# How to write a DMP?

ESHPM RDM Awareness Campaign

Anna Volkova
Research Data Steward, ESHPM

October 21, 2021

**Erasmus University Rotterdam**

# Outline

- Why a Data Management Plan?

  - FAIR

- What's a Data Management Plan?

  - Templates & requirements

- How?

  - EUR DMP Template v4.4
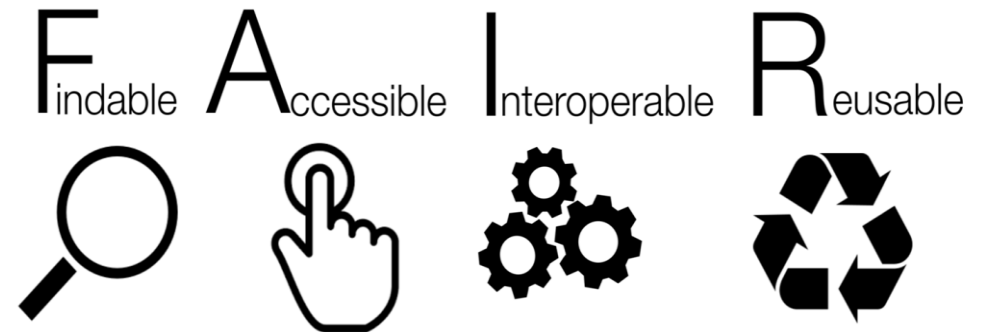
  - DMPOnline

# Why, oh why?

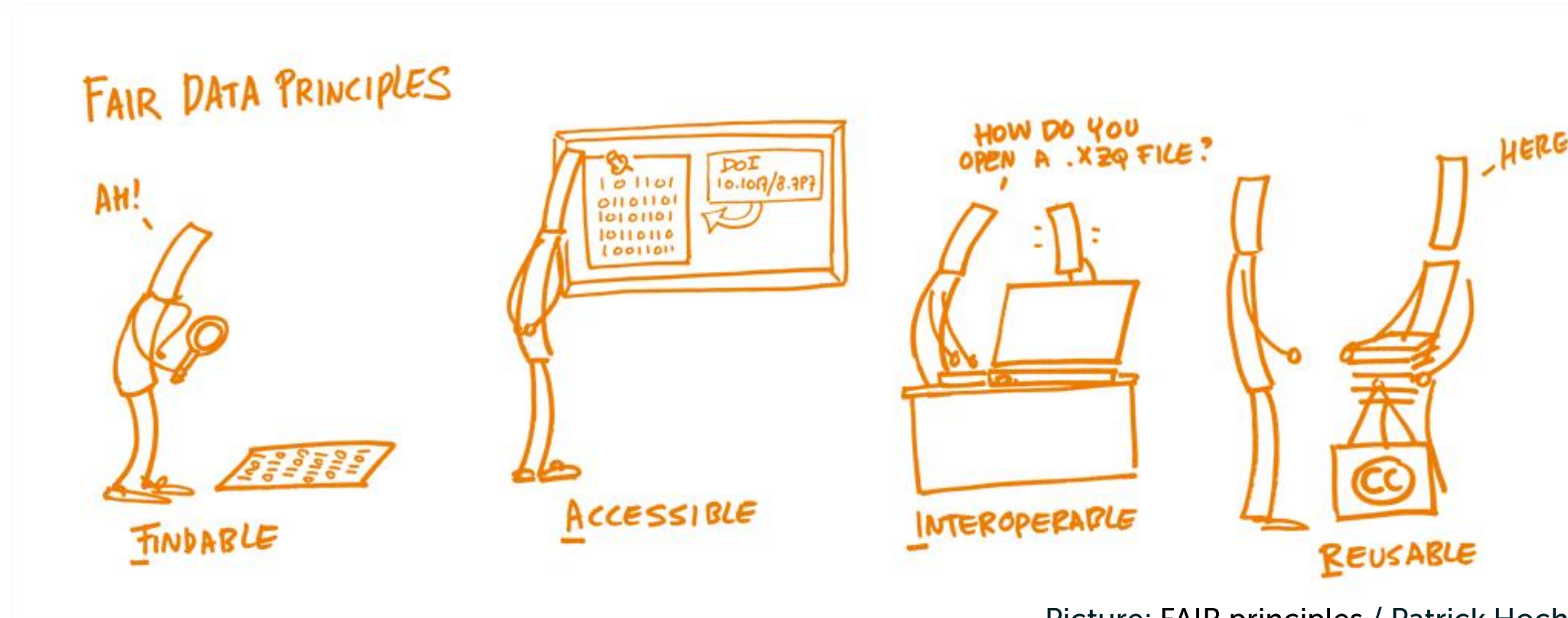# Maximizing the impact of scientific data

## Research: making claims backed up by data
- Data = one of your main ingredients
- Data = one of your main outputs

F indable  A ccessible  I nteroperable  R eusable

# Be FAIR: Findable, Accessible, Interoperable, Reusable



Picture: FAIR principles / Patrick Hochstenbach / CC0 1.0

- FAIR paper: Wilkinson et al., 2016. https://doi.org/10.1038/sdata.2016.18
- More resources at https://www.go-fair.org/

# Researcher's code



Research derives its status from the fact that it is a process governed by standards, which are expressed by a number of guiding principles. Researchers who are not guided by these principles risk harming quality and trustworthiness of research, undermine public trust and mutual trust between individual researchers.

