How to share psychological research data

Recommendations from the German Psychological Society

Open Science Committee of the DGPs:
Mario Gollwitzer, Andrea Abele-Brehm, Christian Fiebach,
Roland Ramthun, Anne Scheel, Felix Schönbrodt, & Ulf Steinberg

Prof. Dr. Felix Schönbrodt
Ludwig-Maximilians-Universität
München

www.nicebread.de
www.researchtransparency.org
@nicebread303

Data sharing is becoming the norm in psychological science

(slowly but steadily)

Funders demand it

„We expect our researchers to maximise the availability of research data, software and materials with as few restrictions as possible. **As a minimum, the data underpinning research papers should be made available to other researchers at the time of publication. [...]**
Wellcome will also consider whether researchers have managed and shared their research outputs in line with our requirements, as a critical part of the end of grant reporting process"

„The NIH expects and supports the **timely release and sharing of final research data** from NIH-supported studies for use by other researchers. [...] ... are expected to include a plan for data sharing or state why data sharing is not possible."

„Ebenso erwartet der SNF, **dass Daten, die während der Forschungsarbeiten produziert wurden, künftig auf öffentlich zugänglichen, digitalen Datenbanken archiviert werden**, sofern dem keine rechtlichen, ethischen, urheberrechtlichen oder andere Klauseln entgegenstehen."

„**It is recommended to make all research data, code and software created within a research project available for reuse**, for example under Creative Commons, GNU, MIT or another relevant licence."
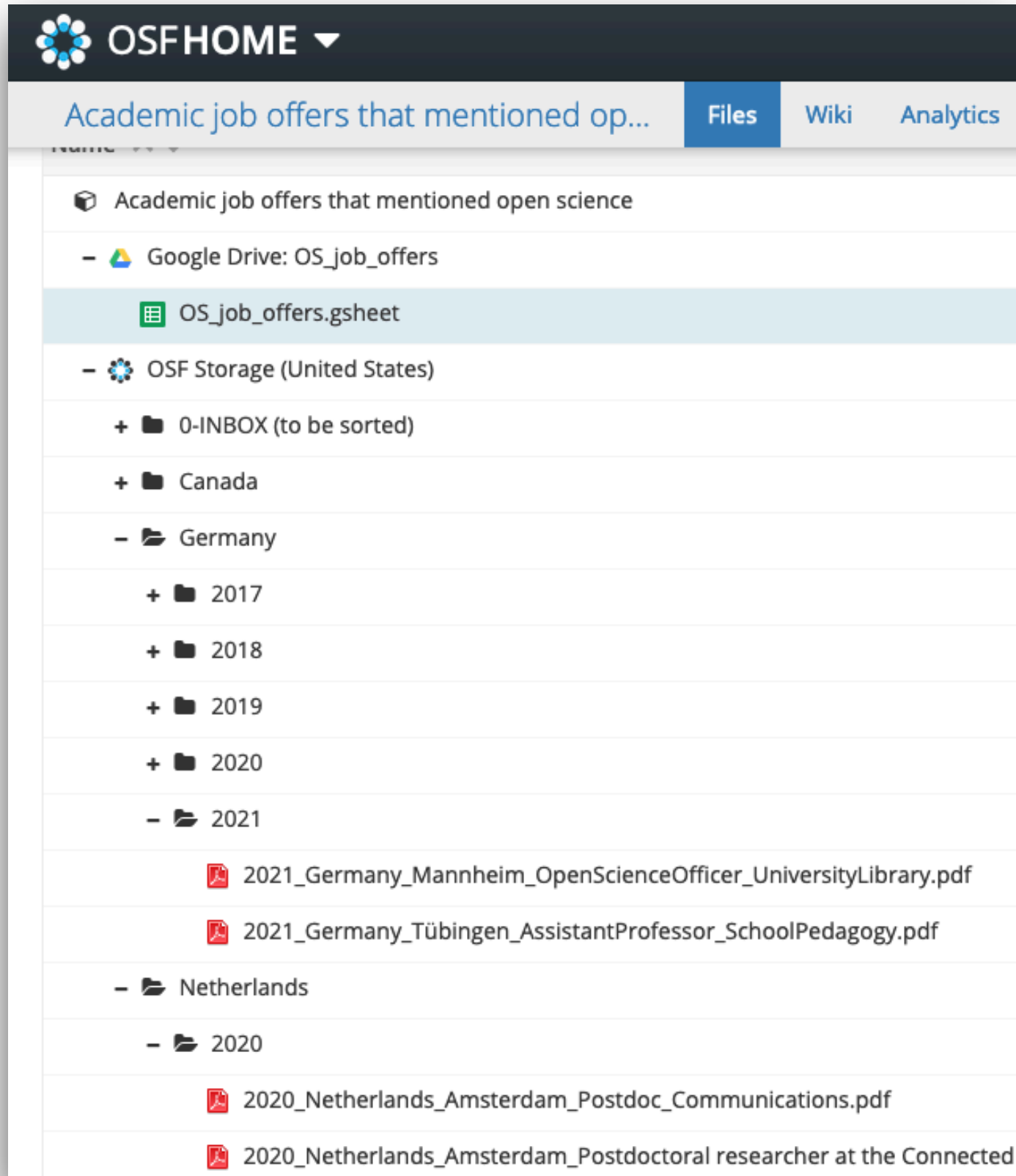
„FAIR (Findable, Accessible, Interoperable and Re-usable data) and open data sharing should **become the default for the results of EU-funded scientific research.**"

