KonsortSWD

Consortium for the
Social, Behavioural, Educational
and Economic Sciences

Daten. Politik. Zukunft. 9|KSWD
Konferenz für Sozial- und Wirtschaftsdaten
27.–28. März 2023 in Berlin

9th Conference on Social and Economic Data - KSWD 2023
Parallel Forums III (L)
29.03.2023

DOI: 10.5281/zenodo.7750031

nfdi

gesis
Leibniz Institute
for the Social Sciences

GESIS Use case
Variable publication and citation
&
Fine granular access to research data

Claus-Peter Klas
Janete Saldanha Bach
Peter Mutschke

GESIS – Leibniz Institute for the Social Sciences

Authors

Claus-Peter Klas

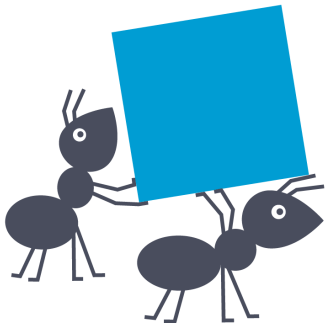
Dr. Claus-Peter Klas, GESIS – Leibniz Institute for the Social Sciences, Team Leader "Data & Service Engineering" and Measure Lead in the NFDI consortium KonsortSWD in the department "Knowledge Technologies for the Social Sciences" . He received his PhD in computer science at the University of Duisburg-Essen and was a postdoctoral researcher in the Department of Multimedia and Internet Applications, Faculty of Mathematics and Computer Science, University of Hagen, Germany. His research focuses on information retrieval, interactive information retrieval, information systems, databases, digital libraries, preservation and grid and cloud architectures. He developed the software Daffodil founded on a nation research project and worked in national and European research projects such as The European Film Gateway, SHAMAN (Sustaining Heritage Access through Multivalent ArchiviNg) and Smart Vortex (Scalable Semantic Product Data Stream Management for Collaboration and Decision Making in Engineering). He is currently responsible for several infrastructure projects within GESIS, such as da|ra, SowiDataNet or Missy, all concerned with providing information and data for social scientists. In addition, he lead the measure PID Services in the national research infrastructure project NFDI. In his team, they are developing an open source DDI suite to support getting DDI into operation.

Janete Saldanha Bach

Dr. Janete Saldanha Bach, GESIS – Leibniz Institute for the Social Sciences. Postdoc in the NFDI consortium KonsortSWD in the department "Knowledge Technologies for the Social Sciences" , team FAIR Data and Human Information Interaction, working in the consortia KonsortSWD Project of the National Research Data Infrastructure (NFDI). She holds a Ph.D. and a Master's degree in Science and Technology Studies (STS) and a bachelor's degree in Information Science. Her research expertise is in Open Science, especially in research data management and data reuse in the Social Sciences. She is currently involved in consortium KonsortSWD, Task Area 5 Measure 1 - developing the conceptual framework for the PID registration service at a variable level and Task Area 5 Measure 2 Enhancing data findability.

Peter Mutschke

Peter Mutschke is deputy head of the department “Knowledge Technologies for the Social Sciences (KTS)" and leader of the team "FAIR Data and Human Information Interaction" of KTS. His research interests include Information Retrieval, Network Analysis and Open Science. He worked in a number of national and international research projects, such as the DFG projects DAFFODIL and IRM and the EU projects WeGov, SENSE4US, OpenMinTeD and MOVING. Peter served as a member of the management committee of the Leibniz research alliance “Science 2.0/Open Science” from 2013-2021. He founded and coordinates the GO FAIR Implementation Network "Cross-Domain Interoperability of Heterogeneous Research Data (Go Inter)", and he is member of the steering committee of the FAIR Digital Objects Forum (fairdo.org) where he also co-chairs a working group on semantics. He is currently involved in consortia KonsortSWD, NFDI4DataScience and BERD@NFDI of the National Research Data Infrastructure (NFDI). ORCID: https://orcid.org/0000-0003-3517-8071.



Agenda

- GESIS Use Case: The PID Registration service for variables
  - GESIS Workflow: Documentation to Dissemination
  - Future use cases for PIDs on variable level

