

gesis

Leibniz-Institut
für Sozialwissenschaften



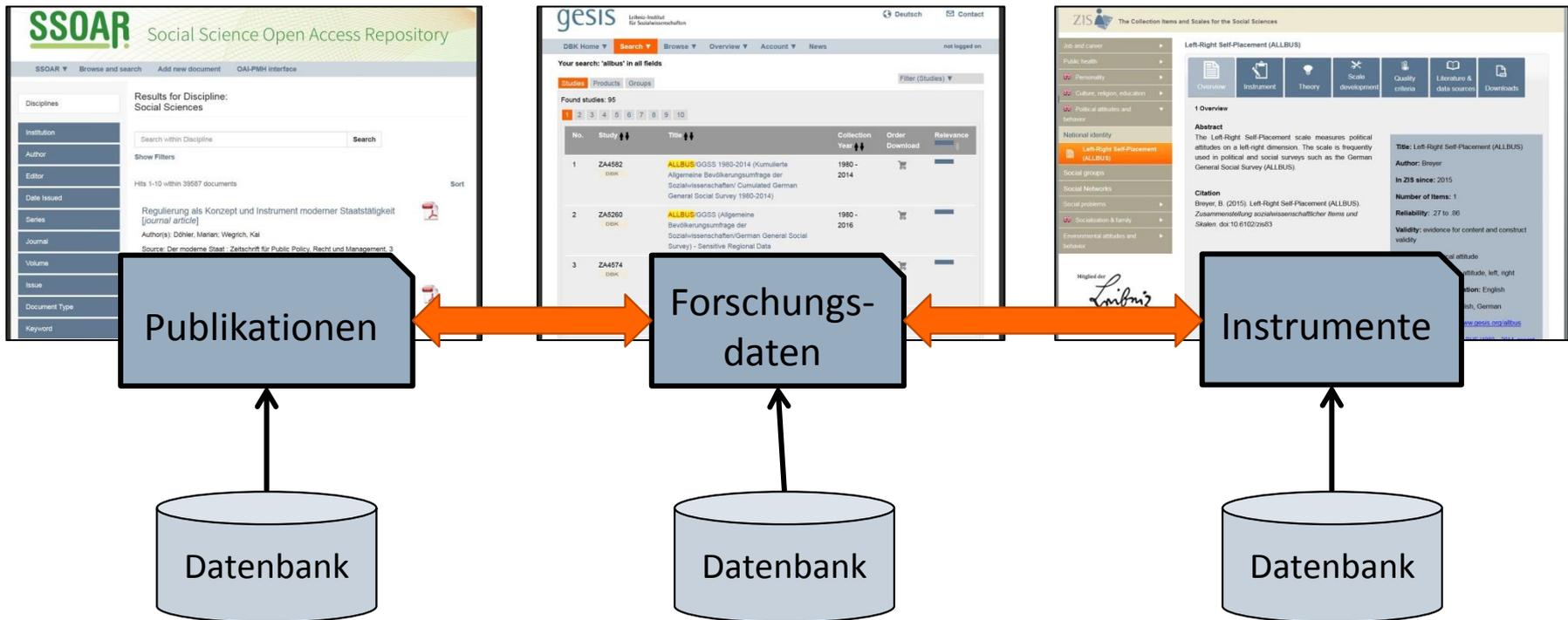
Vernetzte Forschungs- informationen bei GESIS

Benjamin Zapilko

*Workshop Erfolgreiches Journal-
Management, Berlin, 17.01.2019*

Suche nach Forschungsinformationen

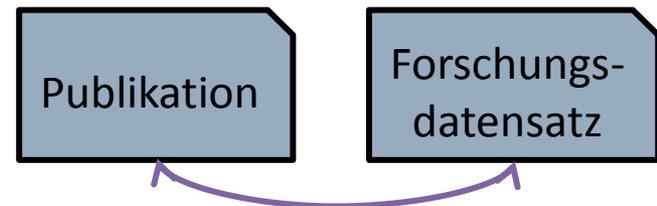
- Heterogene Forschungsinformationen sind in verschiedenen Portalen und Datenbanken getrennt voneinander verfügbar