Journals demand it: https://topfactor.org/

Journal	Total	Data Citation	Data Transparency	Analysis Code Transparency
Meta-Psychology ⧉ LNU Open	27	3	3	3
Archives of Scientific Psychology ⧉ American Psychological Association	3	0	2	0
Journal of Research in Personality ⧉ Elsevier	19	0	2	2
Social Psychological Bulletin ⧉ PsychOpen	18	1	2	1
Collabra ⧉ University of California Press	20	2	2	2
Social Cognition ⧉ Guilford Press	13	2	2	2
Personality Science ⧉	24	3	2	2
Cortex ⧉ Elsevier	23	3	2	2
Royal Society Open Science ⧉ Royal Society Publishing	14	2	2	2
Advances in Methods and Practices in Psychological Science ⧉ SAGE	25	2	2	2
Science ⧉ AAAS	11	2	2	2

- **TOP Level 3** = Data <u>must</u> be posted to a trusted repository, and reported analyses will be reproduced independently prior to publication.

- **TOP Level 2** = Data <u>must</u> be posted to a trusted repository. Exceptions must be identified at article submission.

- As of Dec 2020, the TOP factor website lists for psychology:
  - 1 journal with data transparency level 3, and
  - 14 journals with level 2

4

More and more universities value it

- 85 documented cases where academic job offers (mostly professorship positions) had an „open science" aspect as desirable or required criterion.

- See collection at https:// osf.io/7jbnt/

It's not just carrots and sticks …

Table 1. Items and responses regarding attitudes toward Open Science and public Data Sharing (*N* = 308)

Item	Response categories			
	Do not agree	*Slightly disagree*	*Slightly agree*	*Strongly agree*
5. Science profits from data sharing.	3%	5%	35%	57%
13. Science should be open and transparent.	1%	1%	22%	76%

92%
98%

Abele-Brehm, A. E., Gollwitzer, M., Steinberg, U., & Schönbrodt, F. D. (2019). Attitudes Toward Open Science and Public Data Sharing. Social Psychology, 1–9. http://doi.org/10.1027/1864-9335/a000384

Desirability of Data Sharing (n=600)

"How desirable do you think data sharing is for your research field?" — 6% / 18% / 77%

"How desirable do you think data sharing is for your current research project?" — 9% / 32% / 59%

