

# **Connecting GESIS research data and publication information systems**

**Katarina Boland**

Department Knowledge Technologies for the Social Sciences  
GESIS - Leibniz Institute for the Social Sciences, Cologne, Germany

**OpenAIRE Interoperability Workshop, Braga, Portugal**

**08.02.2013**

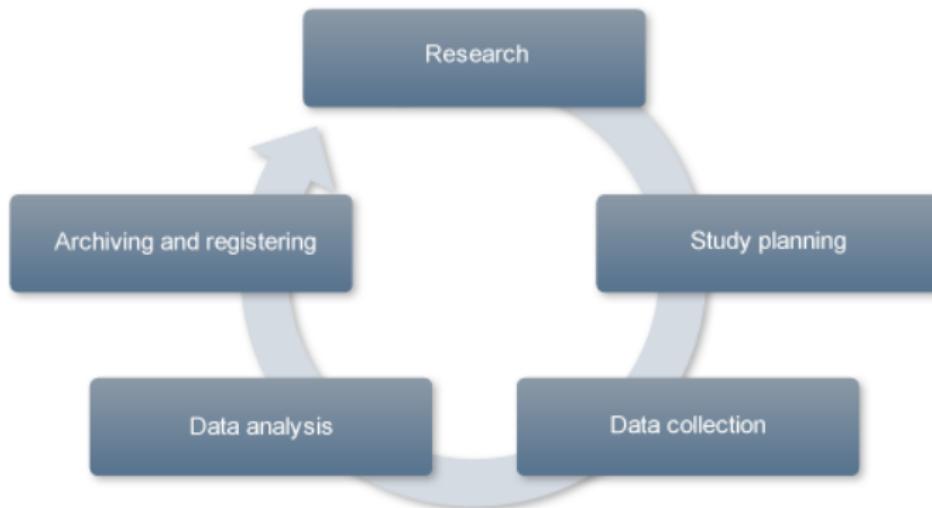
# Outline

- 1 Introduction to GESIS
- 2 GESIS information systems
- 3 Linking publications to datasets
- 4 Connecting research data and publication information systems

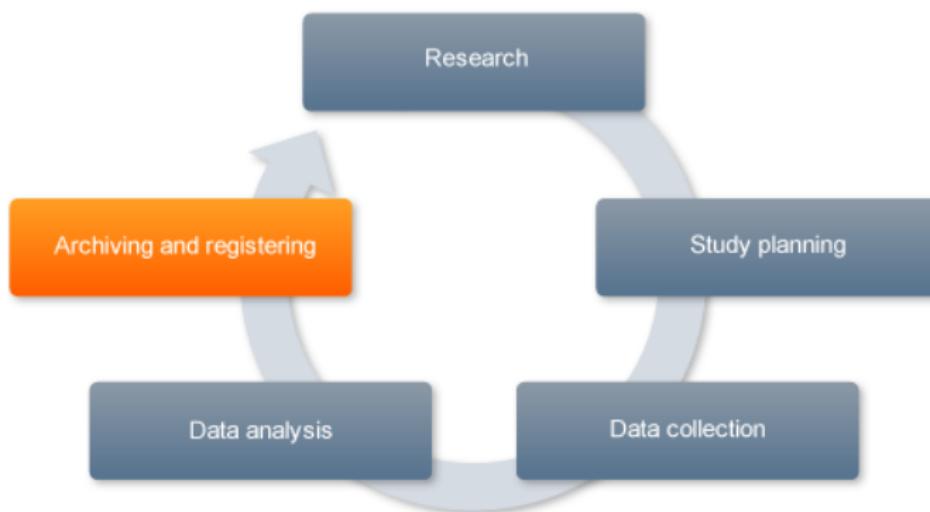
# GESIS

- largest infrastructure institution for the Social Sciences in Germany
- five scientific departments (Mannheim, Cologne, Berlin)

# GESIS



# GESIS



# GESIS



# Publications

- SSOAR - Social Science Open Access Repository
- SOWIPORT - the social science portal

# Publications

## ■ SSOAR - Social Science Open Access Repository

The screenshot shows the homepage of the Social Science Open Access Repository (SSOAR). At the top, there's a navigation bar with links for Home, Browse and Search, Advanced Search, and Add new document. Below this is a search bar with placeholder text "Search our database:" and a "Go" button. To the right of the search bar is a "Share" section with icons for DSpace, DFG, and others. A "Breaking News" section features a link to "SSOAR now with DSpace". At the bottom of the page, there's a logo for "DIN e-Journals" and some footer text about the repository's mission and funding.

