Data Management Plans – one tool with many applications
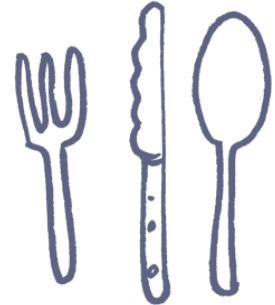
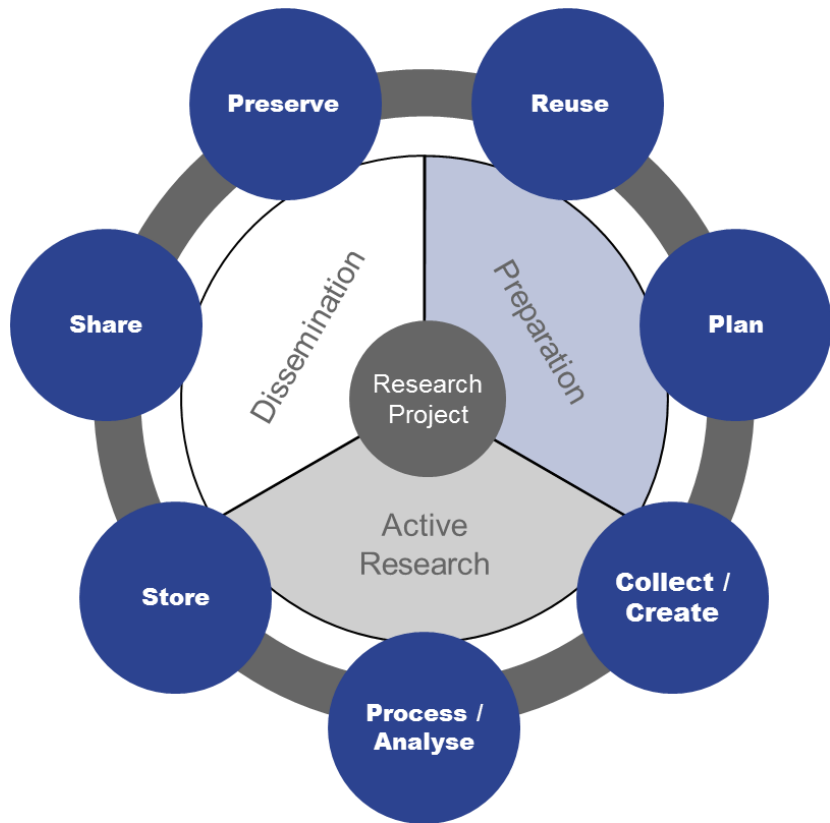Introduction to research data management.

The importance of good research data management.

Plan your research data management.

Write your own data management plan.

Decide what to share and how.

What is research data management?

Diagram showing a research data lifecycle with an outer ring of stages (Preserve, Reuse, Plan, Collect / Create, Process / Analyse, Store, Share) surrounding an inner circle divided into Dissemination, Preparation, and Active Research, centred on "Research Project"
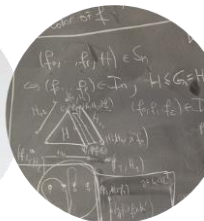
live animals

human participants

published collections

physical samples

theoretical models

textual data

observations

measurements

statistics

Why is good research data management important?

- A proper **plan** reduces risks of problems halfway & aligns expectations

- Thorough **documentation** helps finding & understanding data in the future

- Appropriate **storage** and **security** prevents data loss and leaks

- **Publication** of data makes research more visible

- **Preservation** of data will ensure that you can provide evidence for your published results, if necessary

Retraction notice

Retraction notice to A model study into the effects of light and temperature on the degradation of fingerprint constituents [Science and Justice, 54 (2014) 346 - 350]

Belén González Amorós, M. de Puit

Show more

Share Cite

https://doi.org/10.1016/j.scijus.2015.04.005

Refers to Belén González Amorós, M. de Puit

RETRACTED: A model study into the effects of light and temperature on the degradation of fin...
Science & Justice, Volume 54, Issue 5, September 2014, Pages 346-350

Download PDF

Previous article in issue   Next article in issue

This article has been retracted: please see Elsevier Policy on Article Withdrawal (http://www.elsevier.com/locate/withdrawalpolicy).

This article has been retracted at the request of the authors. The authors identified a inconsistency in the accepted paper and were unable to reproduce the average values that were used for the graphs and tables in the paper, due to the loss of the raw data. This, in turn, means that the authors cannot fulfil the demands of the Association of Dutch Universities and the Royal Dutch Academy of Science in respect to their ethical and research data standards.

View Abstract

Why is good research data management important?

- ❑ **Communities** have traditions for sharing data, code and protocols openly
- ❑ **Funders** might require you to make your results publically available
- ❑ **Publishers** may ask you to provide data to peer reviewers and readers
- ❑ **Institutions** need to ensure information security
- ❑ **Policy makers** want to promote data reuse

⇒ increasing number of **policies and guidelines**

https://allea.org/code-of-conduct/

http://data.europa.eu/eli/dir/2019/1024/oj

L 172/56   EN   Official Journal of the European Union   26.6.2019

DIRECTIVE (EU) 2019/1024 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL

of 20 June 2019

on open data and the re-use of public sector information

(recast)

integrity | in

1 the quality of being h

integrity.

of being w

**The European
Code of Conduct for
Research Integrity**
REVISED EDITION

Why is good research data management important?

