

gesis

Leibniz Institute
for the Social Sciences



Providing high quality data and metadata through the ExploreData search portal

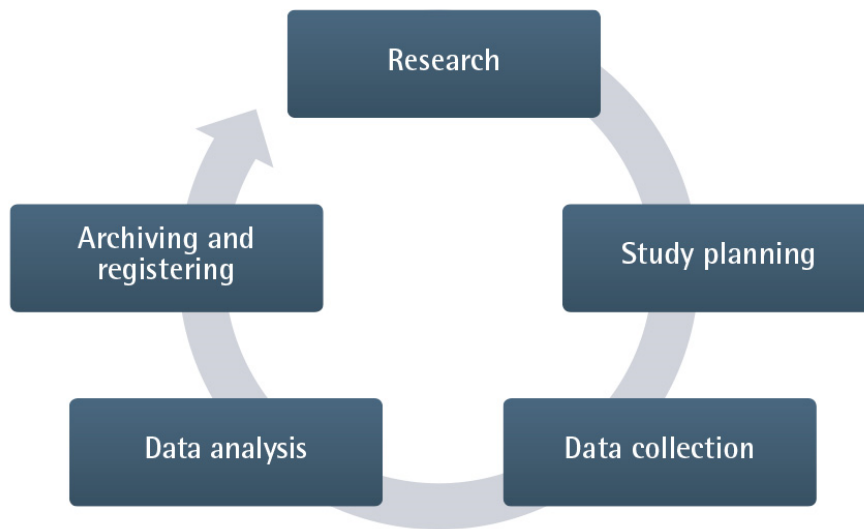
Julia Hermann, Christina Eder, Wolfgang Zenk-Möltgen
May 30th 2018

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



www.gesis.org

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 programmes
- 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](#)

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
 - ▶ interdisciplinary 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



FILTER & BROWSE

CESSDA Topic-
Classification

Series

Study Year

Geography

Interview language

Methodology

[Reset all filters](#)

Search ...



[Reset](#)

[Search help](#)

gesis

Leibniz Institute
for the Social Sciences

Explore Data Project

Access to studies and variables



FILTER & BROWSE

CESSDA Topic-
Classification

Series

Study Year

Geography

Interview language

Methodology

[Reset all filters](#)

Political interest

[Search help](#)



[Reset](#)

Studies (3309) Variables (5722)

Nr.	Study	Study year
1	ZA0035 "Political Interest 1964 "	1964
2	ZA1401 "International Workbook: Political Participation (Australia) "	1967
3	ZA1405 "International Workbook: Political Participation (Netherlands) "	1970-1973
4	ZA1403 "International Workbook: Political Participation (Germany) "	1972
5	ZA1404 "International Workbook: Political Participation (Japan) "	1967
6	ZA1406 "International Workbook: Political Participation (Norway) "	1965
7	ZA1402 "International Workbook: Political Participation (Canada) "	1965-1966
8	ZA1188 "Political Action II "	1979-1981
9	ZA1408 "International Workbook: Political Participation (United States) "	1972-1973
10	ZA1407 "International Workbook: Political Participation (United Kingdom) "	1963-1970

Access to studies and variables

The screenshot shows the gesis website interface. At the top left is the gesis logo and 'Leibniz Institute for the Social Sciences'. At the top right is a 'Language' dropdown menu set to 'EN'. Below the logo is a 'FILTER & BROWSE' section with a list of filter categories: CESSDA Topic-Classification, Series, Study Year, Geography, Interview language, and Methodology. An orange arrow points to the 'CESSDA Topic-Classification' button. To the right of this section is a search bar containing 'Political interest' and a 'Reset' button. A modal window titled 'Filter according to: CESSDA Topic-Classification' is open, showing a list of topics with checkboxes and counts. The topics are: Labour and employment (34), Demography and population (8), Education (42), History (13), Society and culture (31), Law, crime and legal systems (62), Health (40), Trade, industry and markets (55), Information and communication (34), Reference and instructional resources (0), Politics (51), Psychology (14), and Social stratification and groupings (204). At the bottom of the modal are 'Reset selection', 'Apply', and 'Cancel' buttons. In the background, a table of studies is visible with columns for 'Study year' and 'Study year'. The table lists various countries and their corresponding study years: Australia (1964, 1967), Netherlands (1970-1973), Germany (1972), Japan (1967), Norway (1965), Canada (1965-1966), United States (1979-1981, 1972-1973), and United Kingdom (1963-1970).

gesis Leibniz Institute for the Social Sciences

Language EN

FILTER & BROWSE

Political interest Search help Reset

CESSDA Topic-Classification

Series

Study Year

Geography

Interview language

Methodology

Reset all filters

Filter according to: CESSDA Topic-Classification

Select: CESSDA Topic-Classification

Search ...