electronic full texts (Social Sciences) for free access

mainly pursues the Green Way of Open Access

<http://www.ssoar.info/en.html>

## ■ SOWIPORT - the social science portal

# Publications

- SSOAR - Social Science Open Access Repository
- SOWIPORT - the social science portal

The screenshot shows the homepage of the SOWIPORT website. At the top, there is a navigation bar with links for Home, Search, Topics, Service, deutsch, Sitemap, and Contact. Below the navigation bar, there is a search bar with placeholder text "Searching for social science information? You want to have it all in one place; fast and easily available!". To the right of the search bar is a login form with fields for Username and Password, and a "Log in" button. Below the search bar, there is a "Search in SOWIPORT" section with a search input field and a checkbox for "only search for full texts". To the right of this is a "Breaking News" section with several news items. At the bottom of the page, there is an "Event calendar" showing various upcoming events and a "Call for Papers" section.

approximately 7 million references on publications and research projects from 18 databases

additional information on institutions and events

<http://www.gesis.org/sowiport/en/>

## Research Data

detailed information, documentation on variable-level and beyond

- ZACAT - Online Study Catalogue
- MISSY - Microdata Information System

documentation on study-level

- da|ra - Registration agency for social and economic data
- DBK - Data Catalogue

## Research Data

- da|ra - Registration agency for social and economic data
- DBK - Data Catalogue

# Research Data

## ■ da|ra - Registration agency for social and economic data

The screenshot shows the homepage of the da|ra registration agency. At the top, there's a navigation bar with links for 'Help | Contact | Login' and a 'Quick search in metadata' search bar. Below the navigation is a main content area with a 'Welcome to da|ra' message. To the left, there's a sidebar with logos for 'gesis' and 'ZBW' and a 'funded by' section featuring the 'DFG' logo. The main content area includes sections for 'Information for data centers', 'Information for researchers', and 'Information for publishers'. At the bottom, there's a 'Quick search in metadata' input field with a 'Go!' button, a 'Resolve DOI:' field containing '10.4232/1.10028', and a 'fairbank' logo.

German DOI registration  
service for social science and  
economic data

by GESIS and ZBW - German  
National Library of Economics,  
in cooperation with  
DataCite

<http://www.da-ra.de/en/home/>

## ■ DRK - Data Catalogue

Connecting GESIS research data and publication information systems

# Research Data

- da|ra - Registration agency for social and economic data
- DBK - Data Catalogue

## Data Catalogue - Simple Search

The Data Catalogue comprise the study descriptions from all archived studies and empirical primary data from survey research, historical social research and texts for content analyses. The documentation of data from official statistics will be stepwise added.

- Basic Search - Advanced Search
- To highlight search terms and to use multiple selection, please activate JavaScript
- You can browse the complete list of studies available at the Data Archive here: Study List
- You can browse a list of available products here: Product List
- Order & Download is now possible directly from within the data catalogue.

New in this version 1.9

Login

**Data Catalogue - Simple Search**

Please enter search term  
(automatic truncation):

Search term:

in:

Search:

Sort by:

Show:

GESIS Quicklinks

Choose page...

New Data 

Eurobarometer 76.4 15. January 2013 (2011) (ZA5596) available

Demographic Differences in Life Course Dynamics in Eastern and Western Europe (ZA5684) available

GESIS Online Panel 02. January 2013 Pilot: Family life (survey 3) (ZA5584) available

How innovative is Germany? (ZA5639) available

Views on European Unification (ZA5637) available

Terms of use

Charge regulations Usage regulations

study descriptions from survey research, historical social research and texts for content analyses

documentation of data from official statistics will be added successively

[http://www.gesis.org/en/services/research/  
data-catalogue/](http://www.gesis.org/en/services/research/data-catalogue/)

# Outline

- 1 Introduction to GESIS
- 2 GESIS information systems
- 3 Linking publications to datasets
- 4 Connecting research data and publication information systems

## Linking publications to datasets

the InFoLiS project:

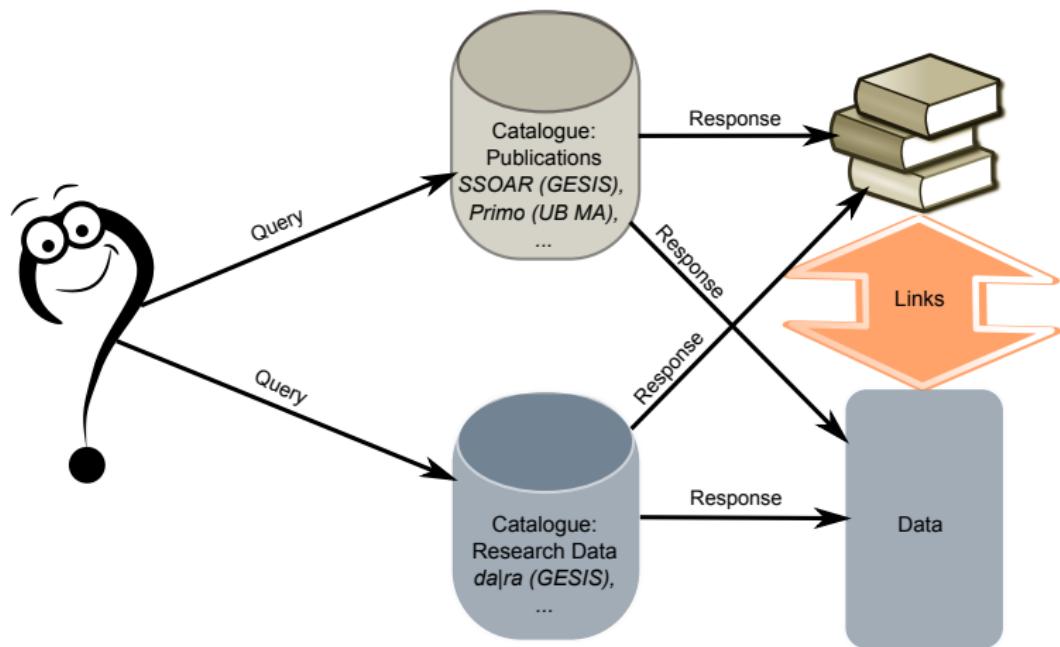
Integration of research data and publications for the Social Sciences



UNIVERSITY OF  
MANNHEIM

InFoLiS is funded by the DFG (SU 647/2-1)

# InFoLiS project goals



# References to datasets

erfolgt die Darstellung und Diskussion der empirischen Ergebnisse. Hierfür werden die Daten des Sozio-oekonomischen Panels (SOEP) aus den Jahren 1990 und 2003 verwendet und für beide Zeitpunkte werden die Einflussfaktoren mittels linearer Regressionsmodelle geschätzt.

presentation and discussion of the empirical findings. For this purpose, data from the Socio-Economic Panel (SOEP) of the years 1990 and 2003 are used and for both periods, the impact factors are estimated using linear regression models.

# References to datasets

erfolgt die Darstellung und Diskussion der empirischen Ergebnisse. Hierfür werden die Daten des Sozio-oekonomischen Panels (SOEP) aus den Jahren 1990 und 2003 verwendet und für beide Zeitpunkte werden die Einflussfaktoren mittels linearer Regressionsmodelle geschätzt.

data from the <title> of the years <year> are used

# References to datasets

*Tabelle 1: Bevölkerungsvorausberechnung für Deutschland nach Altersgruppen - Anteile in Prozent*

(Datenbasis: 10. Bevölkerungsvorausberechnung des Statistischen Bundesamtes, Variante 5)

Table 1: Population forecast for Germany depending on age cohorts - proportion in percent.

Data base: 10th Population Forecast of the Federal Statistical Office , version 5.

# References to datasets

*Tabelle 1: Bevölkerungsvorausberechnung für Deutschland nach Altersgruppen - Anteile in Prozent*

(Datenbasis: 10. Bevölkerungsvorausberechnung des Statistischen Bundesamtes, Variante 5)

(Data base: <number>. <title> of the <data collector>, version <version>)

## References to datasets

- 
- 1 Herangezogen wurden außerdem Allbus, Allensbacher Erhebungen, Eurobarometer, International Social Survey Program, International Social Justice Project, Sozio-ökonomisches Panel, World Values Survey.

