gesis Leibniz Institute for the Social Sciences

Linked Data as a Backend Infrastructure for Scientific Search Portals

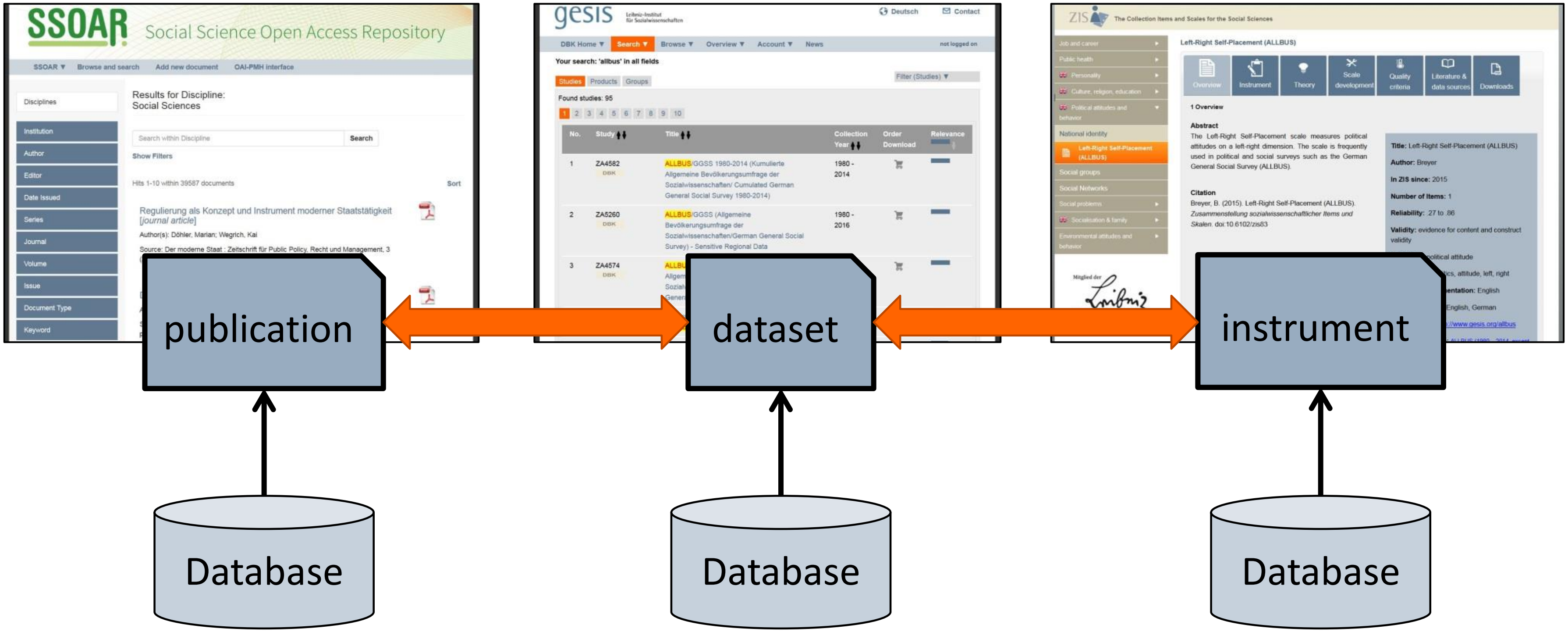
Benjamin Zapilko, Katarina Boland, Dagmar Kern

SWIB 2018, Bonn, Germany, 27.11.2018

Member of Leibniz Leibniz Association

Searching for research information

- Different research information is available in different databases



SSOAR Social Science Open Access Repository

Results for Discipline:
Social Sciences

Hits 1-10 within 39587 documents

Regulierung als Konzept und Instrument moderner Staatstätigkeit [*journal article*]

Author(s): Döhler, Marian; Wegrich, Kai

Source: Der moderne Staat : Zeitschrift für Public Policy, Recht und Management, 3

gesis Leibniz-Institut für Sozialwissenschaften

Your search: 'allbus' in all fields

Found studies: 95

No.	Study	Title	Collection Year	Order Download	Relevance
1	ZA4582 DBK	ALLBUS/GGSS 1980-2014 (Kumulierte Allgemeine Bevölkerungsumfrage der Sozialwissenschaften/ Cumulated German General Social Survey 1980-2014)	1980 - 2014		
2	ZA5260 DBK	ALLBUS/GGSS (Allgemeine Bevölkerungsumfrage der Sozialwissenschaften/German General Social Survey) - Sensitive Regional Data	1980 - 2016		
3	ZA4574 DBK				

ZIS The Collection Items and Scales for the Social Sciences

Left-Right Self-Placement (ALLBUS)

1 Overview

Abstract
The Left-Right Self-Placement scale measures political attitudes on a left-right dimension. The scale is frequently used in political and social surveys such as the German General Social Survey (ALLBUS).

Citation
Breyer, B. (2015). Left-Right Self-Placement (ALLBUS). *Zusammenstellung sozialwissenschaftlicher Items und Skalen*. doi:10.6102/zis83

Title: Left-Right Self-Placement (ALLBUS)

Author: Breyer

In ZIS since: 2015

Number of Items: 1

Reliability: .27 to .86

Validity: evidence for content and construct validity

political attitude

tics, attitude, left, right

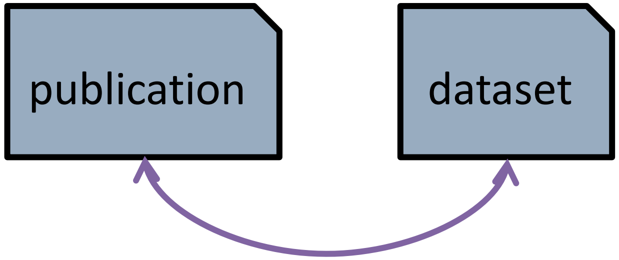
entation: English

English, German

://www.gesis.org/allbus

User survey

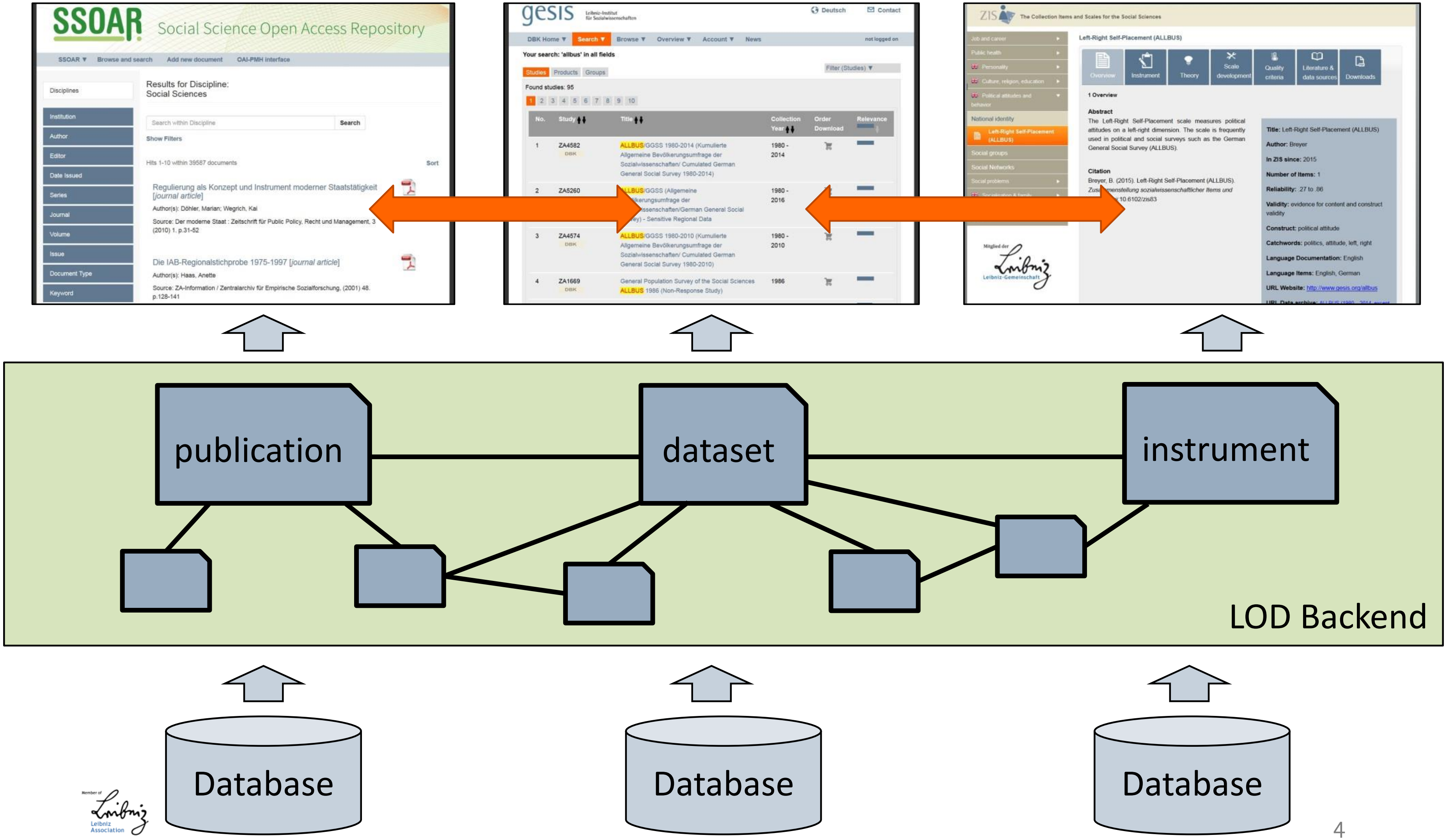
- 337 social science researchers in Germany
- Researchers are interested in links between information of different types and different sources