HORIZON 2020

OPEN RESEARCH DATA IN HORIZON 2020

AS OPEN AS POSSIBLE, AS CLOSED AS NECESSARY

Grantees have the right to opt-out, but need to say why

HORIZON 2020 GRANTEES ARE REQUIRED

take measures to ensure open access to the **data underlying their scientific publications**

provide open access to **any other research data of their choice**

Horizon 2020 grantees are **encouraged to also share datasets** beyond publication

PROJECTS MUST HAVE

DATA MANAGEMENT PLAN (DMP)

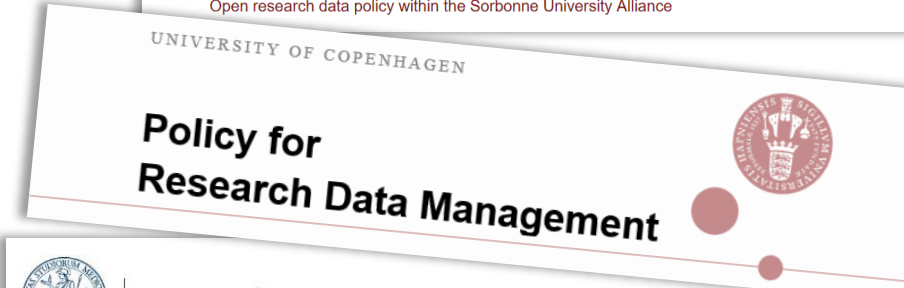Provides information on:

the data the research will generate

how to ensure its curation, preservation and sustainability

what parts of that data will be open (and how)

Open for you! An introduction series to open science

Why is good research data management important?

How do you plan research data management?

- Prepare a data management plan (DMP) at project start.

- Define and classify your research data.

- Identify potential risks and challenges.

- Assign roles and responsibilities.

- Align expectations with your supervisor and collaborators.

Templates available from funders, universities and research organisations.

Horizon Europe

Data Management Plan Template

Version 1.0
05 May 2021

Researcher Guidance for Data Management Plans

TEMPLATE FROM THE SCIENCE EUROPE PRACTICAL GUIDE TO
THE INTERNATIONAL ALIGNMENT OF RESEARCH DATA MANAGEMENT

D | C | C

Checklist for a Data Management Plan, v4.0

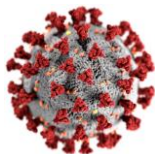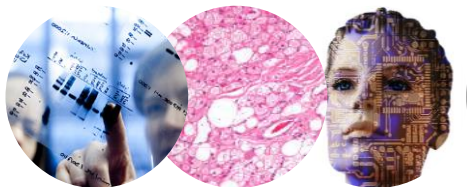**Please cite as:** DCC. (2013). *Checklist for a Data Management Plan*. v.4.0. Edinburgh: Digital Curation Centre. Available online: http://www.dcc.ac.uk/resources/data-management-plans

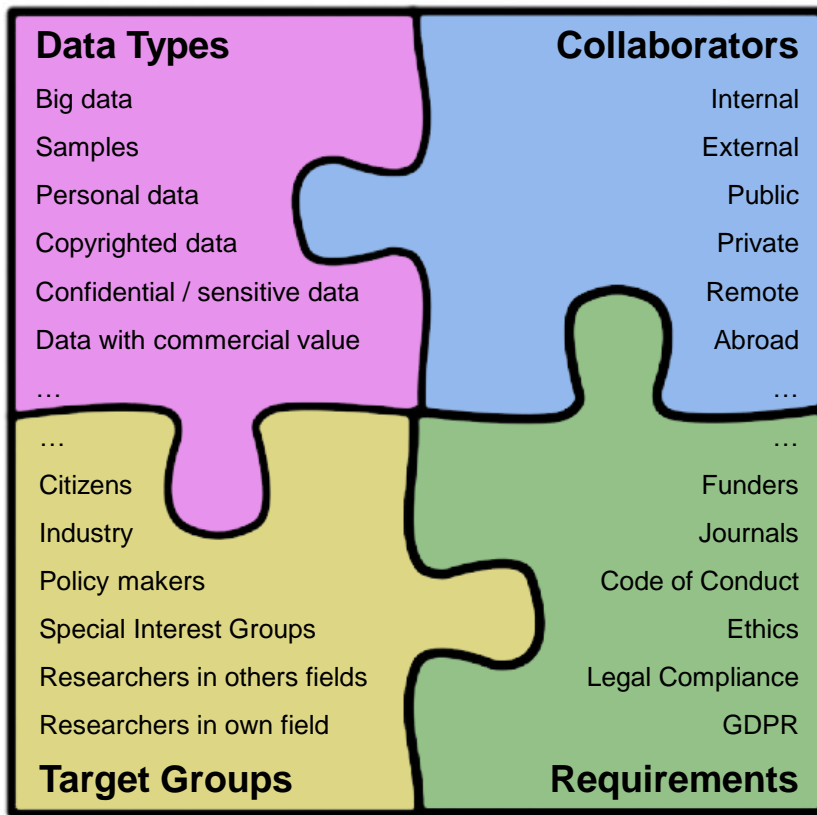| DCC Checklist	DCC Guidance and questions to consider
Administrative Data	
ID	A pertinent ID as determined by the funder and/or institution.
Funder	State research funder if relevant
Grant Reference Number	Enter grant reference number if applicable [POST-AWARD DMPs ONLY]
Project Name	If applying for funding, state the name exactly as in the grant proposal.
Project Description	**Questions to consider:** - What is the nature of your research project? - What research questions are you addressing? - For what purpose are the data being collected or created? **Guidance:** Briefly summarise the type of study (or studies) to help others understand the purposes for which the data are being collected or created.

How do you plan research data management?

- Costs
- Storage & Backup
- Information & IT Security

- Publication
- Communication
- Preservation

Data Types
Big data
Samples
Personal data
Copyrighted data
Confidential / sensitive data
Data with commercial value
…

Collaborators
Internal
External
Public
Private
Remote
Abroad
…

Target Groups
…
Citizens
Industry
Policy makers
Special Interest Groups
Researchers in others fields
Researchers in own field

Requirements
…
Funders
Journals
Code of Conduct
Ethics
Legal Compliance
GDPR

- Workflows
- Access
- Rights & Responsibilities

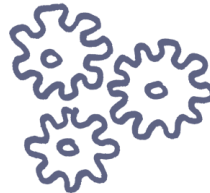- Documentation
- Retention
- Practices

How do you plan research data management?

- Draft a first version of the DMP to outline your strategy.
- Update the DMP along your project with more details.
- Use the DMP as guidelines for data management in your project.
- Discuss and review the DMP regularly with your supervisor and collaborators.
- Keep the DMP along with other project documentation.

