UCLouvain

Introduction to Research Data Management & Data Management Plans

Jonathan Dedonder
Logisticien de recherche IACCHOS, UCLouvain
Joëlle Desterbecq
Research Data Officer, ADRE – SCEB, UCLouvain
Alexis Buckens
Research Data Manager, IACCHOS, UCLouvain
Adeline Grard
Illustrations: Midjourney

1

Presentation

Part 1	Part 2	Part 3
Research Data Management	**Data Management Plans**	**Exercise**
What is it? What are the main steps? How can it help me as researcher?	What is it? What are the main steps? How can it help me as researcher?	Presenting your data & writing a DMP

2

Research Data Management

Get visible and promote

Link your data with papers, spread the news on social media, promote your RDM skills

3

RDM Advantages for researchers

1. Helps planning your research
2. Increase use of data management best practices
3. Get access to data collected by others
4. Share your data with your fellow partner, scientific community, or society
5. ….But keep your sharing in control (legal, ethical)Get cited for your data (DOI)
6. Visibility
7. Transparency (reproducibility)
8. Get more from your data (other researcher): better return on investment
9. Helps YOURSELF to reuse your previoulsy acquired data
10. Store and backup safely
11. Merge datasets and start new research projects
12. Sometimes it's mandatory

4

Get Visible and promote your RDM Skills

Promote your RDM Skills
They are valuable assets for employers (academics or not),
but also to describe your research environment (research proposal)

Some examples:

- ✓ Knowledge in research process (data collection, methods)
- ✓ Knowledge in data curation, coding, IT skills
- ✓ Disciplinary specificities (tools, devices, programs, etc.)
- ✓ Knowledge in ethical and/or commercial use of data in your field
- ✓ Knowledge in the repositories, websites, where you can find/share data in your field
- ✓ Knowledge in data license

5

Research Data Management

Create, organize, make, store and share research data of an institution

RESEARCH DATA: « the recorded factual material commonly accepted in the scientific community as necessary to validate research finding"

6

The context urge to make RDM a reality

International research context

We live in a digital world where data are central

Trust crisis in science - Replicability

Open access

Meet researchers needs

Research support to reach excellence in research

7

Research Data : Definition

What are data ?

- " **Factual records** (numerical scores, textual records, images and sounds) used as **primary sources** for scientific research, and that are commonly accepted in the scientific community as necessary to **validate research findings** ".

OECD, *OECD Principles and Guidelines for Access to Research Data from Public Funding*, OECD Publishing, Paris, 2007, p. 13. https://doi.org/10.1787/9789264034020-en-fr

- Research Data can take a diversity of **forms** and **formats**

- Figures and measurements
- Observational data
- Experimental data
- Geospatial data
- Medical imagery
- Code
- Drawings, maps or plans
- Interviews, surveys
- Texts
- Audiovisual
- Photographs, etc.

→ in all kind of file formats (.png, .mpeg, .svg, .wma, .pdf, .txt, .xml, etc.)

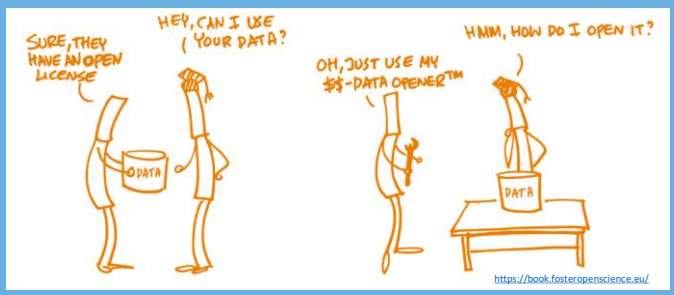
8

Fair vs Open Data

Definition

"Open data is data that can be **freely used, re-used and redistributed** by anyone – subject only, at most, to the requirement to **attribute and share-alike**"

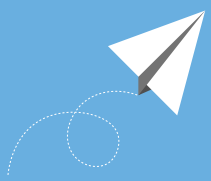
Open Knowledge foundation, Open Data Handbook.
https://opendatahandbook.org/guide/en/what-is-open-data/



9

Fair vs Open Data

- **"Availability and access :** the data must be available as a whole and at no more than a reasonable reproduction cost, preferably by downloading over the internet. The data must also be available in a convenient and modifiable form.
- **Re-use and redistribution :** the data must be provided under terms that permit re-use and redistribution, including the intermixing with other datasets.

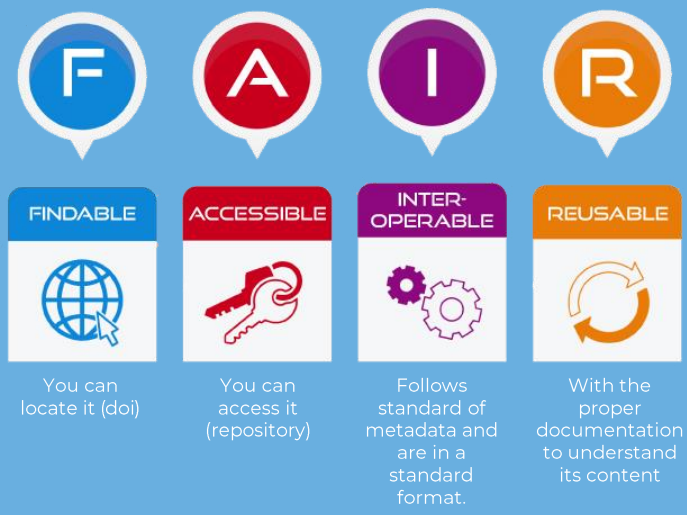


- **Universal participation :** everyone must be able to use, re-use and redistribute – there should be no discrimination against fields of endeavor or against persons or groups :
  For example: restrictions of use for certain purposes (e.g. only in education), are not allowed".

Source: https://opendatahandbook.org/guide/en/what-is-open-data/

10

Fair vs Open Data



11

Fair vs Open Data

Fair data principles

- **Findable**: The first step in (re)using data is to find them. Metadata and data **should be easy to find** for both humans and computers.
- **Accessible**: Once the user finds the required data, **she/he/they need to know how they can be accessed**, possibly including authentication and authorization.
- **Interoperable**: Data can be exploited, exchanged, compared or re-used in a variety of contexts. To achieve this, the data must be able **to be integrated with other data**. In addition, the data must be interoperable with applications or workflows for analysis, storage, and processing.
- **Re-usable**: Optimizing 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. A clear and accessible **license defines the conditions** for re-use.

https://www.go-fair.org/fair-principles/

12

Fair vs Open Data

Fair data principles

Findable	Repository	Persistent identifier	Add Metadata	
Accessible	Standard protocol (http, ...)	Free and Open Protocol (non-proprietary)	Authentication & authorisation (if necessary)	Metadata accessible (even if data are not)
Interoperable	Machine-readable	Vocabulary (FAIR) & ontologies	Interconnexion with other data	
Reusable	Usage license	Metadata with attributes & provenance	Community standards (discipline)	

13

Degrees of data sharing

- The principle : **"As open as possible, as closed as necessary".**
- Open data is not mandatory. But data must be "FAIR"
- There are different levels of Openness
  - It ranges from making data fully open on one end, to keeping them fully closed on the other, with various possible forms of restricted/controlled access in-between.

