

Enabling spatial research using German Administrative Data

Workshop on Integrating Geospatial and Statistical Standards

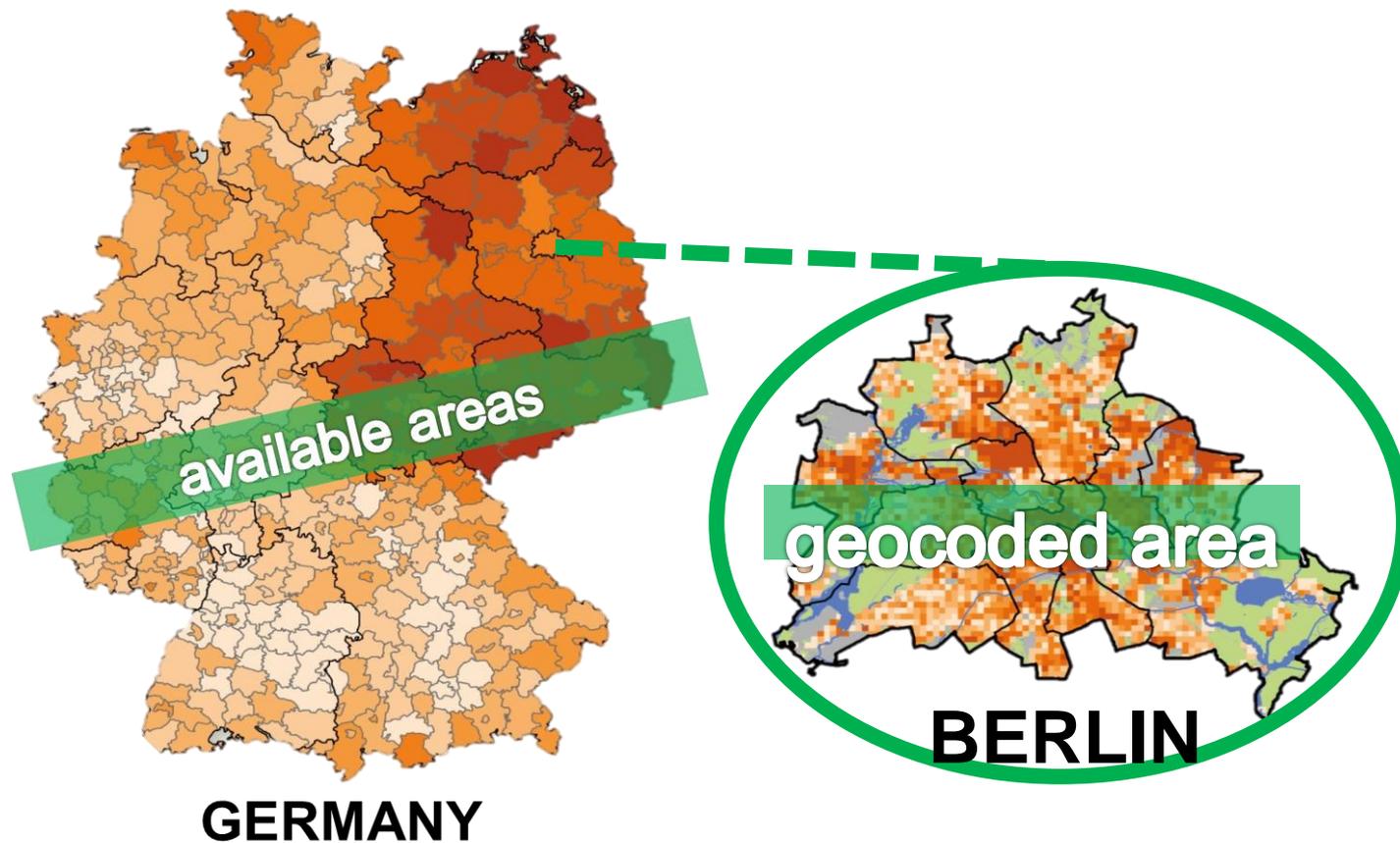
November 8, 2017

Dana Müller
Philipp vom Berge

- Who are we?
 - part of the Institute for Employment Research (IAB), which belongs to the Federal Employment Agency (BA)
- What do we do?
 - access to confidential micro data of BA/IAB for academic purposes
 - standardized data products (data access specific anonymization)
- Spatial research
 - can be done already
 - until now: district level is the lowest level

Why geocoding?