KonsortSWD

Use cases in a nutshell

Institution name	DZHW | German Center for Higher Education Research and Science Research (DZHW)	GESIS | Leibniz Institute for the Social Sciences	GESIS harmonization tools			DIW | German Institute for Economic Research	Qualiservice | University of Bremen
			GESIS QuestionLink	ONBound - Old and new boundaries: National Identities and Religion	Harmonizing and synthesizing partnership histories		
Project / Study	**HEADS - Higher Education Analytical Data System**	**Gesis Data Archive**	**QuestionLink Harmonisation tool**	**ONBound Harmonisation Wizard**	**HaSpaD - Harmonising and synthesising partnership histories**	**German Socio-Economic Panel Study (SOEP-Core)**	**Qualiservice as part of QualidataNet from KonsortSWD**
Attributes	**Survey variables**	**Survey variables**	**Survey variables**	**Survey variables**	**Survey variables**	**Survey variables**	**Qualitative data files**
PIDs uses cases	Variables in datasets; differentiation variables: complex system including one content (dependent variable) plus several independent variables to differentiate this dependent variable by subgroups;	Variables in large datasets from national and international studies	Variables from GESIS and third-parties collections (NEPS and SOEP) for harmonization purposes. Political interest pre-harmonized variables. Some surveys only have one variable name across several years, whereas other surveys have different variable names per wave.	Variables from third-parties collections for harmonization purposes; Religion and Nation in Constitutions Worldwide, Religion and State Project, Church Attendance and Religious; United Nations: Demographic Statistics Database	Variables from third-parties collections for harmonization purpose; The survey programs include Panel Analysis of Intimate Relationships and Family Dynamics s	Assign a PID for each variable from the SOEP-Core v37. SOEP-Core doi:10.5684/soep-core.v37o	Assign a PID for qualitative data, organized in files or dataset, regarding Transcripts, translation, audiovisual and context material for doctor-patient-interaction videos observed
Variables #	Depend on the user selection	507.642	68	750	Depend on the user selection	101.574	N/A

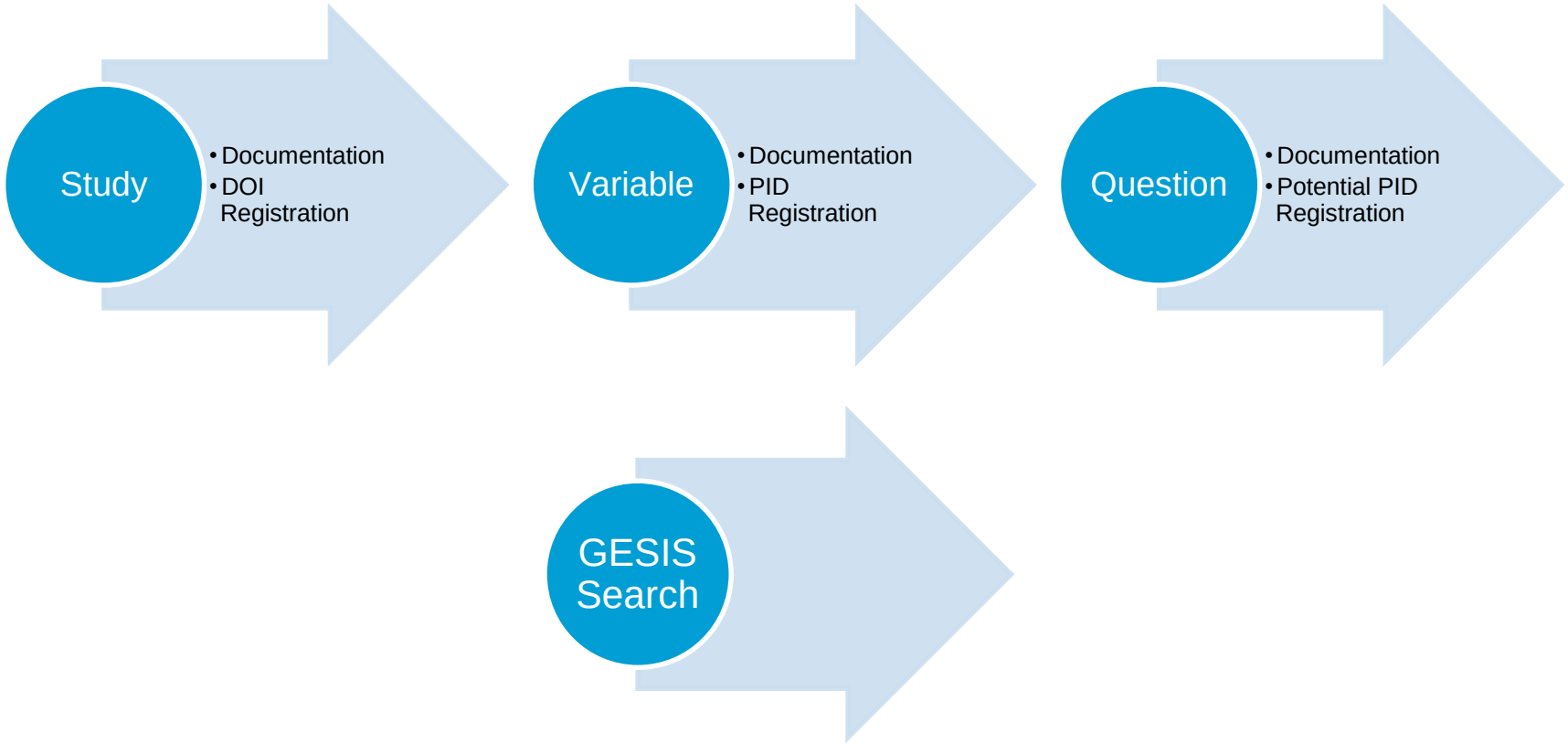
Use Case

gesis
Leibniz Institute
for the Social Sciences

- PIDs for more than 500,000 variables from 6,500 national and international studies covering various topics in the Social Sciences, Economics, and Behaviour Sciences.