14

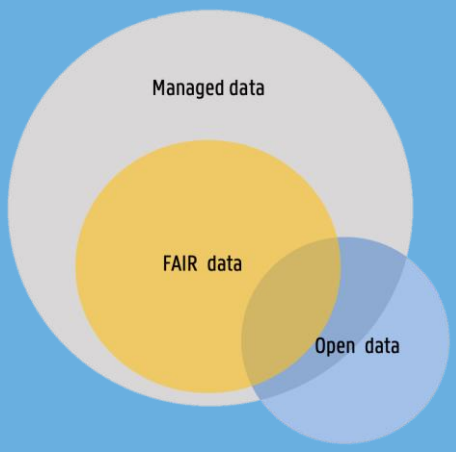
Degrees of data sharing

1. Personal data
2. Sensitive data
3. Confidential data or data subject to secrecy
4. Third-party data
5. Data protected by copyright and / or database right
6. Data with commercial / economic valorisation potential
7. Data that infringe prior commitment related to data sharing (consortium agreement, informed consent forms, …)

https://www.ugent.be/en/research/openscience/datamanagement/after-research/sharing.htm

15

Fair vs Open Data



https://www.ugent.be/en/research/openscience/datamanagement/after-research/fair-data.htm

16

Research Data Management

Search or Collect data

Use secondary data or collect your own

17

RDM in the Research life cycle

18

Why use secondary data?

Using data collected by other researchers is very interesting for you research

Introduce/ discuss	Write a research proposal and build your case on data from several datasets
Save time/ money	Limit the data collection expense in using existing data (and test you hypothesis to them).
Compare/ discuss	Compare or discuss your research results with similar data, collected in other time/places, or with different methods

19

How to use secondary data?

Definition : Data Repository

What is a Data Repository ?

- **A data repository** is an online platform that is used to **deposit completed datasets** with the purpose to **publish, share and/or preserve them**.
  - Share your data over the short or medium term (5, 10, 15 years)
- A data repository is a database infrastructure that compiles, manages and gives access to **data**, **associated metadata** and **documentation**.
- It contributes to **make your data FAIR** : findable, accessible, reusable, interoperable.

20

How to use secondary data?

Check if there are any existing data that you can reuse, by consulting relevant repositories

1. EOSC: https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud
2. Zenodo: https://zenodo.org
3. Mendeley data: https://data.mendeley.com/datasets
4. OpenAire: https://explore.openaire.eu/search/find
5. Re3data: https://www.re3data.org/
6. Zanran: http://www.zanran.com/q
7. Google: https://toolbox.google.com/datasetsearch

Awlays check the quality, read metadata and documantation. Check with an expert.

21

Research Data Management

Data Management

(Support en Méthodologie et Calcul Statistique – SMCS)

Enter, check, clean, organize and document your data

22

Data entry

When data are digitized, entered in a database or spreadsheet, or coded, quality is ensured and error avoided by using standardized and consistent procedures with clear instructions.

- ✓ Using data entry screens
- ✓ Using controlled vocabularies, and choice lists to minimize manual data entry
- ✓ Detailed labelling of variable to avoid confusion

23

Data checking

During data checking, data are edited, cleaned, verified, cross-checked and validated. Checking typically involves both automated and manual procedures. These may include:

- ✓ Verifying random samples of the digital data against the original data
- ✓ Statistical analyses such as frequencies, means, ranges or clustering to detect errors and anomalous values
- ✓ Peer review

24

Data cleaning

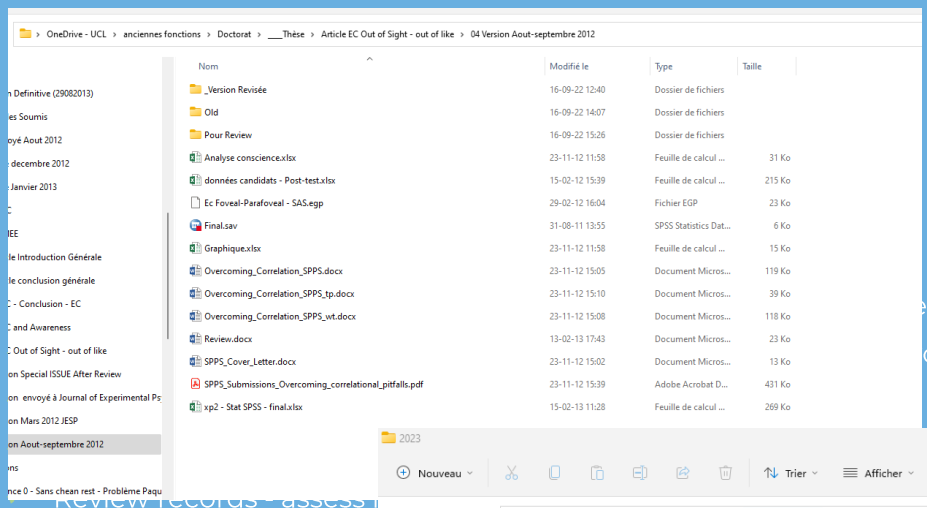
- ✓ How are the missing value encoded ? (several types of missing values)
- ✓ Individuals/observations are in line and not in column.
- ✓ The columns' name should be written on one line (help for importation).
- ✓ Withdraw the useless lines and columns (and avoid leaving empty columns).
- ✓ Data importation: check if there are the same before and after the importation.
- ✓ Look for duplicated observations.
- ✓ Look for consistency between your variables

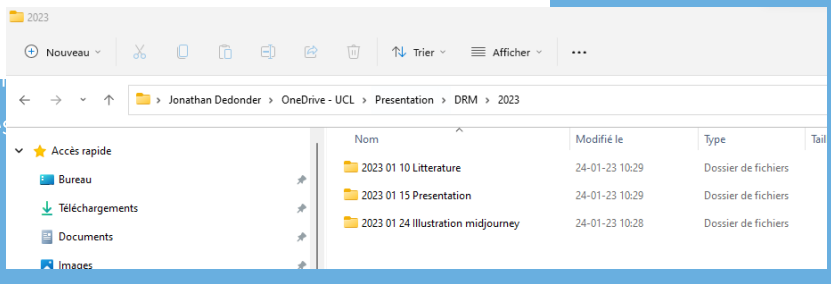
25

Data organization

- ✓ Use folders - and structure folders hierarchically
- ✓ Adhere to existing procedures
- ✓ Name folders appropriately - after the areas of work not after researchers
- ✓ Agree and be consistent – once you have decided on a method, you stick to it.
- ✓ Separate ongoing and completed work - move files regularly
- ✓ Backup
- ✓ Review records - assess materials regularly to ensure
- ✓ Files are not kept needlessly.

26





earchers

ou stick to it.

