Linking Research Data and Literature

Katarina Boland
Knowledge Technologies for the Social Sciences
GESIS - Leibniz Institute for the Social Sciences, Cologne

DASISH Workshop on Persistent Identifier Services

December 9, 2014

Outline

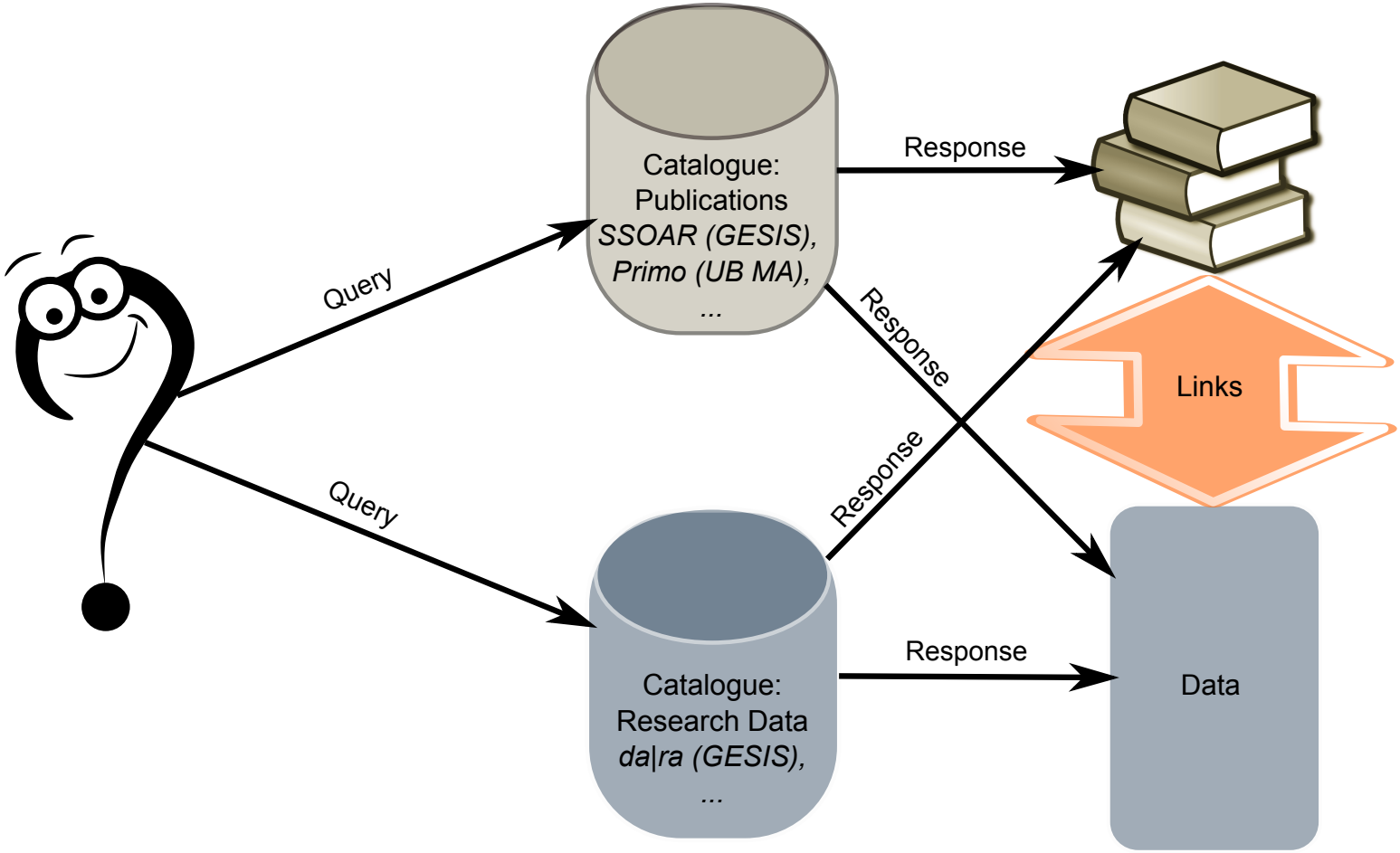
Introduction

the InFoLiS project:
Integration of research data and publications for the social sciences

gesis Leibniz Institute for the Social Sciences
UB MANNHEIM
UNIVERSITY OF MANNHEIM

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

InFoLiS Project Goals



Citation of Research Data

Recommendation:[1]:

Creator (Publication Date): Title. Publication Agent. Identifier

Creator (Publication Date): Title. Version. Publication Agent. Type of Resource. Identifier.