GESIS Use Case: Citation of variables

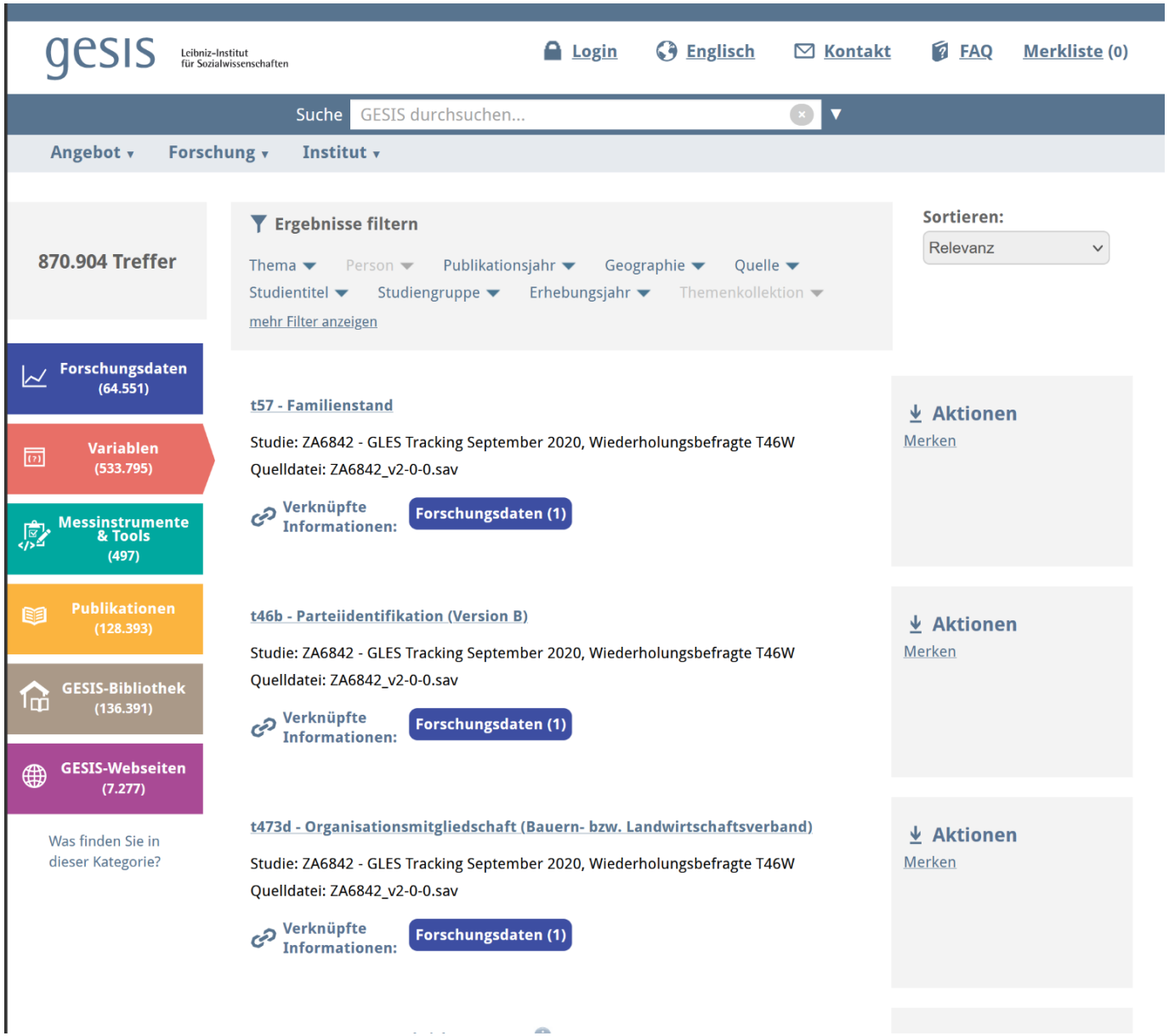
GESIS Data Archive documents surveys incl. variables and questions level in DDI