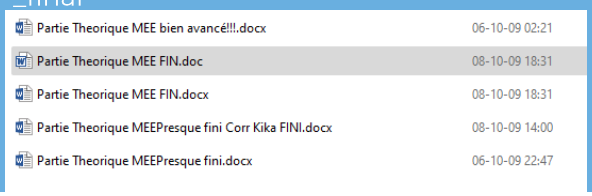
- ✓ files are not kept needles

27

Files naming

Useful file names are consistent, meaningful to you and you colleagues, and allow you to find the file easily

- ✓ Vocabulary, punctuation – everyone uses a common language
- ✓ Dates – agree on a logical use of dates so that they display chronologically
- ✓ Revision procedure: version + reviser (e.g. _V01_AG), change version number for
- ✓ big changes
- ✓ Agree on who is responsible of using "final" _final



28

For UCLouvain researchers

Contacts SMCS if you need help in managing and analyzing your data

https://sites.uclouvain.be/training/smcs/

29

Research Data Management

Store & protect

Storage solutions and best practices, ethical and patents ressources

30

Data storage – best practices

1. Avoid storing your data on your laptop/smartphone
2. Keep 3 copies of important data (2 outside your laptop).
3. Keep at least the raw data, the cured one, and the versions used for publications.
4. use encryption software – Veracrypt
5. Do not forget the physical data (notes, sketches, etc.)

31

Store in practice

1. Use secure files storage like institutional server
2. Transfer large dataset through trusted peers
3. Don't use Dropbox, Google Drive (or any other cloud solution) for your research data (seen by others, data property issue).
4. UCLouvain: One drive can be acceptable solution data are stored and backuped safely in the EU, per UCLouvain agreement.

For UCLouvain, more information at https://uclouvain.be/fr/universite-numerique/rdm/store-uclouvain.html

32

The GDPR

- ✓ Data protection is a fundamental human right (Charter, Art. 8) and a central issue for research ethics.
- ✓ General Data Protection Regulation (2018) GDPR applies to the personal data processing of EU data subjects and processing by a controller/processor located within the EU.
- ✓ Defines personal data as any information relating to an identified or identifiable natural person who can be identified, directly or indirectly, in particular by reference to an identifier, such as a name, an identification number, location data, an online identifier etc

33

The GDPR - principles

- ✓ **Lawfulness** – legitimate basis must be clarified. For research these are most often 'legitimate interest', 'public interest' along with 'consent'.
- ✓ **Fairness** – towards the data subject.
- ✓ **Transparency** – data subjects should be aware of the processing of their personal data.
- ✓ **Purpose limitation** – purpose must be specified, explicit and legitimate. Personal data collected for one purpose should not be used for another purpose unless it is compatible with original purpose.

34

The GDPR - principles

- ✓ **Data minimization and proportionality** – only collect the data you need.
- ✓ **Accuracy** – keep records up to date.
- ✓ **Storage limitation** – assess the purpose and reasoning for storing the data for lengthy periods of time.
- ✓ **Integrity and confidentiality** – protect data from damage and unlawful processing. Information security, encryption, pseudonymisation.
- ✓ **Accountability** – demonstrate responsibility and compliance through documentation

35

Sensitive data for the GDPR

- ✓ Racial or ethnic origin
- ✓ Political opinions
- ✓ Religious or philosophical beliefs
- ✓ Trade union membership
- ✓ Biometric data (where used for ID purposes)
- ✓ Health
- ✓ Sex life
- ✓ Sexual orientation
- ✓ Genetic data
- ✓ Criminal convictions
- ✓ Offences
- ✓ Security measures
- ✓ Data concerning children

36

Informed consent

- ✓ **Ethical and legal process** – 'consent' is now also a legitimate basis under which researchers can process personal data. Under the GDPR **consent needs to be freely given, informed, unambiguous, specific and indicated by a clear affirmative action**.
- ✓ People have a right to **know that they are participating** in research.
- ✓ Informed consent forms should be used in all research with **human subjects**.
- ✓ Informed consent forms state the **reason for data collection, how data will be used, how data is stored and who the responsible contact is**.
- ✓ **Sharing or publishing data should be mentioned** in the informed consent forms – if omitted, it could cause problems later on.
- ✓ **Consent can be withdrawn** at any time…

37

Research Data Management

Publish and share

Select your data for publication, choose a repository, publish a data paper, make a data sharing agreement or license your data, choose an embargo period

38

Definition : Data Repository

What is a Data Repository ?

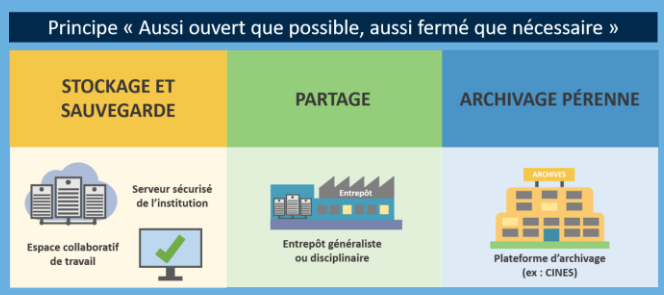
- **A data repository** is an online platform that is used to **deposit completed datasets** with the purpose to **publish, share and/or preserve them**.
  - Share your data over the short or medium term (5, 10, 15 years)
- A data repository is a database infrastructure that compiles, manages and gives access to **data, associated metadata and documentation**.
- It contributes to **make your data FAIR** : findable, accessible, reusable, interoperable.

39

Data Repository ≠ Data storage ≠ Publication Repository

- **Data Repository** : publication and sharing of data in open or FAIR mode at the end of research (ex. UCLouvain Dataverse)
- Different from **storage** during research (Mass storage – CISM, etc.)
- Different from **long-term archiving** (Archive Service)
- Data Repository (Open Data) **≠ Publication Repository** (Open Access)

  ➔Dataverse ≠ DIAL UCLouvain



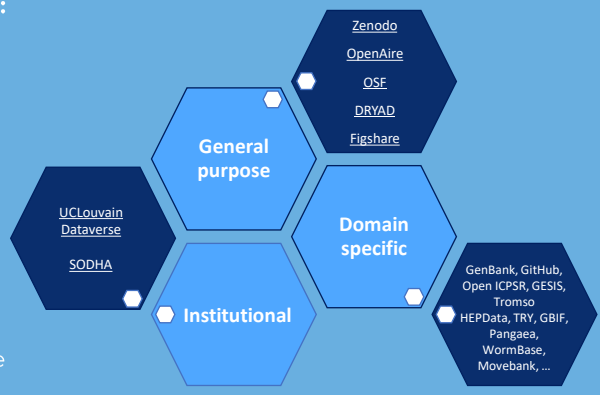
DORANum, https://view.genially.com/60140ceb1bd3060d78c600aa

40

Data Repository categories

There are different types of Data Repositories :

- **General purpose or multidisciplinary** : accept a wide range of data types (and sometimes other research outputs as well) from all disciplines
- **Domain specific / disciplinary :** focus on specific data types or data from specific research domains
- **Institutional :** hold research data outputs from a particular research institution
  ➔ UCLouvain Dataverse, SODHA (Belgian federal data archive for social sciences and the digital humanities)
- Repositories provided by **publishers**



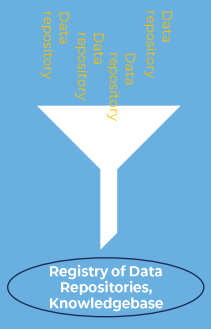
Dedieu L., Barale, M., *Déposer des données dans un entrepôt, en 6 points.* Montpellier, CIRAD, 2020, 4 p. https://doi.org/10.18167/coopist/0070
https://www.ugent.be/en/research/openscience/datamanagement/after-research/data-repositories.htm

41

How to find a Data Repository ?