Legend: "Very Undesirable" | "Undesirable" | "Neutral" | "Desirable" | "Very Desirable"
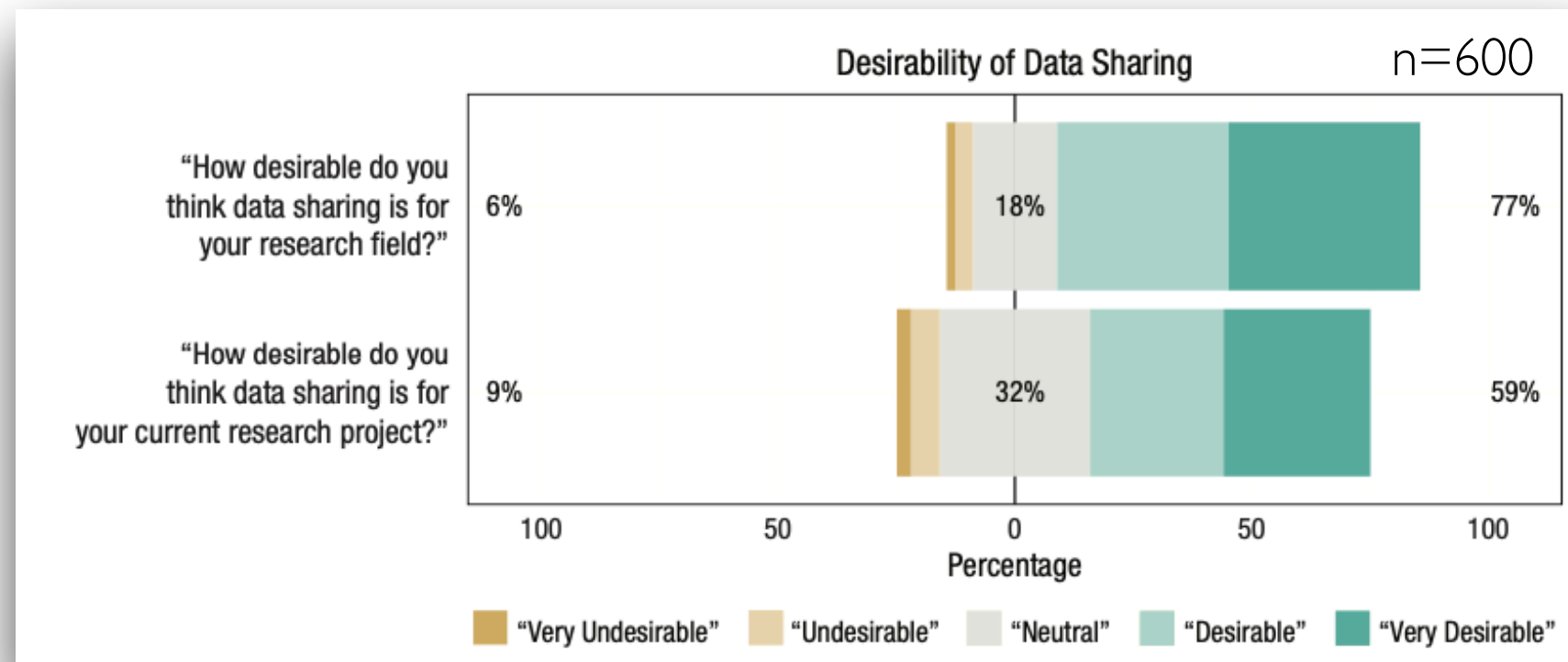
Houtkoop, B. L., Chambers, C., Macleod, M., Bishop, D. V. M., Nichols, T. E., & Wagenmakers, E. J. (2018). Data Sharing in Psychology: A Survey on Barriers and Preconditions. Advances in Methods and Practices in Psychological Science, 63(1), 251524591775188–16. http://doi.org/10.1177/2515245917751886

6

Practical questions

- How can we protect participants' privacy and be as open as possible at the same time?

- In what way can we ensure the legitimate interest of data providers to enact their "right of first usage", in particular when early career researchers invest a lot of time and effort in data collection?

- Is a co-authorship of data sharers warranted, and if so, under which conditions? What about research parasites?

- What happens in the case of misconduct?

For many of these questions, no explicit or implicit norms have been established yet.

The development of the DGPs recommendations on data management and data sharing

Phase 1

- **Dec 2015**: Working group finishes first draft

- **Jan - Feb 2016**: Email to all DGPs (German Psychological Society) members + 6 weeks online discussion forum

- **July 2016**: In-person meeting with several board members, some active discussants from the online discussion, representatives from early career researchers, and a representative from the German Research Foundation (DFG).

