PTOS-open-data

Jürgen Schneider

16 November 2023

Table of contents

Welcome **5**
If you like the workshop... 7

Transparency **8**
Affiliation 8
Cooperation 9

I Reasons **10**

1 Exercise: Your reasons **11**
What are your reasons? 11

2 External Incentives I **12**
2.1 Research funders 12

3 External Incentives II **14**
3.1 Policies scientific societies 14
3.2 Journal policies 15

4 Research I **16**
4.1 No data = not reproducible 16
4.2 No FAIR data = reproducibility tedious 16
4.2.1 Design 17
4.2.2 Results 17
4.2.3 Conclusions 17
4.3 No data = barrier to replication 18

5 Research II **19**
5.1 Reuse 19

6 Researcher **20**
6.1 Researchers' perceptions 20
6.2 Classical metrics 20

6.3 New metrics: Get hired! 21

II Open and FAIR Data **23**

7 Openness **24**

8 FAIRness **25**
8.0.1 Findable 25
8.0.2 Accessible 26
8.0.3 Interoperable 27
8.0.4 Reusable 27

9 Exercise **29**

III Limits **30**

10 Limits **31**
10.1 Not being open may be important 31
10.1.1 Marginalized/vulnerable groups 31
10.1.2 "Closed doors" 31
10.1.3 Costs > benefit 31
10.2 It's not all your responsibility 32

IV Workflow **34**

11 Overview **35**

12 Workflow 1 **38**
12.1 Search for reusable data 38
12.2 Decide between sharing for reuse or reproducibility 38

13 Workflow 2 **40**
13.1 Write a data management plan 40
13.2 Decide for a repository (or research data center) 41

14 Workflow 3 **42**
14.1 Informed consent 42
14.2 Decide for access restrictions 43

15 Workflow 4 **46**
15.1 Create codebook 46

V Reflection **47**

16 Exercise: Barriers **48**

Appendices **50**

References **50**

Welcome

This is a workshop on open and FAIR data.

CC-BY aukeherrema.nl

If you like the workshop...

0.0.0.1 and want to keep it forever, make it yours

For that...

1. Fork the github repo this Quarto book is based on
2. Go to settings of your new repo and go to the "pages" section. Then set the "Branch" option to `gh-pages` (leave the dropdown to the right of this at `/root`)
3. Wait a minute to let the website get deployed. You can check on the status in the "Actions" tab of your repo.
4. Back on the main repo site, click on "About" (top right). In the URL of the website, change "j-5chneider" to your username "[your github username].github.io/PTOS-open-data/" (you might need to activate GitHub Pages for that, by creating a GitHub Pages repo)
5. open your new webpage by clicking on that link in the "About" section

0.0.0.2 give it a star in GitHub

So you get noticed if I update something on the github repo.

And I get that sweet sweet dopamine. Hmm dopamine.

Transparency

Affiliation

DIPF | Leibniz Institute for Research and Information in Education
Project in the *Cooperation Center ShaReD - Sharing and Reusing Data*

See personal webpage

Cooperation

Working closely together, but not at the research data center at DIPF

Part I

Reasons

1 Exercise: Your reasons

What are your reasons?

Now that you are here, there seem to be some drivers for you to share data. What are these?

1. Go to this whiteboard
2. Take one or more of the blue rectangles from the right
3. Write your reason in the rectangle. One reason per shape.

Questions to be answered at the end?
Please put them here!

2 External Incentives I

2.1 Research funders

DFG

DFG Guidelines on the Handling of Research Data

> "Assuming that the publication of research data [...] does not conflict with the rights of third parties (in particular data protection or copyright), research data should be *made available* as soon as possible [...] that allows it to be usefully *reused* by third parties" (DFG, 2015, p. 1)

Guidelines for Safeguarding Good Research Practice. Code of Conduct

> "Where possible and reasonable, this includes making the research data [...] available" (DFG, 2019, p. 19)

BMBF

Typical section in calls for proposals

> "Zuwendungsempfänger sollen, wann immer möglich, die im Rahmen des Projekts gewonnenen Daten [...] in nachnutzbarer Form einer geeigneten Einrichtung [...] zur Verfügung stellen, um [...] Replikationen und

> gegebenenfalls Sekundärauswertungen […] zu ermöglichen. Repositorien sollten aktuelle Standards für Datenveröffentlichungen (FAIR Data-Prinzipien) erfüllen"

z.B. in
Ausschreibung "Wissenschafts- und Hochschulforschung (WiHo)" 10.11.2023
BMBF, 2023
Ausschreibung zu "klinischen Studien mit hoher Relevanz für die Patientenversorgung" 09.11.2023 BMBF, 2023

ERC

Open Research Data and Data Management Plans Information for ERC grantees

> "Grantees are required to deposit their research data in a repository and provide open access at least to those data" (ERC, 2022, p. 4)

Questions to be answered at the end?
Please put them here!

3 External Incentives II

3.1 Policies scientific societies

Examples of societies that have established a policy on open and FAIR data.