- **Honesty**
- **Scrupulousness**
- **Transparency**
- **Independence**
- **Responsibility**

# Ethical research

Research involving one of the following aspects:

- Human beings;
- Potential misuse of research results;
- (Special categories of) personal data;
- Potential conflict of interest;
- Non-EU countries;
- External stakeholders, among which funding organizations;
- Environment, health & safety issues, including potential harm to researchers.
- + if planning to publish in a (medical) journal.

## Ethical review

The Association of Universities in the Netherlands (VSNU) and the Royal Netherlands Academy of Arts and Sciences (KNAW) adopted the Netherlands Code of Conduct for Research Integrity (NCCRI) in 2018. In line with this document, Erasmus University Rotterdam underpins the importance of ethics review of research projects prior to data collection, and as of January 1st, 2021 a number of principles and requirements pertaining to research ethics review at Erasmus University Rotterdam has taken effect.

The aim of the ethics review is to ensure that all research that is conducted by faculty affiliated researchers, complies with good practice in terms of protecting the privacy and security of research participants in the collection, analysis, and dissemination of research data, and with the five principles of proper academic practice as set out in the NCCRI: honesty, scrupulousness, transparency, independence, and responsibility. This will advance the quality and reliability of the research.

# Wait, but what?

# The lifecycle of data in a research project
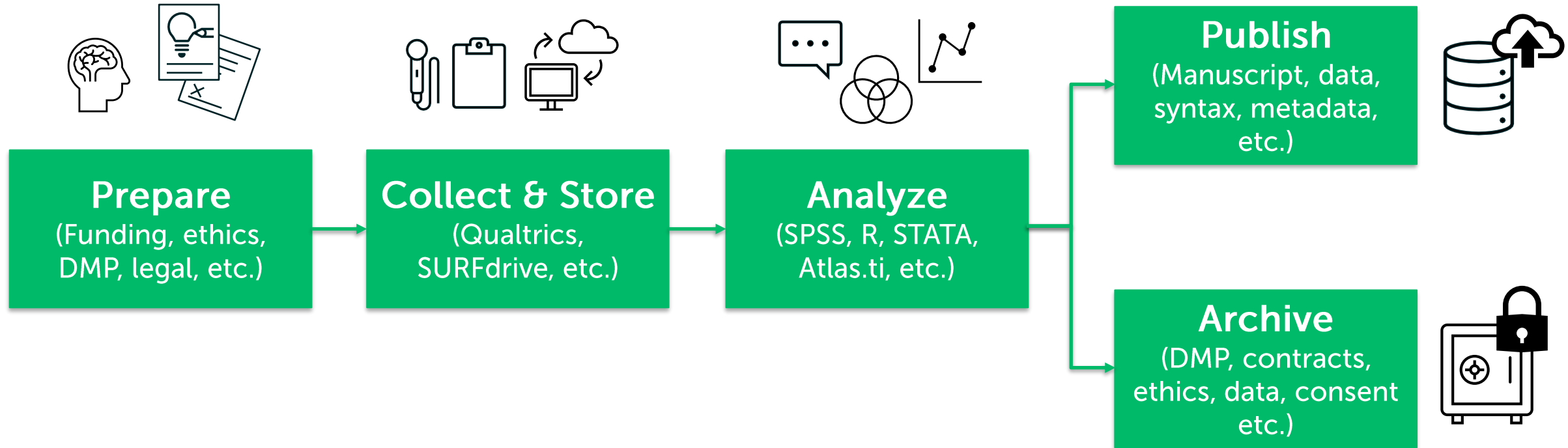
# 10 practices from Briney et al.

- Practice 1: Keep sufficient documentation
- Practice 2: Organize files and name them consistently
- Practice 3: Version the Files

→ **Document & organize**

- Practice 4: Create a security plan, when applicable
- Practice 5: Define roles and responsibilities
- Practice 6: Back up the data

→ **Stay safe**

- Practice 7: Identify tool constraints

→ **Prevent technical issues**

- Practice 8: Close out the project
- Practice 9: Put the data in a repository

→ **Closing & archiving**

- Practice 10: Write these conventions down [in a DMP]

→ **Plan**

# Prepare by writing a data management plan (DMP)
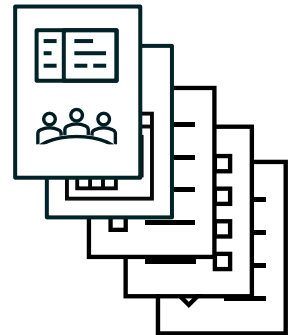
- **What a DMP is:**
  - A way to formulate and record your plans and decisions about your data
  - A "living document", to be revisited every few months
  - Saving you time
  - A tool for you to help with responsible RDM

- **What a DMP is not:**
  - A matter of "one-time approval"
  - Written in stone
  - Wasting your time
  - A box to tick "and then RDM is taken care of"