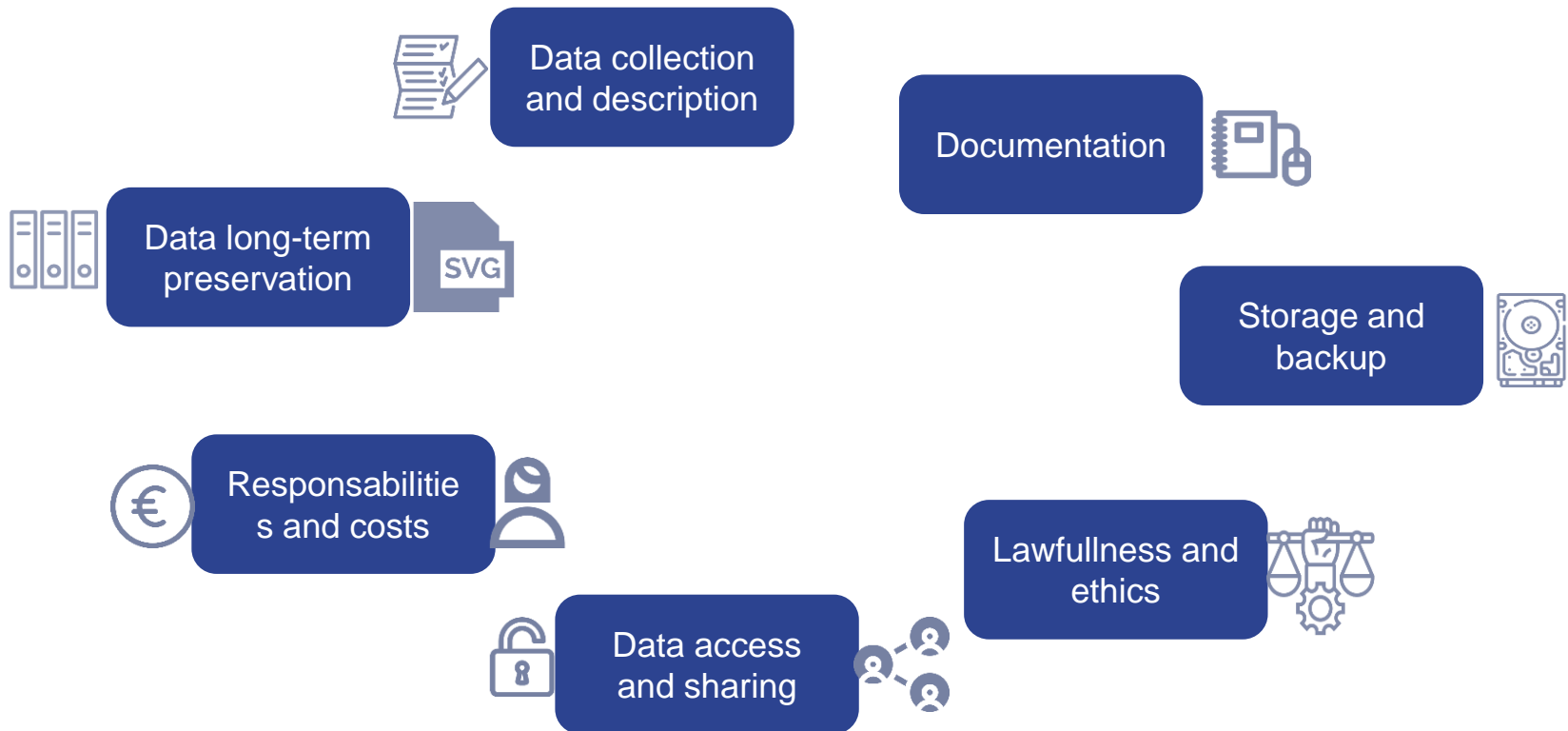**Your DMP will help you to achieve better**   **security**   **quality**   **time-efficiency**   **cost-efficiency**   **impact**

DMP in brief

Data collection and description

Documentation

Data long-term preservation

SVG

Storage and backup

Responsabilities and costs

Lawfullness and ethics

Data access and sharing

Data description
Collection or re-use of existing data

Objectives

- Identify and list the **types of data sets**
- Specify the **purpose of the study, geographic origin, time period**.
- Explain the **purpose of collecting/generating**

In practice

- Explain the **conditions of data collection**
- If possible, use **open file formats**
- **Organize** and **name the files**

Description and collection
Example

✦ Data typology

✦ Files formats

Lumley, Emily. *CompBioMed D3.1_Data Management Plan_v1.0.* 2020, p. 12, https://www.compbiomed.eu/wp-content/uploads/2020/06/D3.1_Data-Management-Plan_v1.0.pdf

"
- **Imaging data** which stem mostly from clinical trials, experiments, or visualize results of the simulations. They may serve as initial data for data analytics and machine learning tasks. Typical file formats used for these data are JPG, PNG, DICOM, MP4, and MOV.
- **Other clinical and experimental data** which serve as reference for simulations. Here a broad variety of file formats are used from formatted/unformatted plain text, PDF and DOCX files, tabulated data formats like CSV and XLSX, as well as (raw) binary data.
- **Musculoskeletal data** which record the motion of joints, bones, muscles, etc. They are the key output of musculoskeletal simulations and use file formats such as C3D and XMDF.
- **Cardiovascular data** resulting from heart and blood flow simulations in the project are mostly recorded in HDF5 and VTK file formats.
- **Molecular modelling data**: Structures of complex biomolecules, assemblies thereof, smaller molecules and molecular dynamics trajectories serve as initial conditions or are output of simulations conducted in the project and may be targets of HDPA approaches. Most frequently used file formats are PDB, PSF, XTC, TRR. A variety of tools is available that are able to read and convert these different formats.
"

Dataset organization
Example

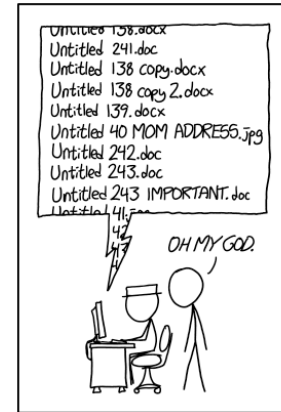Colomb, Julien, Thorsten Arendt, Deepti Mittal, et Keisuke Sehara. 2020. « Folder Structure Template for Research Repositories ». https://doi.org/10.5281/zenodo.4410128