- Increasing need for data at a sufficiently fine regional scale (regional, segregation, transport research)



Approach

- Geocoding of all persons and establishments contained in the Integrated Employment Biographies (IEB)



- Address material: Federal Agency for Cartography and Geodesy (GAB)
 - 22 million building addresses with corresponding coordinate
- Software: infas360 geocoder
 - Reference systems: WGS84, GK3, UTM32
- approx. 95% matches

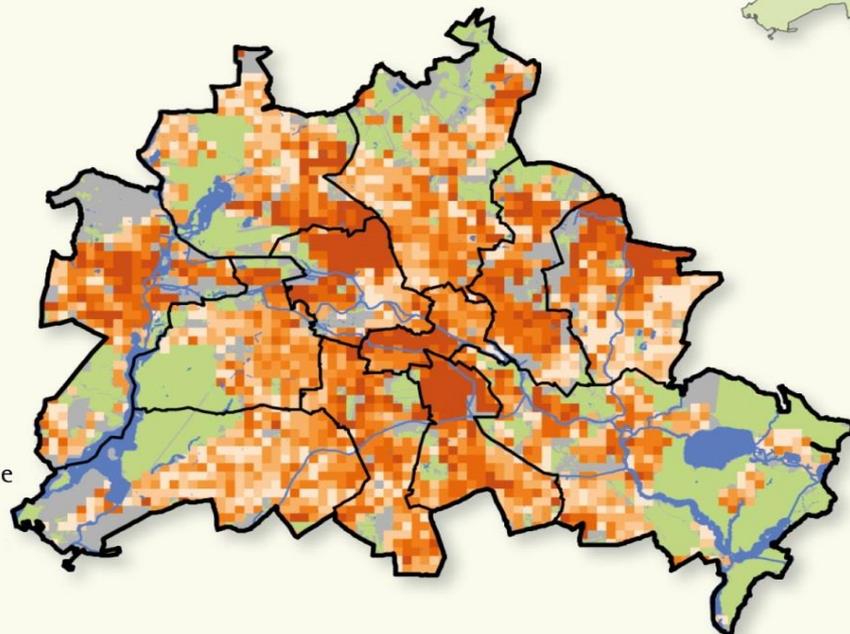
Anonymization

- Problem: exact address information highly sensitive
- Need to find useful aggregation method(s)
 - High research potential without compromising data security
- Our solution so far:
 - Planned access only via on-site use at the RDC
 - Persons and establishments are sorted to its corresponding grid cell
 - Projection ETRS89 LAEA
- Example
 - Y=2999719 m , X=3962799m
 - 500 meter grid cell: N 2999500 E 3962500
- Generate a grid cell identifier

Niedriglohnbeschäftigte in Berlin 2009

Anteile an allen sozialversicherungspflichtig Beschäftigten in Prozent

Anteile in %



Bezirke in Berlin

- 1 Berlin-Mitte
- 2 Friedrichshain-Kreuzberg
- 3 Pankow
- 4 Charlottenburg-Wilmersdorf
- 5 Spandau
- 6 Steglitz-Zehlendorf
- 7 Tempelhof-Schöneberg
- 8 Neukölln
- 9 Treptow-Köpenick
- 10 Marzahn-Hellersdorf
- 11 Lichtenberg
- 12 Reinickendorf

Anmerkung: Die stadtsspezifische Niedriglohnschwelle für Berlin beträgt 1.611 Euro (Bruttomonatslohn).

Quelle: Georeferenzierte Daten des IAB; Geometrien: OpenStreetMap, ThinkGeo, Geofabrik.

© IAB

Challenges

- Select number/details of variables in the research data set
- Select optimal grid size
 - Urban vs. rural areas
 - Flexible vs. Fixed
- How to anonymize cells with view observations?
 - Delete (secondary suppression)
 - Aggregate with neighbouring cells
 - Synthesize/random error
 - Individuals and establishments

If you want to learn more about the RDC:

<http://fdz.iab.de/en>