Consulted were furthermore ...

# References to datasets

- 
- 1 Herangezogen wurden außerdem Allbus, Allensbacher Erhebungen, Eurobarometer, International Social Survey Program, International Social Justice Project, Sozio-ökonomisches Panel, World Values Survey.

Consulted were furthermore <title1>, <title2>, <title3>, ..., <titleN>.

# References to datasets

*Tabelle 3: Stichprobe der Untersuchung in den Jahren 2003 und 2004 sowie Größe der Stichprobe, mit gültigen Daten aus beiden Erhebungen*

(Quelle: [Ditton u.a. 2005a](#))

Table 3: Sample of the surveys conducted in the years 2003 and 2004 as well as size of the sample, with valid data from both surveys

(Source: [Ditton et al. 2005a](#))

## References to datasets

*Tabelle 3: Stichprobe der Untersuchung in den Jahren 2003 und 2004 sowie Größe der Stichprobe, mit gültigen Daten aus beiden Erhebungen*

(Quelle: [Ditton u.a. 2005a](#))

(Source: <citation of descriptive publication>)

## References to datasets

*Grafik 7: Einschätzung der wirtschaftlichen Lage: Einschätzung der eigenen wirtschaftlichen Lage (in Prozent)*

(Quellen: Allbus/Sozialstaatssurvey)

(Sources: Allbus/Sozialstaatssurvey )

## References to datasets

*Grafik 7: Einschätzung der wirtschaftlichen Lage: Einschätzung der eigenen wirtschaftlichen Lage (in Prozent)*

(Quellen: Allbus/Sozialstaatssurvey)

(Sources: <title1>/<title2>)

# Linking publications to datasets

References to datasets are not standardized!

see also...

- Green, Toby (2009). *We Need Publishing Standards for Datasets and Data Tables*. OECD Publishing White Paper.  
doi: 10.1787/603233448430
- Altman, Micah and Gary King (2007). *A Proposed Standard for the Scholarly Citation of Quantitative Data*. In: D-Lib Magazine 13.3.  
url: <http://www.dlib.org/dlib/march07/altman/03altman.html>

# Automatic identification of references

Why not simply search for study titles in publications?

- Studies are referenced using abbreviations, alternative names or literature
- Study titles may be common nouns - ambiguous!
- there is no complete list of all conducted studies

## General idea

How do humans recognize study references?

*Source: Estimations based on SOEP, wave 2002.*

## General idea

How do humans recognize study references?

*Source: Estimations based on xyz, wave 2002.*

## General idea

How do humans recognize study references?

*Source: Estimations based on xyz, wave 2002.*

→ Learn patterns: typical contexts for study references

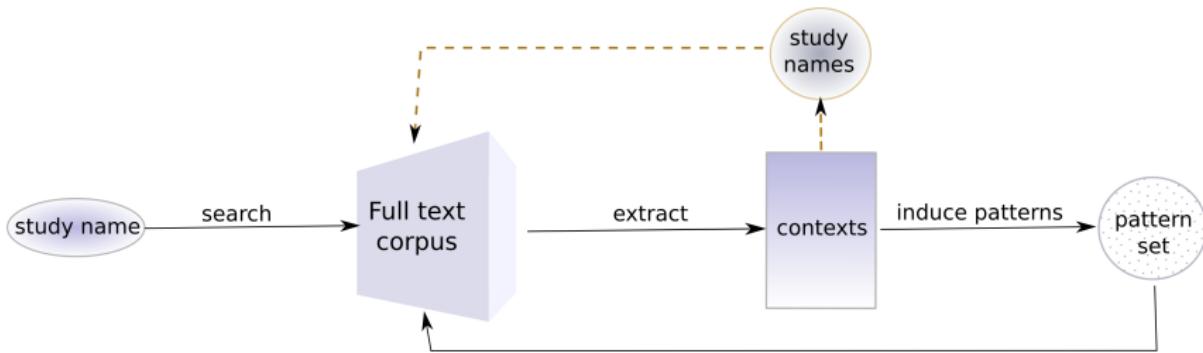
## General idea

How do humans recognize study references?