„I‘m looking for research data mentioned in a paper.“ (134 participants)

„I‘m looking for information which variables are included in a particular research dataset.“ (163 participants)

LOD backend infrastructure



Results for Discipline:
Social Sciences

Hits 1-10 within 39587 documents

Regulierung als Konzept und Instrument moderner Staatstätigkeit [*journal article*]
Author(s): Döhler, Marian; Wegrich, Kai
Source: Der moderne Staat : Zeitschrift für Public Policy, Recht und Management, 3 (2010) 1. p.31-52

Die IAB-Regionalstichprobe 1975-1997 [*journal article*]
Author(s): Haas, Anette
Source: ZA-Information / Zentralarchiv für Empirische Sozialforschung, (2001) 48. p.128-141

Your search: 'allbus' in all fields

Found studies: 95

No.	Study	Title	Collection Year	Order Download	Relevance
1	ZA4582 DBK	ALLBUS/GGSS 1980-2014 (Kumulierte Allgemeine Bevölkerungsumfrage der Sozialwissenschaften/ Cumulated German General Social Survey 1980-2014)	1980 - 2014		
2	ZA5260	LLBUS/GGSS (Allgemeine lkerungsumfrage der ssenschaften/German General Social ey) - Sensitive Regional Data	1980 - 2016		
3	ZA4574 DBK	ALLBUS/GGSS 1980-2010 (Kumulierte Allgemeine Bevölkerungsumfrage der Sozialwissenschaften/ Cumulated German General Social Survey 1980-2010)	1980 - 2010		
4	ZA1669 DBK	General Population Survey of the Social Sciences ALLBUS 1986 (Non-Response Study)	1986		

Left-Right Self-Placement (ALLBUS)

1 Overview

Abstract
The Left-Right Self-Placement scale measures political attitudes on a left-right dimension. The scale is frequently used in political and social surveys such as the German General Social Survey (ALLBUS).

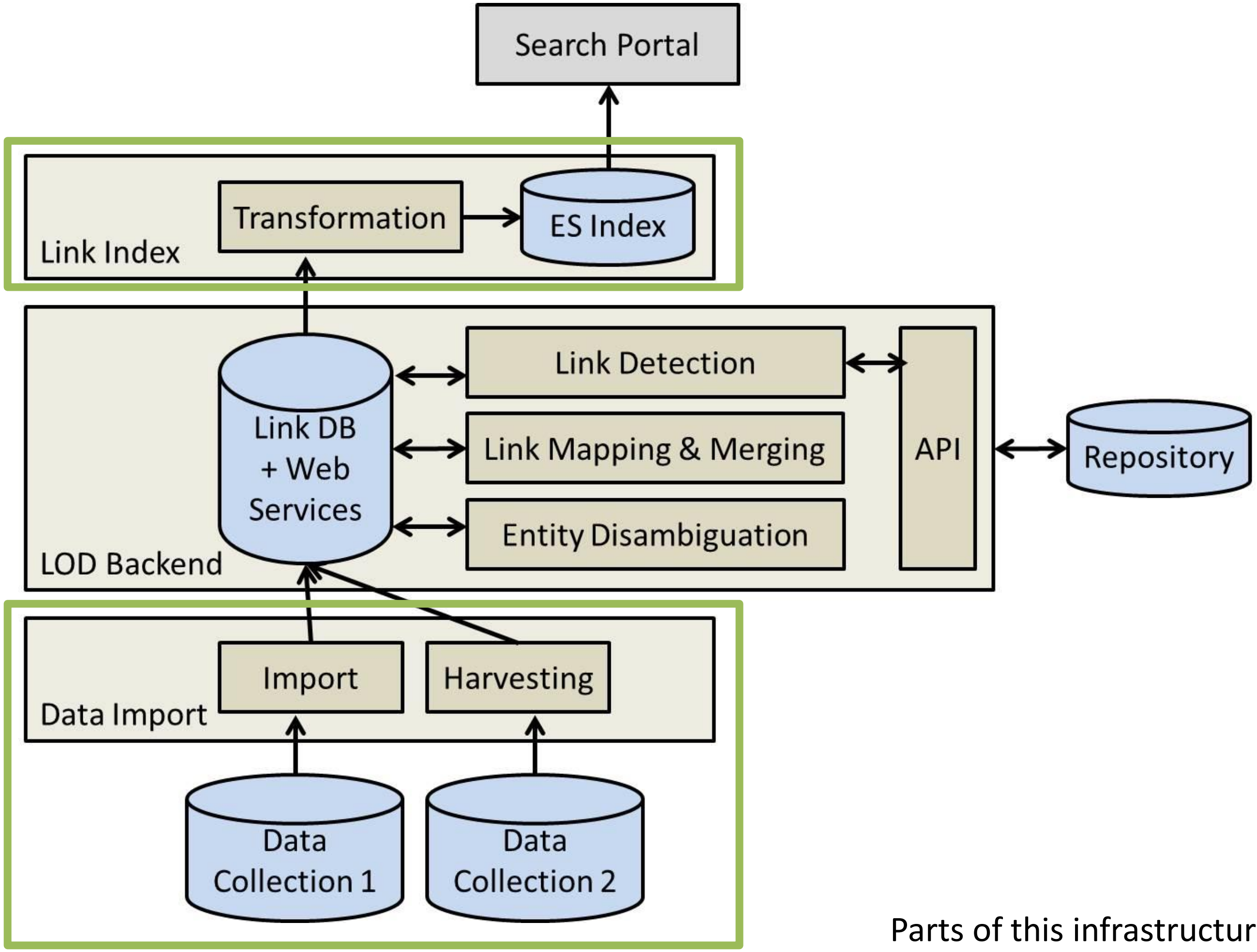
Citation
Breyer, B. (2015). Left-Right Self-Placement (ALLBUS). *Zusa menstellung sozialwissenschaftlicher Items und* i:10.6102/zis83

Title: Left-Right Self-Placement (ALLBUS)
Author: Breyer
In ZIS since: 2015
Number of Items: 1
Reliability: .27 to .86
Validity: evidence for content and construct validity
Construct: political attitude
Catchwords: politics, attitude, left, right
Language Documentation: English
Language Items: English, German
URL Website: http://www.gesis.org/allbus

LOD backend infrastructure

- Features
  - Collecting existing links between research objects from different data sources
  - Generating new links by link detection algorithms
  - Data is modelled as Linked Open Data
  - Links and attached information is available for search portals via a search index
- Existing search portals and their underlying infrastructures are not affected