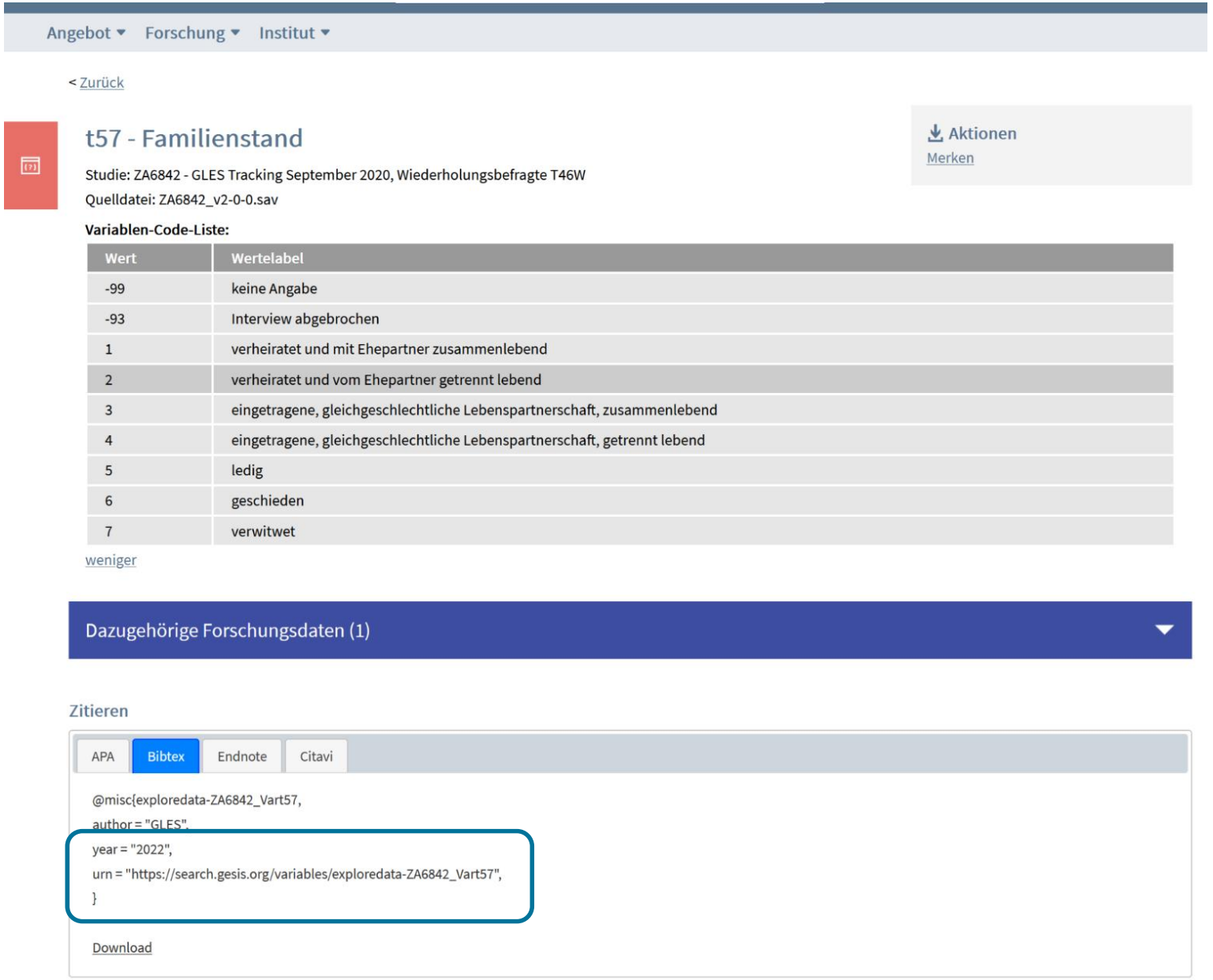
FAIR: Findability, Accessibility, Interoperability, and Reusability

