

Forschungsdaten effizient managen

---

# Mit Metadaten wird alles besser.

Knut Wenzig

Survey Lifecycle Operators Workshop (SLOW2020), München, 9.–11. März 2020

A stylized, handwritten-style logo for 'Libniz' in white script.

- Aufbereitung schwer nachvollziehbar, auch wenn Code vorliegt
- Dokumentation: hoher Aufwand und trotzdem Fehler und Lücken
- Nutzer:innen müssen Variablen finden, richtig verstehen (auch im Längsschnitt: Replikation, Unterschiede identifizieren) und unsere Eingriffe nachvollziehen können
- Daten für Computer/KI verständlich machen (FAIR)

- strukturierte Metadaten verwenden
- Metadaten dann erfassen, wenn die Informationen entstehen
- tatsächlichen Informationsfluss modellieren und Produktionsschritte mit (aktiven) Metadaten steuern
- Single Source of Truth etablieren
- Metadatenobjekte referenzieren, nicht duplizieren
- kontinuierliche Re-Integration
- Standards und Normen im Auge behalten

# Strukturierte Metadaten verwenden

	A	B	C	D	E	F	G	H	I	
1	study	dataset	version	variable	temp	label	label_de	concept	type	de
2	soep-core	ap	v35	hhnr		Original Household Number	Ursprungshaushaltsnumm	cid	byte	
3	soep-core	ap	v35	hhnrakt		Current Wave HH Number (=F	Aktuelle Haushaltsnumme	hid	byte	
4	soep-core	ap	v35	persnr		Never Changing Person ID	Unveraenderliche Person	pid	byte	
5	soep-core	ap	v35	syear		Survey Year	Befragungsjahr	syear	byte	
6	soep-core	ap	v35	cid		Original Household Number	Ursprungshaushaltsnumm	cid	byte	
7	soep-core	ap	v35	ahnr		Current Wave HH Number (=F	Aktuelle Haushaltsnumme	hid	byte	
8	soep-core	ap	v35	hid		Current Wave HH Number (=F	Aktuelle Haushaltsnumme	hid	byte	
9	soep-core	ap	v35	pid		Never Changing Person ID	Unveraenderliche Person	pid	byte	
10	soep-core	ap	v35	welle		Survey Year	Befragungsjahr	jahr	byte	
11	soep-core	ap	v35	abefrper		Person Number Surveyed	Befragte Person	pnr	byte	
12	soep-core	ap	v35	sample1		Subsample Identifier	Stichprobenkennzeichen	sample1	byte	
13	soep-core	ap	v35	apinta		Survey Instrument	Art der Befragung	pform	byte	
14	soep-core	ap	v35	apinda1		Length In Minutes Oral Interv	Dauer muendl.Interv.	pdauer1	byte	
15	soep-core	ap	v35	apinda2		Length In Minutes Written Int	Dauer schrift.Int.	pdauer2	byte	
16	soep-core	ap	v35	ap24_isco88		Current Occupational Classif	aktueller Beruf [pber] (ISO	pgen_pgis88	byte	
17	soep-core	ap	v35	ap24_kldb92		Current Occupational Classif	aktueller Beruf [pber] (Kld	pgen_pgklas	byte	
18	soep-core	ap	v35	ap26_nace		Industry Occupation [pbra] (N	Branche aktueller Beruf [p	1293_p_jcl	byte	

- XML oder Tabellen
- so wenig Information wie möglich in Code oder Text
- ermöglicht Re-Use