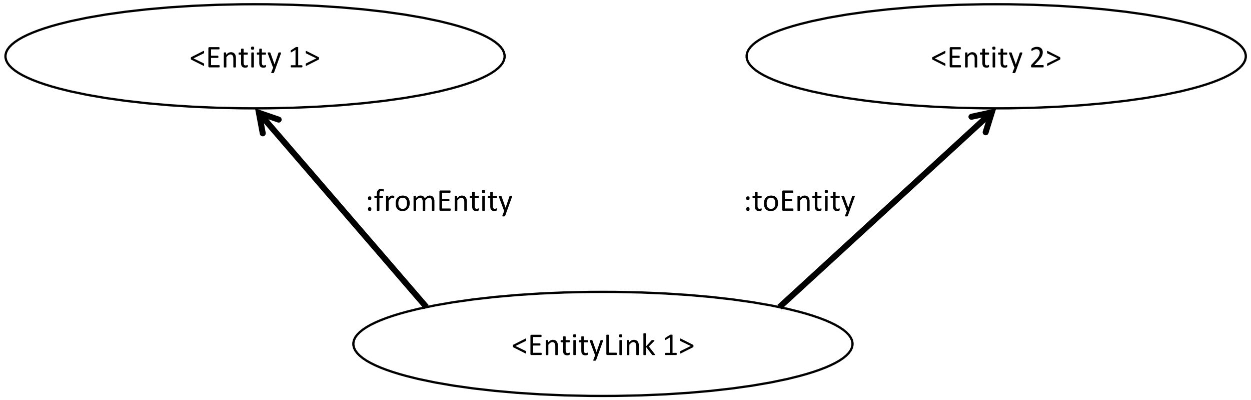
Architecture



Parts of this infrastructure are based on the project InFoLiS funded by DFG: http://www.infolis.gesis.org

Data model

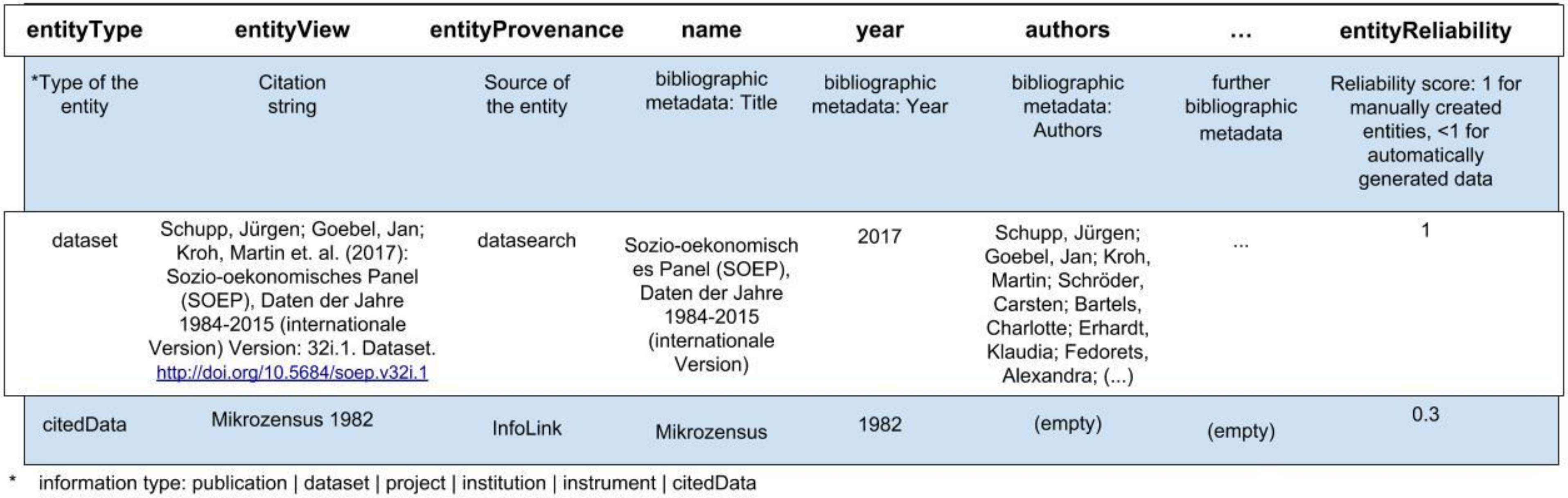
- Basic classes: Entity and EntityLink
- Extension of InFoLiS data model, e.g. additional entity types



Used vocabularies
OWL, RDF/RDFS, DC, SKOS, DCAT, DQM, BIBO, PROV-O

Entities

- Basic metadata about an entity, but also entity type, source, etc.

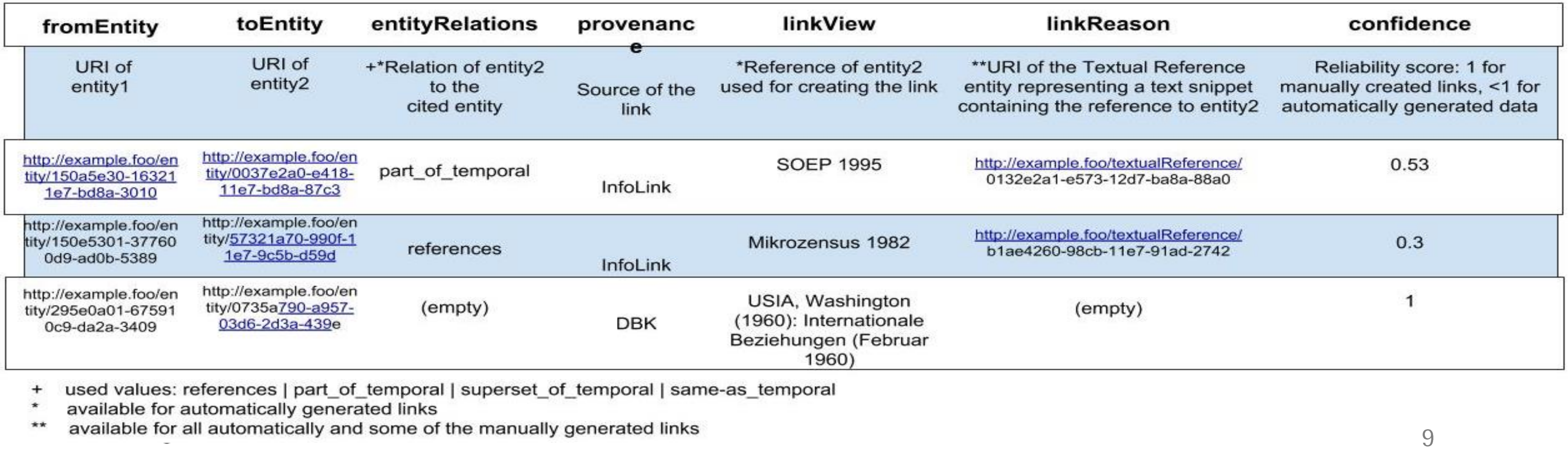


entityType	entityView	entityProvenance	name	year	authors	…	entityReliability
*Type of the entity	Citation string	Source of the entity	bibliographic metadata: Title	bibliographic metadata: Year	bibliographic metadata: Authors	further bibliographic metadata	Reliability score: 1 for manually created entities, <1 for automatically generated data
dataset	Schupp, Jürgen; Goebel, Jan; Kroh, Martin et. al. (2017): Sozio-oekonomisches Panel (SOEP), Daten der Jahre 1984-2015 (internationale Version) Version: 32i.1. Dataset. http://doi.org/10.5684/soep.v32i.1	datasearch	Sozio-oekonomisches Panel (SOEP), Daten der Jahre 1984-2015 (internationale Version)	2017	Schupp, Jürgen; Goebel, Jan; Kroh, Martin; Schröder, Carsten; Bartels, Charlotte; Erhardt, Klaudia; Fedorets, Alexandra; (...)	...	1
citedData	Mikrozensus 1982	InfoLink	Mikrozensus	1982	(empty)	(empty)	0.3

* information type: publication \| dataset \| project \| institution \| instrument \| citedData

EntityLinks

- Source and target of a link
- Type of relation, e.g. “references”
- Provenance information:
  - How was the link created? On which basis? How reliable is the link?