Nutzerumfrage

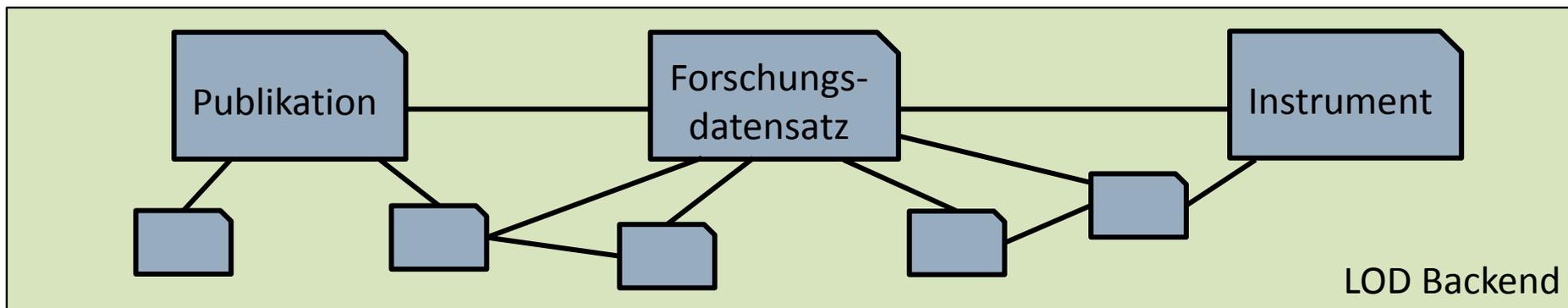
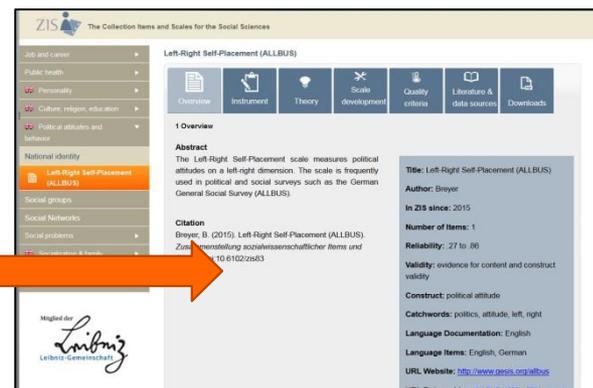
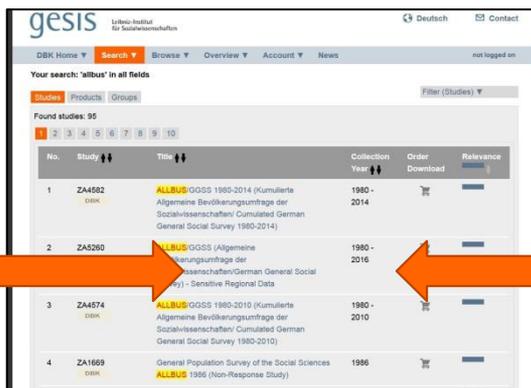
- 337 Sozialwissenschaftler in Deutschland
- Forscher sind an Links zwischen verschiedenen Informationstypen und Informationsquellen interessiert

„Ich suche nach zitierten Forschungsdaten.“
(134 Teilnehmer)



„Ich suche nach Informationen, welche Variablen in einem bestimmten Forschungsdatensatz enthalten sind.“ (163 Teilnehmer)