"Repository registries" or "knowledgebases" – definition

- There are several **"registries" or "knowledgebases"** listing data repositories



- "Also known as secondary databases, knowledgebases **synthetise data** from a number of other data sources including published literature, often via manual **curation**" *
- They offer facilities to **search** and browse descriptions of research data repositories
- They are a good starting point for **identifying** a suitable repository

* Pelletier G., Lister A., Sansone S.-A., *FAIRsharing content: Databases overview,* 26 july 2023. fairsharing.org/educational. https://doi.org/10.5281/zenodo.8186961

42

How to find a Data Repository ?

Registries to identify a data repository suitable for your research :

- **re3data** (registry of research data repositories): directory listing more than 2,000 repositories and allowing you to refine your searches using a range of filters (subjects, countries, certificates, access methods, etc.) : https://www.re3data.org/

- **FAIRsharing**: directory that compares repositories according to their compliance with the FAIR principles : https://fairsharing.org/
- **OAD (Open Access Directory)**: wiki page listing repositories by discipline: https://oad.simmons.edu/oadwiki/Data_repositories
- **Core Trust Seal**: allows you to search for certified repositories. https://amt.coretrustseal.org/certificates/

re3data.org

43

Select a trusted Repository

- Comité pour la Science ouverte – Collège Données de la Recherche (Novembre 2023) :
- The College has defined **a list of exclusion criteria** for selecting **trusted thematic repositories**

Exclusion criteria for a Data Repository

- No moderation of deposits
- No permanent identifier
- No guarantee of infrastructure continuity
- Property Rights transfer
- Excessive pricing policy
- Localisation of data outside the European Union (➔ GDPR)
- Repository restricted by institutional affiliation

Do not select this type of Repository

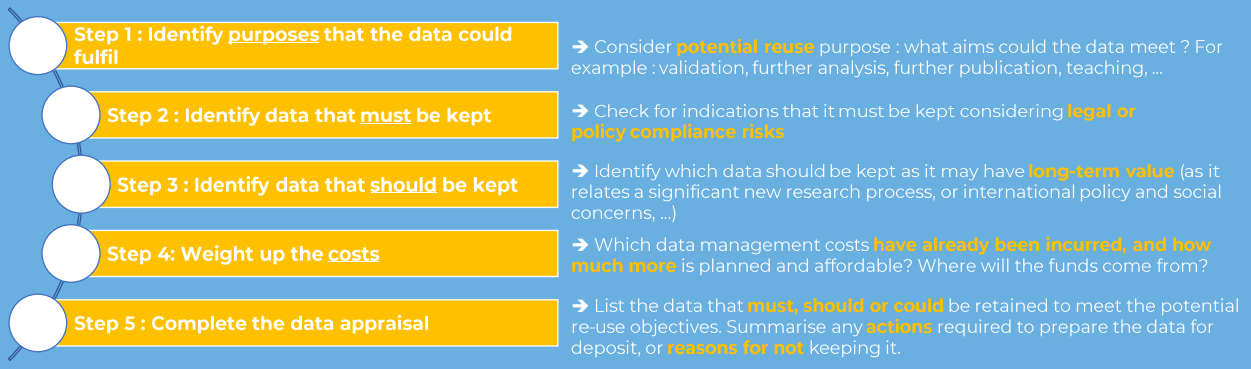
Arnould P.-Y. *et al.*, "Sélectionner un entrepôt thématique de confiance pour la diffusion des données de la recherche : note méthodologique", Ouvrir la Science ! France, Comité pour la Science ouverte – Collège Données de la Recherche, Novembre 2023.

44

Deciding what data to keep

"Digital Curation Center" (DCC) Data appraisal – *Five steps to decide what data to keep* :

- Not all the data from your research is necessarily intended for sharing.
- You have to select the datasets to be shared.



Whyte A., DCC, *Five steps to decide what data to keep: a checklist for appraising research data* (v.1), Edinburgh: Digital Curation Centre, 2014. www.dcc.ac.uk/resources/how-guides. Direct link : https://www.dcc.ac.uk/guidance/how-guides/five-steps-decide-what-data-keep#3

45

Deciding what data to keep

1. Select data you must publish, and delete those you have to (consortum agreement, legal obligations, GDPR requirements).
2. For other data, consider their uniqueness, long-term value and potential of reuse.
3. Keep certain data to validate your publication's results, for future teaching or research.
4. Take also into account the costs (time, software, etc.) and efforts required to preserve these data (preparation, documentation, and storage steps).
5. Depending on these (legal) aspects, you may state a period of preservation: some data will be obsolete in 2, 5, 10 or 50 years.

46

License your data

Creative Commons Licenses : Four components

- Six licenses are possible, combining **four basic elements** : the attribution (BY), the derivatives works (ND), the commercial use (NC), and the "share-alike" principle (SA)

Icon	Name		Meaning
	Attribution	**BY**	You must give credit (mention) to the creator of the work concerned.
	No derivative works	**ND**	No modification : you must use the work unmodified and in its entirety. No derivatives or adaptations of the work are permitted.
	Share alike	**SA**	Share under the same conditions: you must share the modified work using the same license (the same icons).
	No commercial use	**NC**	Reproduction is authorised but you may not use the work (modified or not) for commercial purposes.

Duquesnoy M., « Module 4 : Le droit d'auteur sur internet ». We translate. https://view.genial.ly/634ba0bdd1d0d90012caf04c/presentation-module-4-le-droit-dauteur-sur-internet
https://creativecommons.org/share-your-work/cclicenses/

47

License your data

Creative Commons Licenses : Six Licenses

JoKalliauer; foter, CC BY-SA 3.0, https://creativecommons.org/licenses/by-sa/3.0, via Wikimedia Commons

48

Make a Data Sharing Agreement

What is a Data Sharing Agreement ?

- "A data sharing agreement (DSA) is a **convention between research partners and/or third parties"***.
- It "states several information about **the way data will be formatted**, but also **the way they will be (re)used and shared** in the future (how data can be used and to which purpose).
- The DSA mentions **data sources** and **ownership**, and acknowledge **partners' responsibilities**"*.
- "**DSA can be different** depending on the type of data, the receiving party, and the use that will be made"*.
- You can edit dataset terms on **UCLouvain Dataverse** Repository and add a DSA.