# In a Data Management Plan you describe ...

- What type of data you are going to collect

- Where you are going to store your data

- How you are going to share your data

- How you address security & privacy matters

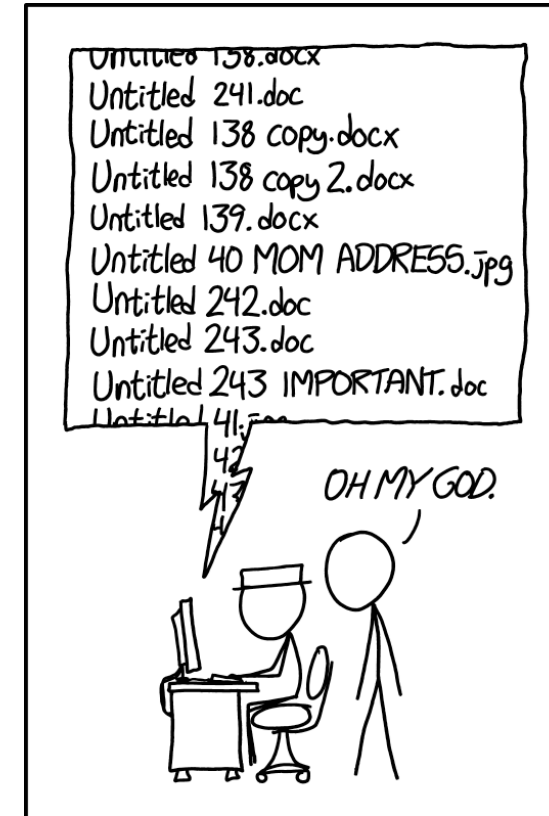A DMP forces you to think through each step of your project making sure you have all you need

# Write a Data Management Plan if ...

- You **care** deeply about research integrity and ...

- You want to **prevent** issues and set-backs

- You want to **comply** with funder and university policy

# Document and organize your research thoroughly

# DMP Templates

- What does a DMP look like?
  - Many different DMP templates exist
  - I recommend to either use EUR template or your funder's template (if applicable)
    - EUR template is approved by NWO and ZonMw

- EUR DMP Template 4.4 (some outdated versions still in circulation!)
  - You can choose between offline version (Word document) or online tool (<u>DMPonline</u>)
  - Loosely structured in chronological order:

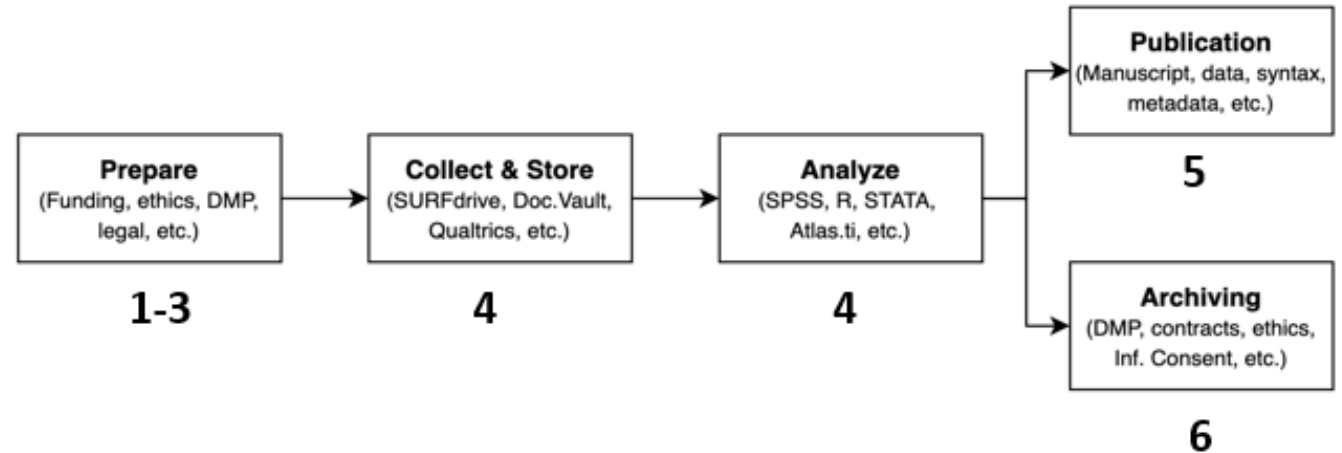Description / Admin → Preparation → During Research → After Research (Reuse / Archiving)
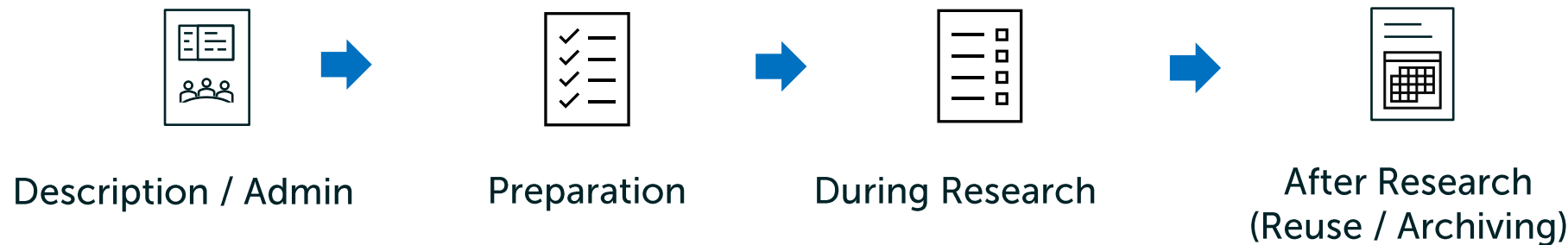
# How?

- **EUR DMP Template**
- **Using DMPOnline**

# EUR DMP Template

# EUR DMP Template

## EUR Data Management Plan

### Version 4.4



The template was created by the RDM specialists at UL/EDSC and the data stewards. It has been approved by NWO and ZonMw, and it is also the default DMP format supported by the EUR.

Table of contents

# General

**Please tick the following boxes if you agree to act according to the following terms:**

☐ I will answer all questions truthfully and to the best of my knowledge

☐ I will discuss the data management plan with my research team

☐ I will check and, if necessary, update my data management plan a minimum of once a year

Support[i] in writing a data management plan is available through the faculty Data Stewards (their contact details can be found on the web page of Erasmus Research Services).