LOD Backend Infrastruktur



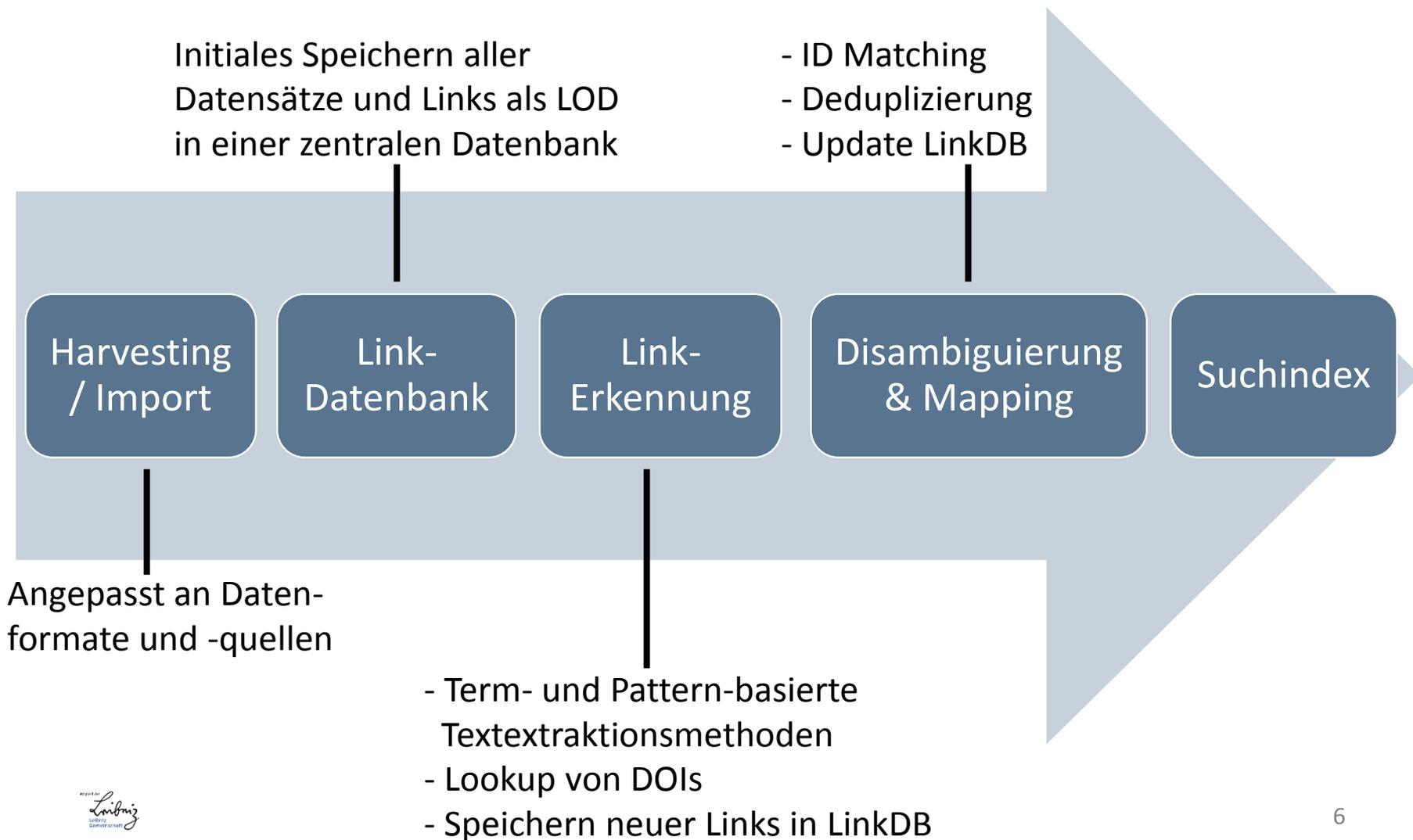
LOD Backend

LOD Backend Infrastruktur

- Ziel
 - ▶ Vernetzung der Forschungsinformationen bei GESIS
 - ▶ Verbesserung der integrierten Suche nach Forschungsinformationen bei GESIS
 - ▶ Beibehaltung der existierenden historisch gewachsenen Infrastrukturen und Portale

- Anforderungen
 - ▶ Sammeln aller existierenden Links zwischen Forschungsobjekten aus verschiedenen Datenquellen
 - ▶ Generierung neuer Links
 - ▶ Effiziente Einbindung in Portale und andere Systeme

Workflow



Enthaltene Datenquellen

- Harvesting- und Importroutinen für verschiedene Datenquellen

Aktuell:
 108435 Dokumente
 277678 Links

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

Verwendete Vokabulare

Dublin Core, SKOS, DCAT, DQM, BIBO, PROV-O

Metadaten für Datensätze

- Grundlegende Metadaten über einen Datensatz, aber auch Informationstyp, Quelle, 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