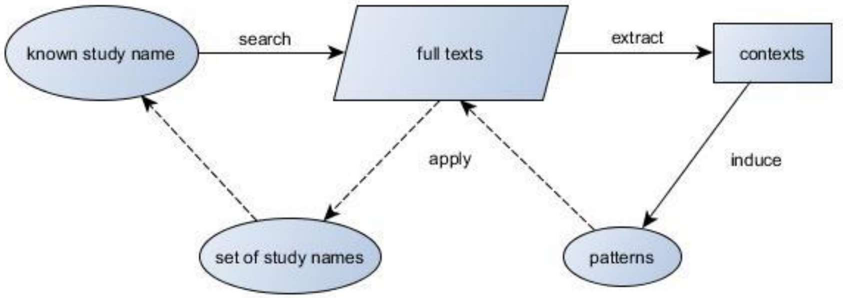


fromEntity	toEntity	entityRelations	provenance	linkView	linkReason	confidence
URI of entity1	URI of entity2	+*Relation of entity2 to the cited entity	Source of the link	*Reference of entity2 used for creating the link	**URI of the Textual Reference entity representing a text snippet containing the reference to entity2	Reliability score: 1 for manually created links, <1 for automatically generated data
http://example.foo/entity/150a5e30-163211e7-bd8a-3010	http://example.foo/entity/0037e2a0-e418-11e7-bd8a-87c3	part_of_temporal	InfoLink	SOEP 1995	http://example.foo/textualReference/0132e2a1-e573-12d7-ba8a-88a0	0.53
http://example.foo/entity/150e5301-377600d9-ad0b-5389	http://example.foo/entity/57321a70-990f-11e7-9c5b-d59d	references	InfoLink	Mikrozensus 1982	http://example.foo/textualReference/b1ae4260-98cb-11e7-91ad-2742	0.3
http://example.foo/entity/295e0a01-675910c9-da2a-3409	http://example.foo/entity/0735a790-a957-03d6-2d3a-439e	(empty)	DBK	USIA, Washington (1960): Internationale Beziehungen (Februar 1960)	(empty)	1

+ used values: references | part_of_temporal | superset_of_temporal | same-as_temporal
* available for automatically generated links
** available for all automatically and some of the manually generated links

Further data processing

- Link detection
  - Extraction and lookup of DOIs
  - Pattern-based reference extraction and linking
  - Term-based reference extraction and linking

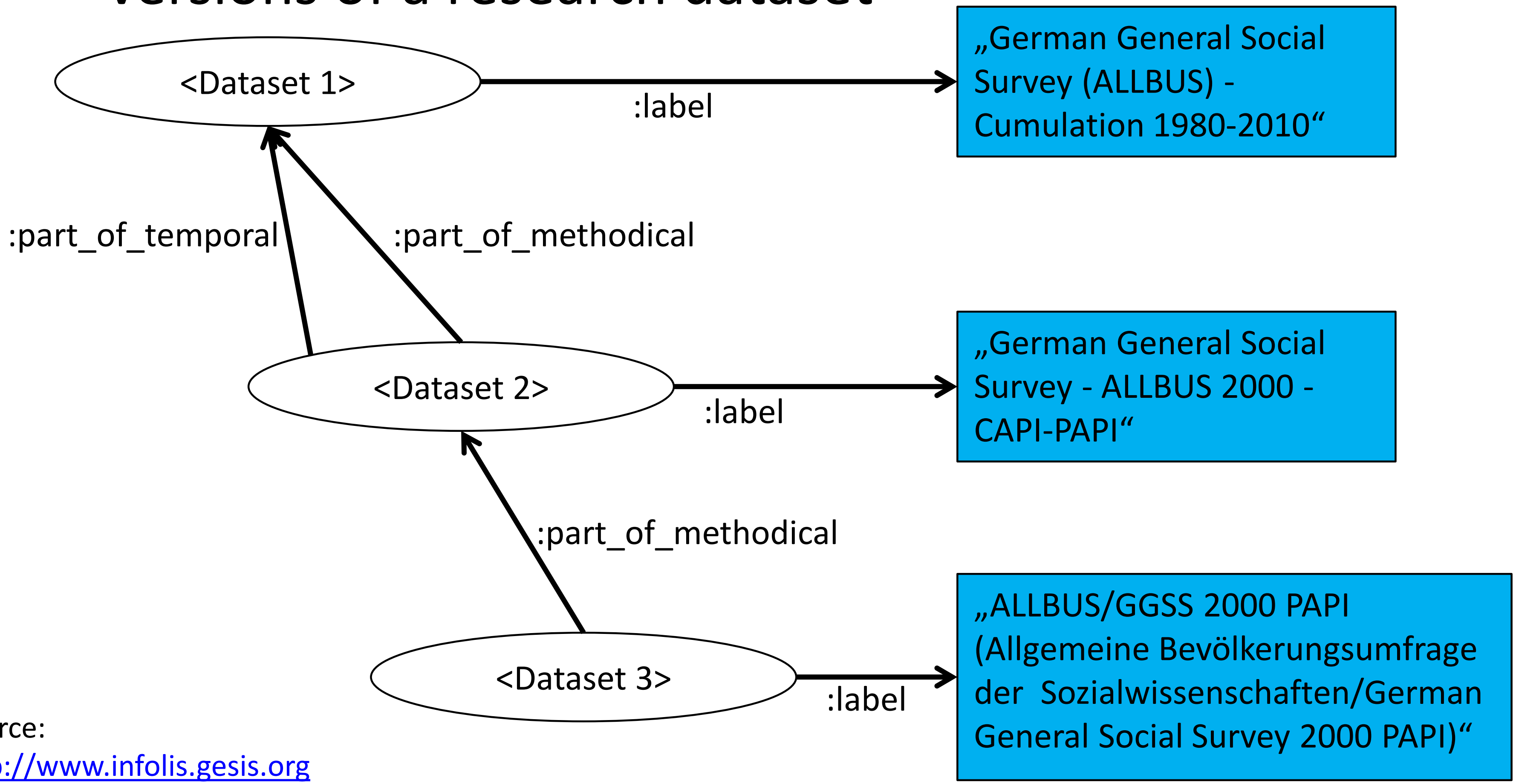


- Entity Disambiguation and link merging
  - ID matching
  - Disambiguation of datasets by modelling relationships with a research data ontology
  - Link merging for duplicate entities

For details, see: Boland et al. (2012). Identifying references to datasets in publications.

Research Data Ontology

- Necessity to generate relations between different versions of a research dataset



<Dataset 1>

:label

„German General Social Survey (ALLBUS) - Cumulation 1980-2010“

:part_of_temporal

:part_of_methodical

<Dataset 2>

:label

„German General Social Survey - ALLBUS 2000 - CAPI-PAPI“

:part_of_methodical

<Dataset 3>

:label

„ALLBUS/GGSS 2000 PAPI (Allgemeine Bevölkerungsumfrage der Sozialwissenschaften/German General Social Survey 2000 PAPI)“