## Metadaten dann erfassen, wenn die Informationen entstehen

**3** Haben Sie das Gefühl, dass das, was Sie in Ihrem Leben machen, wertvoll und nützlich ist?

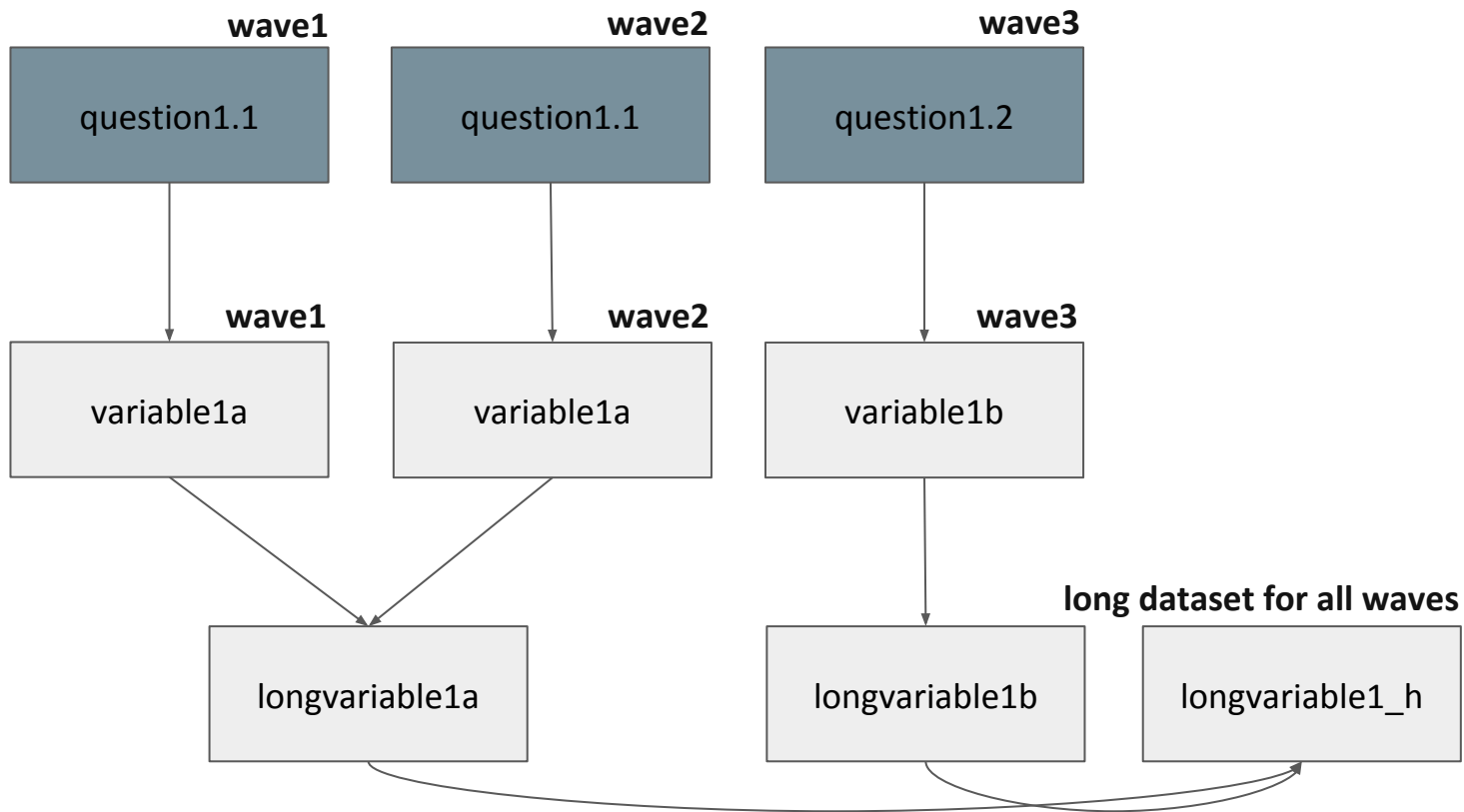
Antworten Sie bitte anhand der folgenden Skala: der Wert 0 bedeutet: "überhaupt nicht wertvoll und nützlich" und der Wert 10: "vollkommen wertvoll und nützlich". Mit den Werten dazwischen können Sie Ihre Einschätzung abstufen.

0	überhaupt nicht wertvoll und nützlich
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	vollkommen wertvoll und nützlich

3:pgwert bip bip\_03 Tätigkeit im Leben wertvoll und nützlich  
3:pgwert pl plh0334 Taetigkeit im Leben wertvoll und nuetzlich

- Fragebögen aus Metadaten sind Gegenstand der Entwicklung nicht mehr nur der Dokumentation

# Informationsfluss mit aktiven Metadaten modellieren



15

Did this dwelling change ownership within the last year?

- Yes  
 No

de

Hat es im letzten Jahr einen Eigentümerwechsel dieser Wohnung gegeben?

- Ja  
 Nein

Related questions

Previous question

Next question

Info

**Name:** 15

**Instrument:** This question is at position **16** in instrument [Household \(PAPI\) 2018](#)

Variables

[Change of ownership](#) [variable: bih\_15]  
[/soep-core/data/bih/bih\\_15](#)

[change in ownership last year](#) [variable: hlhf0006]  
[/soep-core/data/hl/hlf0006](#)

# Single Source of Truth etablieren

The screenshot shows the DIW SOEP web interface. The top navigation bar includes the logo, 'Projekte', 'Gruppen', 'Mehr', and a search bar. The left sidebar contains a navigation menu with 'Dokumentation', 'Repository', 'Dateien', 'Commits', 'Branches', 'Tags', 'Mitwirkende', 'Diagramm', 'Vergleichen', and 'Diagramme'. The main content area displays the breadcrumb 'SOEP Generierungen > Dokumentation > Repository' and the file path 'master / Dokumentation / meta / long\_generations / v35 / generations.csv'. Below this, a commit entry is shown: 'long generations angepasst' by 'Stefan Zimmermann' from 1 month ago. The file 'generations.csv' (4.13 MB) is listed with a 'Bearbeiten' button and a 'Web IDE' link. The file content is displayed as a list of 11 lines of CSV data.