Metadaten für Links

- Quelle und Ziel eines Links
- Linktyp, z.B. “referenziert”
- Provenienzinformation
 - ▶ Automatische oder manuelle Linkgenerierung, Basis, Konfidenzwerte

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

Suchportal: GESIS Suche



Leibniz-Institut
für Sozialwissenschaften

http://search.gesis.org

Suche

Angebot ▾
Forschung ▾
Institut ▾

8.607 Treffer

Ergebnisse filtern

Thema ▾ Person ▾ Jahr ▾ Quelle ▾

Datenkollektionen

nur GESIS (665)

GESIS und weitere (3.810)

Sortieren:

Relevanz ▾

Forschungsdaten
(665)

Publikationen
(5.987)

Instrumente & Tools
(7)

GESIS-Webseiten
(871)

GESIS-Bibliothek
(1.077)

2000 Families: Migration Histories of Turks in Europe

[Guveli, Ayse](#); [Ganzeboom, Harry B.G.](#); [Baykara-Krumme, Helen](#) +

GESIS Datenarchiv, Köln. ZA5957 Datenfile Version 2.0.0 (2016), doi:10.4232/1.12673, Erhebungszeitraum: 2010 - 2012

Abstract: Das Projekt 2000 Families: **Migration**sgeschichten von Türken in Europa untersucht Prozesse der **Migration**, die Mehr-Generationen-Übertragung von sozialen, kulturellen, religiösen und wirtschaftlichen ... mehr

🔗

Verknüpfte Informationen:

Publikationen (1)

Transatlantic Trends Immigration Survey 2011

[German Marshall Fund of the United States \(GMF\)](#), Washington, USA; [Isernia, Pierangelo](#)

GESIS Datenarchiv, Köln. ZA5362 Datenfile Version 1.0.0 (2013), doi:10.4232/1.11522, Erhebungszeitraum: 25.08.2011 - 18.09.2011

Abstract: Wahrnehmung von Einwanderern. Zulassung von Einwanderern. Voraussetzungen zur Einbürgerung. Maßnahmen zur Reduzierung der illegalen bzw. legalen Einwanderung. Einwanderung ... mehr

🔗

Verknüpfte Informationen:

Publikationen (1)

Materialien

[Datensätze](#)

[Codebuch](#)

Aktionen

[Zitieren](#)

Materialien

[Datensätze](#)

[Codebuch](#)

[Andere Dokumente](#)

Aktionen

[Zitieren](#)

Untersuchungsgebiet: Belgien, Dänemark, Deutschland, Frankreich, Niederlande, Norwegen, Österreich, Schweden, Schweiz, Türkei, Vereinigtes Königreich

Anzahl der Einheiten: 48980

Anzahl der Variablen: 499

Analyse-System(e): SPSS, Stata

Grundgesamtheit: Türkische Migrantenfamilien und ihre Nicht-Migranten Pendants in fünf Regionen (Acıpayam, Akçaabat, Emirdağ, Kulu und Şarkışla) in der Türkei mit hoher Entsendung von Migranten

Auswahl: Wahrscheinlichkeitsauswahl: Mehrstufige Zufallsauswahl

Erhebungsverfahren: Persönliches Interview: CAPI (Computerunterstützte persönliche Befragung) Telefonisches Interview: CATI (Computerunterstützte telefonische Befragung)

Datenerhebung: Frekans Research Field and Data Processing Co. Ltd. (2010) Ipsos KMG (2011-2012)

Schlagworte: [Migration](#), [Kulturelle und nationale Identität](#), [Familie und Ehe](#)

Eintrag im Fachportal: [Datenbestandskatalog \(DBK\)](#)

Dazugehörige Publikationen (1)

[Intergenerational consequences of migration](#)

[Guveli, Ayse](#); [Ganzeboom, Harry](#); [Platt, Lucinda](#) 

Basingstoke, Palgrave Macmillan, 2015, Online-Ressource (S. 301 p)