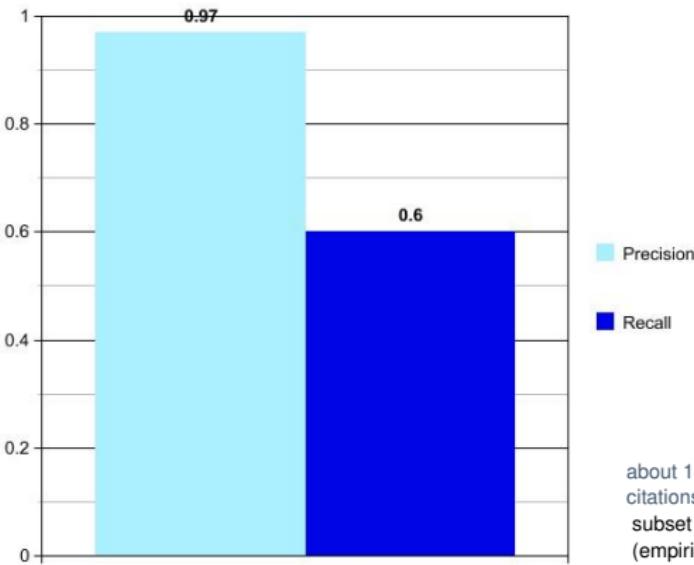
*Source: Estimations based on xyz, wave 2002.*

- Learn patterns: typical contexts for study references
- Sparse Data Problem: use iterative bootstrapping approach

# Algorithm



# Evaluation: Precision & Estimate of Recall



about 14% of the found references are not study names, but citations of publications → not counted as incorrect here  
subset of SSOAR with keyword "empirisch-quantitativ"  
(empirical quantitative)  
German, n = 259  
conversion pdf → txt with automatic correction

## Reference extraction

for details see...

Boland, Katarina, Ritze, Dominique, Eckert, Kai, & Mathiak, Brigitte (2012).  
*Identifying References to Datasets in Publications*. International Conference on Theory and Practice of Digital Libraries (TPDL) (pp. 150-161). Paphos, Cyprus: Springer Berlin Heidelberg. doi:10.1007/978-3-642-33290-6\_17

# Matching to da|ra records

**da|ra** Registration agency for social and economic data

Help | Contact | Login  
Quick search in metadata

My da|ra For data centers For researchers For publishers About us News

Welcome to da|ra

In cooperation with DataCite, the international initiative to establish easier access to digital research data, GESIS Leibniz-Institute for Social Sciences and ZBW - Leibniz Information Centre for Economics offer the DOI registration service in Germany for social science and economic data.

run by:

gesis Leibniz Institute for the Social Sciences  
ZBW Leibniz Information Centre for Economics

Leibniz Association

funded by:

DFG

Quick search in metadata:

ALLBUS 2000

Go!

Advanced search

Resolve DOI: 10.4232/1.10028

Go!

News:

05.02.2013 Podcast about DataCite at D-Radio Wissen more >

29.01.2013 Workshop "More people – more data – more repositories" more >

15.01.2013 Event calendar 2013 more >

our users:

OECD Organisation for Economic Co-operation and Development

based on:

D

# Matching to da|ra records

Availability:

[Download](#)

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

gesis

DOI: [c 10.4232/1.3452](#)

Version: 1.0.0

Principal Investigator:

- Allmendinger, Jutta (Universität Münch ... [see more](#)

Temporal Coverage:

- 2000-01 - 2000-07

Publication Date:

2003

Availability:

[Download](#)

ALLBUS/GGSS 2000 CAPI (Allgemeine Bevölkerungsumfrage der Sozialwissenschaften/German General Social Survey 2000 CAPI)

gesis

DOI: [c 10.4232/1.3451](#)

Version: 1.0.0

Principal Investigator:

- Allmendinger, Jutta (Universität Münch ... [see more](#)

Temporal Coverage:

- 2000-01 - 2000-07

Publication Date:

2003

Availability:

[Download](#)

ALLBUS/GGSS 2000 CAPI-PAPI (Allgemeine Bevölkerungsumfrage der Sozialwissenschaften/German General Social Survey 2000 CAPI-PAPI)

gesis

DOI: [c 10.4232/1.11365](#)

Version: 2.0.0

# Matching to da|ra records

- Precise matching to DOI not always possible!
- Instead: Matchings to **relevant** sources

The screenshot shows a search results page for datasets from the ALLBUS/GGSS 2000 survey. There are two main entries, each with detailed metadata and download links.