KonsortSWD

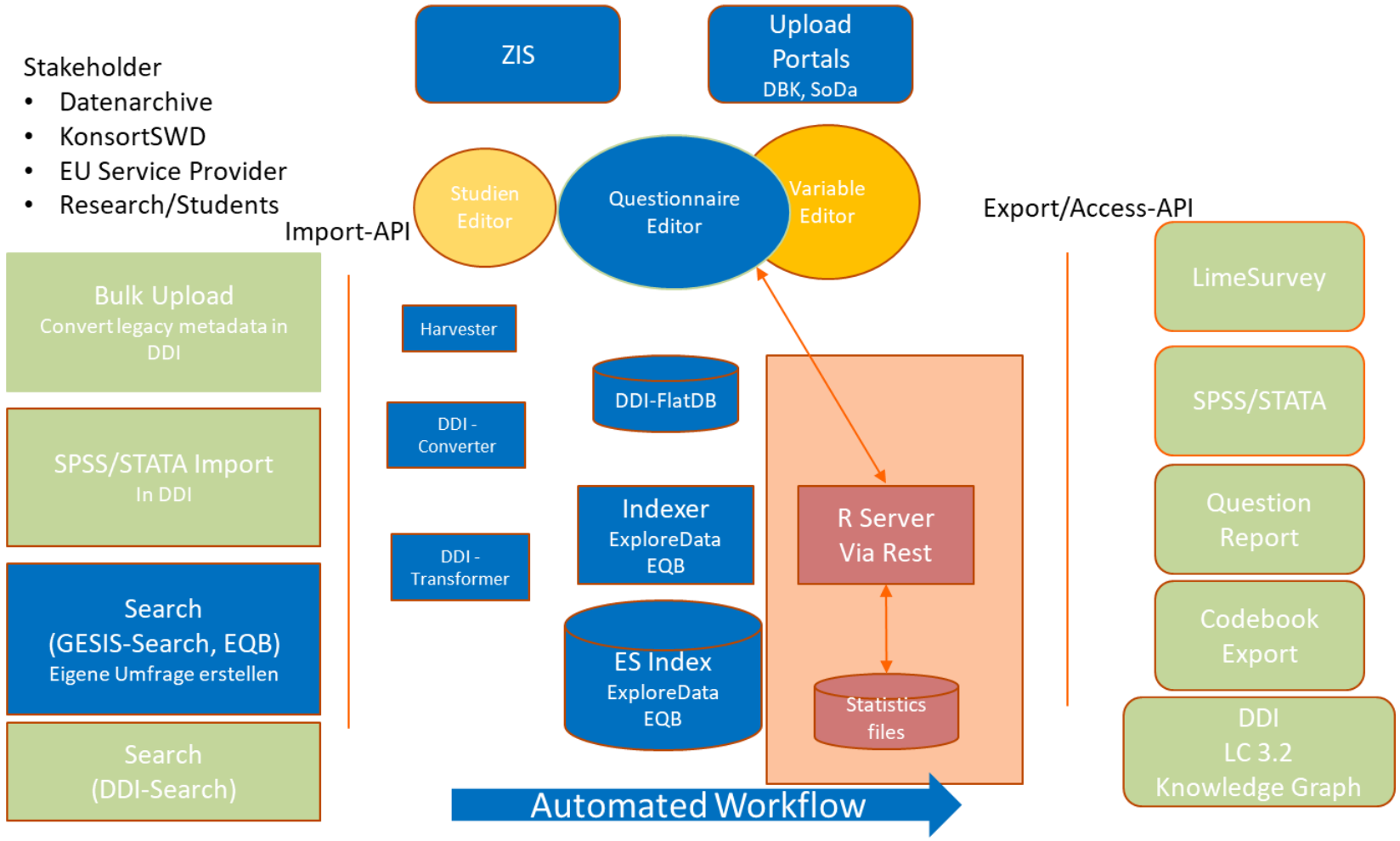
GESIS Search



GESIS USe Case: Citable Variable

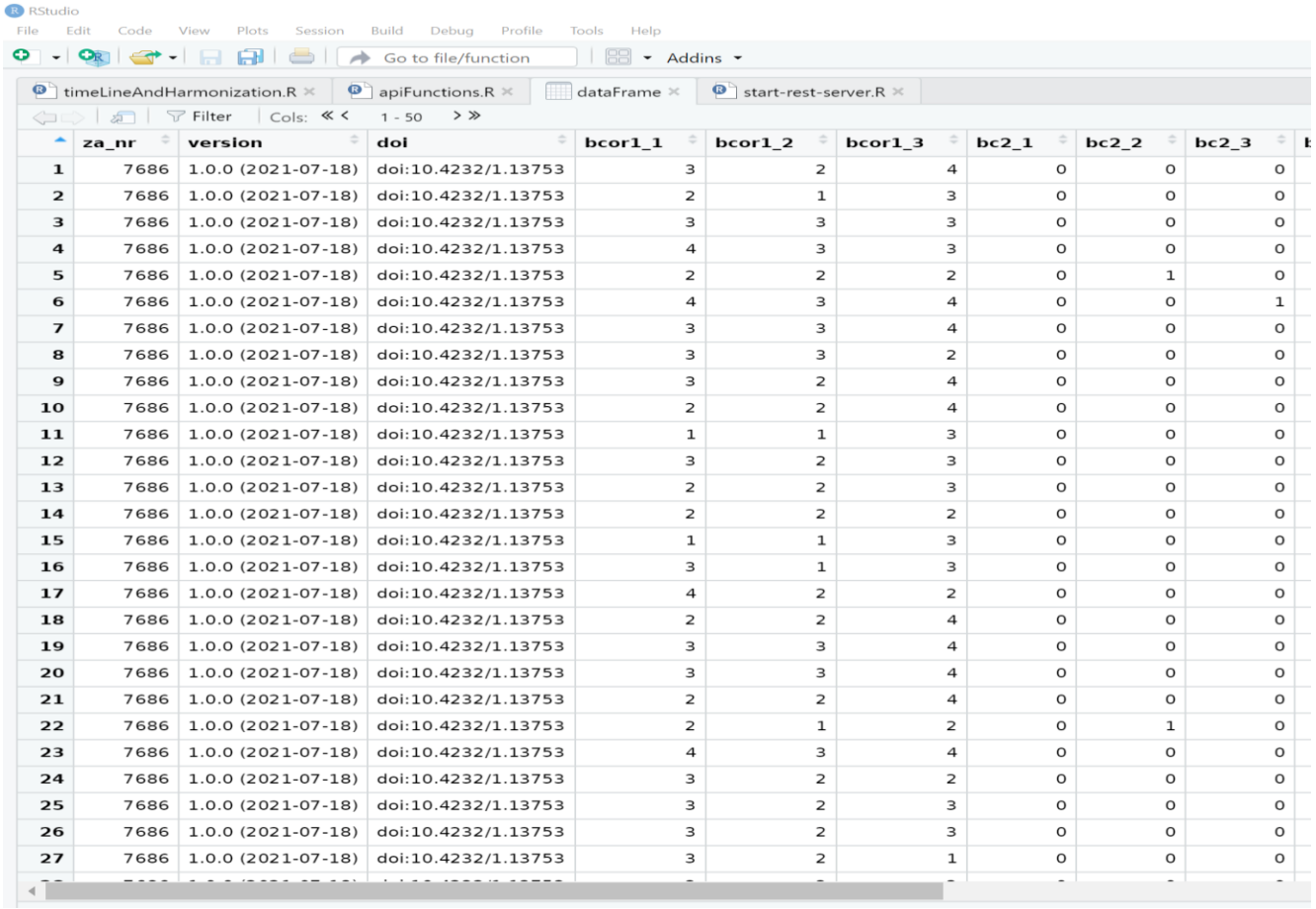


Future uses of variable PIDs



FAIR data access with R: DataFrame

dataFrame <-
read.csv("http://10.6.xx.yy:8100/getDataByStudyId?studyId=https%3A%2F%2F
