

Research Data Management (RDM) in the Social Sciences: How to organize your data and make your life easier



Please note that the webinar will be recorded.

Presenter: Anita Bodlos

Universität Innsbruck

For authors of the slides and license
information, see end of the presentation.

06 November 2024

Version 1.1



Funded by the European Union

CC BY 4.0

Welcome and housekeeping rules

- Please mute your microphone when you are not speaking
- Please turn off your camera during presentations - the session is being recorded
- Post your questions in the chat or raise your hand using the "raise your hand" function in Zoom ("Reactions" button)
- Need help? Write in the chat!
- Tell us how we did in the feedback form (more about that later)
- The materials (including the presentations) will be shared afterwards

Agenda

- Part I: RDM, FAIR data and DMPs
- Part II: Personal data
- Part III: Data curation and repositories

Part I

RDM, FAIR data and DMPs

What is research data management (RDM)?

„Research data management refers to how you handle, organise, and structure your research data throughout the research process.”



CESSDA Training Team (2017 - 2019), p. 9

image: [source](#)

Why RDM?

- Meet the requirements of funding bodies and publishers of academic journals
 - Comply to the policies of many research performing institutions
 - Make your research more efficient
 - Advance research in your discipline and beyond
-
- And do it for... **YOURSELF!**

“As open as possible, as closed as necessary!”

Make your data FAIR!

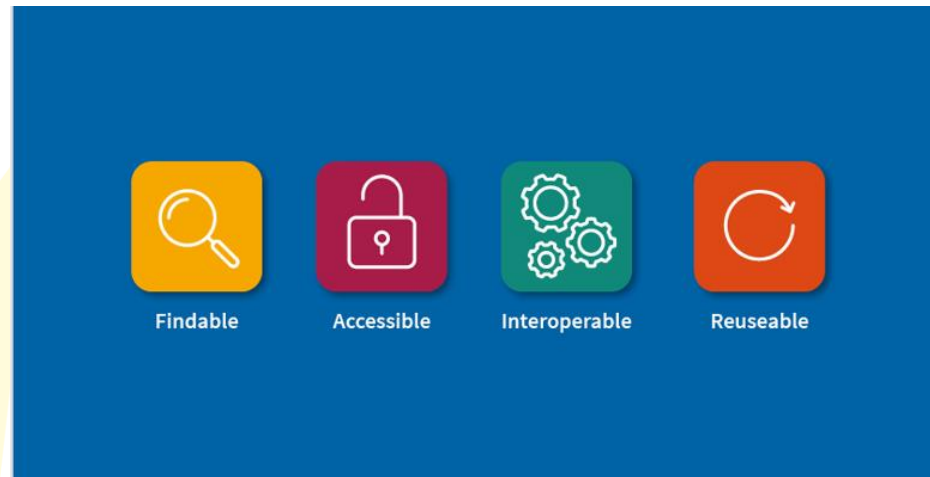


image: [source](#)

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. [The FAIR Guiding Principles for scientific data management and stewardship](#). Sci Data 3, 160018 (2016). [CC BY 4.0](#); see also <https://www.go-fair.org/fair-principles/>

Make your data FAIR!



Findable – Data and metadata should be easy to find by both people and machines. Machine-readable and descriptive metadata enables the discovery of interesting datasets.

e.g.

- assigning PIDs (persistent identifiers like DOIs)
- enriching with metadata
- using standardized vocabularies
- registering data in a searchable catalogue (e.g. in the [CESSDA](#) or [openAIRE](#) metadata catalogue)

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. [The FAIR Guiding Principles for scientific data management and stewardship](#). Sci Data 3, 160018 (2016). <https://www.go-fair.org/fair-principles/>

Make your data FAIR!



Accessible – Data and metadata should be archived and made available in such a way that allows easy retrieval and download by machines and people.

e.g.

- transparent access options
- AAI
- as open as possible
- metadata accessible even if the data are not anymore

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. [The FAIR Guiding Principles for scientific data management and stewardship](https://www.go-fair.org/fair-principles/). Sci Data 3, 160018 (2016). <https://www.go-fair.org/fair-principles/>

Make your data FAIR!