* https://uclouvain.be/fr/universite-numerique/rdm/data-sharing-agreement.html

49

Choose an embargo period

Define an embargo period, if necessary

- An embargo is a **limited period** during which **the data you have deposited** in a data repository **are not open to the public** (closed data).
- In many cases, this embargo period is stated in your consortium agreement, funders' contract, patent, etc.

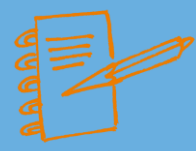
WHY ? There might be several reasons : for example, having a reasonable amount of time during which only you and your partners might publish using your data

HOW ? 1) Consider how long your embargo period will last (if any) and state a clear release date. 2) Edit dataset terms in your chosen repository to specify the embargo period

https://uclouvain.be/fr/universite-numerique/rdm/choose-an-embargo-period-for-your-data.html

50

Document your data



https://book.fosteropenscience.eu/

Data documentation (1/2)

What ? data documentation comprises any **contextual and descriptive information** needed to find, assess, understand, and (re)use research data.

Why ? Data documentation has several objectives :

- it enables you to understand/**interpret** data later;
- it makes data **independently understandable** (i.e. reusable);
- it make results **independently reproducible**, starting from raw data;
- it helps avoid incorrect use/**misinterpretation**.

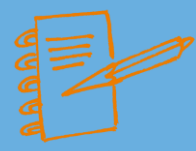
It is an essential step in making your data FAIR :

- Data documentation allows you, other researchers and third parties to (re)use your data adequately.

https://www.ugent.be/en/research/openscience/datamanagement/during-research/documentation.htm
https://uclouvain.be/fr/universite-numerique/rdm/document-your-data.html

51

Document your data



https://book.fosteropenscience.eu/

Data documentation (2/2)

When ? Start gathering information as early as possible & continue as the project progresses

How ? Document 1) the study/the project and 2) the data

➔ Documentation can be captured **in various ways** : in a research paper, project report, **lab notebook**, **codebook**, separate **'readme' file**, database, annotated computer script and/or the data files themselves, etc.

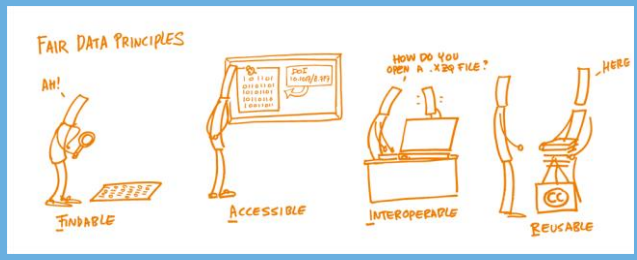
https://www.ugent.be/en/research/openscience/datamanagement/during-research/documentation.htm
https://uclouvain.be/fr/universite-numerique/rdm/document-your-data.html

52

Data sharing check list

Check list to prepare your data for sharing according to FAIR principles :

- ✓ The **datasets** to be shared are defined
- ✓ The **legal, ethical and policy** requirements are respected (GDPR, Funder Grant Agreement, Consortium agreement, …)
- ✓ Files are well **organised** and **named**
- ✓ Files are in sustainable and **open formats**
- ✓ An appropriate **license** has been chosen (re-use)
- ✓ **Access procedures** have been defined
  - Open, closed or restricted access ? Embargo ?
  - What is the authentication procedure?
  - If necessary, has a data sharing agreement been drawn up?
- ✓ Data distribution **rights have been obtained**
- ✓ The data are described and **documented**

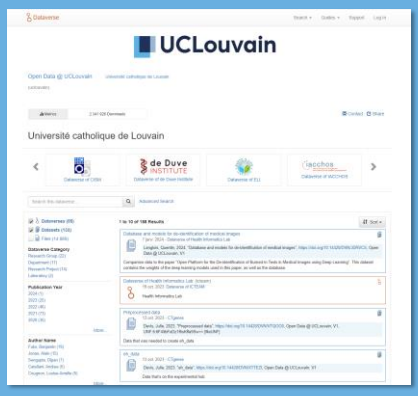


https://book.fosteropenscience.eu/

Sur la base de : DORANum.fr « Où déposer vos données ? », présentation « Le dépôt des données ». DOI : 10.13143/N61E-B629
https://callisto-formation.fr/course/view.php?id=144#section-4

53

Dataverse UCLouvain



https://dataverse.uclouvain.be/

What ?	**UCLouvain Open Data Infrastructure**
Who is it for ?	A service for **UCLouvain's researchers**
Why ?	Preserving and sharing research data in **Open or FAIR mode**
Where ?	Data are stored on a **specific server of UCLouvain** managed by UCLouvain
How much ?	Each research entity that requests it will have a **free space of 1 Tb** (> 1T = financial contribution)
How long ?	All data submitted are guaranteed to last **for 15 years** (> 15 years = extension = financial contribution)

https://uclouvain.be/fr/universite-numerique/opendata.html

54

Dataverse UCLouvain

- **Preserve** data in sustainable way : guarantee to last for 15 years (possibility of extension)
- **Reference** and locate data : DOI
- Allows to **quote** your dataset
- Make data (and researcher !) **visible** : "put a flag" on research and data
- **Share** data in open or FAIR mode
- Specify the **degree of data openness**
- Specify **terms of access** (if restricted access)
- **Creative Commons License** (CC BY-SA by default) or specify **conditions of data re-use**
- Possibility of specifying an **embargo** period
- **Secure** data : specific server of UCLouvain
- Free space of **1Tb** (for each entity)

UCLouvain Dataverse ➔ "trusted repository"

55

Research Data Management

Data
Management
Plan

56

Research Data Management

- **Data Management Plan (DMP)**

✓ The data management plan is a **management tool**. Its purpose is to summarize the **description and evolution of the data sets** in your research project.

✓ The DMP considers **every steps of research data lifecycle** => data management **during and after** the research project.

✓ It prepares your data for sharing, re-use and long-term preservation.

✓ The DMP is continually **updated**, it's a dynamic document.

DoRANum. Données de la recherche : apprentissage numérique [En ligne]. France : DoRANum; 2023 MAJ le 23/08/2023. Le Plan de Gestion de Données pas à pas [consulté le 15/01/2024]. Nous traduisons
https://doranum.fr/plan-gestion-donnees-dmp/le-plan-de-gestion-de-donnees-pas-a-pas_10_13143_t94g-9j96/

57

Data Management plans

DMPonline

1. Open source software
2. Developped by the Digital Curation Center (DCC, UK).
3. Can be shared and edited by our (international) research partners, via ORCid
4. Up to date – Now also RGPD registry.

58

Data Management plans

Data are the **core part** of all research projects : important to manage data carefully

Some examples:

- ✓ Increasingly required by **funders** (Horizon Europe, ERC, FWO, Belspo, FNRS, etc.)
- ✓ Research **proposal**
- ✓ Often seen as an additional administrative load with limited importance

»» But many advantages

59

Data Management plans - Advantages

1. Backbone of projects: will guide all its organization
2. Research partners can always refer to it
   =it provides a common, written understanding of every step of the project.
3. By setting everyone's responsibility it helps to deal with a researcher's leave.
4. Written at the beginnings of the project, but can always been upgraded
5. Save time
   a) refer to it later: procedures just have to be followed
   b) useful basis to write reports, or methods in a paper.

