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

Study planning 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

<table>
<thead>
<tr>
<th>Source</th>
<th>Type of document</th>
<th>Number</th>
<th>Downloads 2017</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data-Archive</td>
<td>Studies</td>
<td>6299</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Study related documents (questionnaires, codebooks, variable reports, method reports, etc...)</td>
<td>28931</td>
<td>869273* (data&amp;docs)</td>
</tr>
<tr>
<td></td>
<td>data sets (spss-, stata-, csv-files)</td>
<td>10587</td>
<td></td>
</tr>
</tbody>
</table>

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 on conceptual Mock-ups were done
Explore Data Search Portal
Access to studies and variables

Explore Data Project
Access to studies and variables

<table>
<thead>
<tr>
<th>Nr.</th>
<th>Study</th>
<th>Study year</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ZA0035 &quot;Political Interest 1964 &quot;</td>
<td>1964</td>
</tr>
<tr>
<td>2</td>
<td>ZA1401 &quot;International Workbook: Political Participation (Australia) &quot;</td>
<td>1967</td>
</tr>
<tr>
<td>3</td>
<td>ZA1405 &quot;International Workbook: Political Participation (Netherlands) &quot;</td>
<td>1970-1973</td>
</tr>
<tr>
<td>4</td>
<td>ZA1403 &quot;International Workbook: Political Participation (Germany) &quot;</td>
<td>1972</td>
</tr>
<tr>
<td>5</td>
<td>ZA1404 &quot;International Workbook: Political Participation (Japan) &quot;</td>
<td>1967</td>
</tr>
<tr>
<td>6</td>
<td>ZA1406 &quot;International Workbook: Political Participation (Norway) &quot;</td>
<td>1965</td>
</tr>
<tr>
<td>7</td>
<td>ZA1402 &quot;International Workbook: Political Participation (Canada) &quot;</td>
<td>1965-1966</td>
</tr>
<tr>
<td>8</td>
<td>ZA1188 &quot;Political Action II &quot;</td>
<td>1979-1981</td>
</tr>
<tr>
<td>9</td>
<td>ZA1408 &quot;International Workbook: Political Participation (United States) &quot;</td>
<td>1972-1973</td>
</tr>
<tr>
<td>10</td>
<td>ZA1407 &quot;International Workbook: Political Participation (United Kingdom) &quot;</td>
<td>1963-1970</td>
</tr>
</tbody>
</table>
Access to studies and variables
Access to studies and variables
## Access to studies and variables

### FILTER & BROWSE

- **CESSDA Topic-Classification**
  - Society and C.

- **Series**
  - OFF

- **Study Year**
  - 2006-2010

- **Geography**

- **Interview language**

- **Methodology**

### Studies (364) & Variables (2040)

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

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

<table>
<thead>
<tr>
<th>Nr.</th>
<th>Study</th>
<th>Study year</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ZA0035 &quot;Political Interest 1964 &quot;</td>
<td>1964</td>
</tr>
<tr>
<td>2</td>
<td>ZA1401 &quot;International Workbook: Political Participation (Australia) &quot;</td>
<td>1967</td>
</tr>
<tr>
<td>3</td>
<td>ZA1405 &quot;International Workbook: Political Participation (Netherlands) &quot;</td>
<td>1970-1973</td>
</tr>
<tr>
<td>4</td>
<td>ZA1403 &quot;International Workbook: Political Participation (Germany) &quot;</td>
<td>1972</td>
</tr>
<tr>
<td>5</td>
<td>ZA1404 &quot;International Workbook: Political Participation (Japan) &quot;</td>
<td>1967</td>
</tr>
<tr>
<td>6</td>
<td>ZA1406 &quot;International Workbook: Political Participation (Norway) &quot;</td>
<td>1965</td>
</tr>
<tr>
<td>7</td>
<td>ZA1402 &quot;International Workbook: Political Participation (Canada) &quot;</td>
<td>1965-1966</td>
</tr>
<tr>
<td>8</td>
<td>ZA1188 &quot;Political Action II &quot;</td>
<td>1979-1981</td>
</tr>
<tr>
<td>9</td>
<td>ZA1408 &quot;International Workbook: Political Participation (United States) &quot;</td>
<td>1972-1973</td>
</tr>
<tr>
<td>10</td>
<td>ZA1407 &quot;International Workbook: Political Participation (United Kingdom) &quot;</td>
<td>1963-1970</td>
</tr>
</tbody>
</table>
ZA-0035 „Political Interest“ 1964

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
- Questionnaire.txt
- Studydescription.pdf
- Studydescription.txt
- Codebook.pdf
- Codebook.txt
- Studynotes.pdf
- Studynotes.txt
ZA-0035 „Political Interest“ 1964

All variables of ZA-0035

- var x1 „Political interest"
- var x2 „label”
- var x2_1 „label”
- var x2_2 „label”
- var x2_3 „label”
- var x1 „label”
- var x2 „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:**

<table>
<thead>
<tr>
<th>value</th>
<th>value label</th>
<th>total</th>
<th>freq (%)</th>
<th>cum.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>yes</td>
<td>300</td>
<td>30</td>
<td>300</td>
</tr>
<tr>
<td>2</td>
<td>no</td>
<td>500</td>
<td>50</td>
<td>800</td>
</tr>
<tr>
<td>98</td>
<td>d.k.</td>
<td>100</td>
<td>10</td>
<td>900</td>
</tr>
<tr>
<td>99</td>
<td>n.a.</td>
<td>100</td>
<td>10</td>
<td>1000</td>
</tr>
</tbody>
</table>

**Statistical values:**

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

**Note:**

The answer categories were ......

**Showcard:**

**Graphs:**
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