→ Extraction based on these patterns?

[1] see http://auffinden-zitieren-dokumentieren.de/zitieren/empfohlene-datenzitation/

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.

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 , variant 5.

(Data base: `number` `title` of the `publication_agent`, variant `variant`)

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 ...

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)

(Source: `citation of descriptive publication`)

References to Datasets

...are hard to detect!

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!

Automatic Identification of References

Why not simply search for study titles in publications?

- Studies are referenced using abbreviations, alternative names or literature
  "ALLBUS/GGSS 1996 (Allgemeine Bevölkerungsumfrage der Sozialwissenschaften/German General Social Survey 1996)"
- Study titles may be common nouns - ambiguous!

Automatic Identification of References

Why not simply search for study titles in publications?

- Studies are referenced using abbreviations, alternative names or literature
  ~~*"ALLBUS/GGSS 1996 (Allgemeine Bevölkerungsumfrage der Sozialwissenschaften/German General Social Survey 1996)"*~~

  "ALLBUS 96"
- Study titles may be common nouns - ambiguous!

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

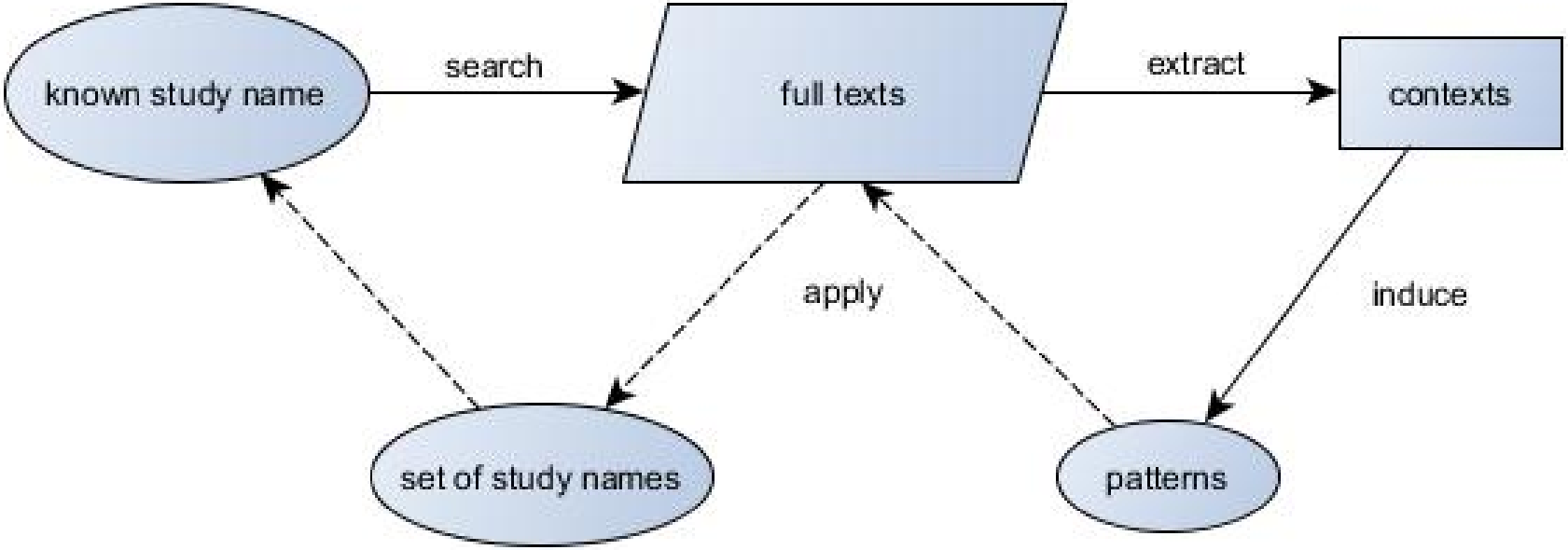
How do humans recognize study references?