(DGfE et al., 2020; DGS, 2019; Gollwitzer et al., 2021)

"Homework": Look up the scientific society most relevant to you and check if they have a *policy/recommendation/guideline* on open science or sharing data. What does it say on openness, FAIRness and limitations?

3.2 Journal policies

Some journals encourage, some journals mandate a **data availability statement** in the manuscript.

	Not Implemented	Level I	Level II	Level III
Data Transparency	Journal encourages data sharing, or says nothing.	Article states whether data are available, and, if so, where to access them.	Data must be posted to a trusted repository. Exceptions must be identified at article submission.	Data must be posted to a trusted repository, and reported analyses will be reproduced independently prior to publication.

E.g.,

- Psychological Bulletin: Data transparency Level 2
- Meta Psychology: Data transparency Level 3
- Psychological Science: Data transparency Level 1

See the TOP Factor website to search for the data transparency rating of your favorite journal.

Evidence that **"Data available upon reasonable request"** often does not keep its promise. Only **6.8%** of data sets were acutally provided in an investigation (Gabelica et al., 2022)

Questions to be answered at the end?
Please put them here!

4 Research I

4.1 No data = not reproducible

Computational reproducibility :=

> "a second investigator (including the original researcher in the future) can recreate the final reported results of the project, including key quantitative findings, tables, and figures, given only a set of files and written instructions" (Kitzes et al., 2018, p. xxii)

➡ *Same data + same analysis = same results*

Why reproducibility?
Allows *independent researchers* to assess the analytic choices, assumptions, and implementations that led to a set of *scientific claims*.
➡ Check for validity and generalizability (Clyburne-Sherin et al., 2019; Obels et al., 2020)

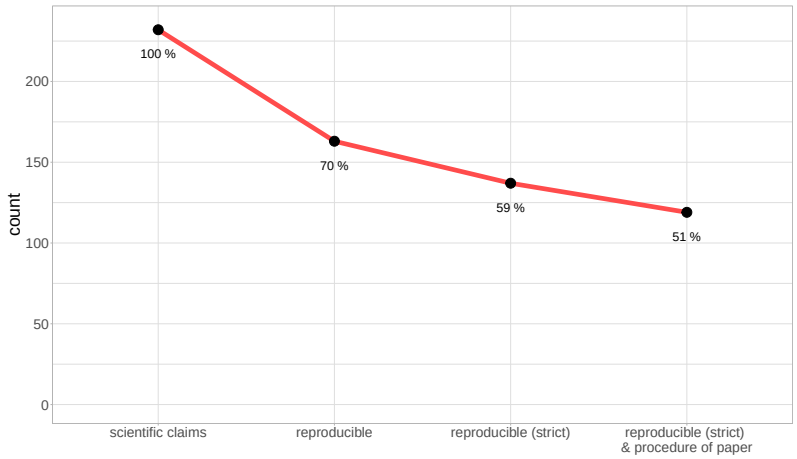
4.2 No FAIR data = reproducibility tedious

But computational reproducibility isn't as easy as it sounds (Artner et al., 2021)

4.2.1 Design

- checked 232 primary statistical claims
- from 3 journals
- after data was provided and accessible (33%, 25%, 26%)

4.2.2 Results



4.2.3 Conclusions

> **Vagueness Makes Assessing Reproducibility a Nightmare**
>
> most successful reproductions are predominantly the result of tedious and time-consuming work
>
> information about the provided raw data was often difficult to understand, and information about the relevant variables, data manipulations, and the used statistical model was often vague or inaccurate

(Artner et al., 2021, p. 12)

4.3 No data = barrier to replication

- Evidence e.g. from replication attempts in cancer biology (Errington et al., 2021)
- Due to various barriers, **50** of the **193** replication experiments could be **conducted** at all
- Missing data = major barrier to compute parameters to replicate

➜ data were open for **4** of **193** experiments

Questions to be answered at the end?
Please put them here!

5 Research II

5.1 Reuse

The reuse of research data can take many forms

Purpose	Advantage	Needs
Answer new research questions	Saves resources	analysis potential of data, good documentation (Logan et al., 2021; Steinhardt et al., 2021)
Teaching / student theses	Real-life-oriented education	good documentation
Meta-analyses	Easier estimation of parameters	Strictly reproducible code (e.g., Burgard et al., 2022)
Historical perspective	Data as historical artifacts	Potential of data varies

Questions to be answered at the end?
Please put them here!

