



Open Data

Die DGPs-Empfehlungen zum Management und
Bereitstellung von Forschungsdaten in der Psychologie

-

How to Codebook

Leitfragen zu den Empfehlungen

Instruktion für die Studierenden:

Bitte lesen Sie die DGPs-Empfehlungen "Management und Bereitstellung von Forschungsdaten in der Psychologie" (<https://psyarxiv.com/hcxtm>) und beantworten Sie die folgenden Leitfragen kurz und präzise.

1. Ein Autor veröffentlicht vollständig Mittelwerte und SDs aller Variablen in einer Tabelle im Ergebnisteil. Sind damit die Primärdaten publiziert und der Dokumentationspflicht entsprochen worden? Wenn nein, warum nicht?
2. Nennen Sie drei Beispiele für Metadaten eines Datensatzes.
3. Was sind "personenbezogene Daten"? Ist das Geschlecht und Alter einer Person ein personenbezogenes Datum?
4. Welche dieser Orte sind geeignet als Repository für die langfristige Archivierung und Bereitstellung von Daten? Wenn "Nein", warum nicht?
(a) Die persönliche Webseite der Forscherin an ihrer Universität; (b) das "Online Supplemental Material (OSM)" einer wissenschaftlichen Zeitschrift des Elsevier Verlags; (c) in einem öffentlichen Projekt des Open Science Frameworks (<https://osf.io>); (d) bei Figshare (<https://figshare.com>); (e) bei Zenodo (<https://zenodo.org>)
5. Beschreiben Sie in jeweils einem Satz die "Datenbereitstellung Typ 1" und die "Datenbereitstellung Typ 2".
6. Was bedeutet ein "Nutzungsembargo"? Wie lange sollte es maximal sein?
7. Welche Zugriffsklassen gelten als "open data"?

- **1. Ein Autor veröffentlicht vollständig Mittelwerte und SDs aller Variablen in einer Tabelle im Ergebnisteil. Sind damit die Primärdaten publiziert und der Dokumentationspflicht entsprochen worden? Wenn nein, warum nicht?**
- Nein - das sind nicht die Primärdaten. Hinter einem Mittelwert und SD können sich verschiedenste Rohdaten verbergen, mit Ausreißern, umplausiblen Werten, etc. Manche Reanalysen sind nur mit den Rohdaten möglich.

- **2. Nennen Sie drei Beispiele für Metadaten eines Datensatzes.**
- Erhebungsdatum/-zeitraum der Daten
- Ort der Erhebung
- Lizenz zur Nachnutzung
- Kontaktdaten der Ansprechperson, die Auskunft über die Daten(erhebung) geben kann
- Zitationsinformationen: Wie soll man diesen Datensatz korrekt zitieren?
- Codebook / Beschreibung der Variablen
- Geldgeber

- **3. Was sind "personenbezogene Daten"?**
- Personenbezogene Daten sind „alle Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person beziehen“ (Art 4 Zif. 1 DSGVO)
- **Ist das Geschlecht und Alter einer Person ein personenbezogenes Datum?**
- Das kommt darauf an: Wenn in dem Datensatz noch hinreichend viele andere (z.B. $k \geq 5$) Personen dieses Geschlecht und Alter haben, dann sind diese Daten nicht auf eine konkrete Person rückführbar. Kann man mit dieser Kombination an Daten jedoch eine konkrete Person identifizieren („der einzige 42-jährige Mann, der gerade Psychologie an der LMU studiert“), dann sind sie personenbeziehbar.

- 4. Welche dieser Orte sind geeignet als Repository für die langfristige Archivierung und Bereitstellung von Daten? Wenn "Nein", warum nicht?
 - (a) Die persönliche Webseite der Forscherin an ihrer Universität: **Nein**, da nicht persistent; kein doi.
 - (b) das "Online Supplemental Material (OSM)" einer wissenschaftlichen Zeitschrift des Elsevier Verlags: **Nein**, da hinter paywall; Verlag kann jederzeit den Service einstellen.
 - (c) in einem öffentlichen Projekt des Open Science Frameworks (<https://osf.io>): **Ja** (non-profit; hat eine „long-term preservation strategy“). Aber besser einen europäischen/deutschen Service nutzen.
 - (d) bei Figshare (<https://figshare.com>): **Jein -> eher Nein**: Die Nutzungsbedingung sind zwar aktuell kompatibel mit „open-science-Anforderungen“, das Repository ist jedoch kommerziell; Nutzungsbedingungen können sich jederzeit ändern.
 - (e) bei Zenodo (<https://zenodo.org>): **Ja** (wird bei CERN betrieben, öffentlich finanziert ohne Gewinnerzielungsabsicht).
- Aktuell optimal: psychArchives beim ZPID.