Interoperable – Data should be available in such a way that people and machines can exchange, interpret and combine these with other data sets in a (semi-)automated way.

e.g.

- allow interoperability with other data tools and systems (through API)
- proper documentation
- different and common formats

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. [The FAIR Guiding Principles for scientific data management and stewardship](https://www.go-fair.org/fair-principles/). Sci Data 3, 160018 (2016). <https://www.go-fair.org/fair-principles/>

Make your data FAIR!



Reusable – A good description of data (metadata) ensures that it can be reused for future research. Data should be permanently citable (such as with a DOI).

e.g.

- clear reuse regulations (licensing)
- additional documentation materials

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. [The FAIR Guiding Principles for scientific data management and stewardship](https://www.go-fair.org/fair-principles/). Sci Data 3, 160018 (2016). <https://www.go-fair.org/fair-principles/>

Data management plans (DMPs)

- ... contain all relevant aspects of RDM
- ... are a formal document
- ... are a „living document“
- ... are a mandatory deliverable for many grants
- ... can vary in their content depending on your research data and requirements by funders
- ... allow you to plan for the future
- ... can guide you through your project

Frequently found DMP sections

- Administrative information
- Description of your data
 - What data do you collect and/or re-use?
- RDM during your research project
 - Storage locations, access regulations...
- How to make your data FAIR
 - In which repository do you publish your data? When?
 - How do you minimize the need for access restrictions?
 - Which license do you choose?
 - What metadata and documentation material accompany your data?
- Legal and ethical considerations

DMPs: Templates and checklists

- Horizon Europe: https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/temp-form/report/data-management-plan_he_en.docx
- Science Europe: <https://doi.org/10.5281/zenodo.4915862>
- CESSDA: <https://doi.org/10.5281/zenodo.3820473>

DMPs: Tools

- Tools
 - Argos: <https://argos.openaire.eu/splash/>
 - Data Stewardship Wizard: <https://ds-wizard.org/>
 - DMPOnline: <https://dmponline.dcc.ac.uk/>
- DMP Collections (not evaluated):
 - Digital Curation Centre (DCC): <https://www.dcc.ac.uk/resources/data-management-plans/guidance-examples>
 - Phaidra: <https://phaidra.univie.ac.at/detail/o:1140797>
- Data Horror Escape Room: <https://sites.google.com/vu.nl/datahorror>

Where to find support on RDM?

- Data Stewards
- Institutional RDM team
- Repositories
- Online Training material:
 - [CESSDA](#)
 - [Open Science practices in the Humanities and Social Sciences](#)
 - [Sensitive Data in the Humanities and Social Sciences](#)
- ...

Questions on RDM basics, FAIR data or DMPs?

Part II

How to work with personal data

Personal data - GDPR

- According to the GDPR: any information relating to an identified or identifiable natural person ("data subject")
- The GDPR only applies to living natural persons
- Anonymous or anonymised data is not considered personal data
- What is not personal data is not covered by GDPR
 - maybe other reasons to protect this data anyway
e.g. ethical, contractual, etc.



Sensitive data – special categories of personal data

- Information that poses a risk to fundamental rights and freedoms
 - ethnic origin
 - political opinions
 - religious or ideological beliefs
 - trade union membership
 - genetic data
 - biometric data
 - data concerning the sex life or sexual orientation of a natural person

Identifiers – direct and indirect

- Article 4 (1) GDPR

*‘personal data’ means any information relating to an identified or identifiable natural person (‘data subject’); an identifiable natural person is one 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** or to **one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity** of that natural person;*

Examples of direct IDs: name, email address, unique identification number (e.g. matriculation number, social security number), IP address, etc.

Examples of indirect IDs: mainly socio-demographic information (e.g. gender, age, profession, place of residence, etc.)

- For examples of how the combination of indirect identifiers can lead to identification, see <https://www.ooa.world/>

Anonymisation

- Best solution: Do not collect identifying information
- Second best solution: Edit information to exclude identification - anonymisation
 - Remove direct IDs: e.g., name, phone number, social security number
 - Remove/replace/aggregate indirect IDs: e.g. combination of occupation, income, age, gender and place of residence
- **If personal data is not needed, delete it!**
If you keep it, have a reason.