**If applicable, please provide the name of the support staff consulted and the date of consultation:**

| Name of support staff | |
|---|---|
| Date of consultation | |

[i]Research funders ZonMw and NWO require all Data Management Plans to be completed in consultation with data management support staff at the home institution of the grant holder in order to be eligible for consideration.

**Scientific research must be conducted in line with existing guidelines on good research practices and integrity. Please tick the boxes if you have read and understand these guidelines and will act accordingly.**

☐ The Netherlands Code of Conduct for Research Integrity (VSNU, 2018)

OR

☐ The European Code of Conduct for Research Integrity (ALLEA, 2017)

| Name | Email | ORCID* | Research Institution |
|------|-------|--------|---------------------|
|      |       |        |                     |
|      |       |        |                     |

# Administration & Project description

| | |
|---|---|
| **Project title** | |
| **Project start date as intended** | |
| **Project duration in months as intended** | |
| **Funding body (if applicable)*** | |
| **Grant number (if applicable)** | |
| **Date of DMP version 1** | |
| **Date of update DMP and version*** | Date:<br>Version: |

Also questions on:

- Project background & research questions

- Research type & methodology

  o Be as detailed as possible!

- Resources needed to implement the DMP

  o 'Standard' services are provided by EUR

# Legal & policy

1. With whom will you need to make legal arrangements?*

   ☐ With nobody / No reason → *Go to Q3*

   ☐ With research participants

   ☐ With third parties

   ☐ For working with multiple research partners

   ☐ I do not know → *Go to Q3*

2. List the agreements that <u>you will initiate</u> and with whom you will make them.

| Who | Type of agreement |
| --- | --- |
|  |  |
|  |  |
|  |  |

| Example: Research Participants | Informed Consent |
| --- | --- |
| Example: Multiple research partners | Consortium Agreement |

4. Do you need to obtain ethical approval for your research project?*

   ☐ No, my project does not require ethical approval

   ☐ Yes, I am preparing to submit my application

   ☐ Yes, I have submitted my application

   ☐ Yes, I have obtained ethical approval

   ☐ I do not know

5. If you have obtained ethical approval, list the reference number:

3. List the agreements or other data management policies that you need to uphold but <u>did not initiate</u>. If you are reusing existing data, list the terms of use under which you may re-use them.

| Who | Type | Version and date |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |

| Example: EUR | RDM policy of Erasmus University Rotterdam (EUR) | Version 1.0 [August 14, 2020] |
| --- | --- | --- |
| Example: Zoom | Terms of Service | Version [August 20, 2020] |
| Example: NWO | Funding Agreement | Version 11.4 [July 11, 2018] |
| Example: EUR | Internet and ICT facilities policy | Version 1.0 [2015] |

# During research: Collecting & Analyzing

6. Specify what data you will be collecting and indicate format, estimated size, and whether this is data that you will be generating or existing data that you will be re-using.*

| Type | Format | Estimated size[i] | Generate or Reuse |
|---|---|---|---|
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

| | | | |
|---|---|---|---|
| Example: Digital survey data | .csv | 1-5 GB | Generate |
| Example: Audio-recorded interviews + Transcripts | .mp3; .docx/.odt | 5-10 GB | Generate |
| Example: Dutch Central Bureau of Statistics (CBS) data | .csv | 10-50 GB | Re-use |

# Collecting & Analyzing

7. **Will you be collecting [sensitive] personal data?***

☐ No – My research does not include [human] participants → *Go to Q9*

☐ No – My research involves human participants, but I will collect fully anonymous data → *Go to Q9*

☐ Yes – Personal data (non-sensitive) → *Consult your faculty's Privacy Officer*

☐ Yes – Personal data (sensitive) → *Consult your faculty's Privacy Officer*

☐ I do not know → *Consult your faculty's Privacy Officer*

8. **If you collect [sensitive] personal data, how will you protect the privacy of participants when sharing your data?***

☐ I will fully anonymize the data

☐ I will pseudonymize the data

☐ I do not know → *Consult your faculty's Privacy Officer*

9. **Will you be collecting non-personal sensitive data?**

☐ Yes [e.g. confidential company data, data related to national security]

☐ No

## Personal data
- Data that can [in]directly identify a natural person
- Name, address, IP-address, student number

## Sensitive personal data
- Personal data that can potentially harm an individual if it were to become public
- Special category data: health, political preferences, sexual preferences, religion, race.

## Sensitive non-personal data
- Data that is otherwise sensitive such a proprietary data, patents, company data

# Collecting & Analyzing

10. Where will you store your data <u>during</u> the project?* You can select multiple options.

☐ I do not know → *Go to Q12*

☐ EUR OneDrive → *Go to Q12*

☐ EUR Dropbox (not for personal data) → *Go to Q12*

☐ EUR SURFdrive → *Go to Q12*

☐ EUR SURF Research Drive (for collaborations) → *Go to Q12*

☐ EUR Document Vault [Blackberry Workspaces] → *Go to Q12*

☐ EUR BEATS → *Go to Q12*

☐ EUR Network Shares → *Go to Q12*

☐ Other – please specify:

**Non-personal data**

**Personal data**

**Sensitive data**

All types of files – EUR employees & students

All types – Researchers NL

All types – Researchers Non-NL & Third parties

Audio & Video

All types of data – EUR employees

# Collecting & Analyzing

11. Is this other software tool supported by the EUR or is it private?

☐ EUR supported

☐ Private

☐ I do not know → *Contact your faculty's Data Steward*

12. What hardware and software do you use?* Select all applicable options.

☐ EUR supported hardware (e.g. @wEURk laptop, @wEURk workstation)

☐ Private hardware (e.g. personal laptop, private external hard drive)

☐ EUR supported software as found in the [software catalog](software catalog)

☐ Private software or freeware (e.g. private Dropbox)

13. If you use private hardware, software, or freeware, please specify what and for what reason:

Example: During fieldwork I have no access to internet, so I temporarily store my data on an encrypted external drive.
Example: I do not have an EUR laptop, so I am using my own device.