- **5. Beschreiben Sie in jeweils einem Satz die "Datenbereitstellung Typ 1" und die "Datenbereitstellung Typ 2".**
 - Datenbereitstellung Typ 1: „Mit Erscheinen einer Publikation sollen alle Primärdaten sowie die dazugehörigen Metadaten bereitgestellt werden, die zur Reproduktion der publizierten Ergebnisse notwendig sind (unabhängig davon, in welchem Kontext (bspw. drittmittelfinanziertes Forschungsprojekt oder studentische Abschlussarbeit) diese entstanden sind)
 - Datenbereitstellung Typ 2: „Gemäß den Empfehlungen der DFG soll der Gesamtdatensatz eines Projekts „unmittelbar nach Abschluss der Forschungen oder nach wenigen Monaten der Öffentlichkeit frei zur Verfügung gestellt werden“. Das beinhaltet auch alle relevanten Daten des Projekts, die zu diesem Zeitpunkt noch nicht Bestandteil einer Publikation sind, einschließlich der entsprechenden Metadaten.

- **6. Was bedeutet ein "Nutzungsembargo"? Wie lange sollte es maximal sein?**
- Die Daten liegen bereits dokumentiert in einem Repository, können aber noch nicht von anderen Forschenden (für alle Zwecke) nachgenutzt werden. D.h., die Datenbereitstellenden haben noch eine zeitlang das exklusive Nachnutzungsrecht.
- Die Freischaltung hängt nicht am Gutdünken der Datenbereitstellenden, sondern wird automatisch vom Repository nach Ablauf des Embargos vorgenommen.
- Das Embargo soll nicht länger als 5 Jahre nach Projektabschluss gelten (und es muss immer begründet werden).

- **7. Welche Zugriffsklassen gelten als "open data"?**
- ZK 0 und ZK 1. Alle restriktiveren Zugriffsklassen sind nicht mehr „open data“.

Bereitstellung von Daten erlaubt?

Einwilligung liegt vor

restriktive Einwilligung:

z.B. Einwilligung sichert Löschung (nach X Jahren) zu; sichert zu, dass Daten nur für bestimmten Zweck genutzt werden; Datenweitergabe explizit in Einwilligung ausgeschlossen; Teilnehmer stimmen einer Nachnutzung/Veröffentlichung explizit *nicht* zu

permissive Einwilligung:

Einwilligung spricht open data und breite Nachnutzung explizit an („Zweck, Art und Umfang dieser Nachnutzung zum gegenwärtigen Zeitpunkt noch nicht absehbar“) & Teilnehmer stimmen zu.

Einwilligung liegt *nicht* vor

Auf eine auf Aufklärung basierende Einwilligung kann unter zwei Bedingungen **verzichtet** werden¹:

1. kein Schaden/ Unbehagen **UND**
2.

(a) gängige Erziehungsmethoden, Curricula oder Unterrichtsmethoden im Bildungsbereich **ODER**
(b) anonyme Fragen/Fragebögen, freie Beobachtung oder Archivmaterial **ODER**
(c) Faktoren, welche die Arbeits- und Organisationseffizienz in Organisationen betreffen

anonym

Datenschutzrechtlich: ✓
Forschungsethisch: ✗

Datenschutzrechtlich: ✓
Forschungsethisch: ✓

Datenschutzrechtlich: ✓
Forschungsethisch: ✓

nicht anonym

Datenschutzrechtlich: ✗
Forschungsethisch: ✗

Datenschutzrechtlich: ✓
(falls Einwilligung auch personenbezogene Daten umfasst)
Forschungsethisch: ✓

Datenschutzrechtlich: ✗
Forschungsethisch: ✗

↓
Diese Variante dürfte es im Regelfall gar nicht geben

¹ siehe DGPs (2018). Ethisches Handeln in der psychologischen Forschung.

How to Codebook

How to Codebook

See the [D-Psy-FAIR standard](#) (developed by Blask et al from the Leibniz-Institute Psychology)

What should a study codebook contain?

To achieve the above outlined purpose, the codebook should contain at least the following attributes for each variable:

Table 2. Content of codebooks for D-Psy-FAIR

Attribute	Description
Variable name	A short, meaningful name of the variable, relating to the variables content; only small letters should be used, spaces should be replaced by an underline (e.g., us_pos)
Variable label	A more comprehensive description of the variable (e.g., unconditioned stimulus, positive valence)
Item text	The exact question of a questionnaire / instruction text
Value labels	Describes the value range of a given variable, in case of paradata (e.g., instructions, material) the relevant file name wherein the information should be indicated
Missing	Specifies missing values (e.g., -99 for omitted questions)
Measure	The measure used for the dependent variable as defined in the study documentation

Save in a human-readable (e.g., .csv) and in a machine-readable (e.g., .json) format!

