

Adeline Grard – Service Central des Bibliothèques

Introduction to Research Data Management & Data Management Plans



MENU

Part 1

Research Data Management

What is it?

What are the main steps?

How can it help me as a researcher?

Part 2

Data Management Plans

What is it?

What are the main steps?

How can it help me as a researcher?

Research Data Management

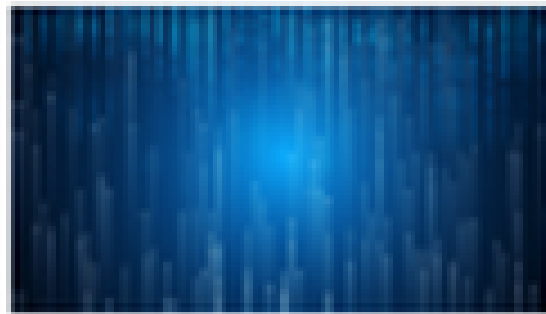
Create, organize, make accessible, store, and share research data of an institution.



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

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 =
reproducibility**



Open access



Meet researchers needs



**Research support to
reach excellence in research**

Fair vs Open Data

FAIR DATA PRINCIPLE

Findable: you can locate it (doi)

Accessible: you can access it (repository)

Interoperable: interoperable: it follows standard of metadata and are in a standard format

Reusable : by researchers, with the proper documentation to understand its content



RDM in the Research life cycle



Search or collect data

Use secondary data or collect your own



Why to use secondary data?

Using data collected by others researchers is very interesting for your 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 your new hypothesis to them).

**Compare /
discuss**

Compare or discuss your research results with similar data, collected in other time/places, or with different methods.



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. Mendeley data*: <https://data.mendeley.com/datasets>
3. OpenAire: <https://explore.openaire.eu/search/find>
4. Re3data: <https://www.re3data.org/>
5. Zanran: <http://www.zanran.com/q>
6. Datasearch*: <https://datasearch.elsevier.com/#/>
7. Google*: <https://toolbox.google.com/datasetsearch>



Data collection: some best practices to ensure quality



1. Calibration of instruments to check the precision, bias and/or scale of measurement
2. Taking multiple measurements, observations or samples
3. Checking the truth of the record with an expert
4. Using standardized methods and protocols

Data Management

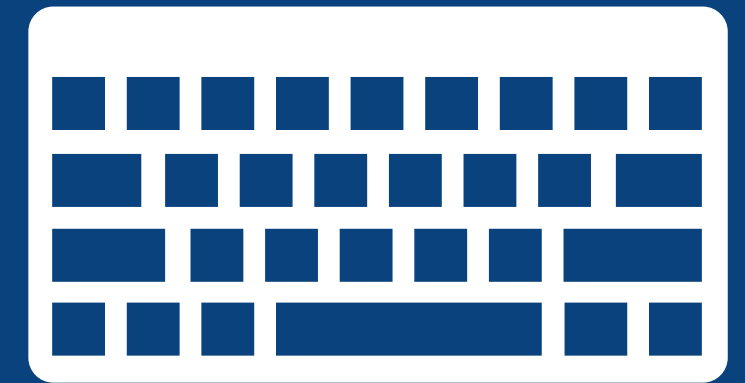
Enter, check, clean, organize,
and document your data

Some best practices from SMCS



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
- Designing a purpose-built database structure to organize data and data files

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
- Double entry of data
- Statistical analyses such as frequencies, means, ranges or clustering to detect errors and anomalous values
- Peer review