DIW SOEP Projekte Gruppen Mehr + v Suche oder springe zu...

D **Dokumentation**

Project overview

**Repository**

**Dateien**

Commits

Branches

Tags

Mitwirkende

Diagramm

Vergleichen

Diagramme

SOEP Generierungen > Dokumentation > **Repository**

master  Dokumentation / meta / long\_generations / v35 / **generations.csv**  Finde Date

**long generations angepasst**  
Stefan Zimmermann verfasste vor 1 Monat

**generations.csv** 4,13 MB

```
1 input_version,input_study,input_dataset,input_variable,output_variable,output_dataset,output_study,output_version
2 v35,soep-core,ap,ap1604,plb0380_v1,pl,soep-long,v35
3 v35,soep-core,bp,bp24b04,plb0380_v1,pl,soep-long,v35
4 v35,soep-core,cp,cp24b04,plb0380_v1,pl,soep-long,v35
5 v35,soep-core,ap,syear,syear,pl,soep-long,v35
6 v35,soep-core,apausl,syear,syear,pl,soep-long,v35
7 v35,soep-core,bap,syear,syear,pl,soep-long,v35
8 v35,soep-core,bbp,syear,syear,pl,soep-long,v35
9 v35,soep-core,bcp,syear,syear,pl,soep-long,v35
10 v35,soep-core,bdp,syear,syear,pl,soep-long,v35
11 v35,soep-core,bep,syear,syear,pl,soep-long,v35
```



## Metadatenobjekte referenzieren, nicht duplizieren

- Wiederverwendung wird identifiziert und spezifiziert

[1] Ja CodeList1

[2] Nein

[-1] Weiß nicht CodeList2

[-1] Weiß nicht CodeList3

[-2] Trifft nicht zu/nicht erreicht

[-8] designbedingt fehlend

- Zielkonflikt: Je mehr Information durch eine einzige Referenz adressiert wird, desto weniger Information muss zusätzlich angegeben werden, desto öfter muss man jedoch neue Objekte anlegen.

- kontinuierliche Re-Integration
  - zunehmende Verwendung schafft neue Bedarfe
  - Erweiterungen nicht zu sehr ausfransen lassen
  - aber auch nicht verhindern
- Standards und Normen im Auge behalten
  - derzeit oft Nutzen niedrig und Kosten hoch, wenn man Standards implementiert
  - Es werden Angebote entstehen, die fremde Metadaten zur Wiederverwendung bereitstellen und eigene Metadaten annehmen. (NFDI)

Vielen Dank für die Aufmerksamkeit.

---



**DIW Berlin — Deutsches Institut  
für Wirtschaftsforschung e.V.**  
Mohrenstraße 58, 10117 Berlin  
[www.diw.de](http://www.diw.de)

**Sozio-oekonomisches Panel**  
[www.soep.de](http://www.soep.de)  
Knut Wenzig, [kwenzig@diw.de](mailto:kwenzig@diw.de)

Mitglied der

*Leibniz*  
Leibniz-Gemeinschaft

Metadaten im Forschungszyklus:

Wenzig, Knut (2019): Metadaten. In: Nina Baur, Jörg Blasius (Hrsg.), Handbuch Methoden der empirischen Sozialforschung, Springer. doi:[10.1007/978-3-658-21308-4\\_92](https://doi.org/10.1007/978-3-658-21308-4_92) (Preprint verfügbar bei ResearchGate)

Detailliertere Use cases mit SOEP-Metadaten:

Knut Wenzig. (2019). Longitudinal Metadata at the Socio-Economic Panel. doi:[10.5281/zenodo.3554859](https://doi.org/10.5281/zenodo.3554859)

Mehr Use cases, Verbindungen zu DDI:

Wenzig, Knut (2018): Same Same But Different v2 – Documenting Panel Data Using DDI.  
doi:[10.5281/zenodo.1745499](https://doi.org/10.5281/zenodo.1745499)

This presentation is offered under license [CC BY-SA 4.0](#).

**The license does not apply to the following copyrighted material used in this presentation:**

the logo of DIW Berlin/SOEP



the logo of the Leibniz Association