6 Researcher

6.1 Researchers' perceptions

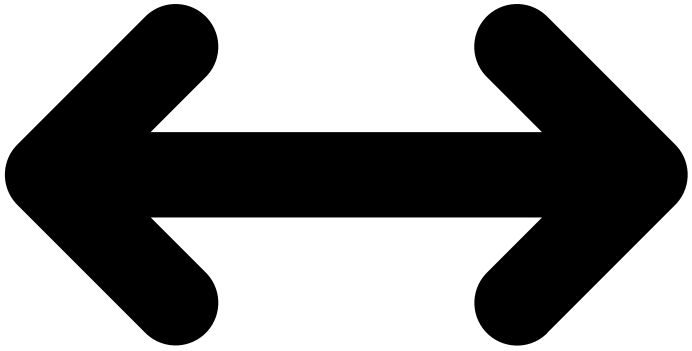
Open Data Survey (Goodey et al., 2022)

- **75%** of researchers say there is too little credit for sharing data
- main drivers:
  - perceived **higher citation** (67%)
  - increased perceived **impact** and **visibility** (61%)

6.2 Classical metrics

- Data for reuse: Additional publication (e.g., data note in F1000 Research)
- Is there higher citation?
  - studies with available data: **9%** more citations (Piwowar & Vision, 2013)
  - studies with link to data in a repository: **25%** higher citation rates (**Colavizza.etal.2020?**)

But:



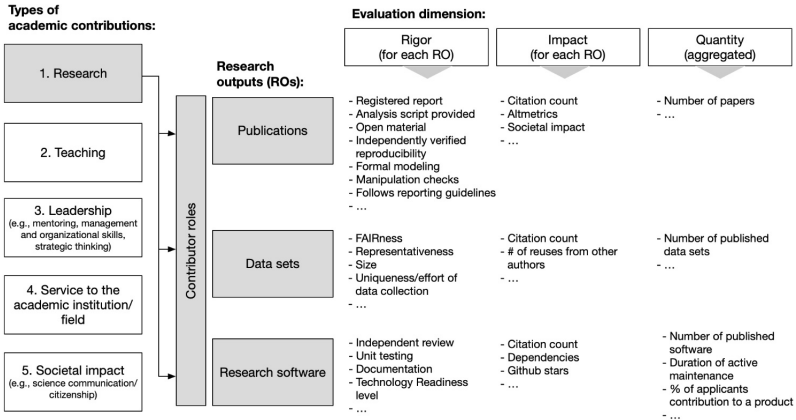
Selection bias: Willingness to share
strength of evidence and quality of reporting (**Wicherts.etal.2011?**)

But:
With higher transparency, researchers have higher trust in authors (Schneider et al., 2022)

6.3 New metrics: Get hired!

new metrics for evaluation evolving

- CoARA: "Value outputs associated with openness (FAIR data sets, […]" (CoARA, 2022, p. 21)
  - signatories: DGPs, ERC, European Commission, DFG, Leibniz Association, …
- Example: DGPs recommendations on hiring and promotion (Gärtner et al., 2022; Schönbrodt et al., 2022)



(Schönbrodt et al., 2022, p. 4)

Questions to be answered at the end?
Please put them here!

Part II

Open and FAIR Data

7 Openness

Definition:

- **anyone**
- can **readily access** the data
- at no more than a **reasonable reproduction cost** (i.e., internet connection)

(Open Knowledge Foundation, 2023)

> Openness is not a dogma and not a dichotomy
>
> "As open as possible as closed as necessary"
> (European Commission, 2023, p. 36)

Questions to be answered at the end?
Please put them here!

8 FAIRness

Purpose:

> "enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals" (Wilkinson et al., 2016, p. 1)

See also go-fair.org

> **FAIRness vs. openness**
>
> "does not necessarily mean that data has to be "open" [...] even highly protected data can be FAIR data" (Kraft, 2023)

8.0.1 Findable

The problem:
Just because we provide data online, doesn't mean that others will find it.

We could have the greatest data set to answer further research questions - if our colleagues don't know it exists or can't locate the data, openness will be of little value.

The solutions:

- Get a persistent identifier (e.g., DOI), where you provided your data

  - search for a research data center that fits your needs: re3data.org
  - recommended research data centers: Verbund FDB (education, Germany), RDC at ZPID (psychology, Germany), ...
  - recommended repositories: Zenodo, psycharchives.org, osf.io, ...

- Mention DOI in publication that builds on this data (e.g., in the "data accessibility statement")
- Describe your data as richly as possible (metadata). *Research data centers* offer form fields tailored to the discipline or data type. With *repositories* use alternative possibilities, such as keyword fields.

  - e.g., which variables does the quantitative data set contain?
  - e.g., which topics does your data cover?
  - e.g., which population did you draw your sample from?

8.0.2 Accessible

The problem:
Just because others find our data doesn't mean the *access barriers* are as low as possible and doesn't mean they know *in which way* they are allowed to access it. Examples:

- Providing a link to the data in the text of a paywalled journal article
- Unclear licensing / use conditions when providing data (e.g., are non-researchers allowed to access the data or is it only open for qualified researchers?)

The solutions:

- Make sure access is free of charge (or as cheap as possible)

  - e.g., by providing link to data in publicly accessible sections of journal articles that are not open access
  - e.g., by using repositories or research data centers that allow access free of charge

- Make sure users know if they can access and under which conditions
  - e.g., *research data centers* ensure that terms of use are clear (who may access under what conditions) and offer different levels of access restriction
  - e.g., on *repositories* provide a readme-file and an open license (e.g., CC0, CC-BY, CC-BY-SA) with data sets for access cases

8.0.3 Interoperable

The problem:
Just because others downloaded our data doesn't mean they can open and manipulate it.

