gesis Leibniz Institute for the Social Sciences

Conceptualizing a Spatial Data Infrastructure for the Social Sciences

An example from Germany

Preserving Transactional Data

Stefan Schweers, 17/03/2016

Member of the Leibniz Association

Content

- Demand for spatial data in the Social Sciences
- The project “GeorefUm“
- Challenges
  - Data protection
  - Metadata
- Outlook

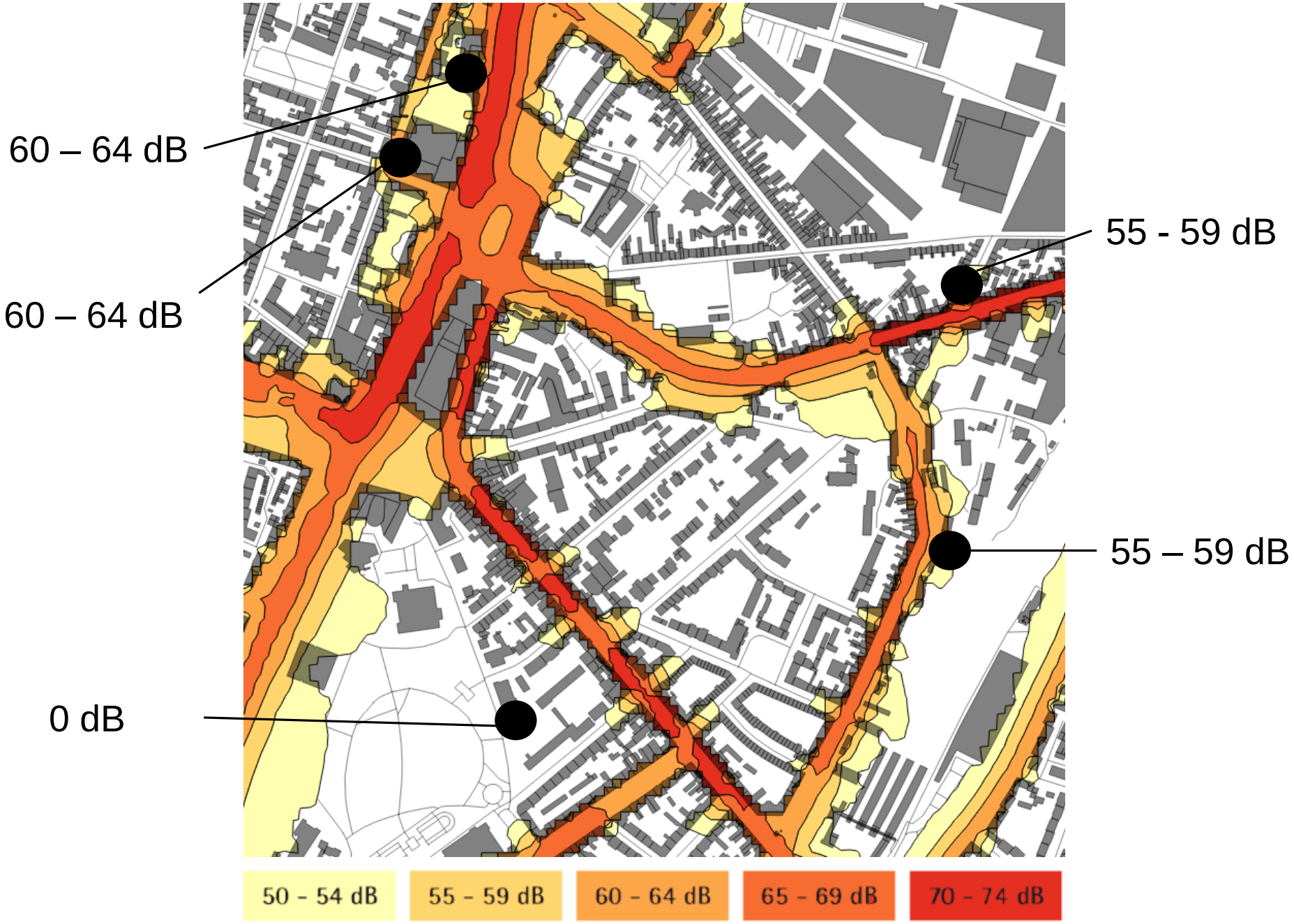
Demand for spatial data in the Social Sciences

- Direct living environment influences individuals' attitudes and behaviors
- Spatial dimension can increase the quality of statistical models for survey data

But yet there is no Spatial Data Infrastructure for the Social Sciences

The project „GeorefUm“

- Funded by the German Research Foundation (DFG)
- Objective: Conceptualizing an SDI, that allows
  - Combine survey data and spatial data
  - Analyze, document, archive and share georeferenced survey data



Traffic Noise Cologne, Source: EIONET Central Data Repository and OpenStreetMap

The project “GeorefUm“

ID	Year	V1 - Vn	Area ID	Population	Pop. over 65	Noise dB
1	2016	…	1	2.100	680	60-64
2	2016	…	2	45	23	0
3	2016	…	3	1.800	550	55-59
4	2016	…	1	2.100	680	60-64
5	2016	…	3	1.800	550	55-59

Challenges: Data protection

A spatial data infrastructure has to provide solutions for using spatial data combined with survey

data in compliance with the legislation in Germany and other European countries.