60

Writing a Data Management Plan

- The data management plan is a **management tool**.
- The DMP is made up of a **series of questions** that works as a checklist of attention points to guide the researcher.
- More and more funders **require a DMP** to be drawn up on the basis of a template (UE Commission, FNRS, …). There is also a **template UCLouvain.**

FNRS Template :
1. Data collection / description
2. Data documentation and data quality
3. Data storage and back-up
4. Ethical and legal requirements
5. Data sharing and preservation
6. Responsibilities and resources

61

fnrs
LA LIBERTÉ DE CHERCHER

DATA MANAGEMENT PLAN (DMP)

The model for the DMP proposed below is based on the model favoured by Science Europe in the document "Practical guide to the international alignment of Research Data Management" published in January 2021, in a simplified form.

None of the questions below are **mandatory**, so researchers are free to answer all or only some of them, if they feel that their project does not require it. They also retain the possibility of using another model of DMP, if they consider it more appropriate for their research project.

1. **Data description and collection or re-use of existing data**
   a. How will new data be collected or produced and/or how will existing data be re-used?
   b. What data (for example the kind, formats, and volumes), will be collected or produced?
2. **Documentation and data quality**
   a. What metadata and documentation (for example the methodology of data collection and way of organising data) will accompany the data?
   b. What data quality control measures will be used?
3. **Storage and backup during the research process**
   a. How will data and metadata be stored and backed up during the research?
   b. How will data security and protection of sensitive data be taken care of during the research?
4. **Legal and ethical requirements, codes of conduct**
   a. If personal data are processed, how will compliance with legislation on personal data and on security be ensured?
   b. How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable?
   c. What ethical issues and codes of conduct are there, and how will they be taken into account?
5. **Data sharing and long-term preservation**
   a. How and when will data be shared? Are there possible restrictions to data sharing or embargo reasons?
   b. How will data for preservation be selected, and where data will be preserved long-term (for example a data repository or archive)?
   c. What methods or software tools are needed to access and use data?
   d. How will the application of a unique and persistent identifier (such as a Digital Object Identifier (DOI)) to each data set be ensured?
6. **Data management responsibilities and resources**
   a. Who (for example role, position, and institution) will be responsible for data management (i.e. the data steward)?
   b. What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?

➔ **Example :**
F.R.S.-FNRS Data Management Plan

62

DMPonline

https://dmponline.be/

Templates

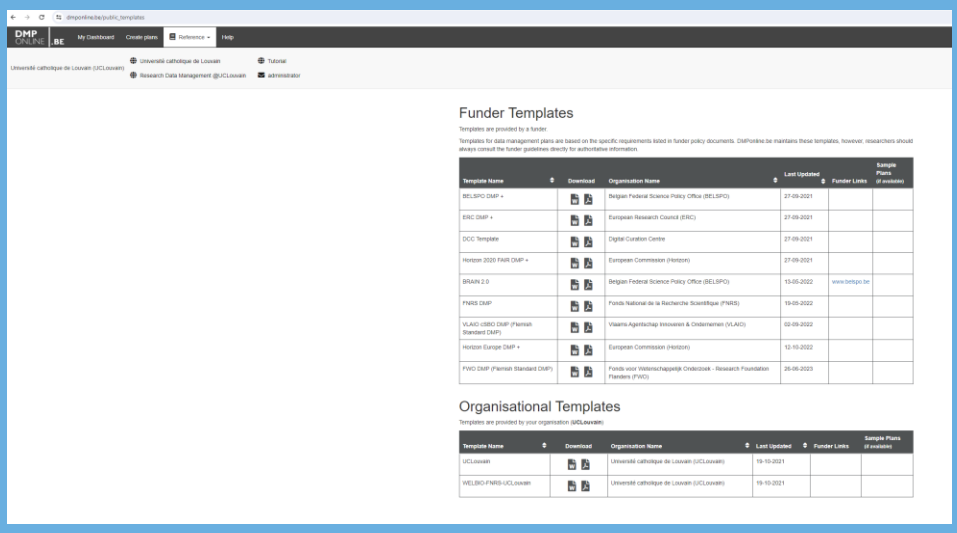
- A set of questions stating on the data processing/data life cycle
- Main templates available are: Horizon europe, ERC, FWO, Belspo, UCLouvain

Guidance

- Helping researchers to answer questions
- Suggesting sample answers
- DCC provides a sample guidance
- Could also be customized by each member for specific uses (UCLouvain guidance)

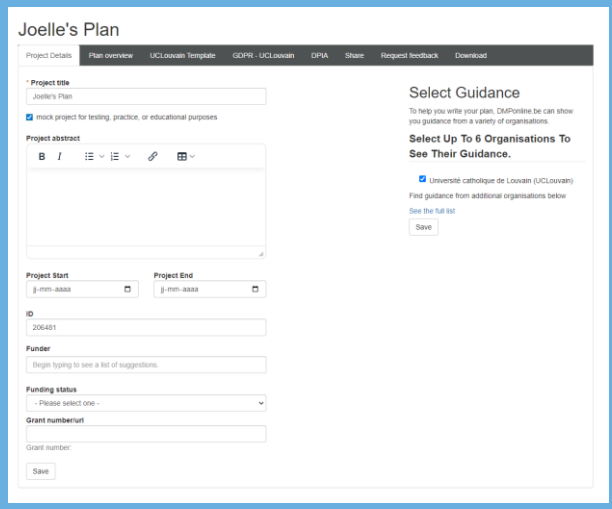
63

DMPonline – Funder Templates



64

DMPonline – Project details



DMP Online "Guidance"

- Help researchers to write the plan
  - Help to answer the questions
  - Suggest examples of answers
- Guidance from a variety of organisations (here UCLouvain)

65

DMPonline – Plan overview

Three parts :

- **Data management plan** (UCLouvain templates here)
- General Data protection Regulation **(GDPR)**
- Data Protection Impact Assessment **(DPIA)**

66

Data management orientation questionnaire (Compass to RDM)

What is it ?

- **Interactive questionnaire** to support researchers drafting their DMP
- Drawn up as part of the **COARA project** in the Wallonia-Brussels Federation (Work package 2 "Open Science", coord. ULiège)
- Currently in a state of development (prototype)

What is the objective ?

- It will help the researcher to **comply with** the institutional and legal **requirements, and recommendations** that apply to his/her data.
- It will the researcher **optimise the overall management** of his/her data throughout the research lifecycle".

67

Data management orientation questionnaire (Compass to RDM)

How does it work ?

- By answering a **few short questions** about your data, you will be offered **personalised recommendations** based on the answers you provide.
- You will be able to :
  - **identify** the data management **requirements** (such as copyright or GDPR) and recommendations that apply to your case,
  - and **discover the resources** at your disposal, such as guides or support staff members.