- **July 2016**: Sent „semi-final" draft to the speakers of the 15 sections of the DGPs, asking (again) for feedback. Roundtable with the „DFG-Fachkollegium Psychologie"

- **Sep 2016**: Formal adoption of the recommendations by the Executive Committee of the DGPs; announce an evaluation and revision of the recommendations in 5 years.

- **Jan 2017**: Publication of the recommendations

Phase 11

- **Nov 2017**: Run an evaluation study on the recommendations (published in *Abele-Brehm et al., 2019: Attitudes toward open science and public data sharing*)

Evaluation study

e.g. „I have more trust into research findings when the respective data are published."

e.g. „I am afraid that I could have a competitive disadvantage when I share my data for usage by others."
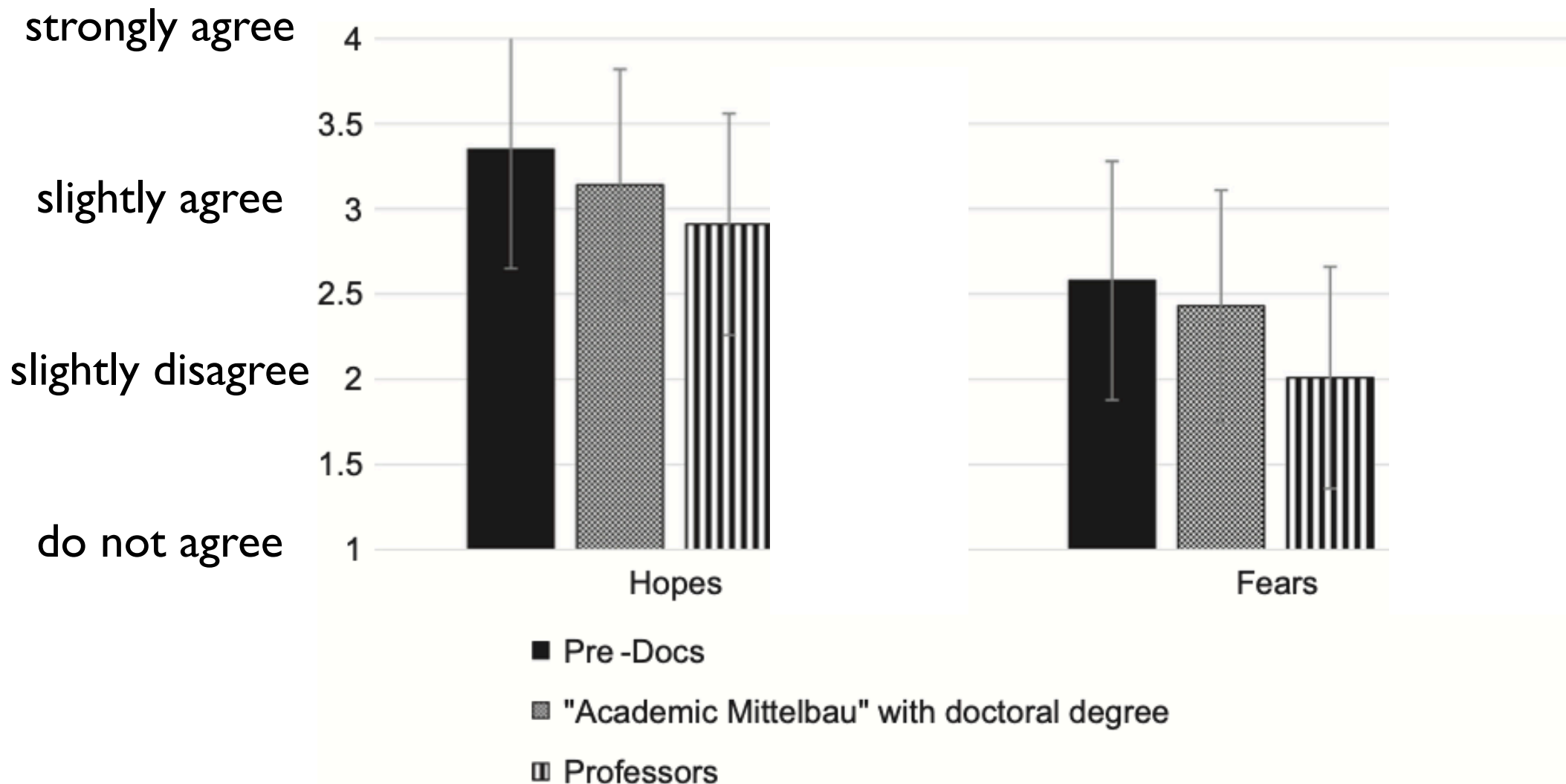
+ open ended questions about hopes and fears associated with data sharing