The solutions:

- Use file formats with open licenses
  - e.g., tabular data: CSV (with additional labelling script), RData
  - e.g., text data: PDF, HTML, ODT, RTF
- Make sure users know how different files are related to one another
  - e.g., define which file contains student data and which teacher data
  - e.g., define which file contains data from cohort 1 and which cohort 2, …

8.0.4 Reusable

The problem:
Just because others opened our data doesn't mean they understand the data and its use-conditions. Examples:

- Others can't understand what the column names of the tabular data set mean: Which columns in the data set relate to which variables in the journal article?

- Can someone from sociology use the data set from psychology they found on osf.io?
- Does someone reusing a data set have to cite the authors?

The solutions:

- Adhere to standards in folder organization
  - e.g., PSYCH-DS (see technical specification draft)
- Rich description/explanation of what user will find *in* the data set (meta descriptions about the data set *as a whole*, as for accessibility)
  - e.g., provide a codebook. How to semi-automatically create a codebook, see the R package `codebook`
- Provide a license for the use-cases
  - again, *research data centers* ensure that terms of use are clear (who may use under what conditions)
  - again, on *repositories* provide a readme-file and an open license (e.g., CC0, CC-BY, CC-BY-SA) with data sets for the use-cases

(FAIR principles and the role of scientists: Kraft, 2023)

Questions to be answered at the end?
Please put them here!

9 Exercise

1. Go to this repository
2. Discuss for **which purposes** you consider this type of sharing to be **suitable** / **less suitable**
3. Discuss what you think makes this type of data sharing **FAIR** and what could be **improved**

Questions to be answered at the end?
Please put them here!

Part III

Limits

10 Limits

10.1 Not being open may be important

When does "as closed as necessary" apply?

The protection of individuals comes first and is more important than the potential reuse of data

10.1.1 Marginalized/vulnerable groups

- e.g., individuals traumatized by war or who experienced sexual abuse
- At the same time: Can sharing data help to protect these groups from being over-researched (possibly re-traumatization)?

10.1.2 "Closed doors"

- cases in which field access is obstructed or denied by the data provision (Prosser et al., 2022)
- cases in which sharing reduces the reduces the willingness to participate

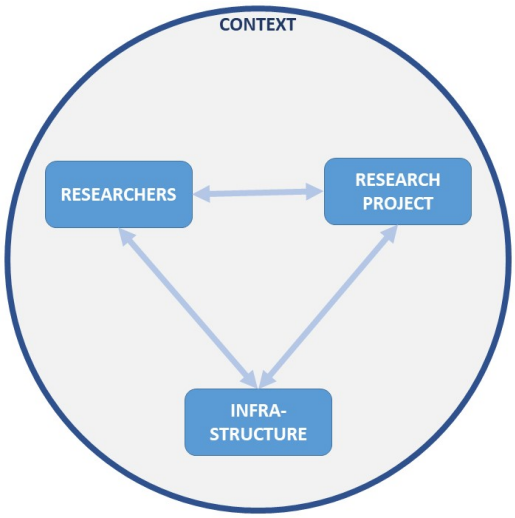
10.1.3 Costs > benefit

- e.g., low reuse potential -> publish for reproducibility
- e.g., when obtaining permission from the school authorities is extremely complicated
- e.g., epistemic problem: context of data collection is highly relevant and other researchers "haven't been there" (Mauthner et al., 1998) -> publish for intersubjective comprehensibility

- e.g., can't anonymize data -> synthpop, create input-output-documents via RMarkdown/Quarto

10.2 It's not all your responsibility

- Responsibility of opening research is a **collective responsibility** in the **"research ecosystem"** (European Commission, 2018; RfII, 2019)
- Researchers are just one part of this



Infrastructure

- Does suitable infrastructure exist?
- Is it "easy to use" and cheap?
- Is it tailored to my needs and type of data?
- Does it allow the implementation of FAIR data?
- Are there resources to support data sharing?

Context

- Do scientific societies, journals, or research funders encourage sharing?
- Is it common practice ("culture") in my field of research to share data? (Bishop, 2006)
- Are there standards established for data sharing?
- Do ethics committees request detailed reasoning for the intent to collect own data as opposed to reusing?

Research project

See reasons above on

- Marginalized/vulnerable groups
- "Closed doors"
- Costs > benefit

Researchers

- Willingness and concerns toward sharing data (Mozersky et al., 2021)
- Knowledge, experiences and skills with relevant processes

Questions to be answered at the end?
Please put them here!