search.gesis.org%2Fresearch_data%2F ZA7686", sep = "\t")


RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins

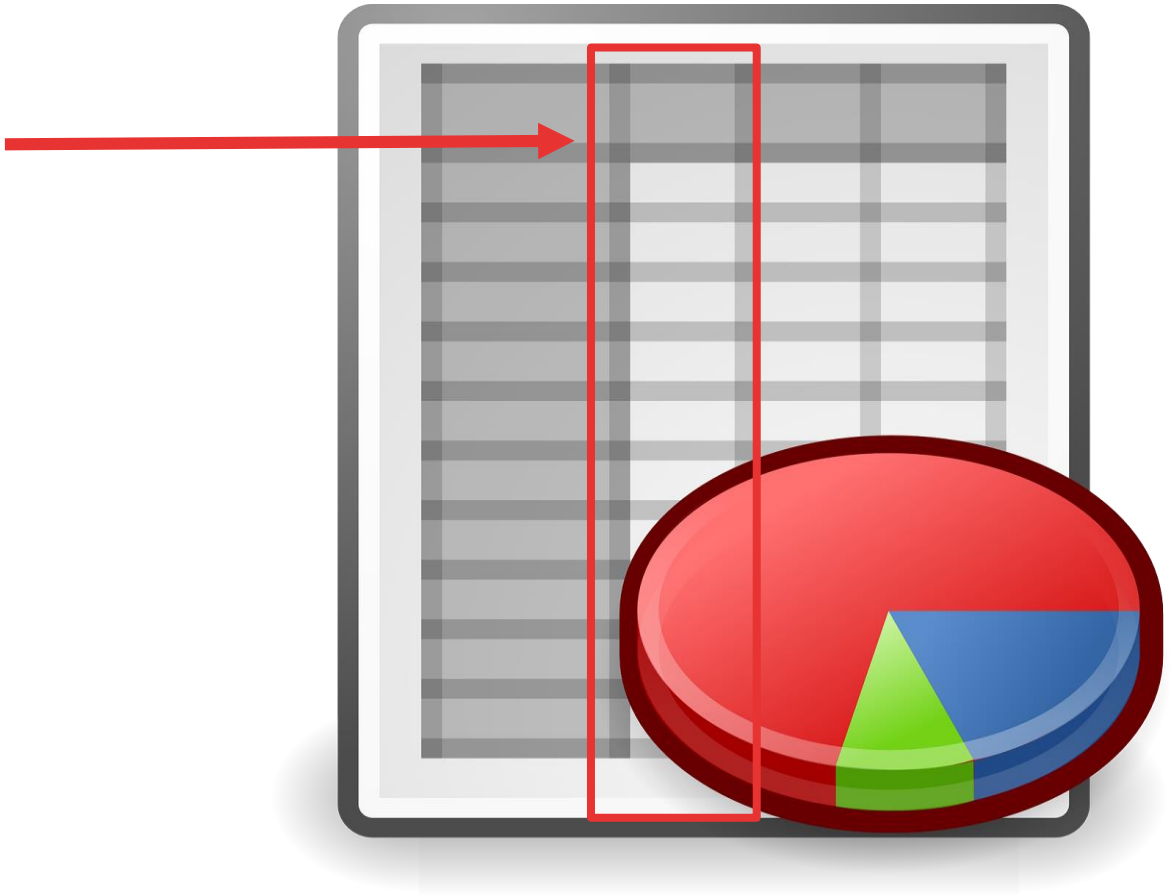
timeLineAndHarmonization.R | apiFunctions.R | dataFrame | start-rest-server.R