Source: Estimations based on xyz, wave 2002.

→ Learn patterns: typical contexts for study references

→ Sparse Data Problem: use iterative bootstrapping approach

Algorithm


known study name
search
full texts
extract
contexts
apply
induce
set of study names
patterns

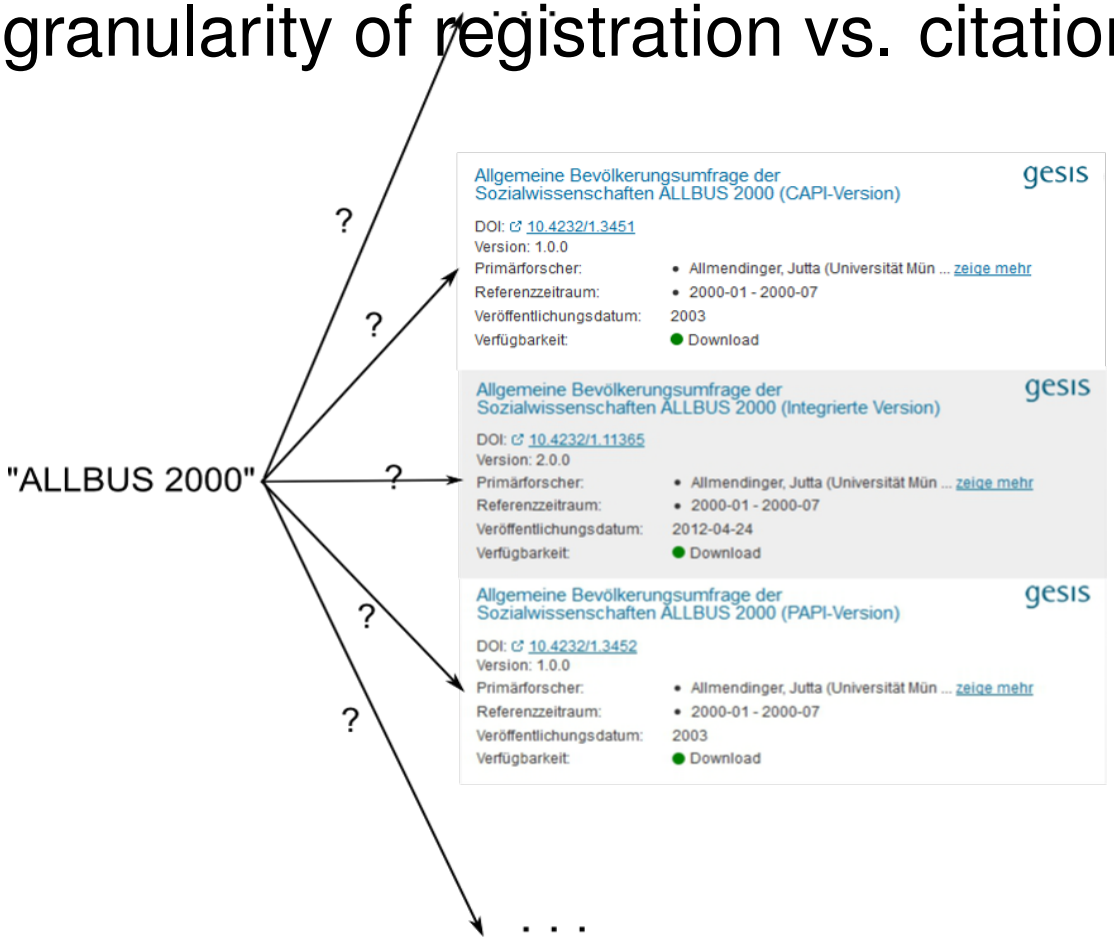
Reference Extraction

for details see...

