gesis
Leibniz Institute
for the Social Sciences

## Providing high quality data and metadata through the ExploreData search portal

Julia Hermann, Christina Eder, Wolfgang Zenk-Möltgen
May 30th 2018

Outline

- About GESIS – Leibniz Institute for the Social Sciences
  - GESIS services
  - Types of data and studies in GESIS Archive
- About Explore Data
  - Project facts
  - Goals of the project
  - Conceptual development
  - Explore Data Search Portal
    - Access to studies and variables
    - Search results
    - Technical features
  - Next steps
  - Technical background

About GESIS – Leibniz Institute for Social Sciences

GESIS services

Research (diagram)
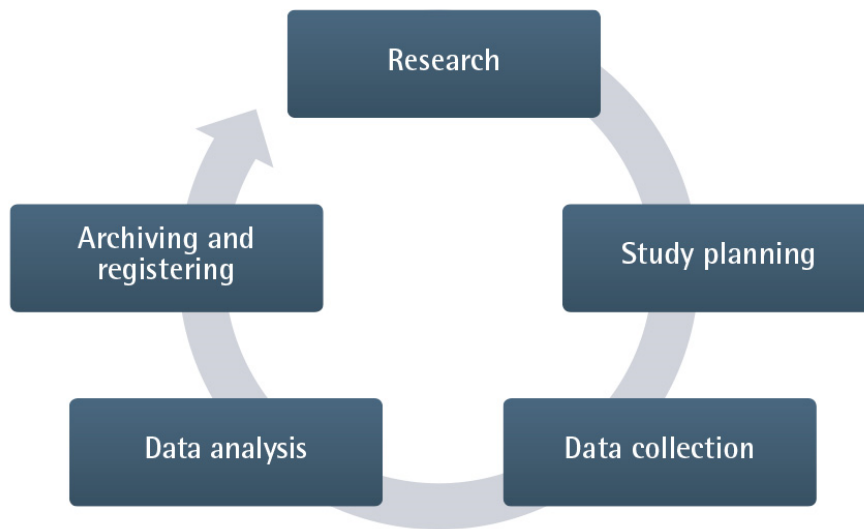
Research

- Search portals for studies (data catalogue DBK)
- Search portal for variables (ZACAT, VariableOverview, ContinuityGuide)
- Search portal for literature (SSOAR)

Study planing and data collection

- Consulting
- Survey Methodology
- Measure Instruments (ZIS, cognitive Pretests)
- Sampling
- Survey Guidelines
- Opportunity to collect survey data (GESIS Panel)

Data analysis: Data for re-use

- National and international survey programms
- Single studies
- Research data centers: FDZ ALLBUS, FDZ German Microdata Lab, FDZ Internationale Umfrageprogramme, FDZ PIAAC, FDZ Wahlen

Archiving and registering

- Long term preservation
- Register data through da|ra

www.gesis.org

4

Types of data and studies in GESIS Archive

Source	Type of document	Number	Downloads 2017
Data-Archive	Studies	6299	
	Study related documents (questionnaires, codebooks, variable reports, method reports, etc...)	28931	869273* (data&docs)
	data sets (spss-, stata-, csv-files)	10587	

Source: ZA6899

Types of data

- Cross-sectional data
- Time series data
- Historical data
- Experimental data

Types of studies

- **National studies**
  - ▶ Archived through GESIS
  - ▶ GESIS projects
  - ⇒ Access via DBK data catalogue
- **National and international study collections**
  - ▶ ALLBUS, Politbarometer, Election Studies (GLES, National Election Studies, Federal State Election Studies), GESIS Panel etc.
  - ▶ ISSP, EVS, Eurobarometer, CSES
  - ⇒ Access via DBK data catalogue
  - ⇒ In Addition: **Value-added products**
    - Continuity Guide (via CD)
    - VariableOverview (Online-Tool)
    - Lists (PDF-, Excel-files)
    - ZACAT (Online-Tool for Variables)

About Explore Data

Project facts

- GESIS internal project
- Project duration
  - April 2016 – August 2019
- Project team
  - interdisciplinairy work between social scientists and computer scientists
  - Associated project partners: GESIS-Teams of large study collections (ALLBUS, Eurobarometer, EVS, GLES, German Federal Election Studies, ISSP, Politbarometer)
- Related projects
  - CESSDA-project „Euro Question Bank"
  - Several GESIS internal projects

Goals of the project

- The existing value-added products of large study collections will be replaced and combined into a unified and state-of-the-art search portal
  - large study collections are easy to manage
  - more user-friendly data usage
  - systematized functions
  - completely webbased

Conceptual development

- Development of persona, user stories and use cases based on study collections' request
- A systematic comparison of similarities and differences in the documentation and the presentation of study information between study collections in order to develop a uniform platform on which the data of all study collections can be presented and distributed in a uniform layout
- Comparing inventory of existing search portals (including ESS, ICPSR, Midus, SOEP)
- GESIS user survey: N=1847, Population= Academic staff, Topics: Knowledge and use of the online GESIS offers, Survey=February-May 2017
- Development of a detailed concept for the research portals functionalities in close cooperation with the representatives of large study collections – based on preexisting work
- Discussion on usability aspects and final design for the search portal with web development experts
- After the conceptual development of the search portal
  - Final consultation with study collections was held and feedback was obtained
  - User tests based of conceptual Mock-ups were done