Open Science and data protection

- Research funding agencies, journals, and research ethics:
 - **Demand: make data as openly accessible as possible (Open Data).**
- Data protection law and research ethics:
 - **Requirement: protect data as much as possible**

How can these requirements be fulfilled?



image: [source](#)

Open Science with personal/sensitive data

Sensible handling of data along the entire research data cycle (from planning to reuse).

Goal: **Making data FAIR!** so that data can be found, accessed, interoperable, and reused.

As open as possible, as closed as necessary.

Data don't have to be completely "open" to be FAIR!

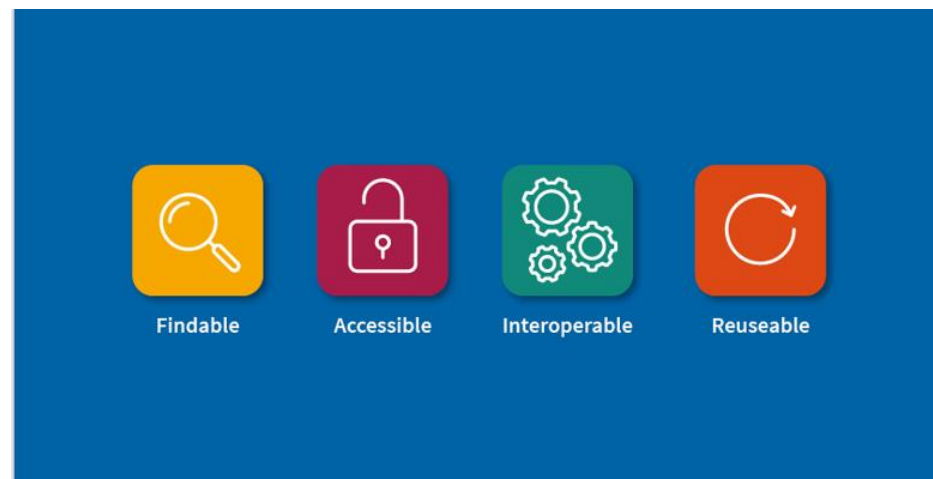


image: [source](#)

Consent

- Be aware of the **legal basis for processing** your data (Article 6 GDPR) and plan necessary measures
- In most cases: get consent!
 - Develop **consent form** – adapt to your project
- Be sure to include: **information about archiving and reuse** of the data
- Consent Form Wizard:
<https://consent.dariah.eu/>

Art. 6 GDPR

Lawfulness of processing

1. Processing shall be lawful only if and to the extent that at least one of the following applies:
 - (a) the data subject has given consent to the processing of his or her personal data for one or more specific purposes;

Data Protection

- Be sure to have the necessary rights to store personal/sensitive data long term (e.g. through consent)
- Be sure to choose a safe environment to store the data
 - Who has access?
 - What other protection measures are in place? E.g. firewall, encryption

Data Protection

- Protect peoples' privacy when publishing personal/sensitive data
- Choose and implement appropriate safety measures
 - Present information in appropriate way – by anonymising data
 - Implement access control – e.g. restrict access
 - Clarify and potentially restrict reuse options – through licenses

Questions on personal data?

Part III

Data curation and how to select a repository

Formats for quantitative & qualitative data

- Working on your own data vs. re-using data: different formats?
- Proprietary vs open formats (e.g. Stata, Sas, R, Python, PDF, Maxqda, Atlas.ti, txt, rtf, xml)

Why is documentation important?

- Easy, more comfortable re-use
- Facilitates re-submission (to journals, supervisors)
- Large quantitative datasets
- Categories and coding procedure in qualitative data
- How? Write it down!