Katarina Boland, Dominique Ritze, Kai Eckert & Brigitte Mathiak (2012). *Identifying References to Datasets in Publications*. In: Proceedings of the Second International Conference on Theory and Practice of Digital Libraries (TPDL), Lecture Notes in Computer Science Volume 7489, pp. 150-161. Berlin: Springer. doi:10.1007/978-3-642-33290-6_17

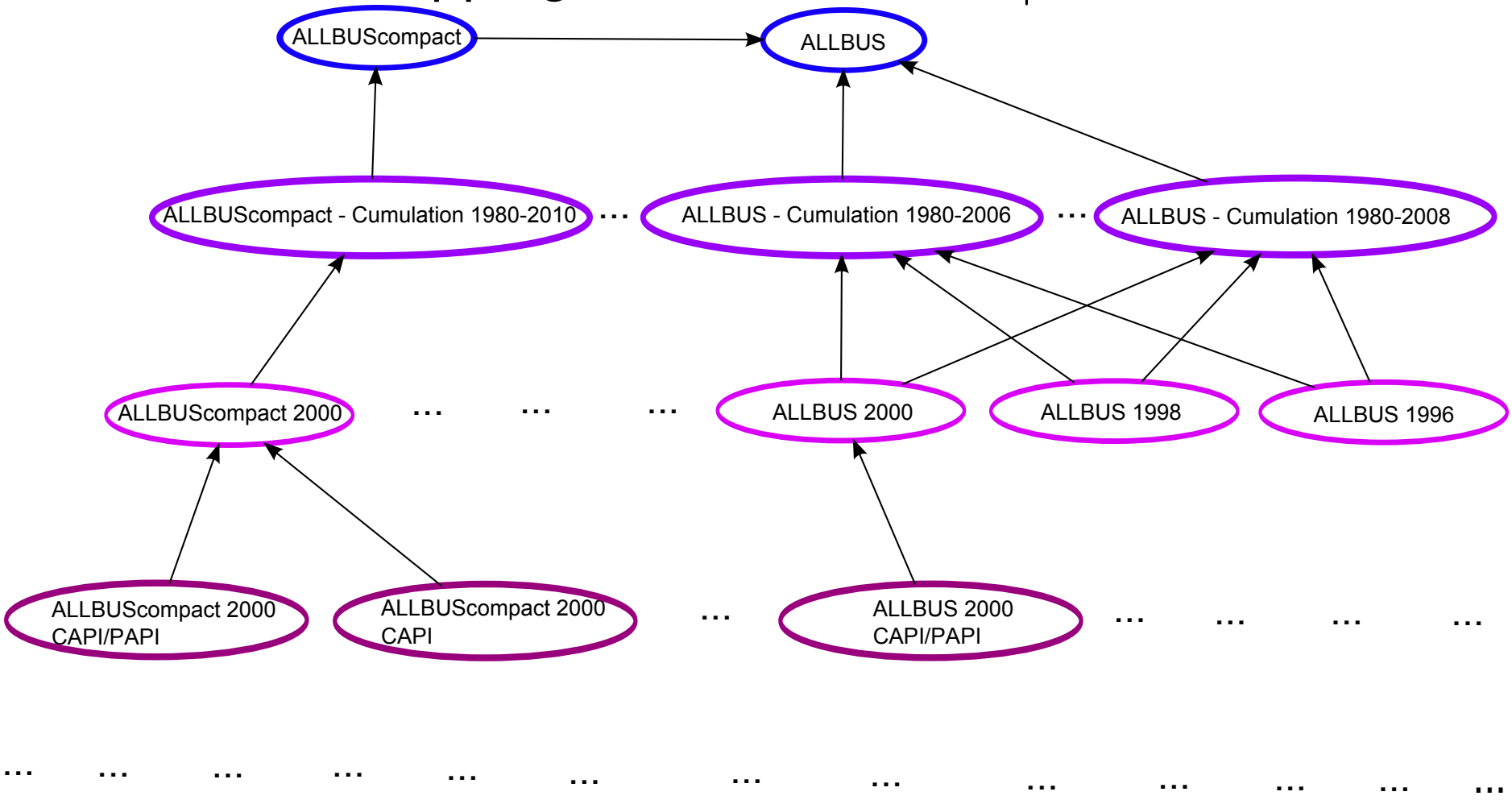
Mapping to Datasets in da|ra

Mapping to Datasets in da|ra: granularity of registration vs. citation



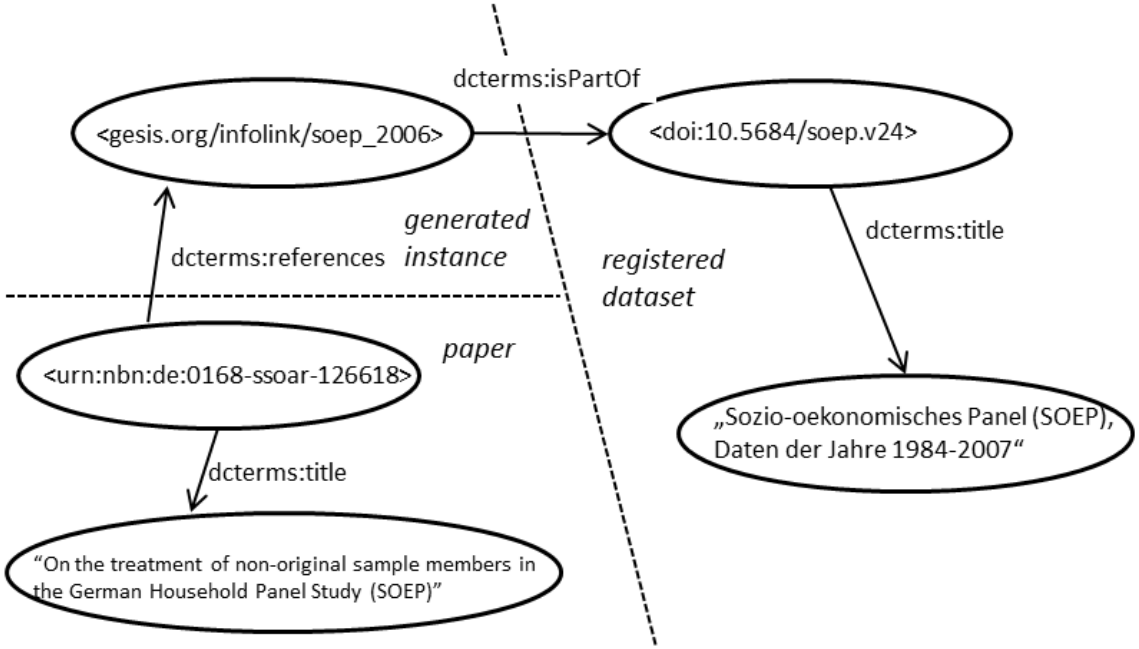
Strategies: 1) greedy; 2) exact; 3) best

