

EGU2020-18956

<https://doi.org/10.5194/egusphere-egu2020-18956>

EGU General Assembly 2020

© Author(s) 2020. This work is distributed under the Creative Commons Attribution 4.0 License.



## From Silos to FAIR Services: Interoperable application of geospatial data for longitudinal surveys in the Social Sciences.

**Peter Löwe**, Tobias Gebel, and Hans Walter Steinhauer

German Institute for Economic Research (DIW Berlin), Berlin, Germany (ploewe@diw.de)

Due to the european INSPIRE directive to establish an infrastructure for spatial information in Europe, the number of national data sources in Europe which are open to the public or at least science continues to grow.

However, challenges remain to enable easy access for society and science to these previously unavailable data silos based on standardized web-services, as defined by the Open Geospatial Consortium (OGC). This is crucial to ensure sustainable data generation and reuse according to the FAIR principles (Findable, Accessible, Interoperable, Reusable).

We report on an interdisciplinary application, using spatial data to improve longitudinal surveys in the social sciences, involving building plans encoded in CityGML, PostGIS, MapServer and R.

The Socio-economic Panel (SOEP) as part of the German Institute for Economic Research (DIW Berlin) provides longitudinal data on persons living in private households across Germany. Lately the SOEP sampled households in certain neighborhoods within cities, areas of the so called „Soziale Stadt“ (social town). Because of restricted area, spatially referenced data has been used. Information on the level of census tiles provided by the Federal Statistical Office was used to form regional clusters.

Within these clusters addresses, spatially referenced by the German Federal Agency for Cartography and Geodesy (BKG), have been sampled. This way, we made sure addresses are within the neighborhoods to be surveyed. As this procedure turned out to reduce organizational burden for the survey research institute as well as for the interviewers and at the same time allows for generating random household samples, it is considered for future use. Yet, addresses can belong to residential buildings as well as cinemas or hotels.

To meet with this obstacle we evaluate the use of 3D Building Models provided by the German Federal Agency for Cartography and Geodesy (BKG).

This data is distributed as compressed data archives for the 16 states of Germany, each containing very large numbers of CityGML files containing LoD1 data sets for buildings. The large storage footprint of these data sets makes their reuse by social scientists using standard statistical software (such as R or Stata) on desktop computers difficult at best. This is overcome by providing flexible access to Areas of Interest (AOI) through OGC Webservices (WMS/WFS) based on a PostGIS

database. The ingestion process is based on the new GMLAS driver of the ogr software project for Complex Features encoded in Geographic Markup Language (GML) based on application schemas.