the guide: **ORBi UMONS: How to Reference Your Datasets**

Data licensing

Data sharing and licenses

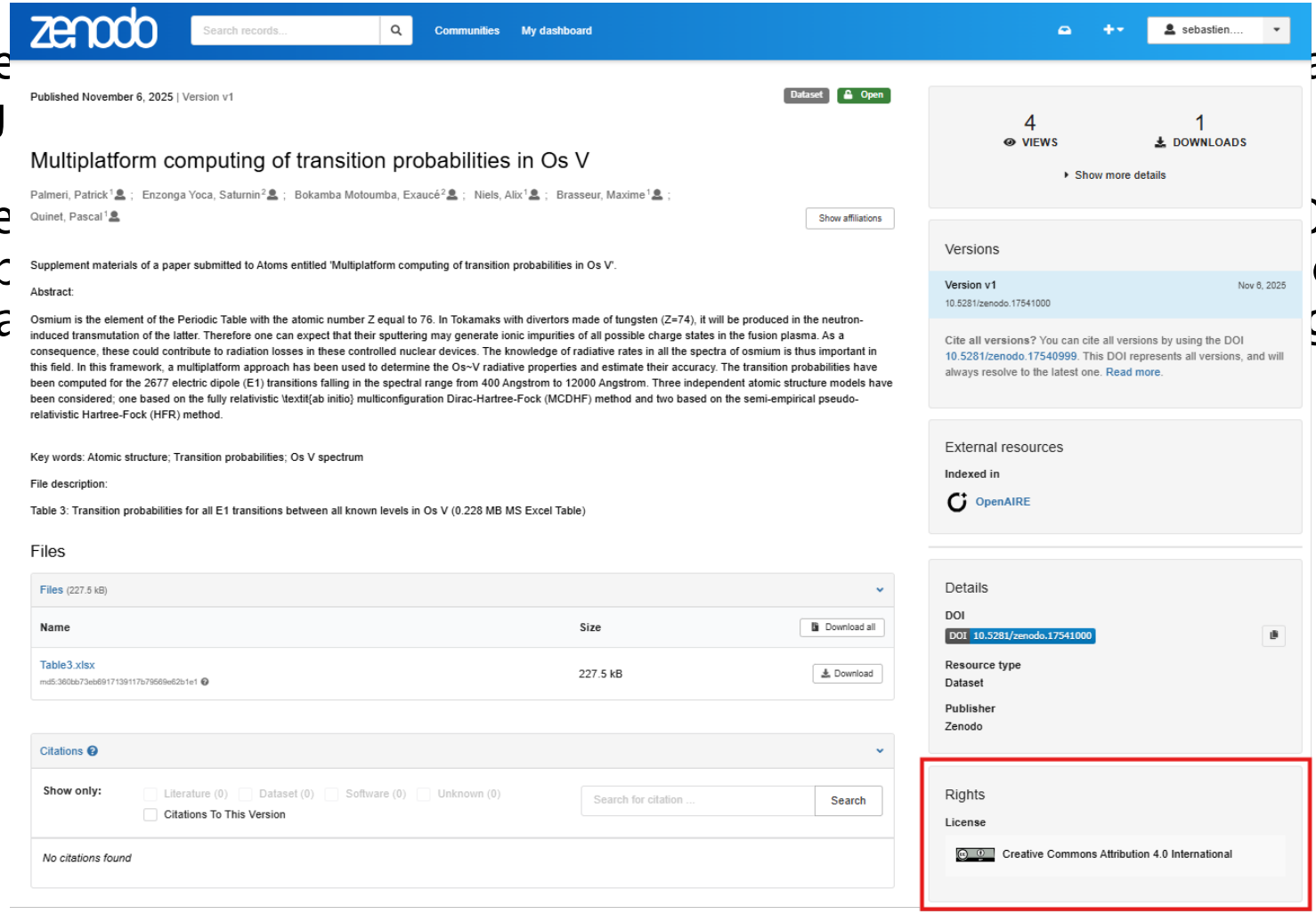
Data licenses are legal agreements defining how others can access, use, share, and distribute a dataset, granting permissions while setting restrictions, crucial for clarity in data sharing.

Common types like Creative Commons (CC-BY) and Open Data Commons (ODbL) offering varying levels of openness for data creators and users. They provide legal certainty, ensuring ethical and compliant use, and are applied by tagging metadata or including license files with the data.

Data sharing and licenses

Data licenses are le…are, and distribute a dataset, granting …in data sharing.

Common types like…ODbL) offering varying levels of op…ertainty, ensuring ethical and compli…g license files with the data.