How to Codebook

Excel file:

Human-readable: ✓✓

Machine-readable: ~

Accessible: 🙅

	A	B	C	D	E	F
1	name	label	item_text	value_label	missings	measure
2	pid	Unique person ID				
3	gender	Gender of participants	Welchem Geschlecht fühlst du dich zugehörig?	1 = male, 2 = female	99 = no answer	
4	BFI2XS1	BFI2-XS Item 1 (Extraversion)	Ich bin eher ruhig.	1 = Stimme überhaupt nicht zu 2 = Stimme eher nicht zu 3 = Teils, teils 4 = Stimme eher zu 5 = Stimme voll und ganz zu	99 = no answer	BFI-2-XS (Rammstedt et al., 2020)
5	BFI2XS2	BFI2-XS Item 2 (Aggreeableness)	Ich bin einfühlsam, warmherzig.	1 = Stimme überhaupt nicht zu 2 = Stimme eher nicht zu 3 = Teils, teils 4 = Stimme eher zu 5 = Stimme voll und ganz zu	99 = no answer	BFI-2-XS (Rammstedt et al., 2020)

How to Codebook

csv file:

Human-readable: ~ (raw) ✓ (after import)

Machine-readable: ✓

Accessible: ✓

```
name;label;item_text;value_label;missings;measure
pid;Unique person ID;;;
gender;Gender of participants;Welchem Geschlecht fühlst du dich zugehörig?;1 = male, 2 = female;99 = no answer;
BFI2XS1;BFI2-XS Item 1 (Extraversion);Ich bin eher ruhig.;"1 = Stimme überhaupt nicht zu
2 = Stimme eher nicht zu
3 = Teils, teils
4 = Stimme eher zu
5 = Stimme voll und ganz zu";99 = no answer;"BFI-2-XS
(Rammstedt et al., 2020)"
BFI2XS2;BFI2-XS Item 2 (Agreeableness);Ich bin einfühlsam, warmherzig.;"1 = Stimme überhaupt nicht zu
2 = Stimme eher nicht zu
3 = Teils, teils
4 = Stimme eher zu
5 = Stimme voll und ganz zu";99 = no answer;"BFI-2-XS
(Rammstedt et al., 2020)"
```

How to Codebook

json file:

Human-readable: ~

Machine-readable: ✓ ✓

Accessible: ✓

```
{
  "name": "pid",
  "label": "Unique person ID",
  "item_text": null,
  "value_label": null,
  "missings": null,
  "measure": null
},
{
  "name": "Gender",
  "label": "Gender of participant",
  "item_text": "Welchem Geschlecht fühlst du dich zugehörig?",
  "value_label": [
    {
      "name": "1",
      "label": "male"
    },
    {
      "name": "2",
      "label": "female"
    }
  ],
  "missings": [
    {
      "name": "99",
      "label": "no answer"
    }
  ],
  "measure": null
},
{
  "name": "condition",
  "label": "Experimental condition",
```

Excursus: json file format

```
{  
  "name": "Jason",  
  "age": 39,  
  "height": 1.92,  
  "gender": "M",  
  "salary": 70000,  
  "married": true,  
  "children": [  
    {"name": "Tom", "age": 9, "gender": "M"},  
    {"name": "Ava", "age": 7, "gender": "F"}  
  ]  
}
```


Excursus: json file format

Opening and
closing brackets

Key
(= variable name)

Value

```
{  
  "name": "Jason",  
  "age": 39,  
  "height": 1.92,  
  "gender": "M",  
  "salary": 70000,  
  "married": true,  
  "children": [  
    {"name": "Tom", "age": 9, "gender": "M"},  
    {"name": "Ava", "age": 7, "gender": "F"}  
  ]  
}
```

Value of type numeric

Value of type string: Use „“

Value of type Boolean: Use *true* and *false* (without „“)

Arrays: encapsulate with []

How to Codebook: Manual

- Step 1: Create your codebook as a csv file
 - e.g., create it in Excel first, then export as .csv file
- Step 2: Import into R and convert it to a json file

```
library(jsonlite)
df <- read.csv("codebook.csv")
json <- toJSON(df, pretty = TRUE)
write_json(json, "codebook.json")
```

- Step 3: Save both codebook files in the /doc subfolder of your project
- Step 4: Upload both the .csv and the .json-file to the repository

How to Codebook: DataWiz2

[← My datasets](#) [View data matrix](#) [Edit codebook](#) [preQ_anon.csv](#) [? Help](#) [Save](#)

	Variable name ?	Variable label ?	Item text ?	Value labels ?	Missings ?	Measure ?
1	pid	Unique person ID		Add labels	Add missings	Select measure
2	Gender	Gender of participant	Welchem Geschlecht fühlst du c	Add labels	Add missings	Select measure
3	condition	Experimental condition		EP = Experimental, po...	Add missings	Select measure
4	AvgWeeklyPartnerMeet			Add labels	Add missings	Select measure
5	BF12XS1	BF12-XS Item 1	Ich bin eher ruhig.	Add labels	Add missings	Select measure
6	BF12XS2			Add labels	Add missings	Select measure
7	BF12XS3			Add labels	Add missings	Select measure
8	BF12XS4			Add labels	Add missings	Select measure
9	BF12XS5			Add labels	Add missings	Select measure
10	BF12XS6			Add labels	Add missings	Select measure

Description of variable 1

Variable name
pid

Variable label
Unique person ID

Item text

Measure Copy to
Select measure Edit measures

Value labels Copy to
 =

+ Add another value label

<https://datawiz2.dev.zpid.de>

How to Codebook

- Resources:
- Video (51 min.): "[PTOS 9: Einführung in das Forschungsdatenmanagement mit D-Psy_FAIR](#)"
 - siehe v.a.: „Schritt 2: Ein Codebuch erstellen“ ab 29:27