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



Data organization

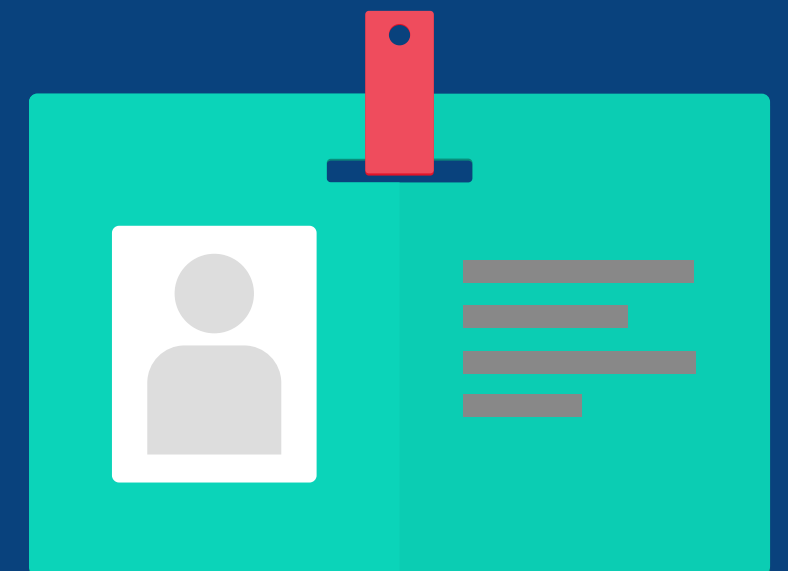
- Use folders - and structure folders hierarchically
- Adhere to existing procedures
- Name folders appropriately - after the areas of work not after researchers
- Agree and be consistent – once you have decided on a method, you stick to it.
- Separate ongoing and completed work - move files regularly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



File naming

Useful file names are consistent, meaningful to you and your 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



Contact smcs if you need help in managing and analyzing your data

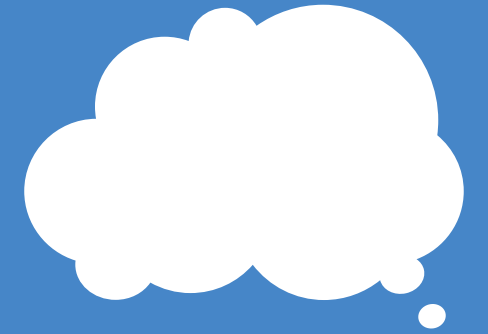
<https://sites.uclouvain.be/smcs-gateway/>

Store & protect

Storage solutions and best practices, ethical and patents resources



Data storage- best practices



1. Avoid to store your data on your laptop
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.

Store in practice

1. Use Service fichiers (OASIS), Nextcloud (with outsiders, 1 GB up to several TB), or SharePoint (<250Mo) + (OneDrive for short time period).
 - a. your data will never be deleted, security guarantees are trustful
 - b. allow all authorized people to access, write, and download your dataset (external on request to SGSI)

2. Transfer large dataset to trusted peers, you can also use transvol

3. Don't use Dropbox, Google Drive (or any other cloud solution) for your research data (seen by others, data property issue).

More information: <https://uclouvain.be/fr/universite-numerique/rdm/store-uclouvain.html>

Patents

- The Louvain Transfer Technology Office (LTTO) helps you managing intellectual property rights and support research knowledge transfer:
<https://uclouvain.be/fr/chercher/litto/services.html>
- Depending on your contract/patent, there might be a joint property rights, and/or a limitation on what you can do with your data (what data, with what embargo period, etc.).



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



Select your data for publication

1. Select data you must publish, and delete those you have to (consortium 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.

Choose a repository

1. Disciplinary repository (re3data.org)
2. UCLouvain dataverse: <https://dataverse.uclouvain.be/>
3. Catch-all repository Zenodo, maintained by CERN



How to select a good repository?

- Persistent identifier (DOI)?
- Is long-term preservation guaranteed?
- Costs per dataset or gigabyte?
- Physical storage location of data?
- License?
- Certified (Data Seal of Approval or CoreTrustSeal)

Publish a data paper

= a scientific article that describe the data you've produced during your research projects, and the management you've done.

- Published on a specific Data Journal, or in disciplinary journals
- check whether this journal is peer reviewed.
- useful to refer to a specific/innovative research design, data collection process or management procedures.



License your data or make a DSA

Creative Commons licenses.



Ready-to-use licenses that are known worldwide. This specifies exactly

- what people can and cannot do with your data (modify, replicate, etc)
- to what purpose data can be used (teaching, research, commercial, etc.)
- if people should give you credit when they use your data (cite your name when used, or wrote about your data).