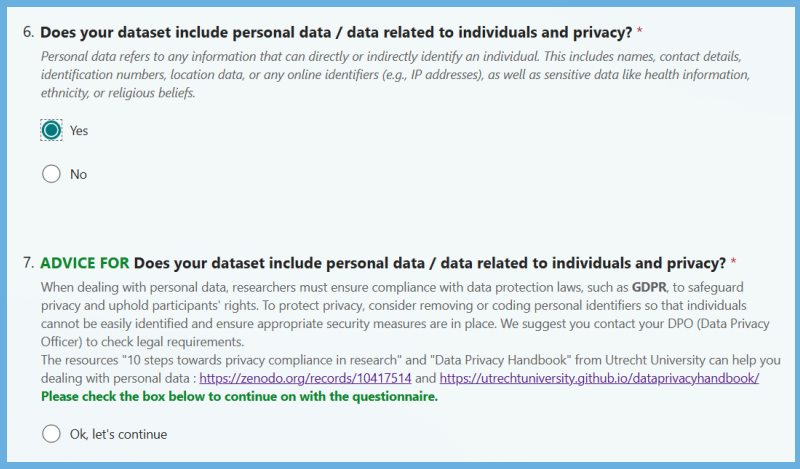
At your disposal @ :
- ULiège
- UCLouvain
- ULB
- UNamur
- UMons

https://tinyurl.com/dataambctrdm

68

Data management orientation questionnaire (Compass to RDM)

- **Identify** the data management **requirements and recommendations** that apply to your case
  - ✓ **Legal requirement (such as GDPR)**



6. **Does your dataset include personal data / data related to individuals and privacy?** *

Personal data refers to any information that can directly or indirectly identify an individual. This includes names, contact details, identification numbers, location data, or any online identifiers (e.g., IP addresses), as well as sensitive data like health information, ethnicity, or religious beliefs.

- (●) Yes
- ( ) No

7. **ADVICE FOR Does your dataset include personal data / data related to individuals and privacy?** *

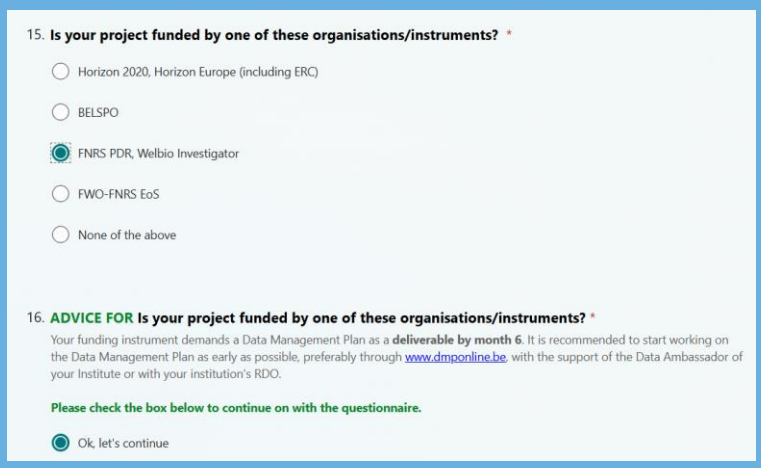
When dealing with personal data, researchers must ensure compliance with data protection laws, such as **GDPR**, to safeguard privacy and uphold participants' rights. To protect privacy, consider removing or coding personal identifiers so that individuals cannot be easily identified and ensure appropriate security measures are in place. We suggest you contact your DPO (Data Privacy Officer) to check legal requirements.
The resources "10 steps towards privacy compliance in research" and "Data Privacy Handbook" from Utrecht University can help you dealing with personal data : https://zenodo.org/records/10417514 and https://utrechtuniversity.github.io/dataprivacyhandbook/
Please check the box below to continue on with the questionnaire.

- ( ) Ok, let's continue

69

Data management orientation questionnaire (Compass to RDM)

- ✓ **Or requirement from your Funder**



15. **Is your project funded by one of these organisations/instruments?** *

- ( ) Horizon 2020, Horizon Europe (including ERC)
- ( ) BELSPO
- (●) FNRS PDR, Welbio Investigator
- ( ) FWO-FNRS EoS
- ( ) None of the above

16. **ADVICE FOR Is your project funded by one of these organisations/instruments?** *

Your funding instrument demands a Data Management Plan as a **deliverable by month 6**. It is recommended to start working on the Data Management Plan as early as possible, preferably through www.dmponline.be, with the support of the Data Ambassador of your Institute or with your institution's RDO.

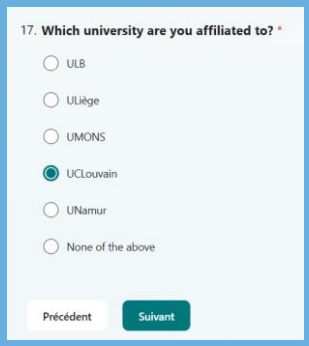
Please check the box below to continue on with the questionnaire.

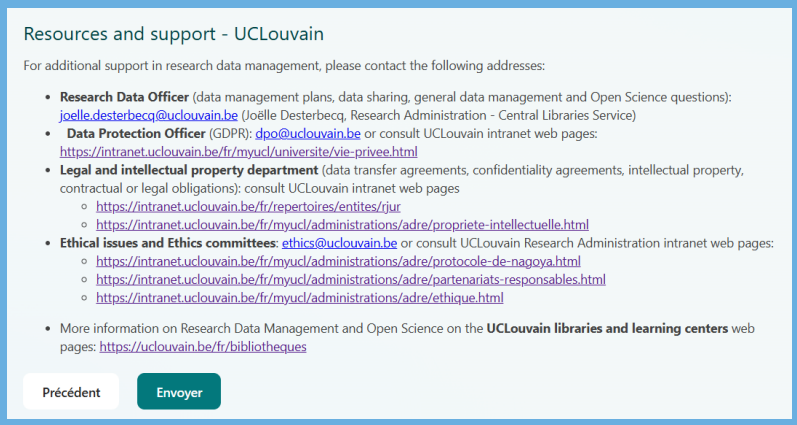
- (●) Ok, let's continue

70

Data management orientation questionnaire (Compass to RDM)

- **Discover the resources** at your disposal, such as guides or support staff members





71

Exercise
Presenting your data & writing a DMP

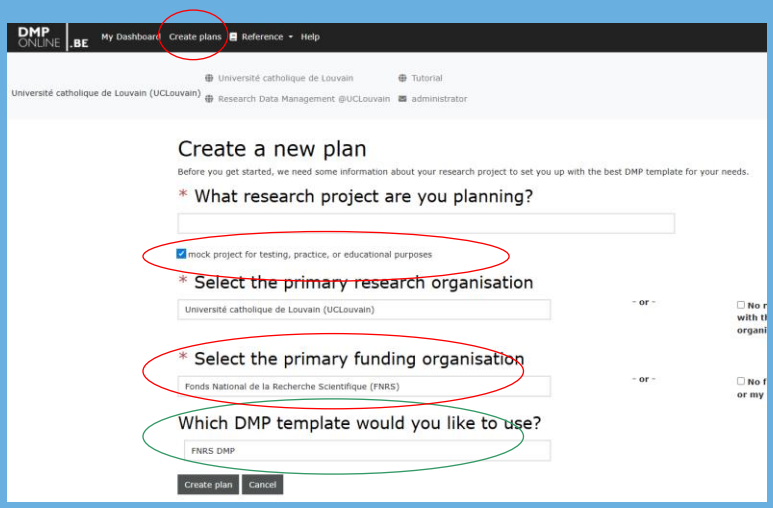
https://book.fosteropenscience.eu/

72

Writing a DMP : it's up to you !