Explore Data Search Portal

Access to studies and variables

Access to studies and variables

Access to studies and variables

Access to studies and variables

Access to studies and variables

Access to studies and variables

- A unified search portal to access studies and their related variables and questions
- Access to studies and variables through keywords-search
  - ▶ Free text search within the search slot (the system scans data files, documents, and metadata)
    - Integrated search: search in documents (codebooks, basic questionnaires, method reports), meta data, survey data
    - Supports bilingual content: Search in English and German is possible
- Access to studies and variables through filters-browse
  - ▶ Filtering the search results according to specific characteristics (CESSDA TC, study collections, years, geography, language, methodological aspects (sample procedure, mode of collection))
    - Multiple selection is possible
    - Dynamic filters
  - ▶ Possibility to browse to get an overview of all specific characteristics

Access to studies and variables

Search results

ZA-0035 „Political Interest" 1964

Study description	View data	Download data & documents

Download:
🛒 **All Files** (8,3 MB)

Dataset:
🛒 **ZA0035_de_v1-0.dta**   🛒 **ZA0035_en_v1-0.dta**
🛒 **ZA0035_de_v1-0.sav**   🛒 **ZA0035_en_v1-0.sav**

Documentation:
🛒 **Questionnaire.pdf**   🛒 **Studydescription.pdf**   🛒 **Codebook.pdf**   🛒 **Studynotes.pdf**
🛒 **Questionnaire.txt**   🛒 **Studydescription.txt**   🛒 **Codebook.txt**   🛒 **Studynotes.txt**

Search results

ZA-0035 „Political Interest" 1964

Study description | **View data** | **Download data & documents**

All variables of ZA-0035

🔍 Political interest [GO]

var x1 „Political interest"
var x2 „label"
var x2_1 „label"
var x2_2 „label"
var x2_3 „label"
var x1 „label"
var x 2 „label"
var x3 „label"
var x4 „label"
var x5 „label"
var x6 „label"
var x7 „label"
var x8 „label"
var x9 „label"
var x10 „label"
var x11 „label"
var x11 „label"
var x12 „label"
var x13 „label"
var x14 „label"
var x15 „label"
var x16 „label"
var x17 „label"
var x18 „label"
var x19 „label"
var x20 „label"
var x21 „label"
var x22 „label"
var x23 „label"
var x24 „label"

🛒 **Variable description.pdf**

🛒 **Variable x1: „Political interest"**
Question wording: „Are you interested in politics…?"
Interview instruction: „please show…"

Response category:

value	value label	total	freq (%)	cum.
1	yes	300	30	300
2	no	500	50	800
98	d.k.	100	10	900
99	n.a.	100	10	1000

Statistical values:

Valid	xx	min.	xx	sum.
Missing	xx	max.	xx	stddev.
Total	xx	mean	xx	variance

Note:
The answer categories were ……

🛒**Showcard:**

SHOWCARD
ja
nein
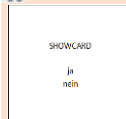
Graphs:

Original Questions/
additional translation | comparable variables

Search results

- Result box displays studies and variables
- Information on studies
  - ▸ Result box shows study information: study-ID, title, study years
  - ▸ User gets access to
    - Study related material: questionnaires, study description, variable report
    - Meta data
    - Survey data (SPSS, STATA, R); OpenSource-Tool with access restriction
- Information on variables
  - ▸ Result box shows: variable-ID, variable label, study information where the variable was taken from
  - ▸ User gets access to
    - variable information: variable name, variable label, question wording, answer categories and codes, distribution of variable (frequencies and statistical values), archive notes

Technical features

- Feature for a customized data set: compile variables via scriptgenerator

- Feature for shopping basket: all documents, data files, graphs, show cards, variable information, individually selected variables can be put into the basket and downloaded

- Possibility for saving and exporting search results

Next steps

- Possibility to create graphs
- Analysis function for variables
- Functionality for "related variables"
  - Related variables from the same study collection
    - Results from existing correspondence lists
  - Related variables from different study collections
    - Results from upcoming correspondence lists
    - Technical solution
      - Information retrieval
      - Settings
      - Algorithms

Technical background

- Programing language: **Java**
- Front-end user interfaces: **Vaadin8**
- Information Retrieval system: **ElasticSearch5.5**
- Content management: **FlatDB** (internal project)
- Integrated search: **Solr-Index technology**
- All contents are based on **DDI Lifecycle format**
  - Presentation will be on Friday the 1st of June, 9:00-10:15
    - „Implementing DDI lifecycle to achieve interoperability between projects at the GESIS - Leibniz-Institute for the Social Sciences Data Archive for the Social Sciences"
    - Esra Akdeniz; Kerrin Borschewski
    - Session block G

Thank you for your attention!

For questions concerning technical contant, please contact:
Wolfgang Zenk-Möltgen: **wolfgang.zenk-moeltgen@gesis.org**

For further questions, please contact:
Julia Hermann: **julia.hermann@gesis.org**
Christina Eder: **christina.eder@gesis.org**