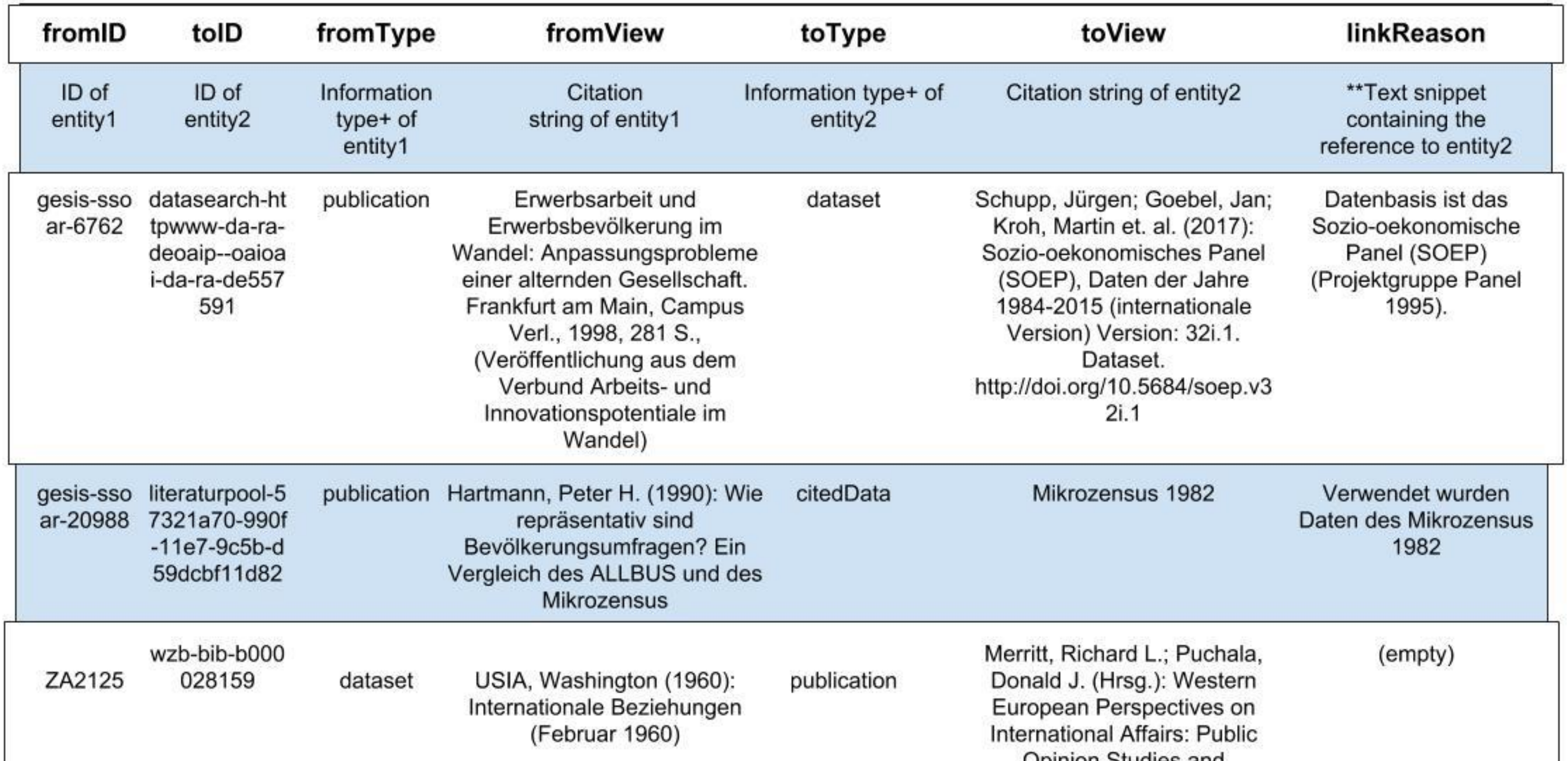
Source: http://www.infolis.gesis.org

gesis Leibniz Institute for the Social Sciences

Link database and search index

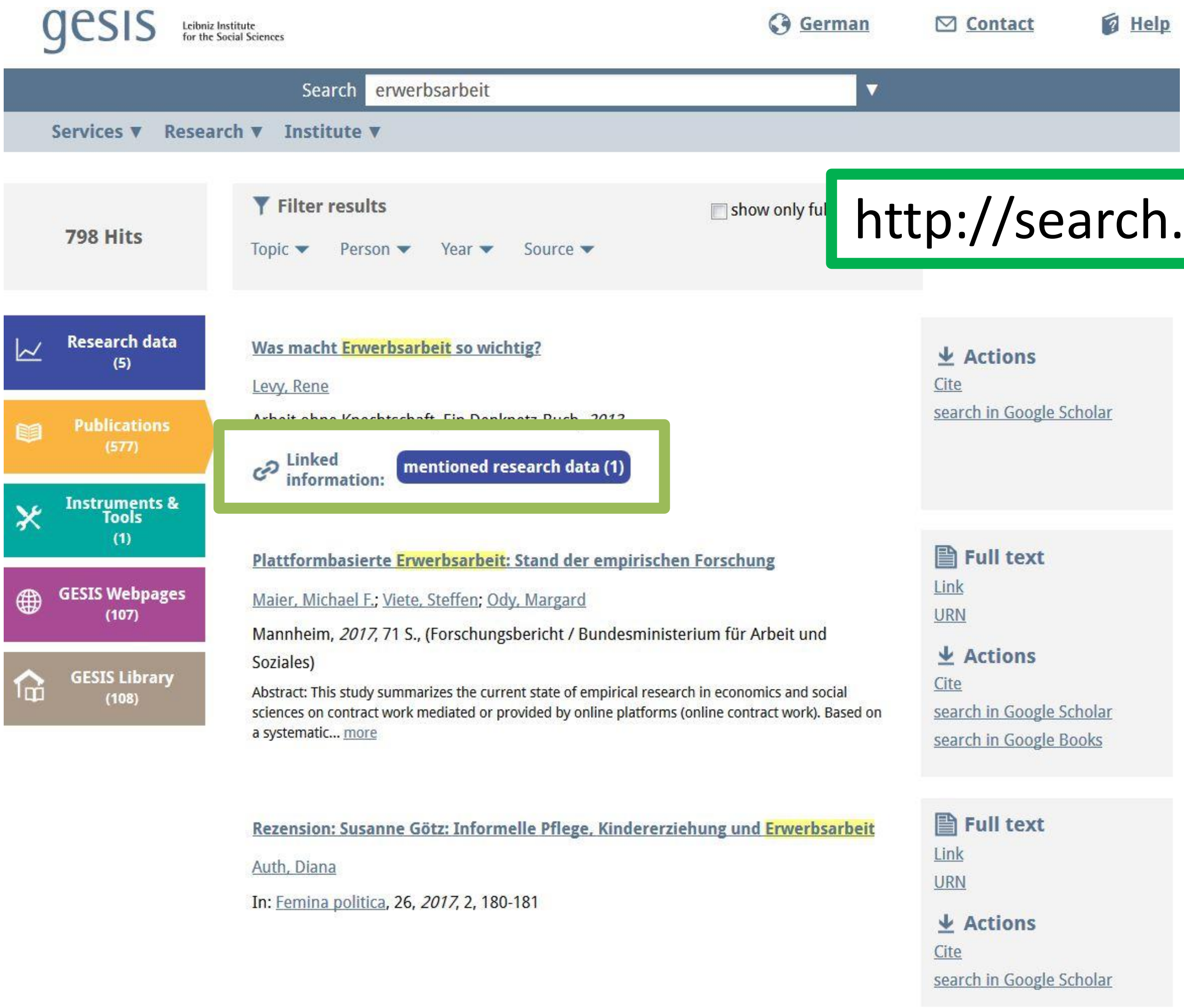
- Database: MongoDB
- Search index: Elasticsearch

108435 documents
277678 links



fromID	toID	fromType	fromView	toType	toView	linkReason
ID of entity1	ID of entity2	Information type+ of entity1	Citation string of entity1	Information type+ of entity2	Citation string of entity2	**Text snippet containing the reference to entity2
gesis-ssoar-6762	datasearch-httpwww-da-ra-deoaip--oaioai-da-ra-de557591	publication	Erwerbsarbeit und Erwerbsbevölkerung im Wandel: Anpassungsprobleme einer alternden Gesellschaft. Frankfurt am Main, Campus Verl., 1998, 281 S., (Veröffentlichung aus dem Verbund Arbeits- und Innovationspotentiale im Wandel)	dataset	Schupp, Jürgen; Goebel, Jan; Kroh, Martin et. al. (2017): Sozio-oekonomisches Panel (SOEP), Daten der Jahre 1984-2015 (internationale Version) Version: 32i.1. Dataset. http://doi.org/10.5684/soep.v32i.1	Datenbasis ist das Sozio-oekonomische Panel (SOEP) (Projektgruppe Panel 1995).
gesis-ssoar-20988	literaturpool-57321a70-990f-11e7-9c5b-d59dcbf11d82	publication	Hartmann, Peter H. (1990): Wie repräsentativ sind Bevölkerungsumfragen? Ein Vergleich des ALLBUS und des Mikrozensus	citedData	Mikrozensus 1982	Verwendet wurden Daten des Mikrozensus 1982
ZA2125	wzb-bib-b000028159	dataset	USIA, Washington (1960): Internationale Beziehungen (Februar 1960)	publication	Merritt, Richard L.; Puchala, Donald J. (Hrsg.): Western European Perspectives on International Affairs: Public Opinion Studies and	(empty)

Scientific search portal



gesis Leibniz Institute for the Social Sciences

German Contact Help

Search erwerbsarbeit ▼

Services ▼ Research ▼ Institute ▼

798 Hits

Filter results

□ show only ful

Topic ▾ Person ▾ Year ▾ Source ▾

Research data (5)

Publications (577)

Instruments & Tools (1)

GESIS Webpages (107)

GESIS Library (108)

Was macht Erwerbsarbeit so wichtig?