Challenges: Data protection

De-anonymization

- Direct identifiers
- Combination of values
  - Example lawyer with 7 children in a small municipality
- Especially small scaled spatial information about the respondent are sensitive, if unique

Challenges: Data protection

Approaches for solutions

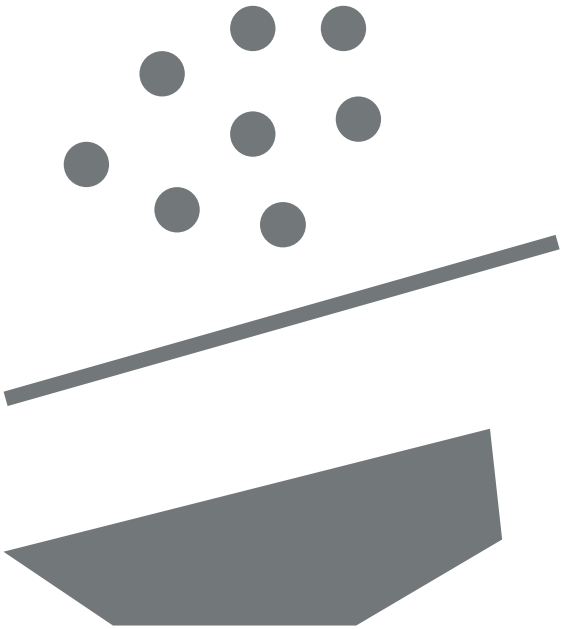
- Coarsening
  - Classification of values
  - Put together spatial information
- Contractual control
- Technical control
  - Access to georeferenced survey data only via on site access

Challenges: Metadata

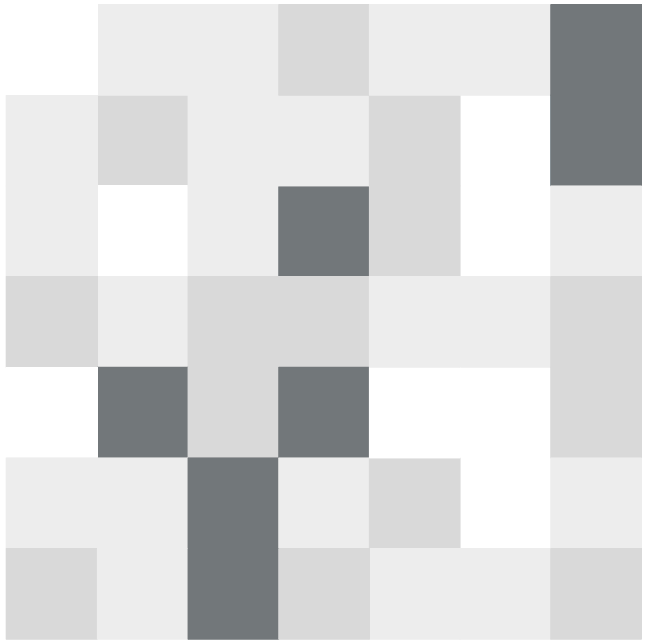
- Standards INSPIRE, ISO 19115, 19118, 3166
- Need for metadata at different levels
  - Catalogue level
  - Variable level
- Different data sources, different variables
  - Different spatial coverage and underlying structures
- DDI has no ‘default’ way to consider these criteria on the variable level

Challenges: Metadata

Different strcutures of spatial data



Vector data



Raster data

Challenges: Metadata

Necessary Metadata (ISO 19115)

- Abstract
- CitedResponsibleParty
- Extent
  - spatial
  - temporal
- SpatialRepresentationType

Outlook

Legal barriers

- Automatic Statistical Disclosure Analysis
- New ways of access to sensitive data

Metadata

- Harvesting via OGC Web Services
- Automatic generating of metadata
- Linked Open Data

Thank you!

Contact details: stefan.schweers@gesis.org

For further information see also:

Schweers, Stefan, Katharina E. Kinder-Kurlanda, Stefan Müller, and Pascal Siegers. 2016. "Conceptualizing a spatial data infrastructure for the social sciences: an example from Germany." *Journal of Map & Geography Libraries*.