Abstract: This book analyzes the impact of [migration](#) on the lives of multiple generations of 2000 Turkish families. Exploring education, marriage, fertility, friends, attitudes and religiosity, it reveals... [mehr](#)

 **Volltext**

[Link](#)

 **Aktionen**

[Zitieren](#)

[in Google Scholar suchen](#)

Dual Citizenship Allowance and Migration Flow

[Alarian, Hannah M.](#); [Goodman, Sara Wallace](#)

Comparative Political Studies, 2017

Datenbank: GESIS Bibliography

↓ Aktionen

[Zitieren](#)

[in Google Scholar suchen](#)

Verwendete Forschungsdaten (2)

Folgende Forschungsdaten wurden zur Erstellung dieser Publikation verwendet - ohne Anspruch auf Vollständigkeit.

Sortieren nach: ▼

[Eurobarometer 70.1 \(Oct-Nov 2008\)](#)

[Papacostas, Antonis](#)

GESIS Datenarchiv, Köln. ZA4819 Datenfile Version 3.0.2 (2012), doi:10.4232/1.10989,

Erhebungszeitraum: 06.10.2008 - 24.11.2008

Abstract: Fragen zur Europäischen Union. Auswirkungen der Globalisierung. Europäische Integration. Aufbau Europas, Europäisches Parlament und Wahlen. Internationale Konflikte: Georgien. Mobilität. EU-Budget... [mehr](#)

📄 Materialien

[Datensätze](#)

[Fragebögen](#)

[Codebuch](#)

↓ Aktionen

[Zitieren](#)

[Eurobarometer 67.3 \(May-Jun 2007\)](#)

[Papacostas, Antonis](#)

GESIS Datenarchiv, Köln. ZA4561 Datenfile Version 2.1.0 (2012), doi:10.4232/1.10985,

Erhebungszeitraum: 25.05.2007 - 27.06.2007

Abstract: Gesundheitliche Einschränkungen im Alltagsleben durch seelische oder körperliche Erkrankungen. Gesundheitswesen. Pflegeleistungen für hilfsbedürftige Personen. Schwarzarbeit. Meinung zu... [mehr](#)

📄 Materialien

[Datensätze](#)

[Fragebögen](#)

[Codebuch](#)

↓ Aktionen

[Zitieren](#)

Labour Migration, Market Competition and Ethnocentrism: Guest workers in Israel and Germany

Schmidt, Peter; Heyder, Aribert

GESIS Datenarchiv, Köln. ZA3974 Datenfile Version 1.0.0 (2010), doi:10.4232/1.3974,
Erhebungszeitraum: 03.1996 - 1999

Abstract: Einstellung zu Gastarbeitern und Ausländern allgemein in Deutschland und Israel.
Kumulierter Datensatz aus einer Erhebung in Israel und Teilen des ALLBUS 1996. Themen: Deutscher
Befragungsteil: 1.... [mehr](#)

 **Verknüpfte
Informationen:**

Publikationen (17)

Instrumente & Tools (1)

Schlagworte: [Berufstätigkeit](#), [Migration](#), [Kulturelle und nationale Identität](#), [Religion und Werte](#), [Politische Verhaltensweisen und Einstellungen / Meinungen](#)

Eintrag im Fachportal: [Datenbestandskatalog \(DBK\)](#)

Dazugehörige Publikationen (17) ▼

Instrumente & Tools (1) ▼

Zitieren:

APA

Bibtex

Endnote

Citavi

Schmidt, Peter, & Heyder, Aribert (2003). Labour Migration, Market Competition and Ethnocentrism: Guest workers in Israel and Germany .
GESIS Datenarchiv, Köln. ZA3974 Datenfile Version 1.0.0 (2010), doi:10.4232/1.3974. doi:10.4232/1.3974.