Part IV

Workflow

11 Overview

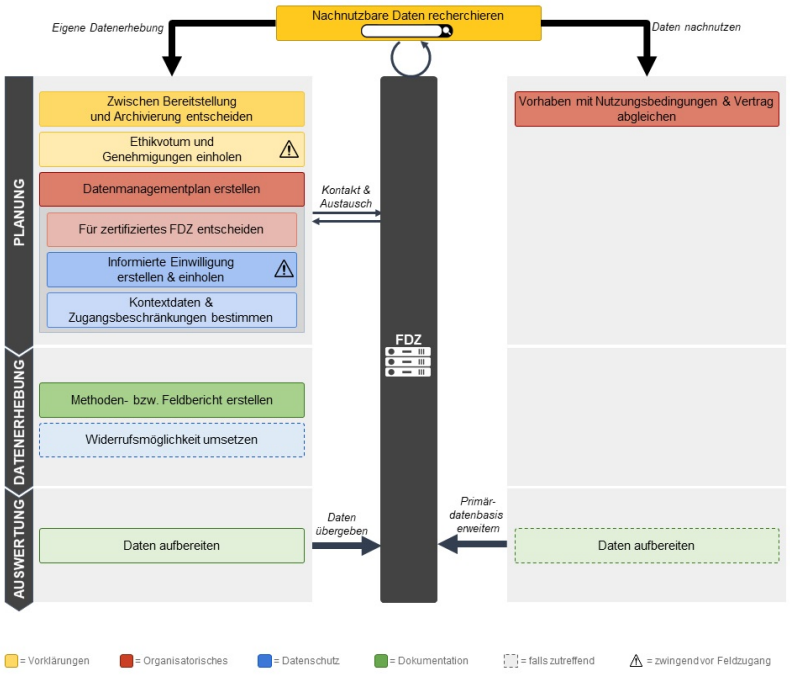


Table: Steps & resources

What	Resources
Search for reusable data	Research data centers have searchable databases re3data.org, Verbund FDB, RDC at ZPID

Sharing for reuse or reproducibility	Costs: How much effort is required? Consent for reuse available?Benefits: Analysis potential, Quality of data	
Data management plan	Templates/ToolsStamp (VerbundFDB)DMP template (ERC)Online tool (Open Aire)Online tool “DataWiz” (ZPID)StandardD-Psy-FAIR (ZPID)	
Decide for a repository or RDC	Research data center:Search: re3data.orgVerbund FDB (education, Germany)RDC at ZPID (psychology, Germany)RepositoriesZenodopsycharchives.orgosf.io	

Informed consent	Checklist, GermanTemplate, German standard languageTemplate, German plain languageTemplate (qualitative data), GermanExplanations + template (DGPs), GermanOverview + links (ZPID), EnglishExplanations + definitions (Michigan Tech), US Examples (VerbundFDB)Examples (DGPs)There are alternatives to restricting access!
Access restrictions	
Create codebook	R Package `codebook`codebook from DataWiz

Questions to be answered at the end?
Please put them here!

12 Workflow 1

12.1 Search for reusable data

Why?
For resource intensive data collections this could save you a lot of time and money

Resources
Research data centers have searchable databases

- re3data.org (Database to search for databases)
- Verbund FDB (Education)
- RDC at ZPID (Psychology)

12.2 Decide between sharing for reuse or reproducibility

Why?

- Providing data at a reseach data center costs time and money for you and the data center
- Typically,
  - sharing for **reuse** purposes is suited for **research data centers**
  - sharing for **reproducibility** purposes is suited for **repositories**

Resources
I am not aware of any standards to make this decision. Here are a couple of guidelines to decide, if your data is fit for reuse

- Costs:
  - How much **effort** is required for well-documented data sharing (e.g., Does a codebook exist? What steps are necessary for data cleaning/editing?)
  - Is there **consent for reuse** available or would it have to be obtained retrospectively?
- Benefits:
  - Analysis **potential** (e.g., not fully analyzed, type of data, connected with other data sources)
  - **Quality** of data (e.g., representative, size, special features of sample)

Questions to be answered at the end?
Please put them here!

13 Workflow 2

13.1 Write a data management plan

Why?

- supports researchers in the process of generating FAIR research data
- ensures good scientific practice

Resources
Templates and online tools for specific applications

- Template Standardized data management plan for educational research "Stamp" (FDZ Bildung): Detailed form with specific instructions and assistance
- For rapid documentation: DMP template (European Research Council) with four open questions
- Online tool for machine-readable DMPs (Open Aire)
- Online tool supporting the creation of a DMP (ZPID)

Standards

- D-Psy-FAIR (Blask et al., 2022)
  - Manual
  - Online Tutorial
  - Slides

13.2 Decide for a repository (or research data center)

Why?
Deciding on a specific repository or research data center early on helps to

- meet its requirements
- identify whether there are templates available
- identify whether the employees can support the sharing process (in the case or RDC)

Resources

- search for a research data center that fits your needs: re3data.org
- recommended research data centers: Verbund FDB (education, Germany), RDC at ZPID (psychology, Germany), ...
- recommended repositories: Zenodo, psycharchives.org, osf.io, ...

Questions to be answered at the end?
Please put them here!