strongly agree — 4
slightly agree — 3
slightly disagree — 2
do not agree — 1

Hopes   Fears

■ Pre-Docs
▨ "Academic Mittelbau" with doctoral degree
▥ Professors

Abele-Brehm, A. E., Gollwitzer, M., Steinberg, U., & Schönbrodt, F. D. (2019). Attitudes Toward Open Science and Public Data Sharing. Social Psychology, 1–9. http://doi.org/10.1027/1864-9335/a000384

Phase II

- **Nov 2017**: Run an evaluation study on the recommendations (published in *Abele-Brehm et al., 2019: Attitudes toward open science and public data sharing*)

- **Dez 2018**: The new DGPs committee „Open Science" starts working; main task: work on revision of data sharing recommendations. Main goals:

  - work with the accumulated feedback of the last years; revise and clarify where necessary

  - add section on GDPR

  - add perspective of a repository (i.e., the ZPID PsychArchives)

- **During 2020**: Multiple feedback loops with representatives from several groups

- **Sep 2020**: Publication of the revised recommendations as preprints.

Core aspects of the recommendations

Definitions

- **Raw data** are the original recordings collected from a source, e.g.

  - marks on a paper questionnaire

  - video recordings

  - physiological recordings in their raw form

- **Primary data** is the first transfer of raw data into a digital format, without any changes.

  - often, raw data and primary data are equivalent

- **Metadata** provide structured information on the primary data set

- **Secondary data** refers to data that has undergone some type of preprocessing

  - e.g., transformations of variables, scale computations, aggregations, exclusions

This should be provided as open data

Privacy considerations

- Too complicated for today … Just a few hints regard data sharing:

- Respect the GDPR!

- Tell participants that their data will be made open and ask explicitly for their consent (see several templates for consent forms)

  - When data are fully anonymized, this consent is not legally required, but it is nevertheless mandatory from a research ethics perspective.

- Do not share data of participants who have refused to give their consent for potential secondary use

  - When legal restrictions to data sharing apply, it should be stated which types of aggregated data or anonymized or pseudonymized partial data *can* be shared.

- Narrow vs. broad consent for personal data (latter is necessary for data reuse)

Sharing Data as Part of a Publication ("Type 1 data sharing")

- With the publication* of a manuscript, the person or group who collected the data should provide all primary data and associated metadata necessary to reproduce the published results.

- Variables that were assessed within the scope of the study but have not been included in the corresponding publication should be reported (but not necessarily shared)

  - e.g., share the full codebook for all variables

* including preprints

Sharing Data after Project Completion ("Type 2 data sharing")

- In accordance with the DFG guidelines, the data that have been collected in a funded research project should be "made available to the public immediately after completion of the research or within a few months"

- This includes **all** relevant data of the project that are not yet part of a publication, including the accompanying metadata

- Primarily applies to projects for which both the scope and completion are properly defined (as is typically the case in third-party funded research projects)

Embargo

- Data sharers can define an embargo for secondary use on the unused part of a type-2 shared data set.

- Typically not more than 5 years after project completion; end date must be public, and ideally lifted automatically by the repository.

- Embargos on type-1 data are only possible in exceptional cases, and should be much shorter. (Sharing upon request for reproducibility purposes must be ensured nonetheless).

No duplicates, please

- Secondary data (i.e. preprocessed data) should never be submitted as a new primary data set

  - Better: Link to original primary data set doi; provide reproducible script that generates the derived data set.