[Download](#)

geschichtete Zufallsauswahl in West- (einschl. West-
en 104 Gemeinden in Westdeutschland und 47
ihrer erwachsenen Einwohne... [mehr](#)

n Fragebogen und schriftliche Zusatzbefragung (drop off)
bogen (CATI)

Nutzerstudie

- Evaluation des Nutzererlebnis
- Szenario: GESIS Suche, <http://search.gesis.org>
- Nutzerstudie
 - ▶ 17 Teilnehmer von deutschen Universitäten
 - ▶ 7 weibliche, 10 männliche Teilnehmer
 - ▶ Durchschnittsalter 33.35 Jahre
 - ▶ 3 Professoren, 4 Postdocs, 9 wissenschaftliche Mitarbeiter, 1 Student
 - ▶ Anwerbung via Email

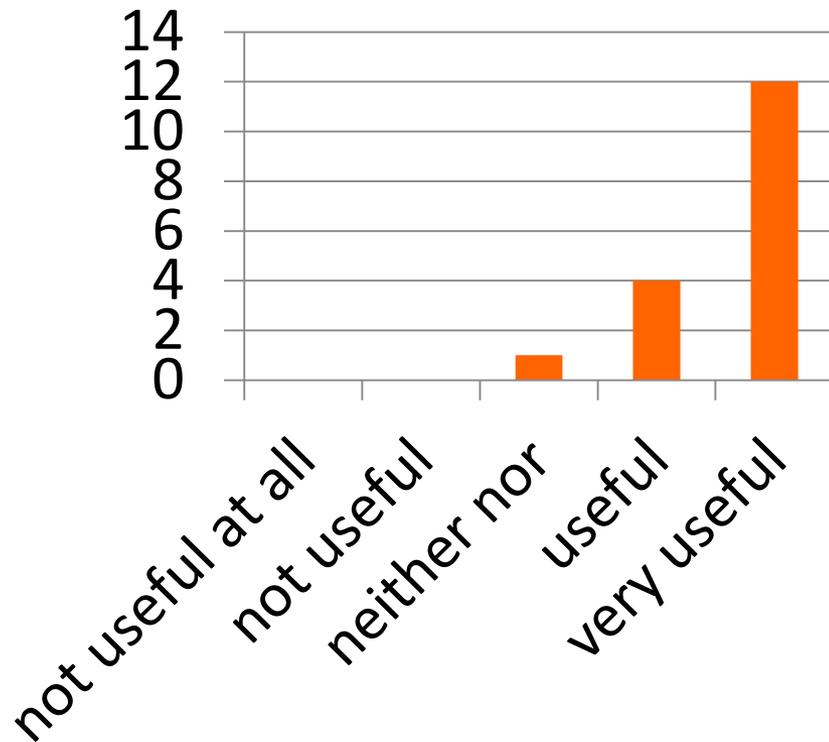
Nutzerstudie

- 2 Schritte (beide Think-Aloud Methode)
 1. Vorgegebenes Evaluationsszenario, um sich mit verlinkter Information vertraut zu machen
 2. Phase zur freien Exploration

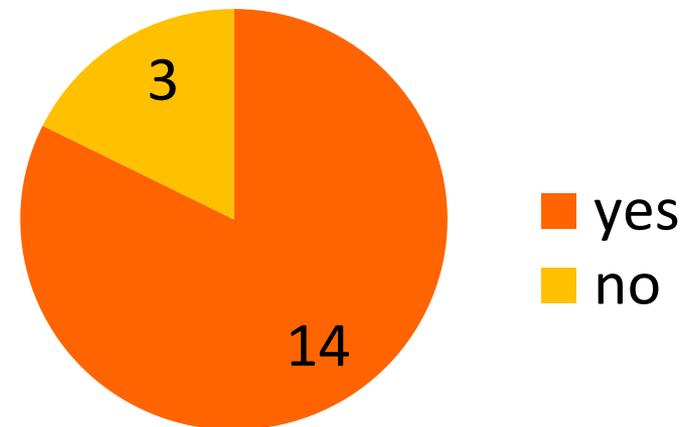
- Umfrage zum Abschluss
 - ▶ Nützlichkeit der Links
 - ▶ Vertrauen in Links
 - ▶ Vollständigkeit der Links
 - ▶ Herkunft der Links