Data sharing and licenses

Key take-away: *To effectively share data, resolve any data ownership or intellectual property rights issues early. Consult LEGAL-AVRE to determine what policies might affect data ownership and sharing.*

- **Understand Your Goals**: What do you want others to be able to do with your data? This can help guide your choice of license.
- **Public Domain**: If you want to give people the most freedom, consider a public domain license like CC0.
- **Attribution**: If you want to allow free use but also want to be credited, consider a license that requires attribution, like CC-BY.
- **Share-Alike**: If you want any derivatives of your work to be licensed under the same terms, consider a share-alike license, like CC-BY-SA.
- **Non-Commercial**: If you want to restrict the commercial use of your data, consider a non-commercial license, like CC-BY-NC.
- **Understand the Implications**: Make sure you understand the implications of the license you choose. Some licenses may have implications for how your data can be used or shared.

UMONS Intellectual property regulation

Budget for storage

Budget for data storage

Data production can be costly (equipment, products, time).

In addition to the costs of data collection, data management, curation, documentation, **storage can be expensive**.

→ Need to think about all costs (IT, server, etc.).

At the moment, **1 TB/person for free** with Microsoft OneDrive, but may not be enough or practical for your project.

These costs may be eligible in some calls for projects.

Examples:

- Cloud storage on Microsoft Azure: 250 GB/month → min. 50€ → 4 years = 2400€ only for your research
- Network attached storage: fixed cost NAS 500€ + UPS 200€ + 4TB disks 120€ * 3 = 1060€ → multiple users, research, etc. (*no support from the university yet*)

* storage, backup, second backup → 3-2-1 rule

Resources from the Data Ambassadors Network

You may need help processing those information or have questions about RDM:

The Data Ambassadors Network is there for you!

It is a inter-university network with representatives in each university with the mission to help and guide you with data in general.

UMONS Data Ambassadors:

VISEUR	Robert	Business and Economics	Robert.VISEUR@umons.ac.be
GALLAS	Mohamed-Anis	Architecture	Mohamed-Anis.GALLAS@umons.ac.be
VILLERS	Agnès	Medicine	Agnes.VILLERS@umons.ac.be
PATRIS	Stéphanie	Medicine	Stephanie.PATRIS@umons.ac.be
GROSJEAN	Philippe	Sciences	Philippe.GROSJEAN@umons.ac.be
DUPONT	Nicolas	Applied Sciences	Nicolas.DUPONT@umons.ac.be
JIMENEZ-SALCEDO	Juan	Languages	Juan.JIMENEZSALCEDO@umons.ac.be
PIEROPAN	Laurence	Languages	Laurence.PIEROPAN@umons.ac.be
RIVIERE LORPHEVRE	Edouard	Applied Sciences	Edouard.RIVIERELORPHEVRE@umons.ac.be
SIMOES LOUREIRO	Isabelle	Psychology	Isabelle.SIMOESLOUREIRO@umons.ac.be

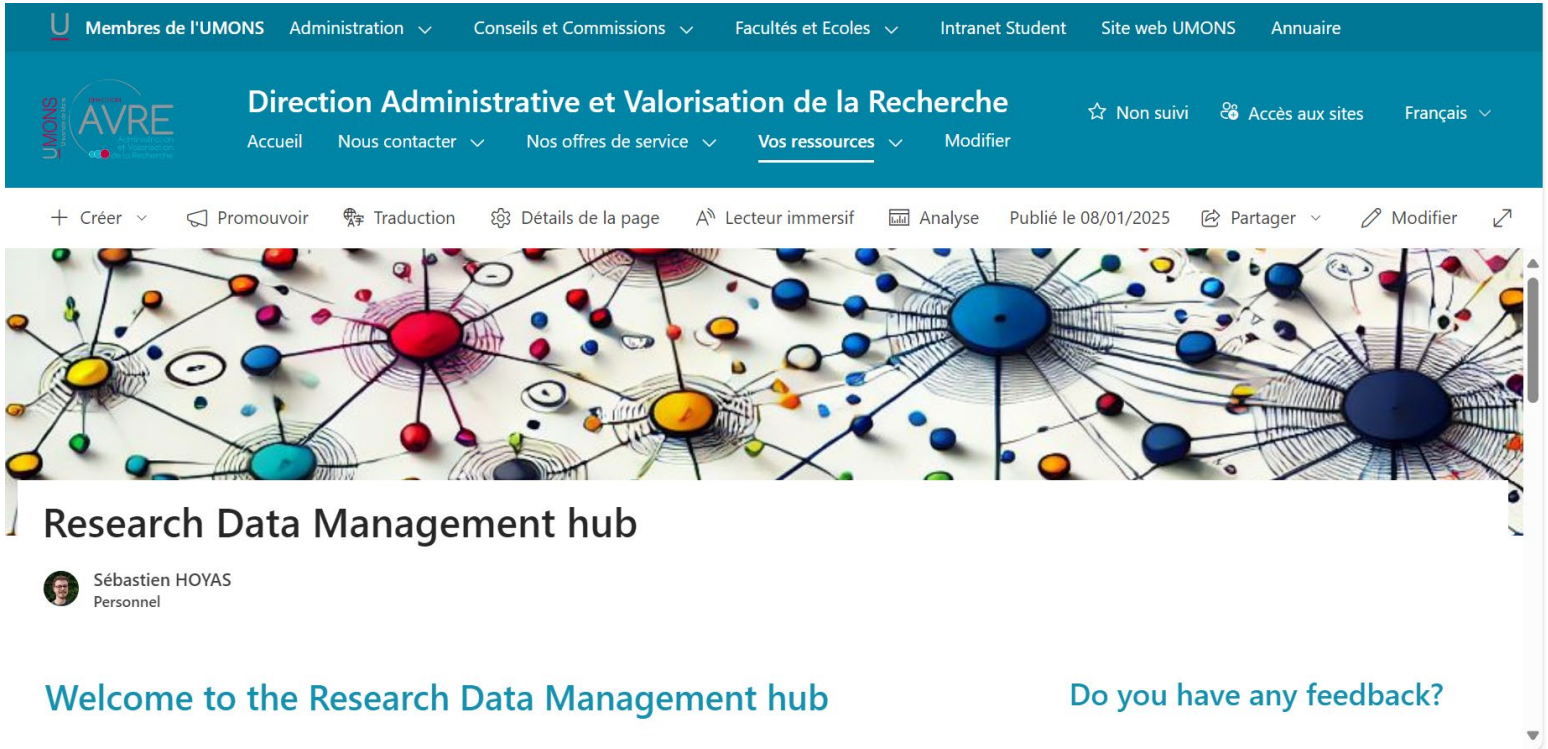
+ webinars (data anonymization, how to archive data, etc.)