14. Are regular backups made of your data?

☐ I do not know

☐ No

☐ Yes, I use only EUR supported tools [as listed in Q12], thus regular backups are made automatically

☐ Yes, manually – please specify:

Note: Include who makes the backups and how often backups are made

# Use EUR supported hardware and software

- University & Faculty provide necessary conditions, such as IT infrastructure

- Your responsibility to use the tools being provided
    - Use EUR supported hardware and software (see <u>software catalog</u>)
    - Use <u>MyApps</u> (Remote Desktop) when working on own device
    - Use <u>Teams/Sharepoint</u>, <u>SurfDrive</u>, <u>ResearchDrive</u>, <u>Document Vault</u>, EUR Network shares to store and process research data



(<u>Note</u>: only with ERNA account)
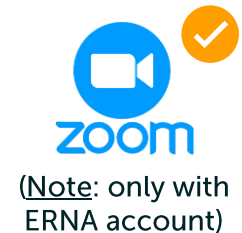
# Collecting & Analyzing

- Access to folder and files
  - Who manages access; who has access; how is access managed in case of staff changes or illness
- Structuring and naming folders and files
  - In X time, would you [or someone else] be able to quickly find the right file and understand the content?
- Versioning of files
  - How many 'definitive' versions do you have of one document

# Organize your data

1. Save your project in a single folder
2. Devise a logical system of subfolders

   o Make sure to have separate folders for data (raw vs. processed), documentation, figures

3. Introduce the project (or subfolder) in a README
4. Use interoperable file types
5. Use descriptive and logical file names

   o Machine readable

   o Human readable

   o Logically sortable (YYYYMMDD)

6. Make your data tidy (each variable is a column, each observation is a row)

# Organize your data and use consistent filenames

- Organize files and name them consistently
  - Choose strategy + convention, and write these down
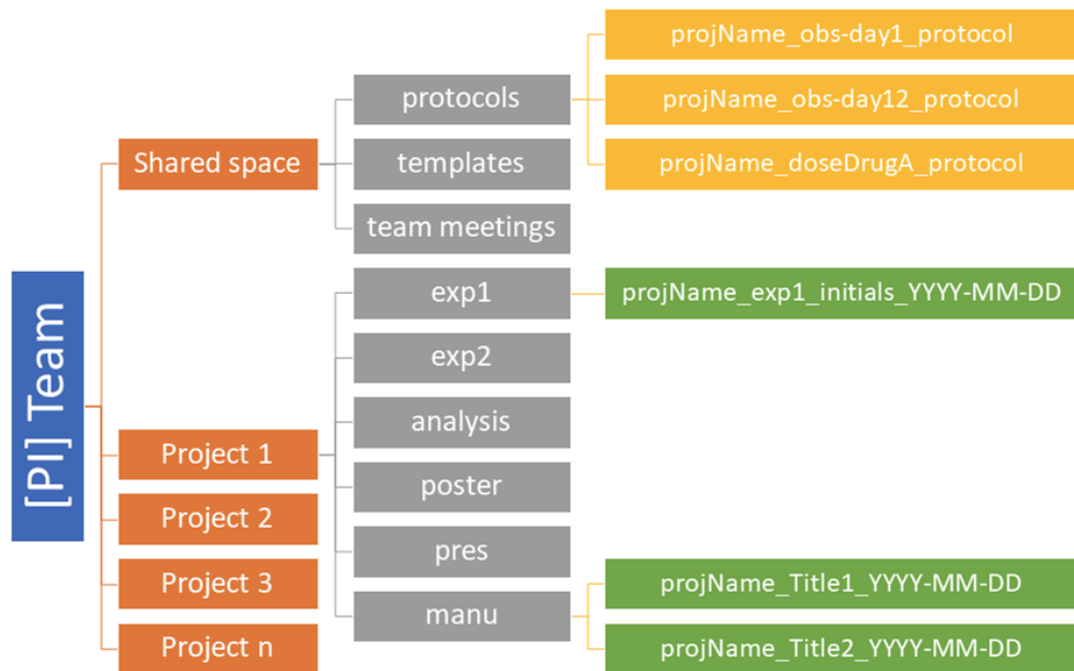