**Entry 1: ALLBUS/GGSS 2000 PAPI (Allgemeine Bevölkerungsumfrage der Sozialwissenschaften/German General Social Survey 2000 PAPI)**

- DOI: [10.4232/1.3452](#)
- Version: 1.0.0
- Principal Investigator: Allmendinger, Jutta (Universität München ... [see more](#))
- Temporal Coverage: 2000-01 - 2000-07
- Publication Date: 2003
- Availability: [Download](#)

**Entry 2: ALLBUS/GGSS 2000 CAPI (Allgemeine Bevölkerungsumfrage der Sozialwissenschaften/German General Social Survey 2000 CAPI)**

- DOI: [10.4232/1.3451](#)
- Version: 1.0.0
- Principal Investigator: Allmendinger, Jutta (Universität München ... [see more](#))
- Temporal Coverage: 2000-01 - 2000-07
- Publication Date: 2003
- Availability: [Download](#)

At the bottom of the page, there is a footer with the text "Allgemeine Bevölkerungsumfrage der Sozialwissenschaften" and the GESIS logo.

# Matching to da|ra records

- Precise matching to DOI not always possible!
- Instead: Matchings to **relevant** sources
- Definition of relevance depends on application

The screenshot displays two entries from the GESIS data archive, each showing a summary of the dataset and download links.

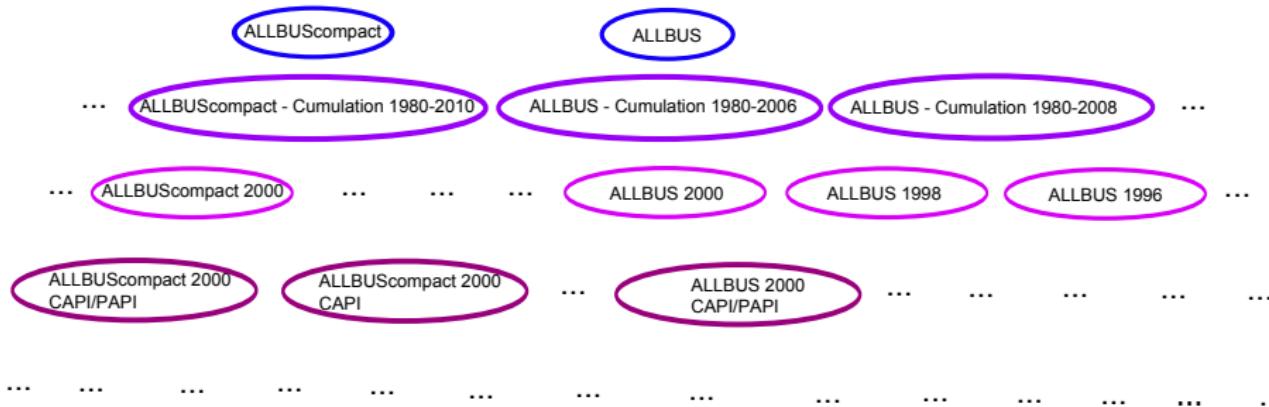
**ALLBUS/GGSS 2000 PAPI (Allgemeine Bevölkerungsumfrage der Sozialwissenschaften/German General Social Survey 2000 PAPI)**

- DOI: [10.4232/1.3452](#)
- Version: 1.0.0
- Principal Investigator: Allmendinger, Jutta (Universität München ... [see more](#))
- Temporal Coverage: 2000-01 - 2000-07
- Publication Date: 2003
- Availability: [Download](#)

**ALLBUS/GGSS 2000 CAPI (Allgemeine Bevölkerungsumfrage der Sozialwissenschaften/German General Social Survey 2000 CAPI)**

- DOI: [10.4232/1.3451](#)
- Version: 1.0.0
- Principal Investigator: Allmendinger, Jutta (Universität München ... [see more](#))
- Temporal Coverage: 2000-01 - 2000-07
- Publication Date: 2003
- Availability: [Download](#)

# Matching to da|ra records



→ semantic web technologies

# Links

**Data Catalogue Home** Search Advanced Search •

**ZA3296: Eurobarometer 53 (2000)**

Bibliographic Citation | Content | Methodology | Data & Documents | Errata & Versions | Further Remarks | Groups

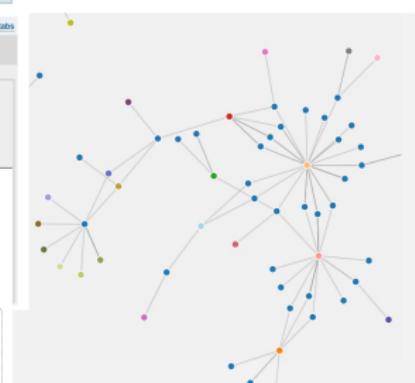
You have requested the DOI for the current version!