- Go to https://dmponline.be/
- Log in (UCLouvain id. Or ORCID)
- Click on "Create plan"
- Select the funding organisation: FNRS
  - The **"FNRS DMP template"** is displayed automatically

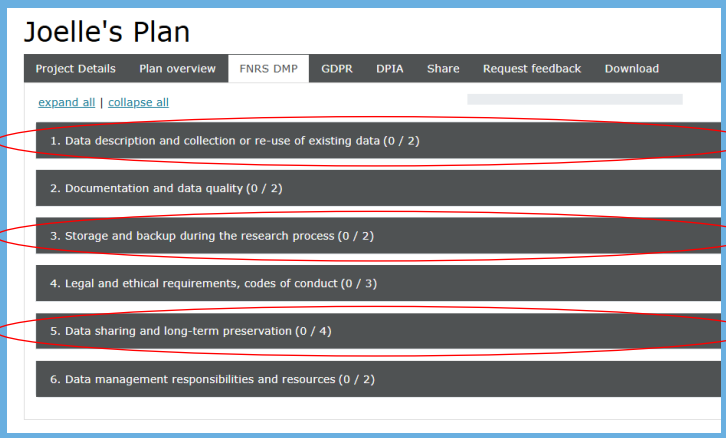
Please **tick the box** "mock project for testing, practice, or educational purposes"



73

Writing a DMP : it's up to you !

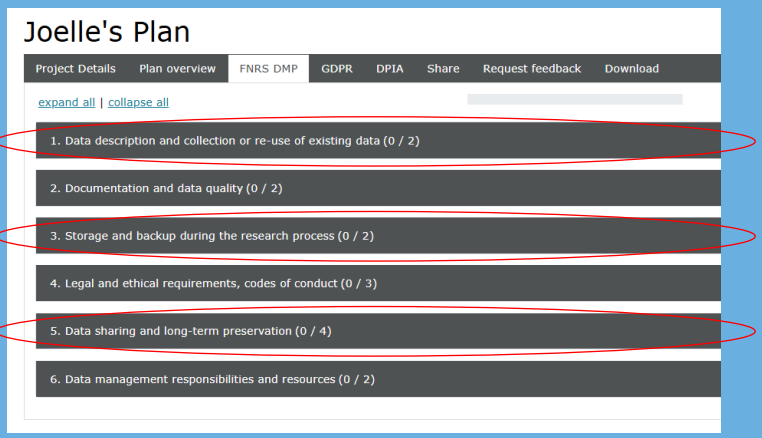
- Think about strategy for managing your PhD Data
- And complete (just a draft) the following parts of the DMP :
  - **1. Data description and collection or re-use**
  - **3. Storage and back-up during the research process**
  - **5. Data sharing and long-term preservation**
- Then, we will have a collective debriefing



74

Writing a DMP : it's up to you !

- **It's up to you ! Let's go !**
- **25 minutes**



75

fnrs
LA LIBERTÉ DE CHERCHER

DATA MANAGEMENT PLAN (DMP)

The model for the DMP proposed below is based on the model favoured by Science Europe in the document "Practical guide to the international alignment of Research Data Management" published in January 2021, in a simplified form.

None of the questions below are **mandatory**, so researchers are free to answer all or only some of them, if they feel that their project does not require it. They also retain the possibility of using another model of DMP, if they consider it more appropriate for their research project.

1. **Data description and collection or re-use of existing data**
   a. How will new data be collected or produced and/or how will existing data be re-used?
   b. What data (for example the kind, formats, and volumes), will be collected or produced?
2. **Documentation and data quality**
   a. What metadata and documentation (for example the methodology of data collection and way of organising data) will accompany the data?
   b. What data quality control measures will be used?
3. **Storage and backup during the research process**
   a. How will data and metadata be stored and backed up during the research?
   b. How will data security and protection of sensitive data be taken care of during the research?
4. **Legal and ethical requirements, codes of conduct**
   a. If personal data are processed, how will compliance with legislation on personal data and on security be ensured?
   b. How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable?
   c. What ethical issues and codes of conduct are there, and how will they be taken into account?
5. **Data sharing and long-term preservation**
   a. How and when will data be shared? Are there possible restrictions to data sharing or embargo reasons?
   b. How will data for preservation be selected, and where data will be preserved long-term (for example a data repository or archive)?
   c. What methods or software tools are needed to access and use data?
   d. How will the application of a unique and persistent identifier (such as a Digital Object Identifier (DOI)) to each data set be ensured?
6. **Data management responsibilities and resources**
   a. Who (for example role, position, and institution) will be responsible for data management (i.e. the data steward)?
   b. What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?

76

Data Management Plan – Debriefing

77

Writing a DMP – Debriefing

- **Who has previously completed a DMP ?**

In the DMP exercise :

- Have you clearly identified **what needs to be explained** in each question (parts 1, 3 and 5)?
- Were you able to provide **answer to each question** ? (Regardless of the time allowed)
- Could you explain **what types of data** are you working with?

78

Writing a DMP – Debriefing

- **How** did you answer the following questions ?
  - 1. DATA DESCRIPTION AND COLLECTION OR RE-USE OF EXISTING DATA
    What data (for example the kind, formats, and volumes), will be collected or produced?
  - 3. STORAGE AND BACKUP DURING THE RESEARCH PROCESS
    How will data and metadata be stored and backed up during the research?
  - 5. DATA SHARING AND LONG-TERM PRESERVATION
    How and when will data be shared? Are there possible restrictions to data sharing or embargo reasons?

79

Writing a DMP – Debriefing

- What did you find **most difficult**? Why ?
  - What did you miss to be able to answer this question ?
- What did you find **most easy** ? Why ?
- Did you use **DMP Online guidance** ?
- Did you find it easy or difficult to **think about a strategy** to manage your data ?
- And now : do you feel **better equipped** to manage your data and complete a DMP ?

80

Useful information

Websites :
https://uclouvain.be/fr/universite-numerique/rdm
https://uclouvain.be/fr/bibliotheques

DMP Online :
https://dmponline.be/

Dataverse :
https://dataverse.uclouvain.be/

Any question ? Feel free to contact :
- The Data Ambassador of your Institute
- The Research Data Officer of your University
  ➔ UCLouvain : joelle.desterbecq@uclouvain.be
- Your Research Administration

81

UCLouvain

Introduction to Research Data Management & Data Management Plans

Thank you !

82