See also: <https://creativecommons.org/licenses/?lang=en>

In some cases a Data Sharing Agreement might be useful (transfer of sensitive data to researchers or third parties): contact ADRE (marie-anne.crijns@uclouvain.be)

Choose an embargo period



= Limited period of time after which your data will be available.

- stated in your consortium agreement, funders' contract, patent, etc.
- Several reason for delaying your data release (only you and your partners might publish using your data).
- Consider how long your embargo period will last (if any) and state a clear release (date).
- Communicate this embargo period with the repository you chose

Get visible and promote

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



Get visible and promote your RDM skills

Link your data with papers: Dial, data citation in papers

Spread the news on social media: include links to data and paper (Dial + repository, never place paper/ data on research social media (Researchgate))

Researchers that will use your data **will cite your dataset** and you will receive credits for it.

Promote your RDM skills: they are valuable assets for employers (academic or not), but also to describe your research environment (research proposal)

Check our webpage to promote them: <https://uclouvain.be/fr/universite-numerique/rdm/checklist-of-rdm-skills-cv-linkedin.html>

RDM Advantage 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)
6. Get cited for your data (DOI)
7. Visibility
8. Transparency (reproducibility)
9. Get more from your data (other researcher): better return on investment
10. Helps YOURSELF to reuse your previously acquired data
11. Store and backup safely
12. Merge datasets and start new research projects
13. Sometimes it's mandatory

Data Management Plans



Data Management Plans

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

- Increasingly required by **funders** (H2020, ERC, FWO, Belspo, FNRS, etc.),
- Research **proposal**.
- Often seen as an **additional administrative** load with limited importance

➔ BUT many advantages

Data Management Plans - Advantages for researcher

(1) **Backbone** of project: 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 beginning of the project, but can always be **upgraded**

(5) **Save time** at later stages.

a) refer to it later: procedures just have to be followed.

B) useful basis to write reports, or methods in a paper.

Data Management Plans

DMPonline

- Open source software
- Developed by the Digital Curation Centre (DCC, UK)
- Can be shared and edited by your (international) research partners, via ORCID
- Recently updated - Now also RGPD registry for UCLouvain

DMPonline

Templates

- A set of questions stating on the data processing/ data life cycle
- Main templates available are : H2020, ERC, FWO, BELSPO, UCLouvain
- the templates can be customized to fit specific/internal needs of institutes @ UCLouvain

Guidance

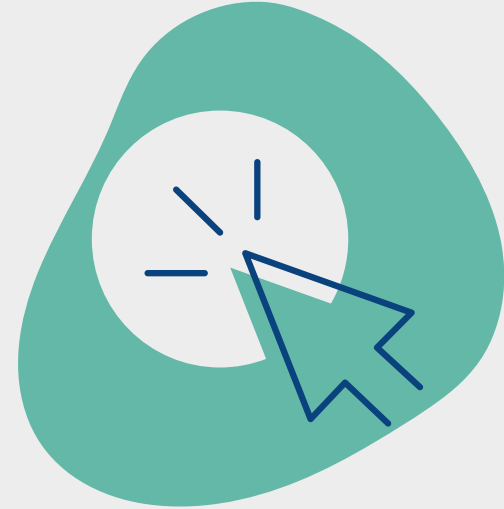
- Helping researchers to answer questions
- Suggesting sample answers
- DCC provides a sample guidance
- Could also be customized by each member for specific uses (UCLouvain guidance)

DMPonline – main steps

<https://dmponline.be>

1. Login with your UCLouvain ID/password
2. Create a plan - select your template (H2020, UClouvain, etc.), select your institution (UCLouvain to receive guidance), Mocj plans
3. Answer question blocks (+ comment for partners)
4. Share with your research partners (enter email) - external log with ORCID
5. Request a feedback
6. Export your plan

Workshops will be organized before UE project deposal,
and can be scheduled too (contact us)



website

<https://uclouvain.be/fr/universite-numerique/rdm>

Useful information



DMPonline

<https://dmponline.be/>
sign with UCLouvain login/password
look at the tutorial

RDM & DMPs @UCLouvain
in a nutshell



Any question?

RDM central email adress: rdm@uclouvain.be
or adeline.grard@uclouvain.be