Mapping to Datasets in da|ra



→ use ontology

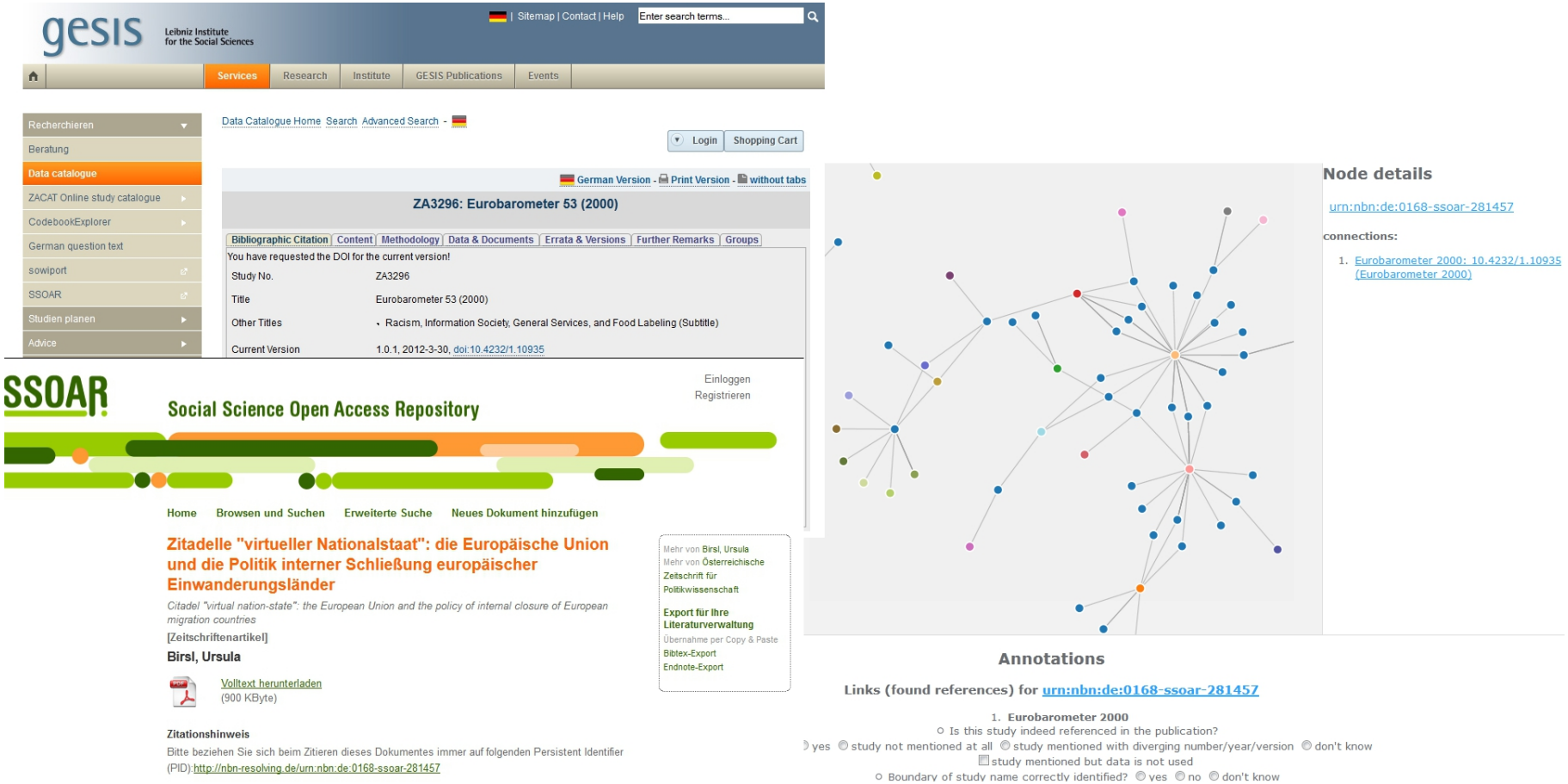
Ontology: Approach



Vocabulary: e.g. DDI-RDF Discovery Vocabulary[2]

[2] Thomas Bosch, Richard Cyganiak, Arofan Gregory, Joachim Wackerow (2013): *DDI-RDF Discovery Vocabulary: A Metadata Vocabulary for Documenting Research and Survey Data.* In: Proceedings of the 6th Linked Data on the Web (LDOW) Workshop at the 22nd International World Wide Web Conference (WWW). CEUR Workshop Proceedings, pp. 46-55

Links


gesis
Leibniz Institute for the Social Sciences
| Sitemap | Contact | Help
Enter search terms...
Services
Research
Institute
GESIS Publications
Events
Recherchieren
Beratung
Data catalogue
ZACAT Online study catalogue
CodebookExplorer
German question text
sowiport
SSOAR
Studien planen
Advice
Data Catalogue Home Search Advanced Search
Login
Shopping Cart
German Version - Print Version - without tabs
ZA3296: Eurobarometer 53 (2000)
Bibliographic Citation | Content | Methodology | Data & Documents | Errata & Versions | Further Remarks | Groups
You have requested the DOI for the current version!
Study 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
SSOAR
Social Science Open Access Repository
Einloggen
Registrieren
Home
Browsen und Suchen
Erweiterte Suche
Neues Dokument hinzufügen
Zitadelle "virtueller Nationalstaat": die Europäische Union und die Politik interner Schließung europäischer Einwanderungsländer
Citadel "virtual nation-state": the European Union and the policy of internal closure of European migration countries
[Zeitschriftenartikel]
Birsl, Ursula
Volltext herunterladen
(900 KByte)
Zitationshinweis
Bitte beziehen Sie sich beim Zitieren dieses Dokumentes immer auf folgenden Persistent Identifier (PID): http://nbn-resolving.de/urn:nbn:de:0168-ssoar-281457
Mehr von Birsl, Ursula
Mehr von Österreichische Zeitschrift für Politikwissenschaft
Export für Ihre Literaturverwaltung
Übernahme per Copy & Paste
Bibtex-Export
Endnote-Export
Node details
urn:nbn:de:0168-ssoar-281457
connections:
1. Eurobarometer 2000: 10.4232/1.10935 (Eurobarometer 2000)
Annotations
Links (found references) for urn:nbn:de:0168-ssoar-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