Resources from the Data Ambassadors Network

Data Ambassador Communities - Peer-to-peer approach for a better research	21/11/2021	Data Ambassador Communities - Peer-to-peer approach for a better research-20211125_130429-Enregistrement de la réunion.mp4
Données sensibles et valorisables	22/04/2022	Données sensibles et valorisables - Webinaire DA FWB-20220422_140413-Enregistrement de la réunion.mp4
Open and FAIR Data - Testimonies	29/11/2022	Open and FAIR Data _ Testimonies-20221129_140517-Enregistrement de la réunion.mp4
Connaissez-vous le Dual Use ?	01/12/2022	Data Ambassadors - Connaissez-vous le Dual Use _ -20221201_123615-Enregistrement de la réunion.mp4
Open Software and Open Data - why and how	20/06/2023	FWB Data Ambassadors - Open Software and Open Data _ why and how _-20230620_123637-Enregistrement de la réunion.mp4
Réutiliser les données de réseaux sociaux pour la recherche	18/10/2023	Data Ambassadors - Réutiliser les données de réseaux sociaux pour la recherche-20231018_090512-Enregistrement de la réunion.mp4
Archivage des Données de Recherche - pourquoi, comment, et pour qui	29/11/2023	Archivage des Données de Recherche _ pourquoi, comment, et pour qui _-20231129_143540-Enregistrement de la réunion.mp4

Resources from AVRE

[Research Data Management hub]



Closing remarks

- Research data management is not an easy task but is not impossible either.
- There is no unique way to properly manage your data, as long as your are ***consistent***, ***descriptive*** and organize in a way that ***anyone can understand***.
- Different tools are available (**Compass to RDM**, **DMP online, RDM Hub**) to help you to plan how to manage your data.
- **Data Ambassadors** and **AVRE** people are also here to help you.

Q&A
If I was not able to answer your question, feel free to reach out by email

sebastien.hoyas@umons.ac.be

UMONS
Université de Mons

UMONS Université de Mons
DIRECTION AVRE
Administration et Valorisation de la Recherche

Thank you for your attention

Resources for data management

Share personal data through a repository

Facilitating FAIR practices in Research Methods, Data, And Software in Natural and Engineering Sciences

Complete training about RDM by Macalester College Library

DocFetcher - Fast Document Search (sourceforge.io) to index your files and quickly find their content

Understanding Research Data Management – University of Pittsburgh

https://libereurope.eu/event/data-management-plans-use-and-reuse-webinar/ ; coming webinar

Challenges in RDM

Completed DMP for « PURE » European Union's Horizon 2020 project Pure Project Data Management Plan (zenodo.org)

Other DMPs for different disciplines: Example DMPs and guidance | DCC

Do's and Don'ts of DMP

Resources to find data for your research

EOSC (Europen Open Science Cloud): https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud

Mendeley Data website: https://data.mendeley.com/datasets

OpenAire : https://explore.openaire.eu/search/find

re3data.org: https://www.re3data.org/: Harvesting several data repositories

Google: https://toolbox.google.com/datasetsearch

FigShare: figshare - credit for all your research

Zenodo: Zenodo - Research. Shared.