- ☐ Labour and employment (34)
- ☐ Demography and population (8)
- ☐ Education (42)
 - ☐ History (13)
- ☐ Society and culture (31)
- ☐ Law, crime and legal systems (62)
- ☐ Health (40)
- ☐ Trade, industry and markets (55)
- ☐ Information and communication (34)
- ☐ Reference and instructional resources (0)
- ☐ Politics (51)
 - ☐ Psychology (14)
- ☐ Social stratification and groupings (204)

Reset selection Apply Cancel

Study year

1964
1967
1970-1973
1972
1967
1965
1965-1966
1979-1981
1972-1973
1963-1970

Page 1 of 331 10 per page

Access to studies and variables

The screenshot shows the gesis website interface. On the left, under 'FILTER & BROWSE', there are several filter categories: CESSDA Topic-Classification (ON), Society and c..., Series (highlighted with an orange arrow), Study Year, Geography, Interview language, and Methodology. A 'Reset all filters' link is also present. The main search area has a search bar containing 'Political interest' and a 'Reset' button. Below the search bar, it shows 'Studies (31)' and 'Variables (164)'. A table lists the studies:

Nr.	Study
1	ZA2570 "Comparative Manifestos Project (Programmatic Profiles of Political Parties in 20 Countries, 1945-1988..."
2	ZA1390 "Dialogs 1 (The Citizen as Partner) "

A modal window titled 'Filter according to: Study Year' is open, showing a search bar and a grid of checkboxes for years from 1981 to 2015, each followed by a count in parentheses. The modal also includes 'Reset selection', 'Apply', and 'Cancel' buttons.

Access to studies and variables

gesis

Leibniz Institute
for the Social Sciences

Language

EN

FILTER & BROWSE

CESSDA Topic-Classification

OFF

Society and C...

Series

Study Year

ON

2006-2010

Geography

Interview language

Methodology

[Reset all filters](#)

Political interest

[Search help](#)

Studies (364) Variables (2040)

Nr.	Study
1	ZA4584 "German General Social Survey (ALLBUS) - Cumulation 1980-2014"
2	ZA5101 "Political Communication Cultures in Western Europe"
3	ZA6796 "Pathways to Power: The Political Representation of Citizens of Immigrant Origin in Belgium (BE-PATHV
4	ZA5638 "Political Education"
5	ZA3950 "International Social Survey Programme: Citizenship - ISSP 2004"
6	ZA5371 "On the Political Situation in Hesse 2009"
7	ZA5227 "Political Online Communication"
8	ZA5836 "EU-Profler: User Profiles in the 2009 European Elections - Respondent-level Data for United Kingdom
9	ZA5805 "EU-Profler: User Profiles in the 2009 European Elections - Respondent-level Data for Croatia"
10	ZA5806 "EU-Profler: User Profiles in the 2009 European Elections - Respondent-level Data for Cyprus"
11	ZA5824 "EU-Profler: User Profiles in the 2009 European Elections - Respondent-level Data for Poland"
12	ZA5813 "EU-Profler: User Profiles in the 2009 European Elections - Respondent-level Data for Greece"

Page 1 of 2 250 per page

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



FILTER & BROWSE

CESSDA Topic-
Classification

Series

Study Year

Geography

Interview language

Methodology

[Reset all filters](#)

Political interest



[Reset](#)

[Search help](#)

Studies (3309)

Variables (5722)

Nr.	Study	Study year
1	ZA0035 "Political Interest 1964 "	1964
2	ZA1401 "International Workbook: Political Participation (Australia) "	1967
3	ZA1405 "International Workbook: Political Participation (Netherlands) "	1970-1973
4	ZA1403 "International Workbook: Political Participation (Germany) "	1972
5	ZA1404 "International Workbook: Political Participation (Japan) "	1967
6	ZA1406 "International Workbook: Political Participation (Norway) "	1965
7	ZA1402 "International Workbook: Political Participation (Canada) "	1965-1966
8	ZA1188 "Political Action II "	1979-1981
9	ZA1408 "International Workbook: Political Participation (United States) "	1972-1973
10	ZA1407 "International Workbook: Political Participation (United Kingdom) "	1963-1970

Search results

Home

Search results

ZA-0035

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

Home

Search results

ZA-0035

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 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

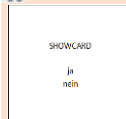
Statistical values:

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

Note:

The answer categories were

Showcard:



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

- Programming 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 content, 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