Integration of Links into Information Systems

Example: da|ra

Integration of Links into Information Systems

UNIVERSITÄT MANNHEIM

Universitätsbibliothek Mannheim
Primo

Univ. Library Holdings | Articles & Univ. Library Holdings | Inter-Library Loan | Research Data

Websearch: lebensstile in der familie — Search | Advanced search | Browse Search

anywhere in the record

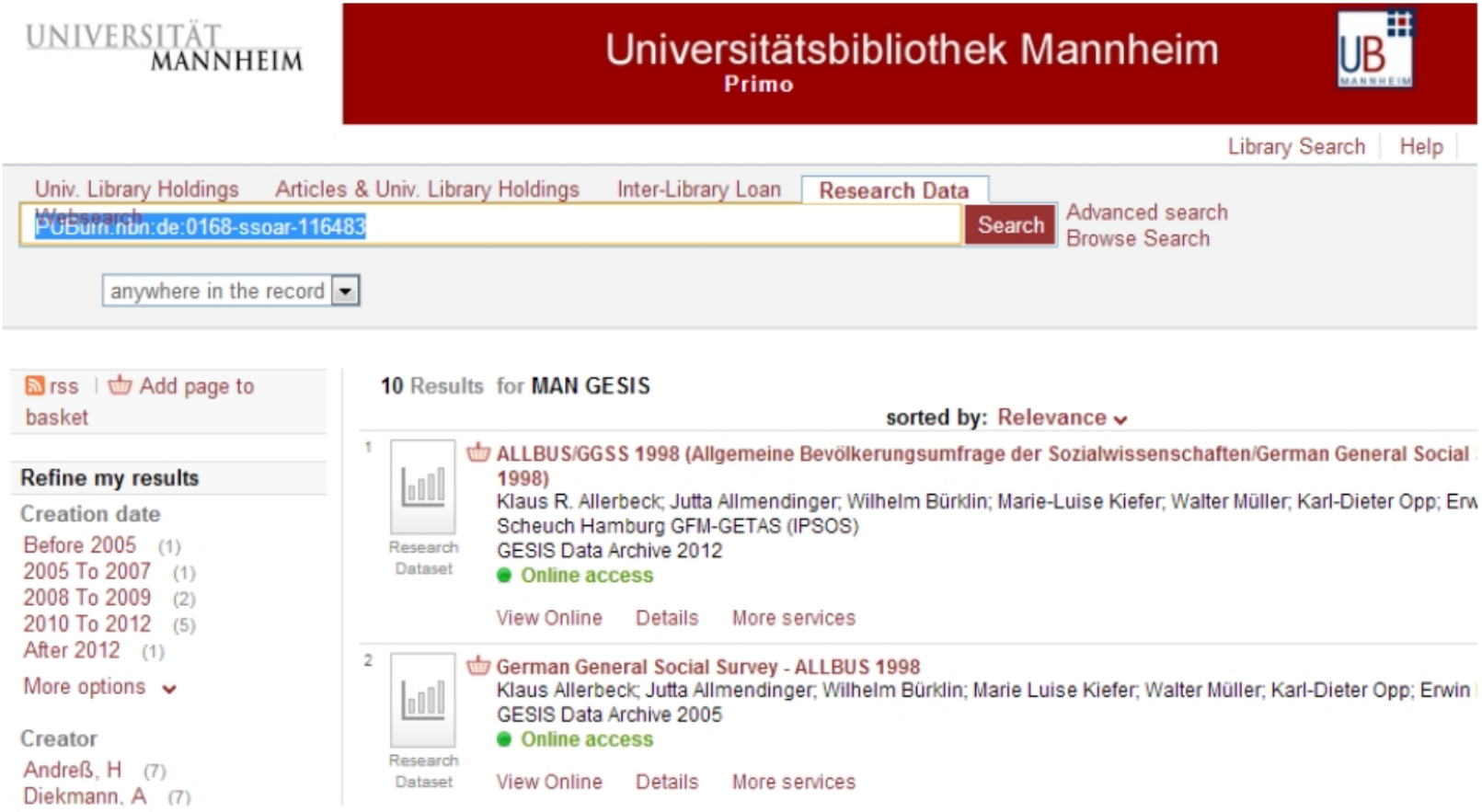
Lebensstile in der Familie; Life styles in the family
Andreas Klocke; Detlev Lück Staatsinstitut für Familienforschung an der Universität Bamberg (ifb)
Nr. 3-01
Deutschland;Germany;Bamberg 2010-07-13T16:30:00Z
● Online access