- Refer to guidelines when designing your system: CESSDA training
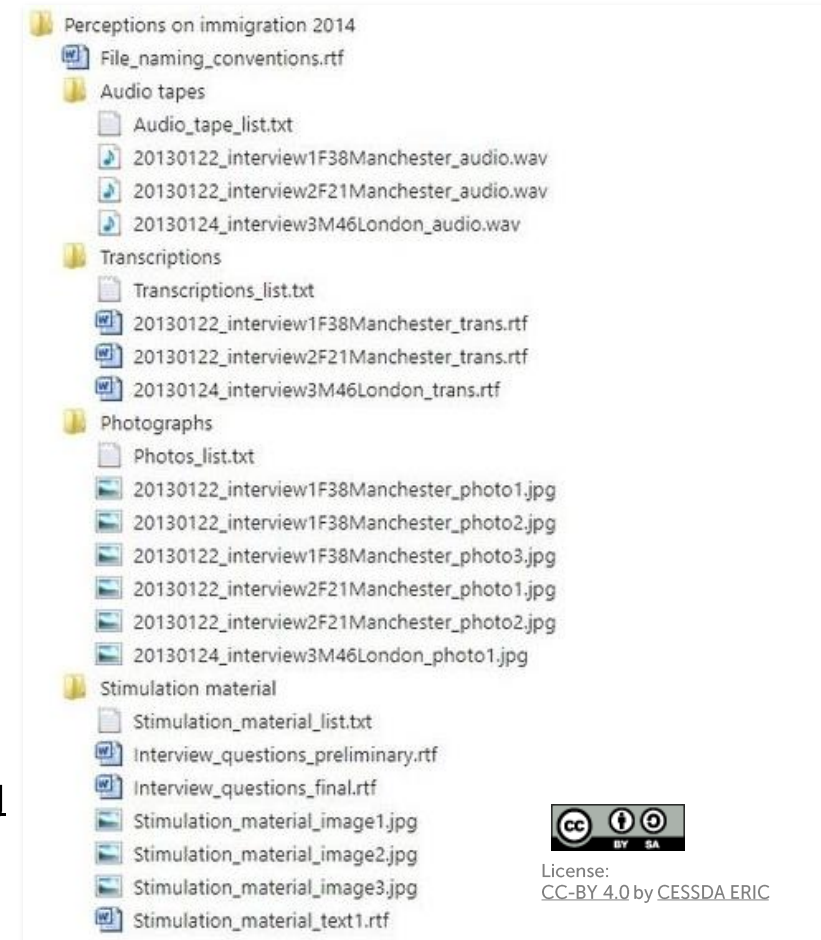
# Document your research thoroughly with a README

- Keep sufficient documentation
  - README.txt 'with enough detail so that the research can be picked up after a year-long hiatus or by a collaborator'
  - Do not trust your memory!

- What is the project about?
- Who are involved?
- How was the data collected?
- When & where is the research conducted?
- Abbreviations, folder structure, file explanations, references & links

# Use consistent filenames



"FINAL".doc

FINAL.doc!

FINAL_rev.2.doc

FINAL_rev.6.COMMENTS.doc

FINAL_rev.8.comments5. CORRECTIONS.doc

FINAL_rev.18.comments7. corrections9.MORE.30.doc

FINAL_rev.22.comments49. corrections.10.#@$%WHYDID ICOMETOGRADSCHOOL????.doc

WWW.PHDCOMICS.COM

JORGE CHAM © 2012

Picture: "Piled Higher and Deeper" by Jorge Cham / www.phdcomics.com

- Version the Files
  - At minimum: separate raw + processed data
  - Use version numbers or dates (manually or automate using Git)

- Resist the temptation to ever use "FINAL" in a file name!

# After research: Data Sharing & Reuse

**20. What data [and code] will be shared in a research data repository?**

☐ I do not know

☐ All data [and code] underlying published papers / reports → *Go to Q22*

☐ All data [and code] produced in the project → *Go to Q22*

☐ A selection of the data [and code] → *Go to Q21*

☐ I cannot share the data [and code]; I will share only the metadata → *Go to Q21*

**21. Please specify why you are unable to share (all) data [and code].**

Example: Contractual obligations, privacy law

- The EUR is an Open Science university
  ○ All publications should be published open access
  ○ All data underlying publications should be shared as openly as possible, as closed as necessary

- Three questions: Why, How, Where
- Why: for your future self, for a possible reuse & collaborations
- How: complete and interpretable metadata and documentation
- Where: EUR Data Repository [EDR] or domain-specific repository

# Tension between what you need to share and what you can share

- From the *Netherlands Code of Conduct for Research Integrity* (2018)
    - https://doi.org/10.17026/dans-2cj-nvwu

Ensure that, in accordance with the FAIR principles, data is open and accessible to the extent possible and remains confidential to the extent necessary.

# Data sharing & reuse

22. List the data [and code] that you plan to share in a research data repository. Also list the information / documentation / metadata that you will include to make the data package self-explanatory and re-usable in the future (for other researchers and yourself)*

| Data | Format | Size |
|---|---|---|
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

| Data | Format | Size |
|---|---|---|
| Example: Anonymized survey data | .csv | <1 GB |
| Example: Codebook, Blank questionnaire | .pdf | <1 GB |
| Example: readme text file (general description of the data, incl. date of collection, selection procedure of participants, tools used to collect the data, etc.) | .txt | <1 GB |

- Information needed to interpret, understand, and reuse data
    - Codebook, methods, date of collection, definitions
    - Metadata standard: if we all used the same e.g. coding and definitions, life would be much easier!

# Data sharing & reuse

23. In which repository will you place the [meta]data and/or code associated with your paper?*

☐ I do not know

☐ EUR Data Repository (EDR)

☐ Other – please specify:

|  |
|--|
|  |

Example: DANS EASY, Zenodo, Dryad, Dataverse, 4TU, Open Science Framework (OSF)

24. What metadata standard will you use to document your research?*

☐ I do not know

☐ None

☐ DCMI (Dublin Core Metadata Initiative) Note: Default within the EUR Data Repository

☐ DDI (Data Documentation Initiative)

☐ SDMX (Statistics Data and Metadata Exchange)

☐ Other – please specify:

|  |
|--|
|  |

Erasmus University Rotterdam

# Discover research from Erasmus University Rotterdam (EUR)

+ Follow

ALL    CATEGORIES    GROUPS    SEARCH

⌄ Hide footer

# EUR Data Repository

## Metadata (Dublin Core):

Title;
Authors;
Categories;
Group;
Item Type,
Keywords;
Description;
Funding;
Resource title;
Resource DOI;
References;
License