Ergebnisse

- Nützlichkeit der Links

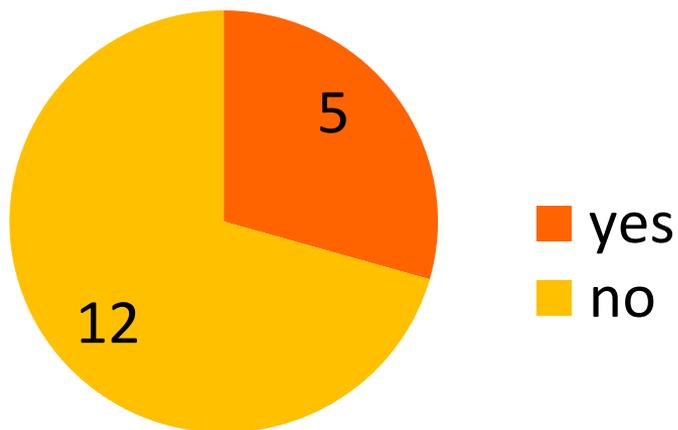


- Vertrauen in Links

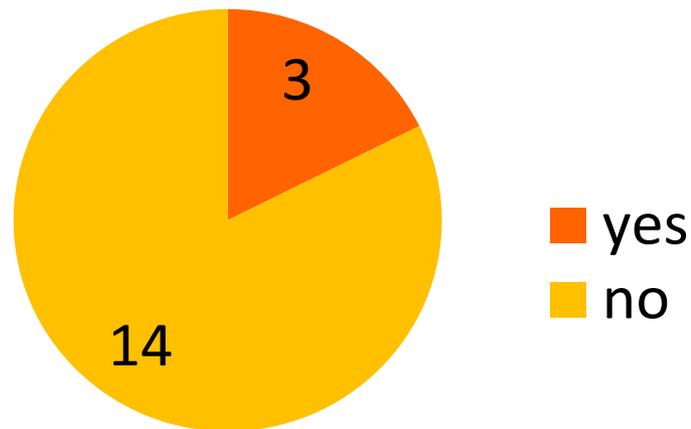


Ergebnisse

- Vollständigkeit der Links



- Herkunft der Links



Ergebnisse

- Nutzende verlieren sich nach einigen gefolgten Links und haben Schwierigkeiten, den Ausgangspunkt zu finden
- Semantische Verbindung zum Ausgangspunkt wird beim Folgen mehrerer Links immer schwächer

