



Introduction to Research Data Management & Data Management Plans



24 janvier 2023

Jonathan Dedonder
Logisticien de recherche IACCHOS, UCLouvain
Adeline Girard
Illustrations: Midjourney

Presentation

Part 1

Research Data Management

What is it?
What are the main steps?
How can it help me as researcher?

Part 2

Data Management Plans

What is it?
What are the main steps?
How can it help me as researcher?

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 »

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

Fair vs Open Data

- F** FINDABLE: You can locate it (so)
- A** ACCESSIBLE: You can access it (repository)
- I** INTER-OPERABLE: Follows standard of metadata and are in a standard format.
- R** REUSABLE: With the proper documentation to understand its content

RDM in the Research life cycle

Research Life Cycle: Create the project, Collect the data, Analyse the data, Publish, Promote your work

Data Management Cycle: Plan the data management, Search/collect & manage, Store & protect, Publish & share, Get visible & promote

Research Data Management

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 you reserach

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



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

Always check the quality, read metadata and documentantion. Check with an expert.

Research Data Management

Data Management


Enter, check, clean, organize and document your data



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



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



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.

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

For UCLouvain researchers

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

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


Research Data Management

Store & protect



Storage solutions and best practices, ethical and patents ressources

Data storage – best practices



1. Avoid to store 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.)

Store in practice

- 1. Use secure files storage like institutional server (OASIS and Nextcloud in UCLouvain)
- 2. Transfer large dataset to trusted peers (Belnet filesender, UCLouvain: transvol)
- 3. Don't use Dropbox, Google Drive (or any other cloud solution) for your research data (seen by others, data property issue).
- 4. Rq: One drive is a good solution data are stored and backedup safely in the EU, which corresponds to GDPR recommendation.

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

Ethics in research

- ✓ **Beneficence (do good):** research should be conducted for the benefit of individuals taking part in your research, and for broader society and the natural environment.
- ✓ **Non-maleficence (cause no harm):** research should not increase discrimination or expose people to risk - therefore their identity should be protected as far as possible.
- ✓ **Accountability:** An accountable person must be assigned for each research project. This person is answerable to research participants and others, regarding the research conduct. Accountable researchers establish processes and documentation to ensure privacy and confidentiality for research participants.
- ✓ **Transparency:** Research should be conducted in a transparent manner. Research participants should be aware of their participation and how their data is used within it.

The GDPR



- ✓ Data protection is a fundamental human right 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

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.

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

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.

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
- ✓ Data relating to:
- ✓ Criminal convictions
- ✓ Offences
- ✓ Security measures
- ✓ Data concerning children

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

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.

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

Research Data Management

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

Spread the news on social media: include links to data and paper
Careful, link to the repository do not put the paper on the social media (Researchgate)


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

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



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 previously acquired data
10. Store and backup safely
11. Merge datasets and start new research projects
12. Sometimes it's mandatory



Research Data Management

Data Management Plan




Data Management plans

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

Some exemples:

- ✓ 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

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 be upgraded
5. Save time
 - a) refer to it later: procedures just have to be followed
 - b) usefull basis to write reports, or methods in a paper.

Data Management plans



DMPonline

1. Open source software
2. Developed 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.

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

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

Templates

1. Login with institutional ID/password
2. Create a plan - select your template (H2020, UCLouvain, etc.), select your institution (UCLouvain to receive guidance), Mocij 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 deposit, and can be scheduled too (contact us)




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

Useful information

RDM & DMPs @UCLouvain
in a nutshell



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



Any question?
RDM central email adress: rdm@uclouvain.be