Title No. ZA3296

Title Eurobarometer 53 (2000)

Other Titles ▾ Racism, Information Society, General Services, and Food Labeling (Subtitle)

Current Version 1.0.1, 2012-3-30, doi:10.4232/1.10935



## Node details

[urn:nbn:de:0168-soar-281457](#)

## connections:

1. [Eurobarometer 2000](#): 10.4232/1.10935  
([Eurobarometer 2000](#))

**SSOAR**  
Social Science Open Access Repository

Einloggen Registrieren

Home Browse und Suchen Erweiterte Suche Neues Dokument hinzufügen

**Zitadelle "virtueller Nationalstaat": die Europäische Union und die Politik interner Schließung europäischer Einwanderungsänder**

Citadel "virtual nation-state": the European Union and the policy of internal closure of European migration courses [Zeitschriftenartikel]

Birsi, Ursula

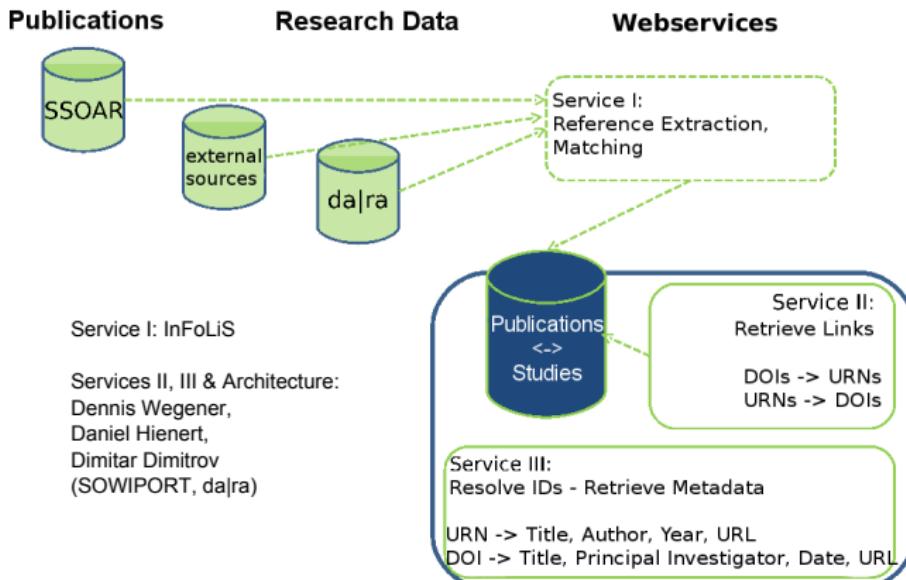
Volltext herunterladen (900 kB/byte)

Wein von Best: Urheber  
Wein von Österreichische Zeitschrift für Politikwissenschaft  
Export für Ihre Literaturverwaltung (Bibliographie per Copy & Paste)  
Basis-Export  
Endnote-Export

## Links (found references) for [urn:nbn:de:0168-soar-281457](#)

1. [Eurobarometer 2000](#)
  - Is this study indeed referenced in the publication?  yes  study not mentioned at all  study mentioned with diverging number/year/version  don't know
  - study mentioned but data is not used
  - Boundary of study name correctly identified?  yes  no  don't know

# Connecting information systems



## Connecting information systems

Demo: da|ra test system with SOWIPORT links

## Connecting information systems

Demo: da|ra test system with SOWIPORT links (offline version)

# Conclusion: our aim

- interlink our own repositories and information systems
- provide services for reference extraction and matching to all interested institutions (free access to webservices)
  - domain- and language-independent
- link to publications and data stored in external repositories

Thank you for your  
attention!

[katarina.boland@gesis.org](mailto:katarina.boland@gesis.org)