- 📁 template_par
  - 📄 ._LICENSE-CC-BY
  - 📁 01_project_management
    - 📄 ._05_data_management_plans
    - 📄 ._06_notebook
    - 📁 05_data_management_plans
      - 📄 ._DMP_main.txt
    - 📁 06_notebook
      - 📄 ._.gitkeep
  - 📁 04_data_analysis
    - 📄 ._LICENSE-MIT
- 📁 template_par
  - 📁 01_project_management
    - 📁 01_administration_files
      - 📄 .gitkeep
    - 📁 02_accepted_grants
      - 📄 .gitkeep
    - 📁 03_meeting_minutes
      - 📄 .gitkeep
    - 📁 04_related_literature
      - 📄 .gitkeep
    - 📁 05_data_management_plans
      - 📄 DMP_main.txt
    - 📁 06_notebook
      - 📄 .gitkeep
  - 📁 02_material_and_methods
    - 📁 01_protocols
      - 📄 .gitkeep
    - 📁 02_code
      - 📄 .gitkeep
    - 📄 Readme_MM.md
  - 📁 03_data
    - 📁 001_defaultexp
      - 📄 LICENSE
      - 📄 README.md
    - 📁 990_processed_data
      - 📄 .gitkeep
      - 📁 001_defaultexp_processeddata
        - 📄 .gitkeep
    - 📄 Readme_data_store.md
  - 📁 04_data_analysis
    - 📄 000_Readme_data_analysis.md
    - 📁 001_defaultexp_analysis
      - 📄 .gitkeep
    - 📁 990_code_libraries
      - 📄 .gitkeep
    - 📁 991_preregistration
      - 📄 .gitkeep
    - 📄 LICENSE-MIT
  - 📁 05_figures
    - 📁 001_defaultexp_fig
      - 📄 .gitkeep
    - 📁 990_shared_figures
      - 📄 Readme_sharedfig.md
  - 📁 06_disseminations
    - 📁 01_report_conf

Naming conventions
Examples

✦ Identify topics for your files

✦ Abbreviate

✦ Use versioning

✦ Do it as simple as possible

✦ Write down your naming conventions

✦ Examples
  ✦ Project_Topic-yyyy-mm-dd_version
  ✦ Date_Project_identifier-experiment_version

Untitled 138.docx
Untitled 241.doc
Untitled 138 copy.docx
Untitled 138 copy 2.docx
Untitled 139.docx
Untitled 40 MOM ADDRESS.jpg
Untitled 242.doc
Untitled 243.doc
Untitled 243 IMPORTANT.doc
Untitled 41...
42
43
OH MY GOD.

PROTIP: NEVER LOOK IN SOMEONE ELSE'S DOCUMENTS FOLDER.
https://xkcd.com/1459/

Documentation and data quality

Objectives
- Give **all the information** to someone who did not participate in the project so that he can reuse your data

In practice
- Describe your data with **metadata**
- Assign **keywords**
- Add a **readme file** to your dataset

Metadata, readme file and keywords
Example

Open for you! An introduction series to open science

17

Storage and backup during the research process

✦ **Objectives**
- **Secure your data**, especially if they are **sensitive**
- **Prevent the risks** of data loss

✦ **In practice**
- **Keep several copies**
- Use robust storage systems with **automatic backup**
- **Manage access to data**
- **Encrypt** data if it is sensitive
- Use a **secure cloud**
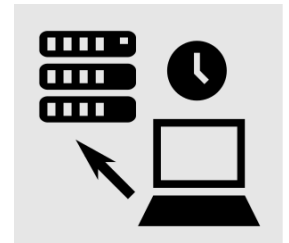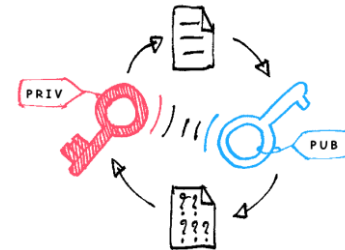
Storage and backup
Example

> The data on the WUR One Drive that a platform are **protected by the data security politic of WUR**. In the same way, the data on INRA dataverse are secured by INRA.
>
> Concerning the LANDMARK data warehouse**, a backup of the SQL databases is made every night** to allow reinstatement following system failure. The outward facing server at the INRA Institute is a virtual machine llows rapid reinstatement following catastrophic failure.

> The data will be stored throughout the project at different scales:
>
> **- first of all on the researchers' computers**.
>
> **- on 4 external hard disks (1 TB)** – one for each WP;
>
> **- in a Sharedocs storage** space provided by TGIR Huma-Num (MONACORALE account opened on 13/04/2021); - in a Network Attached Storage (NAS), which can be consulted at long distance from Europe (access with password).
>
> The External Hard Drives are used for **daily backups**. The Sharedocs is regularly fed by sets. The server and the NAS are **mirrors of the Sharedocs**.
>
> During the data validation meetings, **backups of homogeneous and compiled data batches are stored, after renaming**.

Marchand, Julie. *MONACORALE - MONAsteriorum CORpus Adriaticorum et Locorum Ecclesiasticorum v1(june2021) : Plan de gestion de données.* 2021, https://dmp.opidor.fr/plans/9691/export.pdf
Saby, Nicolas. LANDMARK LAND : FINAL DATA MANAGEMENT PLAN D 2.1. 2019, https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5c9723849&appId=PPGMS

Legal and ethical requirements, codes of conduct

Objectives

- Clarify the **legal framework** and anticipate possible risks
- Verify if **the exploitation of pre-existing data** is possible
- Define if the **re-use of data is: possible, under conditions, or prohibited**

In practice

- Be particularly careful with these data types:
    - **Personal** data, **health** data,
    - Data that raise **ethical issues**
    - Data on which **intellectual and commercial** property rights apply
    - Data that are classified as **defense secrets**
    - Collected **data from websites**
    - …

Ethical and legal framework
Example

> InSPIRES project will perform research in seven different countries: Spain, France, Italy, The Netherlands, Hungary, and, outside the European Union, Bolivia and Tunis. This broad spectrum of countries leads to a deep reflection on ethical issues concerning the research performed. The consortium is aware of the international legislation, guides and codes that regulate management of data:

- The **Nuremberg Code (1947)** addressing volunteer consent and proper acting
- The **Revised Declaration of Helsinki** in its last version of 2013
- The **charter of Fundamental rights of the EU Directive** 95/46/EC (…), protection of individuals (…)
- **Opinions of the European Group of Advisers on the Ethical Implications of Biotechnology** (1991-1997) and the **European Group on Ethics in Science and New Technologies** (as from 1998)
- The **New Brunswick Declaration**: A Declaration on Research Ethics, Integrity and Governance resulting (…), Canada (2013)
- The **Respect Code focused in socio-economic research**.