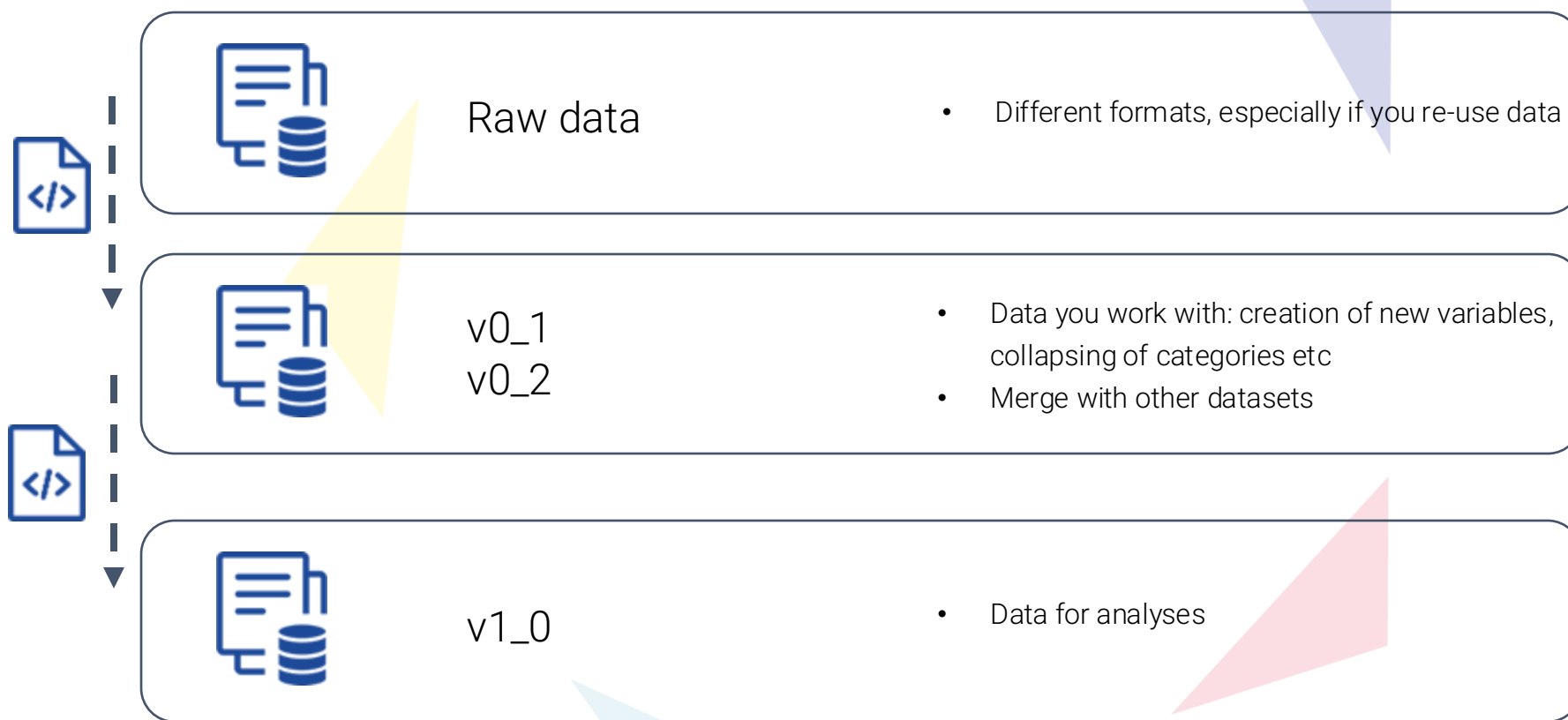
Data processing: Good practice for quantitative data - scripts

- Scripts are files in which all steps to wrangle with data are documented
- Script-based working is recommended: reuseability and documentation!

```
1  * =====  
2  * A U S S D A   D A T A   C H E C K S  
3  *  
4  * Archival#:    10#XXX#  
5  * Study Title:  #TITLE#  
6  * Date:        202#B#-#CC#-#DD#  
7  * Data Curated: #Name#  
8  *  
9  * data_checks-template version: 2.6  
10 *  
11 * *all elements enclosed by "#" have to be filled in to make the code run  
12 * =====  
13  
14
```

Source: [link](#)

Data processing: Good practice for quantitative data - versioning



Data processing: Good practice for quantitative data – variable names

- Best practice
 - 1_visit vs. visit_1
 - Var1 vs. age
 - thisisareallylongvariablename vs thisIsAReallyLongVariableName
- Codebooks help!