Levy, Rene

Arbeit ohne Knechtschaft. Ein Denknetz-Buch, 2013

Linked information: mentioned research data (1)

Actions
Cite
search in Google Scholar

Plattformbasierte Erwerbsarbeit: Stand der empirischen Forschung

Maier, Michael F.; Viete, Steffen; Ody, Margard

Mannheim, *2017*, 71 S., (Forschungsbericht / Bundesministerium für Arbeit und Soziales)

Abstract: This study summarizes the current state of empirical research in economics and social sciences on contract work mediated or provided by online platforms (online contract work). Based on a systematic... more

Full text
Link
URN

Actions
Cite
search in Google Scholar
search in Google Books

Rezension: Susanne Götz: Informelle Pflege, Kindererziehung und Erwerbsarbeit

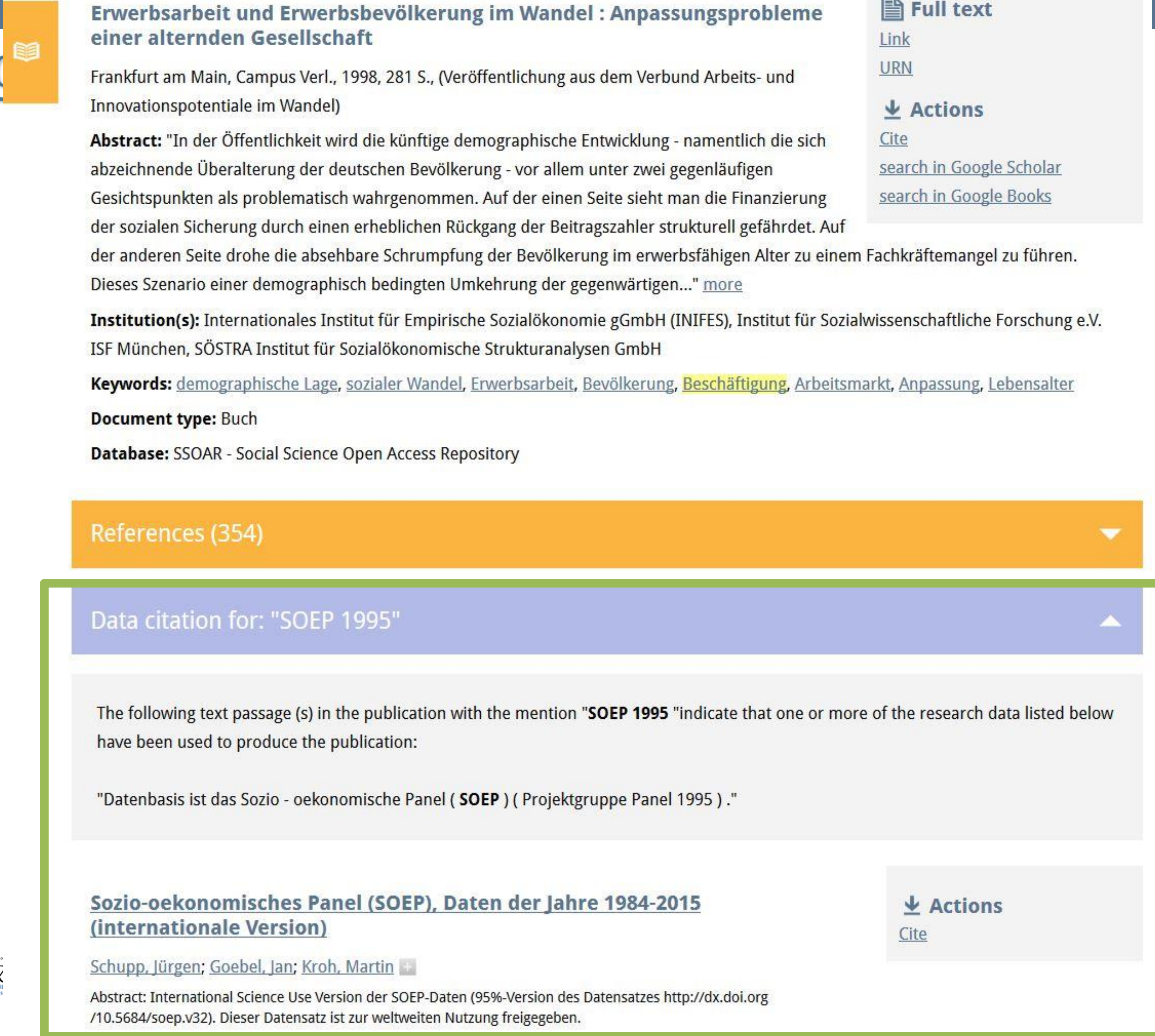
Auth, Diana

In: Femina politica, 26, *2017*, 2, 180-181

Full text
Link
URN

Actions
Cite
search in Google Scholar

http://search.gesis.org



Erwerbsarbeit und Erwerbsbevölkerung im Wandel : Anpassungsprobleme einer alternden Gesellschaft

Frankfurt am Main, Campus Verl., 1998, 281 S., (Veröffentlichung aus dem Verbund Arbeits- und Innovationspotentiale im Wandel)

Abstract: "In der Öffentlichkeit wird die künftige demographische Entwicklung - namentlich die sich abzeichnende Überalterung der deutschen Bevölkerung - vor allem unter zwei gegenläufigen Gesichtspunkten als problematisch wahrgenommen. Auf der einen Seite sieht man die Finanzierung der sozialen Sicherung durch einen erheblichen Rückgang der Beitragszahler strukturell gefährdet. Auf der anderen Seite drohe die absehbare Schrumpfung der Bevölkerung im erwerbsfähigen Alter zu einem Fachkräftemangel zu führen. Dieses Szenario einer demographisch bedingten Umkehrung der gegenwärtigen..." more

Institution(s): Internationales Institut für Empirische Sozialökonomie gGmbH (INIFES), Institut für Sozialwissenschaftliche Forschung e.V. ISF München, SÖSTRA Institut für Sozialökonomische Strukturanalysen GmbH

Keywords: demographische Lage, sozialer Wandel, Erwerbsarbeit, Bevölkerung, Beschäftigung, Arbeitsmarkt, Anpassung, Lebensalter

Document type: Buch

Database: SSOAR - Social Science Open Access Repository

Full text

Link

URN

Actions

Cite

search in Google Scholar

search in Google Books

References (354)

Data citation for: "SOEP 1995"

The following text passage (s) in the publication with the mention "**SOEP 1995** "indicate that one or more of the research data listed below have been used to produce the publication:

"Datenbasis ist das Sozio - oekonomische Panel (**SOEP**) (Projektgruppe Panel 1995) ."

Sozio-oekonomisches Panel (SOEP), Daten der Jahre 1984-2015 (internationale Version)

Schupp, Jürgen; Goebel, Jan; Kroh, Martin

Abstract: International Science Use Version der SOEP-Daten (95%-Version des Datensatzes http://dx.doi.org /10.5684/soep.v32). Dieser Datensatz ist zur weltweiten Nutzung freigegeben.

Actions

Cite

Evaluation

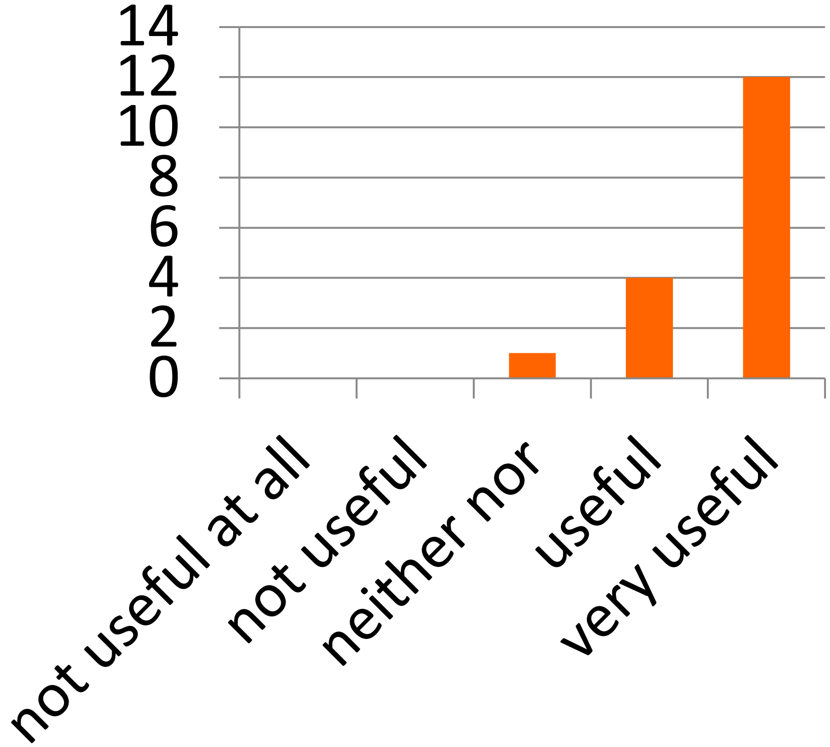
- Evaluation of user experience
- Scenario: GESIS search portal, http://search.gesis.org
- User study
  - 17 participants from German universities
  - 7 female, 10 male
  - Average age 33.35 years
  - 3 professors, 4 postdocs, 9 research associates, 1 student assistant
  - Recruitment by email