Villanueva Baselga, Sergio, Anne-Sophie Gresle, et María-Jesús Pinazo. « INSPIRES: Ingenious Science Shops to promote Participatory Innovation, Research and Equity in Science - D8.1: Data Management Plan », 2017. https://inspiresproject.com/wp-content/uploads/2018/03/D8.1_v02_FINAL-VERSION.pdf.

Data sharing and long-term preservation

✦ **Objectives, for data that can be opened:**
- Make data **easy to find and accessible**
- Ensure that data is **reusable**

✦ **In practice**
- Use an **open license**
- Identify a **trusted data repository**
- Use a **persistent identifier (DOI)**
- Share **other research products**: **codes** and **softwares**, **protocols**, etc.

Sharing
Example

" All the data used in the project is made openly available by **publishing it to Github and Zenodo**. Specify how the data will be made available
(...)
The **scripts, documentation and results** are also readable offer the browser or a simple text editor. To **reproduce the results** either docker or python (including pandas and plotly) is needed (pip is recommended to install pandas and plotly).
As descibed before the data including the metadata, documentation and code is deposited to a code repository (Github) as well as a data respository (Zenodo) and **will be available even after the end of the project**. "

Hinterndorfer, David. *Correlation between number of marriages in the EU/austria and students of public universities in austria*. avril 2019, https://zenodo.org/record/2634933#.XdG8lL9Cc0p

- This example uses a multidisciplinary repository.
- If your institution has a repository, use it!

UNIVERSITÀ
DEGLI STUDI
DI MILANO
LA STATALE

heiDATA | Heidelberg Open Research Data

UNIMI Dataverse

Data management responsabilities and resources

✦**Objectives for a team project**

- Define the roles of **each person** in data management
- Anticipate the **costs of data management**
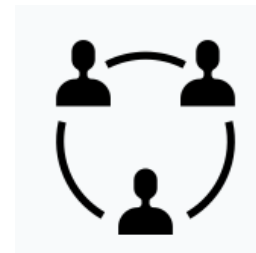
✦**In Practice**

- Appoint a **data management steward**
- Appoint a **person to write the data management plan**
- Estimate the **costs** of **storage**, **sharing** and **archiving**
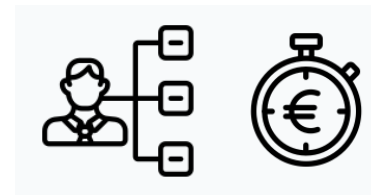
The DMP: a living document

The DMP is a deliverable that evolves:
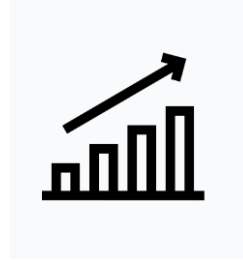
1 expected 6 months after contracting

2 at mid-project

3 at the end of the project

Two most used templates
Horizon Europe

1. Data summary

2. FAIR data

2. 1. Making data <u>findable</u>, including provisions for metadata

2.2. Making data openly <u>accessible</u>

- Repository
- Data
- Metadata

2.3. Making data <u>interoperable</u>

2.4. Increase data <u>re-use</u> (through clarifying licences)

3. Other research outputs

4. Allocation of resources

5. Data security

6. Ethics

Two most used templates
Science Europe

Template

General information

1. DATA DESCRIPTION AND COLLECTION OR RE-USE OF EXISTING DATA

2. DOCUMENTATION AND DATA QUALITY

3. STORAGE AND BACKUP DURING THE RESEARCH PROCESS

4. LEGAL AND ETHICAL REQUIREMENTS, CODE OF CONDUCT

5. DATA SHARING AND LONG-TERM PRESERVATION

6. DATA MANAGEMENT RESPONSIBILITIES AND RESOURCES

How to choose a DMP?

University	Requirements	Template
UMIL	DMP Pilot for PhD students and tutors	Simplified H2020 Funder templates
UHEI	DMP recommended	Templates in RDMO and DMPonline
UW	National Funder: National Science Center	Funder templates
CU	National Funder: Technology Agency of the Czech Republic (KAPPA programme)	None
UCPH	DMP mandatory National Funder: Innovation Fund Denmark (Grand Solutions)	UCPH template
SU	National decree for DMP's	Funder templates

dmponline.dcc.ac.uk

argos.openaire.eu

ds-wizard.org

rdmo.forschungsdaten.info

dmponline.deic.dk

dmp.opidor.fr

Contact us

- Sorbonne University: data-bsu@sorbonne-universite.fr
- Milano: dataverse@unimi.it
- Copenhagen: forskerservice@kb.dk
- Heidelberg: data@uni-heidelberg.de
- Warsaw: oa.buw@uw.edu.pl
- Charles University: researchdata@cuni.cz

What to share?

Research data? Data sets? Data records? Scientific information?

OECD: "research data" are defined as 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. A research data set constitutes a systematic, partial representation of the subject being investigated.

Open for you! An introduction series to open science

30

Research data and DMPs
Funders' requirements

- Europe, Horizon Europe :
  - **Mandatory DMP**
  - FAIR data management
  - **Open data** if possible

- Europe, H2020 :
  - If pilot, mandatory DMP
  - If pilot, open data if possible

What data will be collected, processed and/or generated?