Data processing: Good practice for quantitative data – value labels

- Best practice
 - Label all values (categorical variables)
 - Label at least the extreme points for scales
 - Re-use values and labels (i.e. keep 0 = “no”, 1 = “yes” consistently)
 - Represent missing values consistently (i.e. “don’t know” = 99)
- Codebooks help!

Data processing: Good practice, qualitative approaches: anonymisation

- Identify identifiers!
- Document, document, document

Data processing: Good practice, qualitative approaches: anonymization plan

- Discuss the when (will data be anonymised), how and by whom!
- Read interview transcripts carefully
- Find direct and indirect identifiers
 - <Max Mustermann> <McDonalds>
- Find possible terms to replace identifiers
 - [Name] [BigCompany]
- Write it down!
- For example, see [Kalova & Fürst \(2023\)](#)

How to select a repository?

- Rankorder of findability of your data:
 1. Discipline specific repository
 2. Institutional repository
 3. Generic repository ([Zenodo](#), [The Harvard Dataverse](#), [Dryad](#), [Figshare](#))
- List of repositories:
 - www.re3data.org (self-selection)
 - [CESSDA archives](#)
 - [OpenAIRE data providers](#)
- Check your founder's requirements

How to select a repository?

- Certificates
- A persistent identifier (e.g. DOI, handle) is a minimum criteria
- Terms of use and license information online?
- If you work with personal data: Does the repository offer the necessary infrastructure to protect your research subjects privacy?
 - Which access mechanisms does the repository offer?
 - Which licenses does the repository offer?
- Metadata standards? (minimum Dublin core)

Questions on data curation and how to select a repository?

Q & A

Please give feedback:

<https://umfrage.uibk.ac.at/limesurvey/allgemein/index.php/791927?lang=en>



Sources

- CESSDA Training Team (2017 - 2019). CESSDA Data Management Expert Guide. Bergen, Norway: CESSDA ERIC. DOI: [10.5281/zenodo.3820473](https://doi.org/10.5281/zenodo.3820473).
- GO FAIR (2022). FAIR Principles. accessed: 11 April 2023. URL: <https://www.go-fair.org/fair-principles/>.
- Hönegger, Lisa (2021). Research Data Management in the Social Sciences: RDM basics, challenges and support [Presentation. Webinar for the event Research Data Management in Social Sciences]. URL: <https://hdl.handle.net/11353/10.1168878>.
- Kalová, T., Fürst, E. & Gergely, E. (2024, February 12). Research Data Management at the University of Vienna Training Course. Zenodo. <https://doi.org/10.5281/zenodo.8334152>
- Wilkinson, M. D. et al. (2016). “The FAIR Guiding Principles for scientific data management and stewardship”. In: Scientific Data 3.1. DOI: 10.1038/sdata.2016.18.

License information

The content in this presentation is licensed under a Creative Commons Attribution 4.0 International license.

This presentation is largely based on the following presentations:

- “Research Data Management in the Social Sciences” by Hirsch, Lisa and Hönegger, Lisa (2024). Zenodo. <https://doi.org/10.5281/zenodo.14277047>
- “Workshop on Research Data Management and data protection in the Social Sciences” by Hönegger, Lisa, Bodlos, Anita and Heider, Veronika (2021). Zenodo. <https://doi.org/10.5281/zenodo.5526146> (version 1.1)
- “Research Data Management at the University of Vienna Training Course” by Kalová, Tereza, Fürst, Elena and Gergely, Eva (2024, February 12). Zenodo. <https://doi.org/10.5281/zenodo.8334152> [especially the section „RDM, FAIR data and DMPs“]
- “Workshop on writing data management plans (DMP)” by Schönherr, Annemarie and Bodlos, Anita (2023).
- “Research Data Management in the Social Sciences” by Hönegger, Lisa (2021). Phaidra. <https://hdl.handle.net/11353/10.1168878>.

Contact

 researchdata@uibk.ac.at

 infra4nextgen.com