Nachnutzbarkeit

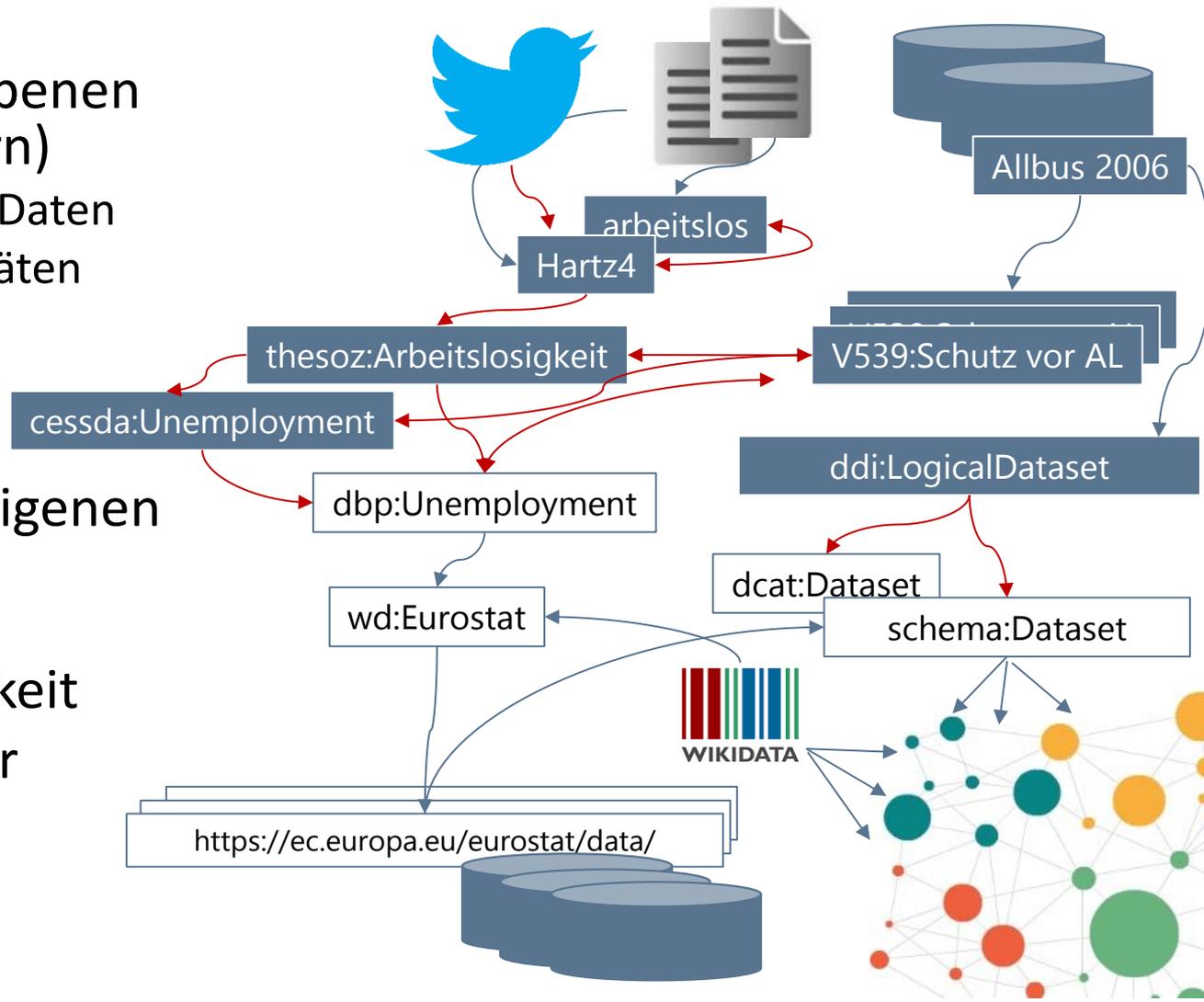
- Entwicklung unabhängig von Portalen und Metadaten
 - ▶ Wiederverwendbarkeit einzelner Komponenten unabhängig voneinander als Webservices über API
- Erweiterbare Architektur
 - ▶ Neue Datenquellen -> Neue Importer / Harvester
- Erweiterbares Datenmodell
 - ▶ Einbindung neuer Link- und Informationstypen
- Source Code: <http://github.com/infolis>

Herausforderungen

- Forschungsdaten und -informationen lassen sich zunehmend im Web finden
 - ▶ Interdisziplinäre Forschungsdatenrepositorien wie Zenodo, etc.
- (Social) Web Daten gewinnen in den Sozialwissenschaften an Relevanz
 - ▶ z.B. Twitter
- Konvergenz von Forschungsdateninfrastrukturen, Web Standards und proprietären Web-Plattformen
 - ▶ FAIR-Prinzipien für Forschungsdaten
 - ▶ Open Science

Perspektive

- Verlinkung auf verschiedenen Ebenen (intern und extern)
 - ▶ Metadaten und Daten
 - ▶ Terme und Entitäten
 - ▶ Schemata und Vokabulare
- Anreichern der eigenen Metadaten
- Erhöhung der Datenauffindbarkeit
- Verbesserung der Interoperabilität



Nur mit Verlinkung ist es nicht getan

- Extraktion von Entitäten
 - ▶ z.B. Datenzitationen, Umfragevariablen, etc.
 - ▶ Einsatz von Named Entity Recognition und Classification Methoden
- Disambiguierung und Auflösung von Entitäten
 - ▶ z.B. Personennamen, Variablen, Datensätze, etc.
 - ▶ Einsatz von Entity / Coreference Resolution Methoden
- Evolution von Daten, Schemata, Vokabularen, Links

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

Vielen Dank für Ihre Aufmerksamkeit!

gesis

Leibniz-Institut
für Sozialwissenschaften

Mitglied der
Leibniz
Leibniz
Gemeinschaft