14 Workflow 3

14.1 Informed consent

Why?

- *Personal data* is subject to General Data Protection Regulation (GDPR)
- Informed consent must therefore fulfill a number of requirements. E.g.,
  - purpose of data collection (includes *sharing the data* and *future use*)➡ therefore often "broad consent"
  - participation is voluntary and without disadvantages
  - revocation is possible at any time (until anonymized)

Resources

- Checklist of legally compliant consent forms, German (VerbundFDB, 2019)
- Template for informed consent, German standard language (VerbundFDB, 2018)
- Template for informed consent, German plain language (VerbundFDB, 2018)
- Template for informed consent, German (Qualiservice)
- Explanations including template (DGPs), German (not specific for reuse)
- Overview and links concerning informed consent (ZPID), English
- Explanations and definitions around informed consent (Michigan Tech), US (not specific for reuse)

14.2 Decide for access restrictions

Why?

- Some data *cannot or should not be anonymized* (e.g., losing their reuse potential)
  - Therefore access needs to be restricted to certain groups (as defined in consent form)
- Some researchers fear being *scooped* (**Laine.2017?**)

With repositories...

restriction levels are usually limited to

- public (everybody sees everything)
- private (only you and your collaborators see everything)

With research data centers...

there are different restriction levels possible for different files (*in the same project*). Restriction levels depend on what the research data center offers.

Level	Prerequesite	For what
Public Use-file		anonymized data, codebooks, transcription rules
Student Use-file	Short application states use purpose	non-anonymized data with right to use for teaching

Scientific Use-file	Longer application states use purpose, handling of data, and data analyses; identification via PostIdent	non-anonymized data with right to use for research
Remote Access	… + access only via virtual machine	non-anonymized sensible data with right to use for research
Safe room	… + access only in person at research institute	non-anonymized very sensible data with right to use for research

An example: Project DESI, where

- codebooks are publicly accessible (files on the right side)
- video data are restricted for scientific use (files on the bottom of page)

Resources

- Examples of restriction levels (VerbundFDB) (Meyermann & Porzelt, 2019, p. 30f)
- Examples of restriction levels (DGPs) (DGPs, 2021, p. 141ff)

Alternatives

- *Embargo period*
  - Specify a time period, before data go public
  - Possible with research data centers and some repositories
- *Exclude certain research questions* from reuse
  - Specify these research questions in the terms of use

  - Usually only possible with research data centers, except you are writing a very good license yourself

- Create *synthetic data* (e.g., with R package `synthpop`)
  - Mimics the properties of your data
  - Then possible to share this synthetic data set

Questions to be answered at the end?
Please put them here!

15 Workflow 4

15.1 Create codebook

Why?
Remember?

> Vagueness Makes Assessing Reproducibility a Nightmare
> most successful reproductions are predominantly the result of tedious and time-consuming work information about the provided raw data was often difficult to understand, and information about the relevant variables, data manipulations, and the used statistical model was often vague or inaccurate (Artner et al., 2021, p. 12)

Resources

- R Package `codebook`
  - semi-automated creation of a codebook (depending on how well prepared/labelled your data set is)
  - in combination with the formr survey framework, this package saves you a ton of time
  - still has some minor bugs, be prepared to mingle with it
- Codebook as a result of using DataWiz

Questions to be answered at the end?
Please put them here!

Part V

Reflection

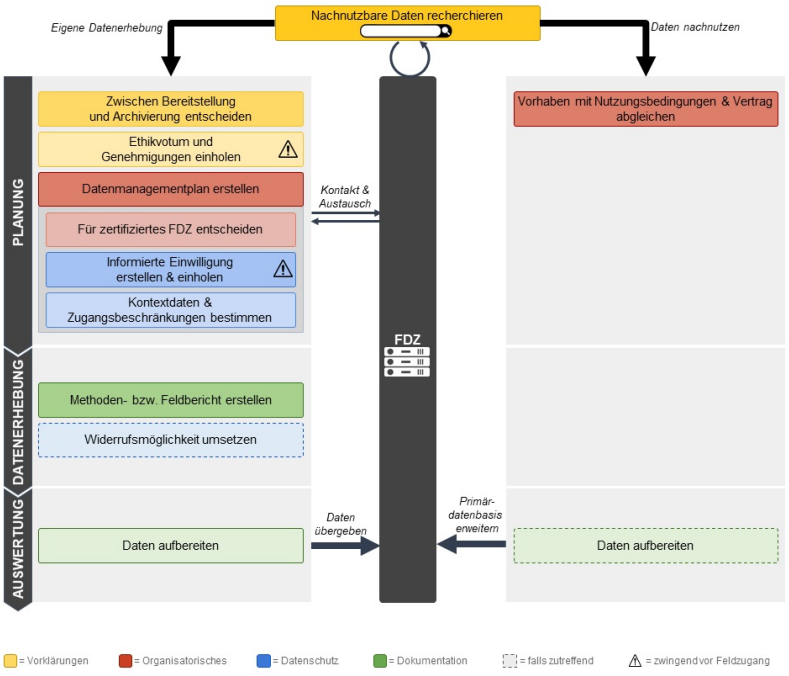
16 Exercise: Barriers

Let's assume: Sharing data is possible. In one way or the other.