View Online | Details | More services

Title:	Lebensstile in der Familie; Life styles in the family
Author:	Andreas Klocke ; Detlev Lück Staatsinstitut für Familienforschung an der Universität Bamberg (ifb)
Is Part Of:	Nr. 3-01
Description:	"Lebensstile bezeichnen persönliche Arrangements, die die Bereiche Arbeit, Familie, Freizeit, Kultur und Lebensorientierung umspannen. Sie sind damit u.a. in den Kontext der
Publisher:	Deutschland;Germany;Bamberg
Creation Date:	2010-07-13T16:30:00Z
Identifier:	http://www.ssoar.info/ssoar/handle/document/11648; http://www.ifb.bayern.de/imperia/md/content/stmas/ifb/materialien/mat_2001_3.pdf; urn:nbn:de:0168-ssoar-116483
Subjects:	Social sciences, sociology, anthropology ; Sozialwissenschaften, Soziologie ; Familie ; Bundesrepublik Deutschland ; Lebensstil ; Typologie ; Ehepartner ; Eltern ; Kind ; Geschwister ; of Sexual Behavior ; ; theory application ;

Links
- Volltext
- Volltext
- Show related research data

Integration of Links into Information Systems



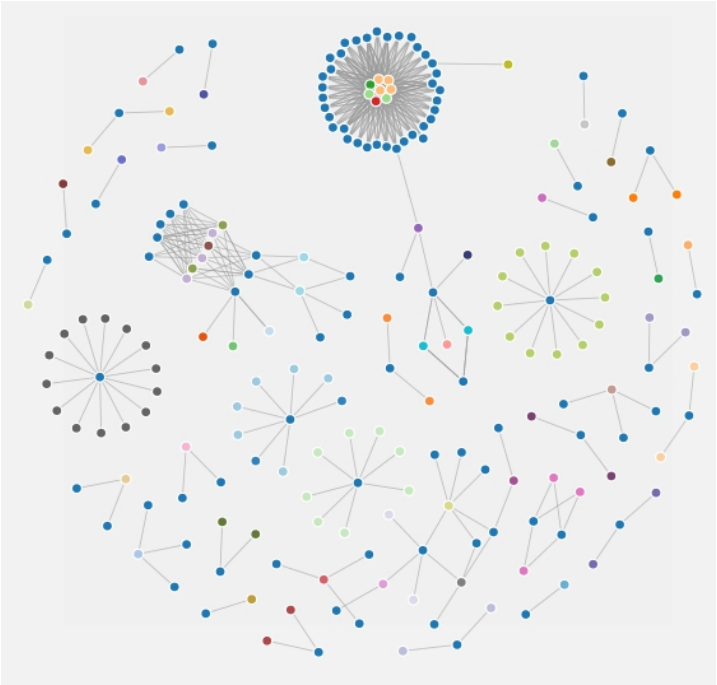
InFoLiS Follow-Up Project: InFoLiS II

1. enhance approach and extend to other domains and languages
2. tackle the granularity problem - ontology, linking and enhanced information retrieval
3. build re-usable, open source infrastructure

own research:

- classify data citations w.r.t. citation purpose

Thank you for your attention!



katarina.boland@gesis.org