Filter | Cols: « < 1 - 50 > »

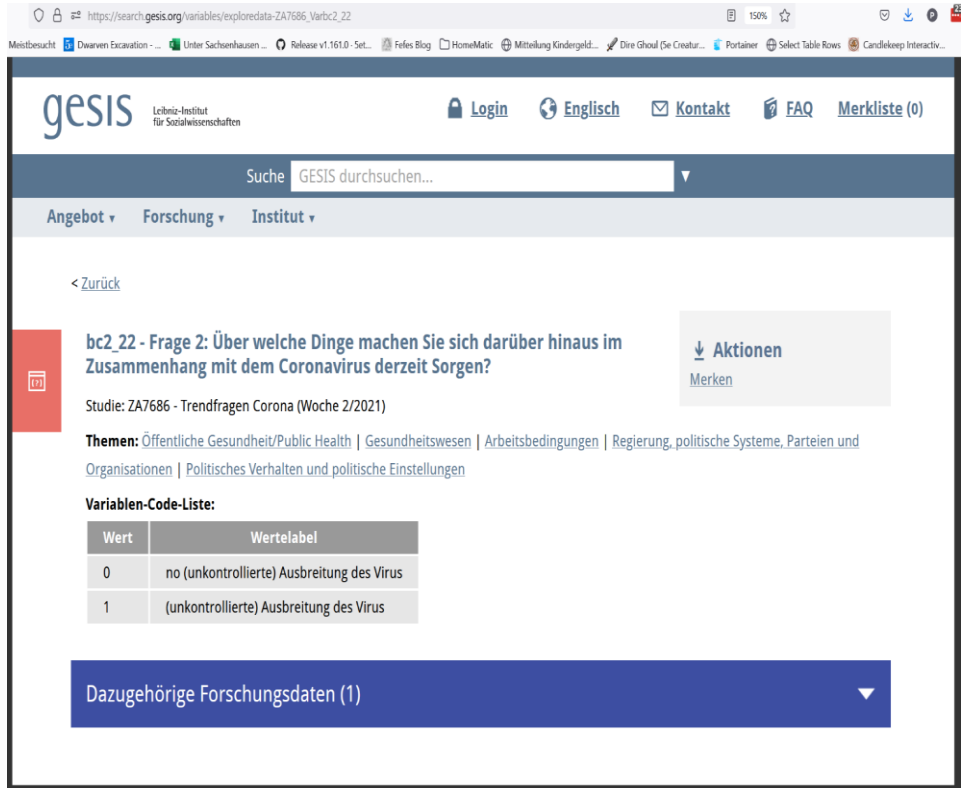
	za_nr	version	doi	bcor1_1	bcor1_2	bcor1_3	bc2_1	bc2_2	bc2_3
1	7686	1.0.0 (2021-07-18)	doi:10.4232/1.13753	3	2	4	0	0	0
2	7686	1.0.0 (2021-07-18)	doi:10.4232/1.13753	2	1	3	0	0	0
3	7686	1.0.0 (2021-07-18)	doi:10.4232/1.13753	3	3	3	0	0	0
4	7686	1.0.0 (2021-07-18)	doi:10.4232/1.13753	4	3	3	0	0	0
5	7686	1.0.0 (2021-07-18)	doi:10.4232/1.13753	2	2	2	0	1	0
6	7686	1.0.0 (2021-07-18)	doi:10.4232/1.13753	4	3	4	0	0	1
7	7686	1.0.0 (2021-07-18)	doi:10.4232/1.13753	3	3	4	0	0	0
8	7686	1.0.0 (2021-07-18)	doi:10.4232/1.13753	3	3	2	0	0	0
9	7686	1.0.0 (2021-07-18)	doi:10.4232/1.13753	3	2	4	0	0	0
10	7686	1.0.0 (2021-07-18)	doi:10.4232/1.13753	2	2	4	0	0	0
11	7686	1.0.0 (2021-07-18)	doi:10.4232/1.13753	1	1	3	0	0	0
12	7686	1.0.0 (2021-07-18)	doi:10.4232/1.13753	3	2	3	0	0	0
13	7686	1.0.0 (2021-07-18)	doi:10.4232/1.13753	2	2	3	0	0	0
14	7686	1.0.0 (2021-07-18)	doi:10.4232/1.13753	2	2	2	0	0	0
15	7686	1.0.0 (2021-07-18)	doi:10.4232/1.13753	1	1	3	0	0	0
16	7686	1.0.0 (2021-07-18)	doi:10.4232/1.13753	3	1	3	0	0	0
17	7686	1.0.0 (2021-07-18)	doi:10.4232/1.13753	4	2	2	0	0	0
18	7686	1.0.0 (2021-07-18)	doi:10.4232/1.13753	2	2	4	0	0	0
19	7686	1.0.0 (2021-07-18)	doi:10.4232/1.13753	3	3	4	0	0	0
20	7686	1.0.0 (2021-07-18)	doi:10.4232/1.13753	3	3	4	0	0	0
21	7686	1.0.0 (2021-07-18)	doi:10.4232/1.13753	2	2	4	0	0	0
22	7686	1.0.0 (2021-07-18)	doi:10.4232/1.13753	2	1	2	0	1	0
23	7686	1.0.0 (2021-07-18)	doi:10.4232/1.13753	4	3	4	0	0	0
24	7686	1.0.0 (2021-07-18)	doi:10.4232/1.13753	3	2	2	0	0	0
25	7686	1.0.0 (2021-07-18)	doi:10.4232/1.13753	3	2	3	0	0	0
26	7686	1.0.0 (2021-07-18)	doi:10.4232/1.13753	3	2	3	0	0	0
27	7686	1.0.0 (2021-07-18)	doi:10.4232/1.13753	3	2	1	0	0	0