# EUR Data Repository

## Metadata (Dublin Core):

Title;
Authors;
Categories;
Group;
Item Type,
Keywords;
Description;
Funding;
Resource title;
Resource DOI;
References;
License

**Funding**

Search grant by name/number or add your own

**+ Add another grant**

**Resource title**

Add the title of the resource you want to link to

**Resource DOI**

Add the DOI you want to link to

**References**

Link to references or related content

**Licence** (what's this?)

CC BY 4.0

**This item is a draft** (metadata required for publication missing)

Apply embargo

Generate private link

DOI  Reserve Digital Object Identifier

Delete item

**Tips**

you can still drag
more file(s) on the page or
browse

Preview item (private)
Edit timeline
Last edited on 30 Nov 2020 - 11:35

Cancel          DOI          Publish     Save changes

# Data sharing & reuse

After uploading data, you stay in control and can choose your data to be:

- Open to all

- Placed under embargo (temporary vs. permanent)

- Made restricted access:
    - Sharing restrictions: specified person(s), faculty, IP-range, etc.
    - Confidential: accessible only upon request, user agreement can be set up

**28. Under which license will you make your data available for re-use?***

☐ I do not know

☐ Creative commons (e.g. CC-BY)

☐ License for specific types of data (e.g. software license)

☐ Other – please specify:

**29. Please specify which license:**

# After research: Archiving

- Closed archive 'cold storage', unlike e.g. EUR Data Repository

  ○ Doubles as emergency backup; original files for auditing and verification

- Typical files for archiving include:

  ○ Informed Consent; Contracts; DMP; Ethics documents

  ○ Complete raw data; Code and Software; Codebook

- Per research domain the term may vary between 5 to 15+ years

  ○ The EUR recommends a minimum of 10 years

# Archiving

- Select what needs to be destroyed
  - Exemption for personal data under GDPR
- Determine what you want to archive
- Use durable formats

| Type | Preferred format(s) | Non-preferred format(s) |
|---|---|---|
| Text documents | PDF/A (.pdf)<br>ODT (.odt) | Microsoft Word (.doc)<br>Office Open XML (.docx)<br>Rich Text File (.rtf)<br>PDF other than PDF/A (.pdf) |
| Plain text | Unicode text (.txt) | Non-Unicode text (.txt) |
| Markup language | XML (.xml)<br>HTML (.html)<br>Related files: .css, .xslt, .js, .es | SGML (.sgml)<br>Markdown (.md) |
| Programming languages | MATLAB<br>NetCDF<br>TextFabric | |
| Spreadsheets | ODS (.ods)<br>CSV (.csv) | Microsoft Excel (.xls)<br>Office Open XML Workbook (.xlsx)<br>PDF/A (.pdf) |

30. **You may be obliged to destroy some data before archiving. Do any of such obligations apply to you?**

☐ I do not know

☐ No

☐ Yes - Contractual obligation (e.g. licenses)

☐ Yes - Commercial objectives

☐ Yes - Privacy law (e.g. personal data of participants)

☐ Yes - Other - please specify:

31. **List the data you will be archiving. These data constitute your archival package.**

| Data | Format | Size |
|---|---|---|
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |

| | | |
|---|---|---|
| Example: Informed Consent forms (signed) | .pdf | <1 GB |
| Example: Raw data | .csv, .docx, mp3, mp4 | 5-10 GB |
| Example: Processed data | .csv, .odt, .mka | 1-5 GB |
| Example: Questionnaires | .pdf | <1 GB |
| Example: Contracts & Terms of use | .pdf | <1 GB |
| Example: Data Management Plan | .pdf | <1 GB |
| Example: Grant application & project description | .pdf | <1 GB |
| Example: Ethical review application & approval document | .pdf | <1 GB |

DMPOnline

# DMPOnline

# DMPOnline: Edit profile

Erasmus University Rotterdam

🌐 RDM policy    ✉ RDM enquiries

🌐 RDM webpages                                                                        🔒 Admin ▾

# Create a new plan

Before you get started, we need some information about your research project to set you up with the best DMP template for your needs.

## * What research project are you planning?

☐ mock project for testing, practice, or educational purposes

## * Select the primary research organisation

Erasmus University Rotterdam ⊗

- **or** -  ☐ No research organisation associated with this plan or my research organisation is not listed

## * Select the primary funding organisation

Begin typing to see a filtered list

- **or** -  ☐ No funder associated with this plan or my funder is not listed

**Create plan**    **Cancel**

**Support in writing a data management plan is available through the faculty Data Stewards. If your faculty has no Data Steward, support is available through the Research Data Management Specialist of the Erasmus Data Service Centre (EDSC is part of the University Library). Which research support professional is available for you?**

☐ Data Steward of my own faculty - ERIM / RSM
☐ Data Steward of my own faculty - ESE
☐ Data Steward of my own faculty - ESSB
☐ Data Steward of my own faculty - ESHCC
☐ Research Data Management Specialist of the EDSC

Save

| Guidance | Comments |

EUR

Support at the faculty RSM / ERIM & Faculty ESE

Support faculty ESSB

Support faculty ESHCC

Support by the University Library / EDSC

---

**Scientific research must be conducted in line with existing guidelines on good research practices and integrity.**

**Please select which guidelines you are familiar with.**

☐ The European Code of Conduct for Research Integrity (ALLEA, 2017)
☐ The Netherlands Code of Conduct for Research Integrity (VSNU, 2018)

Save

| Guidance | Comments |

EUR

You can find The European Code of Conduct for Research Integrity (ALLEA, 2017) here.