Examples of research data types:

PRIMARY SOURCE	Data that constitute the basic materials of the research (before any processing and analysis).
EXTERNAL SOURCE DATA	Data not produced by the researcher (open access, on demand or reuse). Example: External quantitative databases; corpus and and textual archives collected by other people/institutions, audio/audiovisual archives audio/audiovisual archives; statistics produced by public/private actors, website content
SOURCE DATA COLLECTED	Data collected and/or compiled by the researcher directly or remotely. Example: Database developed from multiple sources, questionnaire/survey database, field survey data (interviews, transcripts, observations, photos, field notes...), textual corpus developed from web data
RESULTS DATA	Data produced after exploitation and processing of source data and which constitute the result data. Example: statistical analysis data, qualitative analysis data, network analysis data, interview extracts, maps, graphs, models and indicators, aggregated numerical data, audiovisual productions, syntheses, notes, annotations
DOCUMENTATION ON THE DATA	Data that provide information on the data itself (methodology, analysis materials used, treatments applied, ownership and protection of data, data valuation). Example: Survey procedure/methodology guide, program and application, analysis grid, dictionary of variables, request for authorization from the confidentiality clauses, metadata (title, keywords, DOI) ...

32

What data will be collected, processed and/or generated?

List (not exhaustive) of shareable data:

- laboratory notebooks, studies, experimental protocols, experimental reports, test reports,
- softwares, codes,
- databases,
- reading notes, notebooks,
- interviews,
- expert's report,
- conferences/seminars: announcement, program, list of speakers, press kits, posters,
- photo, film, 3D, surveys,
- documentation of scientific equipments, methodological guide,...

⟶ Data may be **numerical**, **descriptive**, **visual** or **tactile**. It may be raw, cleaned or processed, and may be held in any format or media.

Methodology and standards

Metadata standards and Research Data Management guidelines:

- 📬 The FAIRsharing portal with information and resources on data standards, databases, and policies in the life sciences and other scientific disciplines.
- 📬 DM guidelines and good practices for the Life Sciences, the Social Sciences and the Humanities provided by relevant research infrastructures, ELIXIR, CESSDA and DARIAH, respectively along with relevant data resources and repositories/databases.
- 📬 For more information on disciplinary metadata standards, visit Digital Curation Centre and Research Data Alliance Metadata Standards Directory

Examples of disciplinary metadata standards : PACTOLS, Generic Earth Observation Metadata Standard (GEOMS), VRA Core, Darwin Core, Ecological Metadata Language (EML)...

Overview on FAIR principles

FAIR principles:
https://www.force11.org/group/fairgroup/fairprinciples

FAIR logo: SangyaPundir. CC: BY-SA 4.0.
Image : ANDS. CC: BY 4,0,

"as open as possible, as closed as necessary"

Access to, and use of, certain research data will necessarily be limited by various types of legal requirements, which may include restrictions for reasons of:

- national security
- privacy and confidentiality
- trade secrets
- legal process

It is always possible to define rules of communicability making the data open after a certain period (depending on national laws), ex: in France, documents related to national security may be opened after 50 years.

Minimum standards are defined in the Tromsø Convention (CETS No. 205)

Open for you! An introduction series to open science

36

Opening up data: what's in my data?

Some ways to secure the collect of personal/health data:

Questions to be ask before collecting personal data

- How long do I really need the data to achieve the objective?
- Do I have legal obligations to keep the data for a certain period of time?
- Do I have to keep certain data in order to protect myself against possible litigation? Which ones?
- What information must be archived? For how long?
- What are the rules for deleting data?
- What are the rules for sharing
- What are the rules for digital preservation?

Some ways to secure the collect of personal/health data:

Document the conformity of data acquisition methods

The documents formalising the internal procedures:
- the register in which the processing activities and the related retention periods must be recorded,
- the various actions undertaken, including whether these actions are still ongoing,
- written instructions to the processor regarding time limits.

Internal procedures:
- Any procedures for sharing, storing or archiving data,
- the procedure for destroying data, if applicable
- describe security measures, including where paper records are stored, to protect data from destruction, loss, alteration, unauthorised disclosure or access. These measures should be appropriate to the risks and the nature of the data.

Open for you! An introduction series to open science

39

Some ways to secure the collect of personal/health data:
Tools

Anonymisation is often preferred to pseudonymisation - > A tool to raise awareness of the risks of re-identification after anonymisation: https://cpg.doc.ic.ac.uk/observatory/

- ✦ Sort through the data,
- ✦ Anonymize,
    - ✦ https://amnesia.openaire.eu
    - ✦ https://github.com/SGMAP-AGD/anonymisation
- ✦ Pseudonymization,
    - ✦ https://github.com/etalab-ia/pseudo_app
- ✦ Delete,
    - ✦ Redaction of a secure data deletion procedure in accordance with the recommendations of the CISO.
    - ✦ Use dedicated software for data deletion without physical destruction that has been audited or certified.
- ✦ Secure through encryption
    - ✦ GnuPG, https://gnupg.org/
- ✦ Name and re-name,
    - ✦ ReNamer, https://renamer.fr.softonic.com/
    - ✦ Ant Renamer, http://www.antp.be/software/renamer/fr

How to share despite access restrictions?

❾ anonymize/ pseudonymize,
❾ share samples,
❾ archive with access rights retention

If data are restricted, **the restriction should not be for an indefinite period**
❾ better to have a specific agreement about the number of years of the closure, e.g. ten or fifty years, or during the lifetime of the individual concerned.

Open for you! An introduction series to open science

41

Licenses
Original works

International :
- ⑨ Creative commons

Europe:
- ⑨ Open Licence Assistant

Where to share your data?

- Multi-disciplinary repository (EU / CERN): Zenodo
- Commercial repositories: Figshare, GigaDB, Mendeley data
- Disciplinary repositories: DRYAD, NCBI, Nakala, Gbif, Seanoe
- Organizations repositories: Edinburgh Datashare, INRAE Datapartage, The Heidelberg Open Research Data , UNIMI Dataverse

DMP as efficient tool for :

- ❾ Improve finding, understanding and using data,
- ❾ Define and classify your research data,
- ❾ Identify potential risks and challenges
- ❾ Align expectations with your supervisor and collaborators

Questions ?

