Georeferenced Survey Data at the GESIS Data Archive

Stefan Müller | Wolfgang Zenk-Moeltgen | Stefan Schweers
GESIS – Leibniz Institute for the Social Sciences
Data Archive for the Social Sciences
{stefan.mueller|wolfgang.zenk-moeltgen|stefan.schweers}@gesis.org

Content

1. Background
2. DDI and Georeferenced Survey Data
3. The Future? Archiving 'real' spatial data
4. Outlook
gesis

ONE DOES NOT SIMPLY

GEOREFERENCE SURVEY DATA
High efforts

Conceptual
High efforts

Conceptual
gesis

High efforts

Conceptual  Technical
High efforts

Conceptual  Technical  Documentary
The project: GeorefUm

- Geocoding and georeferencing
- Building a database of available spatial data
- Disseminating in a secure environment
- Consult and guide
Road traffic noise in Cologne, Source: EIONET Central Data Repository (CDR) and OpenStreetMap
Road traffic noise in Cologne, Source: EIONET Central Data Repository (CDR) and OpenStreetMap
Road traffic noise in Cologne, Source: EIONET Central Data Repository (CDR) and OpenStreetMap
Road traffic noise in Cologne, Source: EIONET Central Data Repository (CDR) and OpenStreetMap

60 – 64 dB

50 – 54 dB

50 – 54 dB
Road traffic noise in Cologne, Source: EIONET Central Data Repository (CDR) and OpenStreetMap
New data types, formats, contents etc.

...new documentary demands!

Source: pixabay.com
Actually, not that different data

<table>
<thead>
<tr>
<th>ID</th>
<th>...</th>
<th>ROAD_DEN</th>
<th>ROAD_N</th>
<th>DIST_ROAD_DEN</th>
<th>DIST_ROAD_N</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>...</td>
<td>55</td>
<td>50</td>
<td>24.56</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>...</td>
<td>75</td>
<td>60</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>3146</td>
<td>...</td>
<td>0</td>
<td>0</td>
<td>6.23</td>
<td>10.76</td>
</tr>
</tbody>
</table>
Still, we need new metadata

Different standards or frameworks:
- ISO 19118, 19115, 3166 and INSPIRE

Different purposes:
- cataloging
- usage
- structure and meaning
- preservation

Source: pixabay.com
ISO 19115 wishlist

- Abstract
- CitedResponsibleParty
- Extent
  - spatial
  - temporal
- SpatialRepresentationType
GESIS DDI Lifecycle Snippet

- DDIInstance
  - StudyUnit
  - Creator/affiliation
  - SpatialCoverage
    - LogicalProduct
      - Variable
        - Category Scheme
        - GeographicStructureScheme
    - ConceptualComponent
Study Level (as is)

- **SpatialCoverage**
  - **Description**: ISO 3166 as well as 'home grown' city codes, e.g. DE-NW-K for the city of Cologne
  - **TopLevelReference**: ISO 3166-1 alpha-2
  - **LowestLevelReference**: ISO 3166-1 alpha-2
Study Level (as could be)

SpatialCoverage
  - Description
    - CountryCode — ISO 3166 codes
    - TopLevelReference — ISO 3166-1 alpha-2
    - LowestLevelReference — ISO 3166-1 alpha-2
    - BoundingBox — xmin, ymin, xmax, ymax
Structure of spatial data

Vector data

Raster data
Hello. I am raster data, based on 1km INSPIRE compliant grids. This is important as I am not serving straight to the point information. Use me when georeferenced variables were produced with data from the same grid structure.
Variable Level ('whole' story I)

- LogicalProduct
  - Variable
    - ... (omitted)
    - Codelist
- ConceptualComponent
- GeographicStructureScheme
- Universe
- UniverseReference
  - R
Hello. I am raster data, based on 1km INSPIRE compliant grids. This is important as I am not serving straight to the point information. Use me when georeferenced variables were produced with data from the same grid structure.
A shortened example

<l:Variable>
  ...
  <l:VariableName>
    <r:String>road_lden</r:String>
  </l:VariableName>
  <r:Label>
    <r:Content>Road Traffic Noise (DEN) in decibels</r:Content>
  </r:Label>
  <r:Description>
    <r:Content>Road traffic noise that was measured at the respondent’s dwelling in context of EG (2002).</r:Content>
  </r:Description>
  <r:UniverseReference>
    <r:ID>INSPIRE_grid_1km</r:ID>
  </r:UniverseReference>
  ...
</l:Variable>
'Whole' story II

- Create one study unit for each geographic structure
  - e.g. one for every 1km grid structure

SU I: 'Ordinary' survey variables

SU II: Variables extracted from rasters

SU III: Variables extracted from polygons
What's left?

- Actual implementation
- Final mapping to ISO 19115
- Using DDI-RDF and controlled vocabulary
- What about 'real' spatial data?
'Real' spatial data: base data

Different understandings of these data

- Administrative borders (e.g. Switzerland)
- Shapes of cadasters, buildings, roads as well as administrative borders (e.g. Germany)
Archiving 'real' spatial data

Great, because:

- Important to document data collection process or to visualize results
- Makes data linking easier

However:

- Pitfall of redundant archiving
- DDI uncommonly used format
Outlook

- Georeferencing survey data is costly
  - This is why we do it
- New documentary demands
  - e.g. ISO 19115 and INSPIRE compliance
  - Indeed, we use DDI for that
    - Variable level is tricky
- 'Real' spatial data
  - Future is uncertain, depends on demand
Thanks to:

Wolfgang Zenk-Möltgen and Stefan Schweers as well as to Julia Klinger, Katharina Kinder-Kurlanda and Pascal Siegers