Evaluation

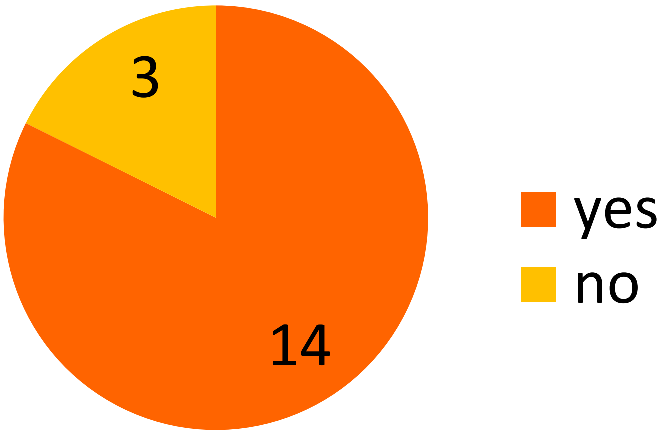
- 2 steps (both think-aloud method):
  - 1. Prescribed evaluation scenario to familiarize participants with interlinked information
  - 2. Free exploration phase
- Survey at the end regarding
  - Usefulness
  - Trust in provided links
  - Completeness of linked information
  - Origin of linked information

Results

- Usefulness

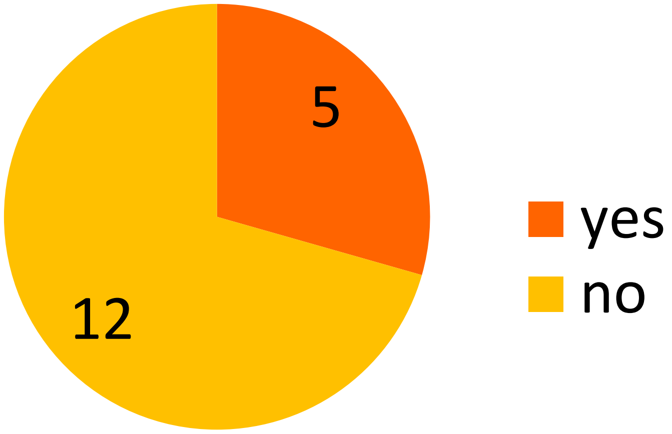


- Trust in provided links

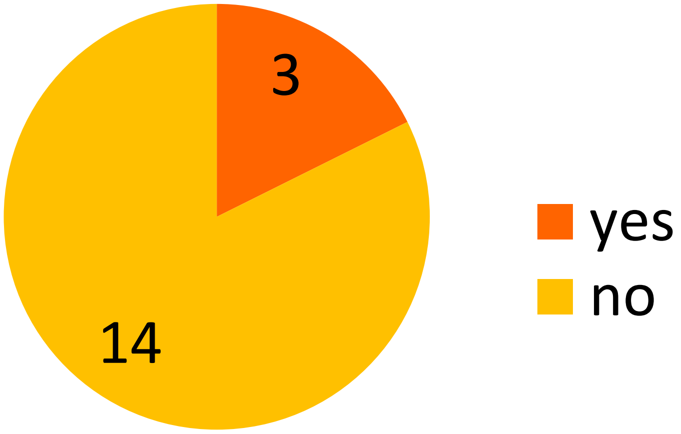


Results

- Completeness



- Origin of links



Challenges

- After following a couple of links
  - Users may get lost and have difficulties to find their starting point
  - Relation to original information gets lower

General applicability

- All components have been developed independently of any specific portal or metadata
  - All components can be reused independent from each other as web service via the API
- Extensible architecture
  - New data sources = new importers / harvesters
- Extensible data model
  - For including new information types
- Source code: http://github.com/infolis

Future Work

- Switching from MongoDB to a triple store
- Linking with thesauri, authority data and external knowledge graphs
- Author disambiguation

Acknowledgements

- Parts of the infrastructure, the data model, and the Research Data Ontology have been developed jointly with **University Library Mannheim**, **University Mannheim**, and **Stuttgart Media University** in the project InFoLiS funded by DFG: http://www.infolis.gesis.org

LOD infrastructure at GESIS: http://search.gesis.org
Source code: http://github.com/infolis

Contact: Dr. Benjamin Zapilko
benjamin.zapilko@gesis.org

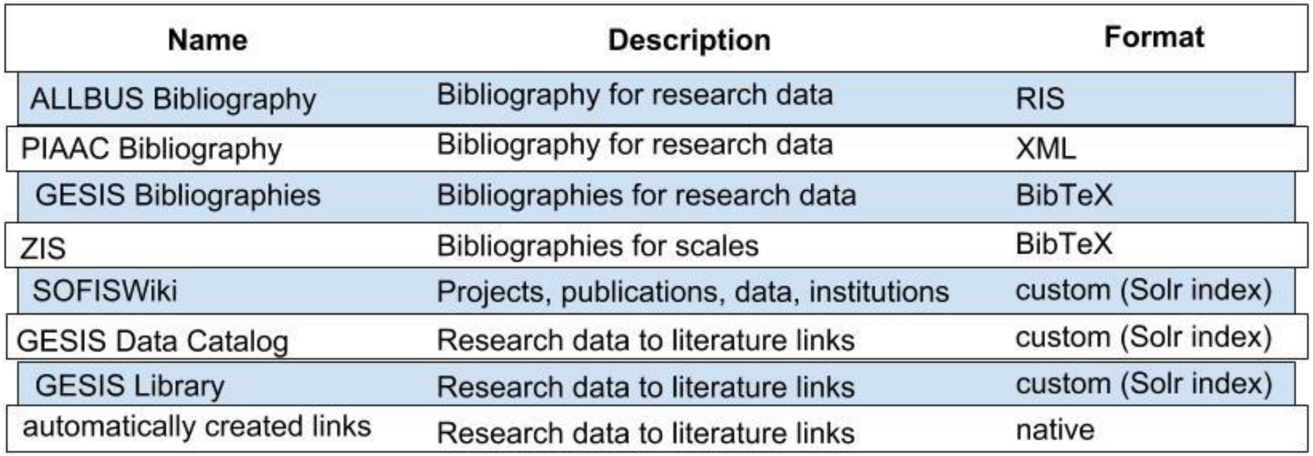
Thank you for your attention!

gesis
Leibniz Institute
for the Social Sciences

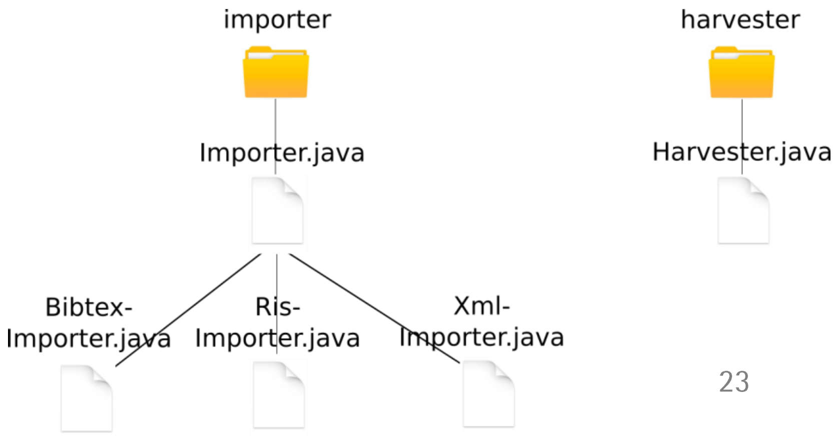
Member of Leibniz
Leibniz
Association

Data import

- Different importers and harvesters for different sources and formats



Name	Description	Format
ALLBUS Bibliography	Bibliography for research data	RIS
PIAAC Bibliography	Bibliography for research data	XML
GESIS Bibliographies	Bibliographies for research data	BibTeX
ZIS	Bibliographies for scales	BibTeX
SOFISWiki	Projects, publications, data, institutions	custom (Solr index)
GESIS Data Catalog	Research data to literature links	custom (Solr index)
GESIS Library	Research data to literature links	custom (Solr index)
automatically created links	Research data to literature links	native



Why a Research Data Ontology?

