Searching beyond datasets in the Social Sciences

Philipp Mayr

International Workshop on Data Search

SIGIR 2018

July 12th, 2018

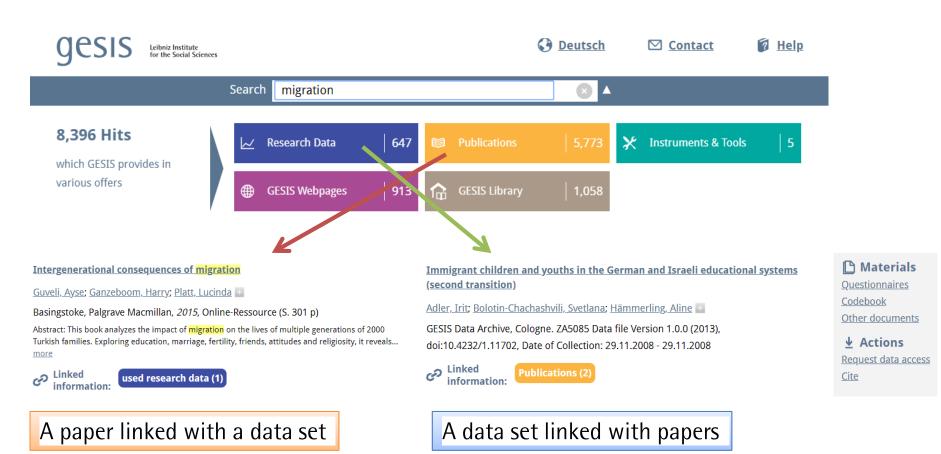https://datasearch-ws.github.io/2018/

State of the art @GESIS

- Curated datasets (surveys, longitudinal data, …)
  - Metadata/documentation, keywords
  - Additional materials (questionnaires, codebooks, …)
  - Raw data
- Harvested datasets
- **Linking infrastructure** (links between items)
- ElasticSearch for searching

Integrated Searching

A paper linked with a data set

A data set linked with papers

A social science dataset

🌐 **German**   ✉ **Contact**

Search | migration | ▼

Services ▼   Research ▼   Institute ▼

< Back

https://search.gesis.org/research_data/ZA5957

2000 Families: Migration Histories of Turks in Europe

Guveli, Ayse; Ganzeboom, Harry B.G.; Baykara-Krumme, Helen

GESIS Data Archive, Cologne. ZA5957 Data file Version 2.0.0 (2016), doi:10.4232/1.12673

Abstract: "The 2000 Families: Migration Histories of Turks in Europe project explores migration processes, the multi-generational transmission of social, cultural, religious and economic resources, values and behavior. The research is targeted Turkish migrant and non-migrant families, their members in European countries and those who did not migrate to European countries or returned to Turkey, and involves survey interviews with approximately 6000 family members across three generations. The study consists of three parts: Family Tree (Pilot and Main), Proxy interviews (Pilot and Main) and Personal..." more

Study number: ZA5957

Current Version: 2.0.0, 2016-11-11, doi:10.4232/1.12673

DOI: doi:10.4232/1.12673

Date of Collection: 2010 - 2012

Content: Belgium, Denmark, Germany, France, Netherlands, Norway, Austria, Sweden,

Number of Units: 48980

Number of Variables: 499

📄 **Materials**
Datasets
Codebook

⬇ **Actions**
Cite

Download Datensätze   ✖

The **download** of datasets generally requires a **login** to the data catalog (DBK).

ZA5957_v2-0-0.sav.zip SPSS (Datensatz) 6.25 MB
ZA5957_v2-0-0.dta.zip Stata 14 (Datensatz) 2.9 MB
ZA5957_v2-0-0_stata12.dta.zip Stata 12 (Datensatz) 2.9 MB

Availability: A - Data and documents are released for academic research and teaching. te our terms of use.

🔗 **Linked information:**   **Publications (1)**

Beyond

- Linking studies with papers and visa versa is helping the user
- But social scientists often want to reuse certain parts of a dataset (e.g. are interested in the data of a set of questions , e.g. „belief")
- There is a need for **more fine-grained retrievability and linkages** of datasets

Step 1: Extraction of questions and variables

Question: Q.30B - Eurobarometer 59.2 (May-Jun 2003)

Variables: v330

Study title: Eurobarometer 59.2 (May-Jun 2003)

Question text: What was the approximate total value of what you bought or ordered?

Fragenlabel: Q.30B

Categories: Migration, Information society, Mass media, International politics and organisation, Minorities, Transport, travel and mobility, Rural economics, Consumption / consumer behaviour

Variables: v330

Item categories:

Name	Code
On the Internet	Q.30B_3

Answer categories:

Name	Code
Less than 50 Euro	1
From 50 to 99 Euro	2
From 100 to 199 Euro	3
From 200 Euro or more	4
DK	5
Inap. (not coded 1 in V321; not coded 1 in V324)	9

And make questions etc. searchable! Citable!

Step 2: Linking of questions to papers and vice versa

- E.g. a researcher is planing a study of „migrants and their educational backgrounds" and wants to reuse certain questions of an existing study

The Relations between Education and Migration in Ukraine

Vakhitova, Ganna; Coupe, Tom

2013

Linked information: used research data (1)

Challenge: Finding the questions which have been used in a certain study

Metastudy of the utilization of a set of questions across papers

Step 3: Providing more advanced search facilities

- An advanced search interface is helping to find right datasets

Discussion

- There might be dozens of such specific use cases in DATA SEARCH

- A Google for data can help finding dataset but more advanced applications are requested

Thank you

Contact:

Dr Philipp Mayr

GESIS – Leibniz Institute for the Social Sciences, Germany

Email: philipp.mayr@gesis.org

Twitter: @philipp_mayr