You can find The Netherlands Code of Conduct for Research Integrity (VSNU, 2018) here.

# If you need anything …

- If you have questions or need support

  - Let me know!

- If you have written a DMP…

  - Send it to me for review!

- If you used DMPOnline to write your DMP

  - Click 'Request Feedback' and it will be sent to me for review

volkova@eur.nl
datasteward@eshpm.eur.nl

# Questions

# Useful resources

- Erasmus University Rotterdam:
    - Research Data Management pages
    - Legal Support pages
    - MyEUR Privacy pages
    - MyEUR Software pages & Software catalog
    - MyEUR Cybersecurity pages
    - EUR DMPonline (for drafting data management plans)
    - EUR Data Repository (for sharing data)
    - Infographics on Privacy: Why, What and How
    - EUR Privacy & Security app: Apple App Store

# Useful resources

- Research Data Management information:
  - 23 things for RDM
  - Research Data Management – the Turing Way
  - Harvard Data Management pages
  - UK Data Service research data management
  - Go FAIR – FAIR principles
  - National Coordination Point Research Data Management
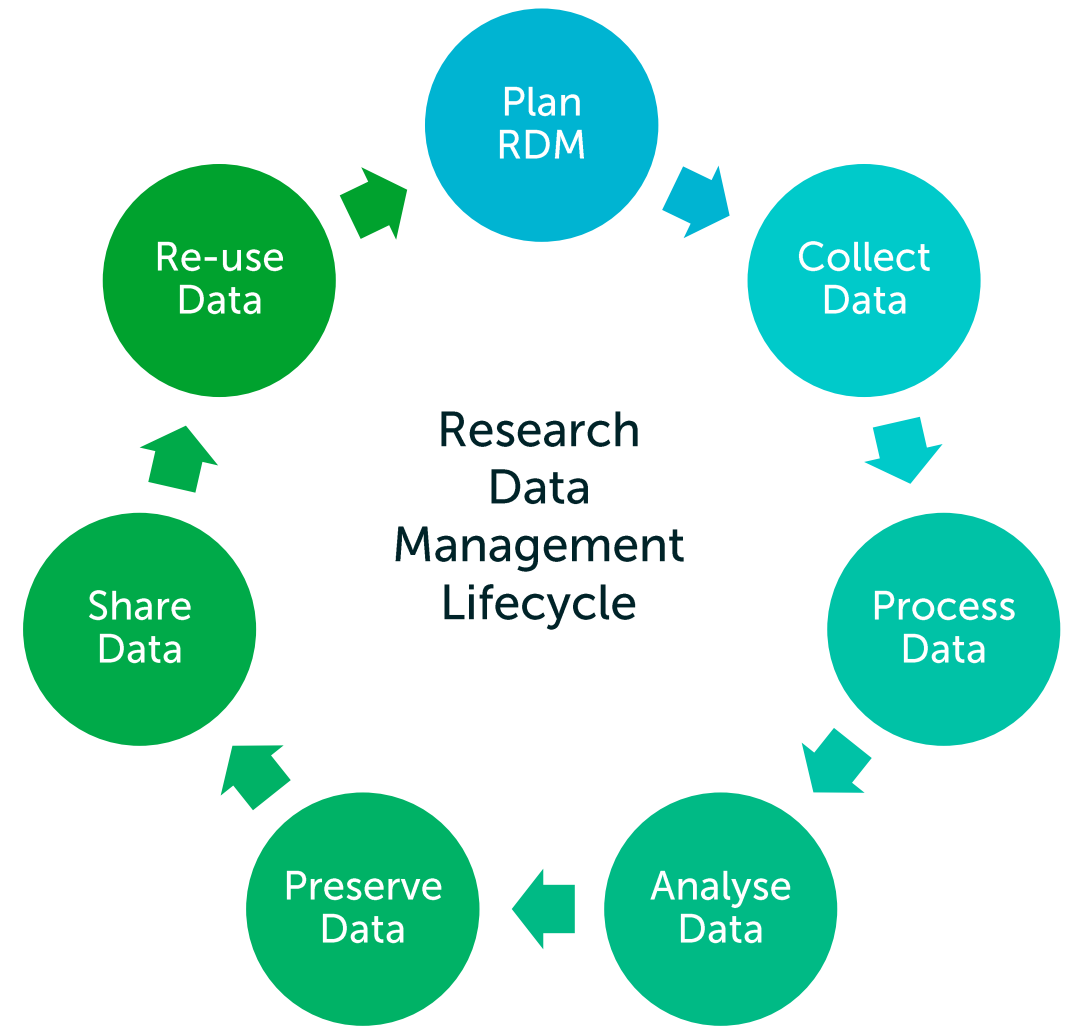
# Making your data FAIR(er)

- Findable
  - Assign persistent IDs, provide rich metadata, register in a searchable resource
- Accessible
  - Retrievable by their ID using a standard protocol, metadata remain accessible even if data aren't
- Interoperable
  - Use formal broadly applicable languages, use standard vocabularies, qualified references
- Reusable
  - Rich, accurate metadata, clear licenses, provenance, use of community standards

# Enters RDM…

**Research data management** (RDM) is a term that describes the

- organization,
- storage,
- preservation, and
- sharing of **data** collected and used in a **research** project.



Research Data Management Lifecycle

- Plan RDM
- Collect Data
- Process Data
- Analyse Data
- Preserve Data
- Share Data
- Re-use Data