- A research dataset can be available in different aggregations and versions with different IDs

„German General Social Survey (ALLBUS) - Cumulation 1980-2010“

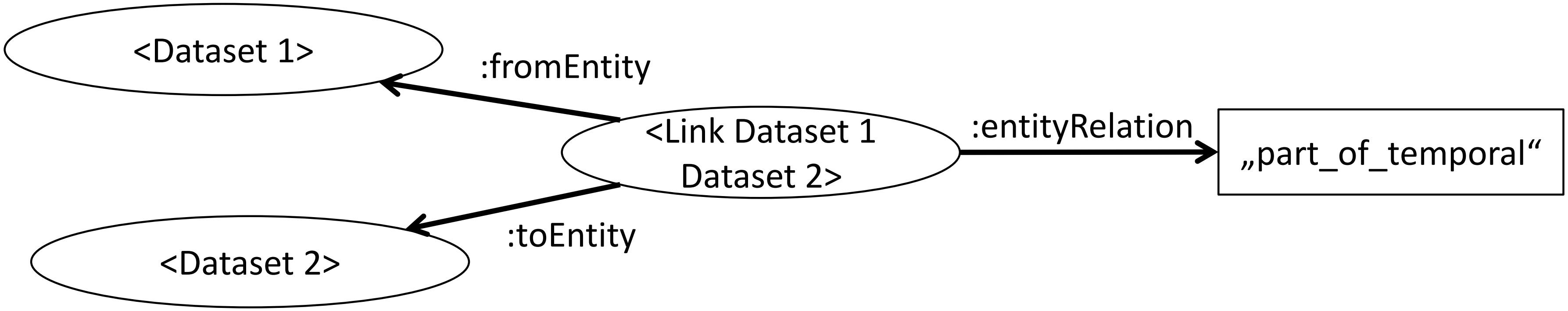
„German General Social Survey - ALLBUS 2000 - CAPI-PAPI“

„ALLBUS/GGSS 2000 PAPI (Allgemeine Bevölkerungsumfrage der Sozialwissenschaften/German General Social Survey 2000 PAPI)“

- Necessity to generate relations between different versions of a research dataset
  - The detected target of an EntityLink is often unprecise, e.g. “German General Social Survey 2000”

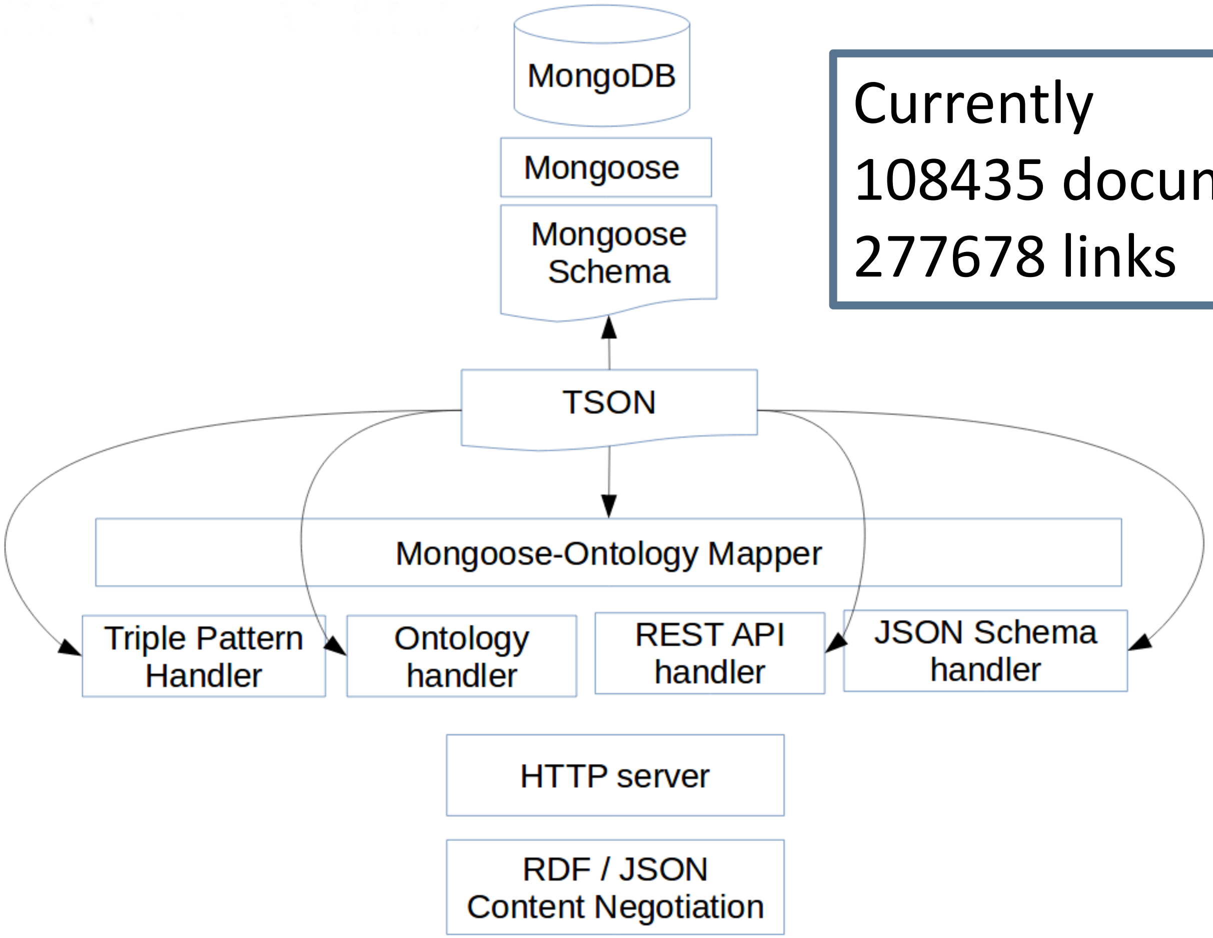
Research Data Ontology

- Adds new properties to the data model



:part_of_ / :superset_of_	Example
temporal	Cumulated over time
spatial	Different countries
methodical	Different collection methods
sample	Subsamples
confidential	Different privacy restrictions

Link database

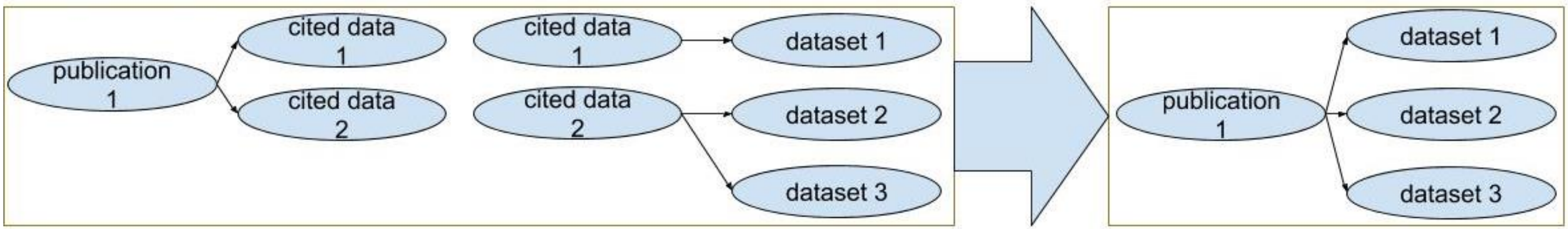


Currently
108435 documents
277678 links

Source: Baierer et al (2015): A RESTful JSON-LD Architecture for Unraveling Hidden References to Research Data

Link transformation

- Flattening of indirect links for efficient queries


publication 1
cited data 1
cited data 2
cited data 1
cited data 2
dataset 1
dataset 2
dataset 3
publication 1
dataset 1
dataset 2
dataset 3