Register for our upcoming events: https://4euplus.eu/4EU-273.html

Next sessions:
- "Predatory publishers and identity fraud – how to identify dubious providers" | 27 January 2022, 15:00 - 16:30
- "Publications strategies for monographs in humanities and social sciences" | 7 Februrary 2022, 10:00 - 11:30
- "Research data management – introduction to fair and open data" | 7 March 2022, 15:00 - 16:30
- 9 further session in 2022

Open for you! An introduction series to open science

Cécile Arènes, Océane Valencia (Sorbonne Université)
Falco Hüser (University of Copenhagen

Open for you! An introduction series to open science

How to choose a DMP?

University	Requirements	Template
UMIL	DMP Pilot for PhD students and tutors	Simplified H2020 Funder templates
UHEI	DMP recommended	Templates in RDMO and DMPonline
UW	National Funder: National Science Center	Funder templates
CU	National Funder: Technology Agency of the Czech Republic (KAPPA programme)	None
UCPH	DMP mandatory National Funder: Innovation Fund Denmark (Grand Solutions)	UCPH template
SU	National decree for DMP's	Funder templates

dmponline.dcc.ac.uk

argos.openaire.eu

ds-wizard.org

rdmo.forschungsdaten.info

dmponline.deic.dk

dmp.opidor.fr

Contact us

- Sorbonne University: data-bsu@sorbonne-universite.fr
- Milano: dataverse@unimi.it
- Copenhagen: forskerservice@kb.dk
- Heidelberg: data@uni-heidelberg.de
- Warsaw: oa.buw@uw.edu.pl
- Charles University: researchdata@cuni.cz

What to share?

Research data? Data sets? Data records? Scientific information?

OECD: "research data" are defined as 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. A research data set constitutes a systematic, partial representation of the subject being investigated.

Research data and DMPs
Funders' requirements

- Europe, Horizon Europe :
  - **Mandatory DMP**
  - FAIR data management
  - **Open data** if possible

- Europe, H2020 :
  - If pilot, mandatory DMP
  - If pilot, open data if possible

What data will be collected, processed and/or generated?

Examples of research data types:

PRIMARY SOURCE	Data that constitute the basic materials of the research (before any processing and analysis).
EXTERNAL SOURCE DATA	Data not produced by the researcher (open access, on demand or reuse). Example: External quantitative databases; corpus and and textual archives collected by other people/institutions, audio/audiovisual archives audio/audiovisual archives; statistics produced by public/private actors, website content
SOURCE DATA COLLECTED	Data collected and/or compiled by the researcher directly or remotely. Example: Database developed from multiple sources, questionnaire/survey database, field survey data (interviews, transcripts, observations, photos, field notes...), textual corpus developed from web data
RESULTS DATA	Data produced after exploitation and processing of source data and which constitute the result data. Example: statistical analysis data, qualitative analysis data, network analysis data, interview extracts, maps, graphs, models and indicators, aggregated numerical data, audiovisual productions, syntheses, notes, annotations
DOCUMENTATION ON THE DATA	Data that provide information on the data itself (methodology, analysis materials used, treatments applied, ownership and protection of data, data valuation). Example: Survey procedure/methodology guide, program and application, analysis grid, dictionary of variables, request for authorization from the confidentiality clauses, metadata (title, keywords, DOI) ...

51

What data will be collected, processed and/or generated?

List (not exhaustive) of shareable data:

- laboratory notebooks, studies, experimental protocols, experimental reports, test reports,
- softwares, codes,
- databases,
- reading notes, notebooks,
- interviews,
- expert's report,
- conferences/seminars: announcement, program, list of speakers, press kits, posters,
- photo, film, 3D, surveys,
- documentation of scientific equipments, methodological guide,...

→ Data may be **numerical**, **descriptive**, **visual** or **tactile**. It may be raw, cleaned or processed, and may be held in any format or media.

Open for you! An introduction series to open science

52

Methodology and standards

Metadata standards and Research Data Management guidelines:
- The FAIRsharing portal with information and resources on data standards, databases, and policies in the life sciences and other scientific disciplines.
- DM guidelines and good practices for the Life Sciences, the Social Sciences and the Humanities provided by relevant research infrastructures, ELIXIR, CESSDA and DARIAH, respectively along with relevant data resources and repositories/databases.
- For more information on disciplinary metadata standards, visit Digital Curation Centre and Research Data Alliance Metadata Standards Directory

Examples of disciplinary metadata standards : PACTOLS, Generic Earth Observation Metadata Standard (GEOMS), VRA Core, Darwin Core, Ecological Metadata Language (EML)...

Open for you! An introduction series to open science

53

Overview on FAIR principles

FAIR principles:
https://www.force11.org/group/fairgroup/fairprinciples

FAIR logo: SangyaPundir. CC: BY-SA 4.0.
Image : ANDS. CC: BY 4,0,

"as **open** as possible, as **closed** as necessary"