Why isn't everybody sharing all data "as open as possible as closed as necessary"?

1. Check out the flow chart again (see below)
2. Individually [3min]: Reflect on
   - What is a barrier/challenge to you?
   - What might be a powerful barrier/challenge for others?
3. In the breakout rooms [12min]: Discuss
   - What are the biggest barriers for you/others?
   - What would be different in an "ideal world" that would lead to you/others overcoming these barriers?
   - Document the barriers & your needs in this sheet

Questions to be answered at the end?
Please put them here!


Eigene Datenerhebung
Nachnutzbare Daten recherchieren
Daten nachnutzen
PLANUNG
DATENERHEBUNG
AUSWERTUNG
Zwischen Bereitstellung und Archivierung entscheiden
Ethikvotum und Genehmigungen einholen
Datenmanagementplan erstellen
Für zertifiziertes FDZ entscheiden
Informierte Einwilligung erstellen & einholen
Kontextdaten & Zugangsbeschränkungen bestimmen
Methoden- bzw. Feldbericht erstellen
Widerrufsmöglichkeit umsetzen
Daten aufbereiten
Kontakt & Austausch
Daten übergeben
FDZ
Primär-datenbasis erweitern
Vorhaben mit Nutzungsbedingungen & Vertrag abgleichen
Daten aufbereiten
= Vorklärungen
= Organisatorisches
= Datenschutz
= Dokumentation
= falls zutreffend
= zwingend vor Feldzugang

References

Artner, R., Verliefde, T., Steegen, S., Gomes, S., Traets, F., Tuerlinckx, F., & Vanpaemel, W. (2021). The reproducibility of statistical results in psychological research: An investigation using unpublished raw data. *Psychological Methods*, *26*(5), 527–546. https://doi.org/10.1037/met0000365

Bishop, L. (2006). A Proposal for Archiving Context for Secondary Analysis. *Methodological Innovation Online*, *1*(2), 10–20. https://doi.org/10.4256/mio.2006.0008

Blask, K., Latz, M., Müller, M.-L., & Gellert, S. (2022). *D-PsyFAIR: Four simple steps to sustainable data documentation in psychology*. PsychArchives. https://doi.org/10.23668/PSYCHARCHIVES.12180

Burgard, T., Bosnjak, M., & Studtrucker, R. (2022). PsychOpen CAMA: Publication of community-augmented meta-analyses in psychology. *Research Synthesis Methods*, *13*(1), 134–143. https://doi.org/10.1002/jrsm.1536

Clyburne-Sherin, A., Fei, X., & Green, S. A. (2019). Computational Reproducibility via Containers in Psychology. *Meta-Psychology*, *3*. https://doi.org/10.15626/MP.2018.892

CoARA. (2022). *Agreement on Reforming Research Assessment*.

DFG. (2015). *DFG Guidelines on the Handling of Research Data*.

DFG. (2019). *Guidelines for Safeguarding Good Research Practice. Code of Conduct*.

DGfE, GEBF, & GFD. (2020). *Gemeinsame Stellungnahme der Deutschen Gesellschaft für Erziehungswissenschaft (DGfE), der Gesellschaft für Empirische Bildungsforschung (GEBF) und der Gesellschaft für Fachdidaktik (GFD) zur Archivierung, Bereitstellung und Nachnutzung von Forschungsdaten in den Erziehungs- und Bildungswissenschaften und Fachdidaktiken*.

DGPs. (2021). Management und Bereitstellung von

Forschungsdaten in der Psychologie: Überarbeitung der DGPs-Empfehlungen: DGPs-Kommission „Open Science“ (beschlossen durch den Vorstand der DGPs am 26. 06. 2020). *Psychologische Rundschau*, *72*(2), 132–146. https://doi.org/10.1026/0033-3042/a000514

DGS. (2019). *Stellungnahme: Bereitstellung und Nachnutzung von Forschungsdaten in der Soziologie*.

ERC. (2022). *Open Research Data and Data Management Plans. Information for ERC grantees*.

Errington, T. M., Denis, A., Perfito, N., Iorns, E., & Nosek, B. A. (2021). Challenges for assessing replicability in preclinical cancer biology. *eLife*, *10*, e67995. https://doi.org/10.7554/eLife.67995

European Commission. (2018). *OSPP-REC: Open Science Policy Platform Recommendations.* Publications Office.

European Commission. (2023). *Horizon Europe (HORIZON). HE Programme Guide. Version 4.0*. Publications Office.