- Consecutive releases of type-1 data slices via publications, however, conflict with this principle. Compromise solution:

  - number of partial data sets for each data collection should be kept as low as possible

  - clearly indicate in repository that this data set has been derived from a separate primary data set

  - disclose relationship between "related" partial and primary data sets as transparent as possible

  - Show partial data sets (and the final type-2 release of the full data set) in a bundled (or linked) project in the repository.

Mechanisms for keeping a balance of values

A balance of values
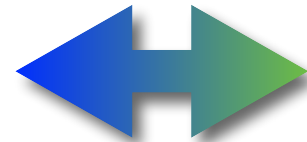
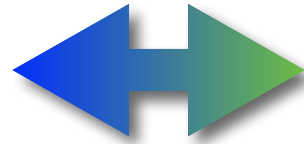|		
public money = public data	⬌	privacy rights of participants
Right of primary use of data, incentives for collecting data	⬌	Efficient gain of knowledge by secondary data usage
Verifiability and reproducibility of scientific claims	⬌	Protect original authors and society from misuse

public money = public data	privacy rights of participants

- As open as possible, as closed as necessary

- Solution:

  - If sharing of primary data is legally not possible, aim to share anonymized, synthetic, derived, or aggregated data instead. (E.g., not the videos but behavioral codings; not the interview texts but content categories).

  - Levels of data access: Share as scientific use file with restricted access.
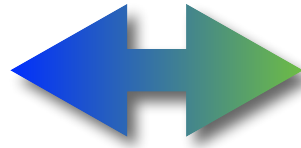
Right of primary use of data, incentives for collecting data	⬌	Efficient gain of knowledge by secondary data usage

- Data sharing and secondary data use is intended to be a "win-win situation" for the entire scientific community

  - Contra: „research parasites" discussion

- We need incentives for collecting data in the first place; in particular for early career researchers (ECRs); at the same time we want as much reusability as possible.

- Solution:

  - Type 1 vs. type II data sharing: Keep unused variables

  - Right of first use; increase the period with an embargo

  - But: At the end of day, research data is open and freely reusable for any purpose (of course respecting legal constraints)

  - Appreciate data sharing in job applications and tenure committees
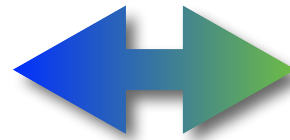
| Verifiability and reproducibility of scientific claims | ⬌ | Protect original authors and society from misuse of „bad faith actors" |

- Evaluations of past analyses should apply the standards of that time
  - Evaluations of the *evidence* should apply the current best practice
- „Share-alike principle": Reanalyses must comply to the same high standards of openness and scientific integrity as the original research (e.g., should publish reproducible code).
- Levels of data access, for example when data can be weaponized:
  - Restrict access to scientific audience
  - Granting of access ideally be done by repository (or any other third party), based on predefined rules, not by original authors.
- Resolving dispute
  - Many of the recommendations are not legally binding
  - If restrictions apply, clarify them upfront in written and signed sharing agreements.
    Violation = scientific misconduct
  - To deal with disputes between secondary users and data sharers that do not fall into the category of "scientific misconduct," but rather represent different points of view, we recommend the establishment of an ombuds committee (ideally authorized by election) that can be consulted by the involved parties, provided they are members of the DGPs.

Spread the word

- Be as open as possible in your own research.

- When you review a DFG grant proposal:
  Refer to the recommendations and require open data sharing (or an explanation why it is not possible)

- When you are a member of a hiring / tenure committee: Appreciate data sharing as a valid scientific achievement.

Get the recommendations (revised version)

- **German version**: Gollwitzer, M., Abele-Brehm, A., Fiebach, C., Ramthun, R., Scheel, A. M., Schönbrodt, F. D., & Steinberg, U. (2020, September 10). Management und Bereitstellung von Forschungsdaten in der Psychologie: Überarbeitung der DGPs-Empfehlungen. https://doi.org/10.31234/osf.io/hcxtm

- **English version**: Gollwitzer, M., Abele-Brehm, A., Fiebach, C., Ramthun, R., Scheel, A. M., Schönbrodt, F. D., & Steinberg, U. (2020, September 10). Data Management and Data Sharing in Psychological Science: Revision of the DGPs Recommendations. https://doi.org/10.31234/osf.io/24ncs