Access to, and use of, certain research data will necessarily be limited by various types of legal requirements, which may include restrictions for reasons of:
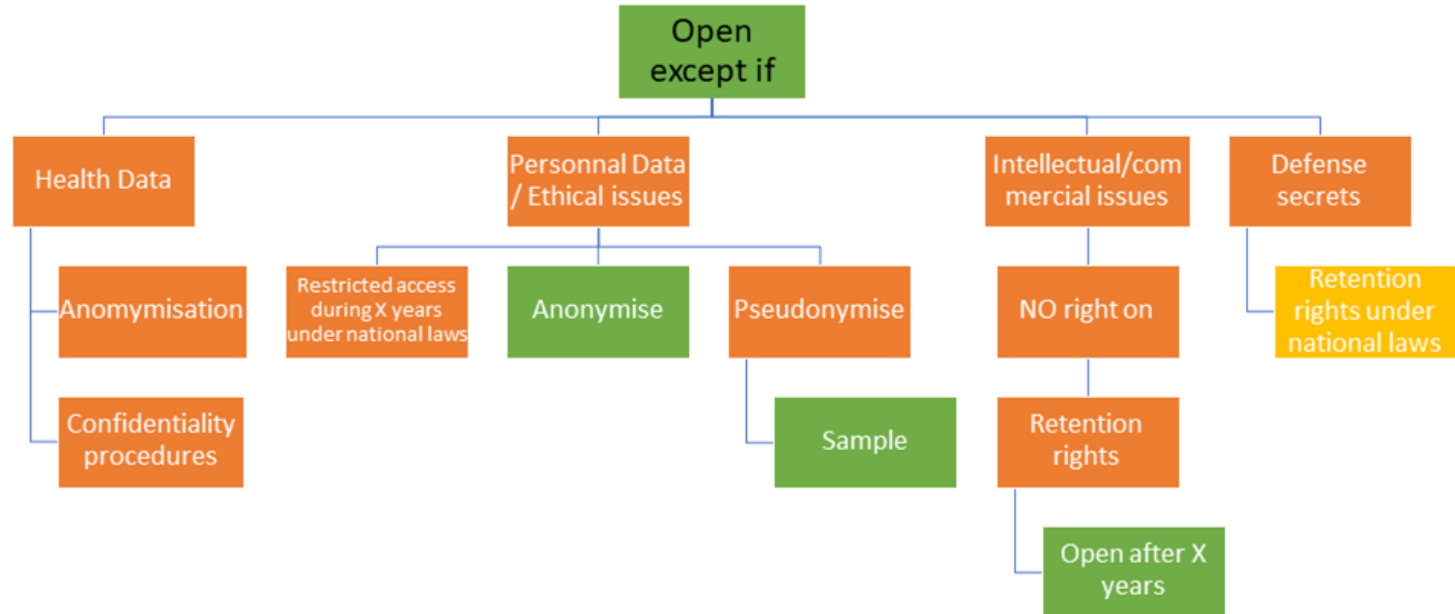
- national security
- privacy and confidentiality
- trade secrets
- legal process

It is always possible to define rules of communicability making the data open after a certain period (depending on national laws), ex: in France, documents related to national security may be opened after 50 years.

Minimum standards are defined in the Tromsø Convention (CETS No. 205)

Opening up data: what's in my data?

Open for you! An introduction series to open science

56

Some ways to secure the collect of personal/health data:

Questions to be ask before collecting personal data

- How long do I really need the data to achieve the objective?
- Do I have legal obligations to keep the data for a certain period of time?
- Do I have to keep certain data in order to protect myself against possible litigation? Which ones?
- What information must be archived? For how long?
- What are the rules for deleting data?
- What are the rules for sharing
- What are the rules for digital preservation?

Some ways to secure the collect of personal/health data:

Document the conformity of data acquisition methods

The documents formalising the internal procedures:
- the register in which the processing activities and the related retention periods must be recorded,
- the various actions undertaken, including whether these actions are still ongoing,
- written instructions to the processor regarding time limits.

Internal procedures:
- Any procedures for sharing, storing or archiving data,
- the procedure for destroying data, if applicable
- describe security measures, including where paper records are stored, to protect data from destruction, loss, alteration, unauthorised disclosure or access. These measures should be appropriate to the risks and the nature of the data.

Open for you! An introduction series to open science

58

Some ways to secure the collect of personal/health data:
Tools

- Sort through the data,
- Anonymize,
    - https://amnesia.openaire.eu
    - https://github.com/SGMAP-AGD/anonymisation
- Pseudonymization,
    - https://github.com/etalab-ia/pseudo_app
- Delete,
    - Redaction of a secure data deletion procedure in accordance with the recommendations of the CISO.
    - Use dedicated software for data deletion without physical destruction that has been audited or certified.
- Secure through encryption
    - GnuPG, https://gnupg.org/
- Name and re-name,
    - ReNamer, https://renamer.fr.softonic.com/
    - Ant Renamer, http://www.antp.be/software/renamer/fr

Anonymisation is often preferred to pseudonymisation - > A tool to raise awareness of the risks of re-identification after anonymisation:
https://cpg.doc.ic.ac.uk/observatory/

How to share despite access restrictions?

➔ anonymize/ pseudonymize,
➔ share samples,
➔ archive with access rights retention

If data are restricted, **the restriction should not be for an indefinite period**
➔ better to have a specific agreement about the number of years of the closure, e.g. ten or fifty years, or during the lifetime of the individual concerned.
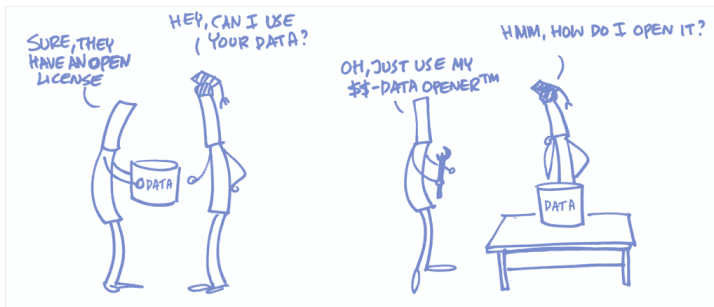
Licenses
Original works

International :
➜ Creative commons

Europe:
➜ Open Licence Assistant

Where to share your data?

- Multi-disciplinary repository (EU / CERN): [Zenodo](#)
- Commercial repositories: [Figshare](#), [GigaDB](#), [Mendeley data](#)
- Disciplinary repositories: [DRYAD](#), [NCBI](#), [Nakala](#), [Gbif](#), [Seanoe](#)
- Organizations repositories: [Edinburgh Datashare](#), [INRAE Datapartage](#),[The Heidelberg Open Research Data](#), [UNIMI Dataverse](#)

DMP as efficient tool for :

➔ Improve finding, understanding and using data,

➔ Define and classify your research data,

➔ Identify potential risks and challenges

➔ Align expectations with your supervisor and collaborators

Questions ?

Register for our upcoming events: https://4euplus.eu/4EU-273.html

Next sessions:
- "Predatory publishers and identity fraud – how to identify dubious providers" | 27 January 2022, 15:00 - 16:30
- "Publications strategies for monographs in humanities and social sciences" | 7 Februrary 2022, 10:00 - 11:30
- "Research data management – introduction to fair and open data" | 7 March 2022, 15:00 - 16:30
- 9 further session in 2022

Open for you! An introduction series to open science