Gabelica, M., Bojčić, R., & Puljak, L. (2022). Many researchers were not compliant with their published data sharing statement: A mixed-methods study. *Journal of Clinical Epidemiology*, *150*, 33–41. https://doi.org/10.1016/j.jclinepi.2022.05.019

Gärtner, A., Leising, D., & Schönbrodt, F. D. (2022). *Responsible Research Assessment II: A specific proposal for hiring and promotion in psychology*. PsyArXiv. https://doi.org/10.31234/osf.io/5yexm

Gollwitzer, M., Abele-Brehm, A., Fiebach, C. J., Ramthun, R., Scheel, A., Schönbrodt, F., & Steinberg, U. (2021). Management und Bereitstellung von Forschungsdaten in der Psychologie: Überarbeitung der DGPs-Empfehlungen: DGPs-Kommission „Open Science“ (beschlossen durch den Vorstand der DGPs am 26. 06. 2020). *Psychologische Rundschau*, *72*(2), 132–146. https://doi.org/10.1026/0033-3042/a000514

Goodey, G., Hahnel, M., Zhou, Y., Jiang, L., Chandramouliswaran, I., Hafez, A., Paine, T., Gregurick, S., Simango, S., Peña, J. M. P., Murray, H., Cannon, M., Grant, R., McKellar, K., & Day, L. (2022). *The State of Open Data 2022*. https://doi.org/10.6084/m9.figshare.21276984.v5

Kitzes, J., Turek, D., & Deniz, F. (Eds.). (2018). *The practice*

of reproducible research: Case studies and lessons from the data-intensive sciences. University of California Press.

Kraft, A. (2023). *The FAIR Data Principles*. https://doi.org/10.23668/PSYCHARCHIVES.13577

Logan, J. A. R., Hart, S. A., & Schatschneider, C. (2021). Data Sharing in Education Science. *AERA Open*, *7*, 233285842110064. https://doi.org/10.1177/23328584211006475

Mauthner, N. S., Parry, O., & Backett-Milburn, K. (1998). The Data are Out there, or are They? Implications for Archiving and Revisiting Qualitative Data. *Sociology*, *32*(4), 733–745. https://doi.org/10.1177/0038038598032004006

Meyermann, A., & Porzelt, M. (2019). *Datenschutzrechtliche Anforderungen in der empirischen Bildungsforschung. Eine Handreichung. Version 2.1*. https://doi.org/10.25656/01:21990

Mozersky, J., McIntosh, T., Walsh, H. A., Parsons, M. V., Goodman, M., & DuBois, J. M. (2021). Barriers and facilitators to qualitative data sharing in the United States: A survey of qualitative researchers. *PLOS ONE*, *16*(12), e0261719. https://doi.org/10.1371/journal.pone.0261719

Obels, P., Lakens, D., Coles, N. A., Gottfried, J., & Green, S. A. (2020). Analysis of Open Data and Computational Reproducibility in Registered Reports in Psychology. *Advances in Methods and Practices in Psychological Science*, *3*(2), 229–237. https://doi.org/10.1177/2515245920918872

Open Knowledge Foundation. (2023). What is Open Data? In *Open Data Handbook*. https://opendatahandbook.org/guide/en/what-is-open-data/.

Piwowar, H. A., & Vision, T. J. (2013). Data reuse and the open data citation advantage. *PeerJ*, *1*, e175. https://doi.org/10.7717/peerj.175

Prosser, A. M. B., Hamshaw, R. J. T., Meyer, J., Bagnall, R., Blackwood, L., Huysamen, M., Jordan, A., Vasileiou, K., & Walter, Z. (2022). When open data closes the door: A critical examination of the past, present and the potential future for open data guidelines in journals. *British Journal of Social Psychology*, 12576. https://doi.org/10.1111/bjso.12576

RfII. (2019). *Herausforderung Datenqualität – Empfehlungen zur Zukunftsfähigkeit von Forschung im digitalen Wandel*.

Schneider, J., Rosman, T., Kelava, A., & Merk, S. (2022). Do Open-Science Badges Increase Trust in Scientists Among Undergraduates, Scientists, and the Public? *Psychological Science*, *33*(9), 1588–1604. https://doi.org/10.1177/09567976221097499

Schönbrodt, F., Gärtner, A., Frank, M., Gollwitzer, M., Ihle, M., Mischkowski, D., Phan, L. V., Schmitt, M., Scheel, A. M., Schubert, A.-L., Steinberg, U., & Leising, D. (2022). *Responsible Research Assessment I: Implementing DORA for hiring and promotion in psychology.* https://doi.org/10.23668/PSYCHARCHIVES.8162

Steinhardt, I., Fischer, C., Heimstädt, M., Hirsbrunner, S. D., İkiz-Akıncı, D., Kressin, L., Kretzer, S., Möllenkamp, A., Porzelt, M., Rahal, R.-M., Schimmler, S., Wilke, R., & Wünsche, H. (2021). *Opening up and Sharing Data from Qualitative Research: A Primer: Results of a workshop run by the research group „Digitalization and Science“ at the Weizenbaum Institute in Berlin on January 17, 2020.* https://doi.org/10.34669/WI.WS/17

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*(1), 160018. https://doi.org/10.1038/sdata.2016.18