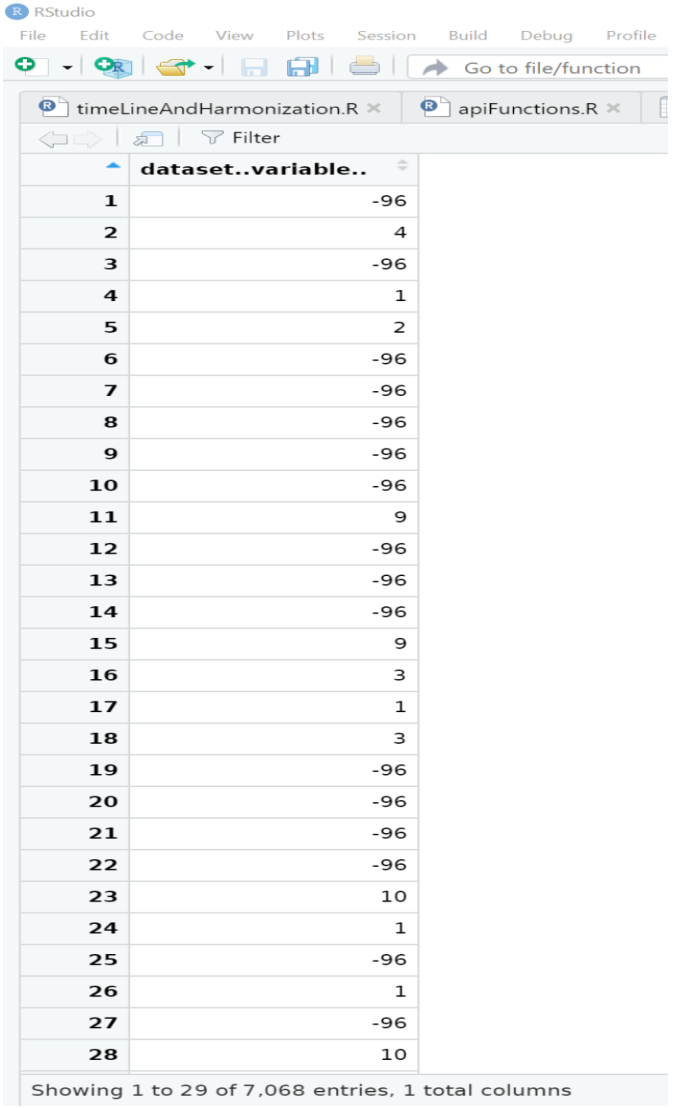
Data formats: initial approach

Variables in tabular data format



FAIR data access with R: Variable





vec <- read.csv("http://10.6.xx.yy:8100/getDataByListOfVariablePIDs?variables=https://search.gesis.org/variables/exploredata-ZA7703_Varpre010j")

KonsortSWD

What do you need todo?

- Document your variables
  - Create a new metadata field for PIDs
  - Register the variable with the mandatory metadata
  - Store the PID in the metadata
- Make your variables searchable incl. a landing page
- Create a citation format
- Future: Open up your data

Citation

Claus-Peter Klas, Janete Saldanha Bach and Peter Mutschke. 2023. GESIS Use case Variable publication and citation & Fine granular access to research data . In 9th Conference on Social and Economic Data - KSWD 2023, 28 - 29 March. 2023. 15 slides. DOI: 10.5281/zenodo.7750031.

Acknowledgments

PID Service report
https://doi.org/10.5281/zenodo.6397367

PID use cases extended report
https://doi.org/10.5281/zenodo.7588944

PID metadata schema extended report
https://doi.org/10.5281/zenodo.7588902

The service is part of KonsortSWD project deliverable, NFDI funding number 442494171

nfdi

KonsortSWD

gesis
Leibniz Institute
for the Social Sciences