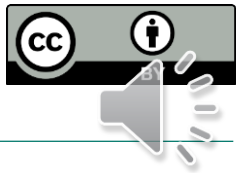
DIW BERLIN

From Silos to FAIR Services

Interoperable application of geospatial data for longitudinal surveys in the Social Sciences.

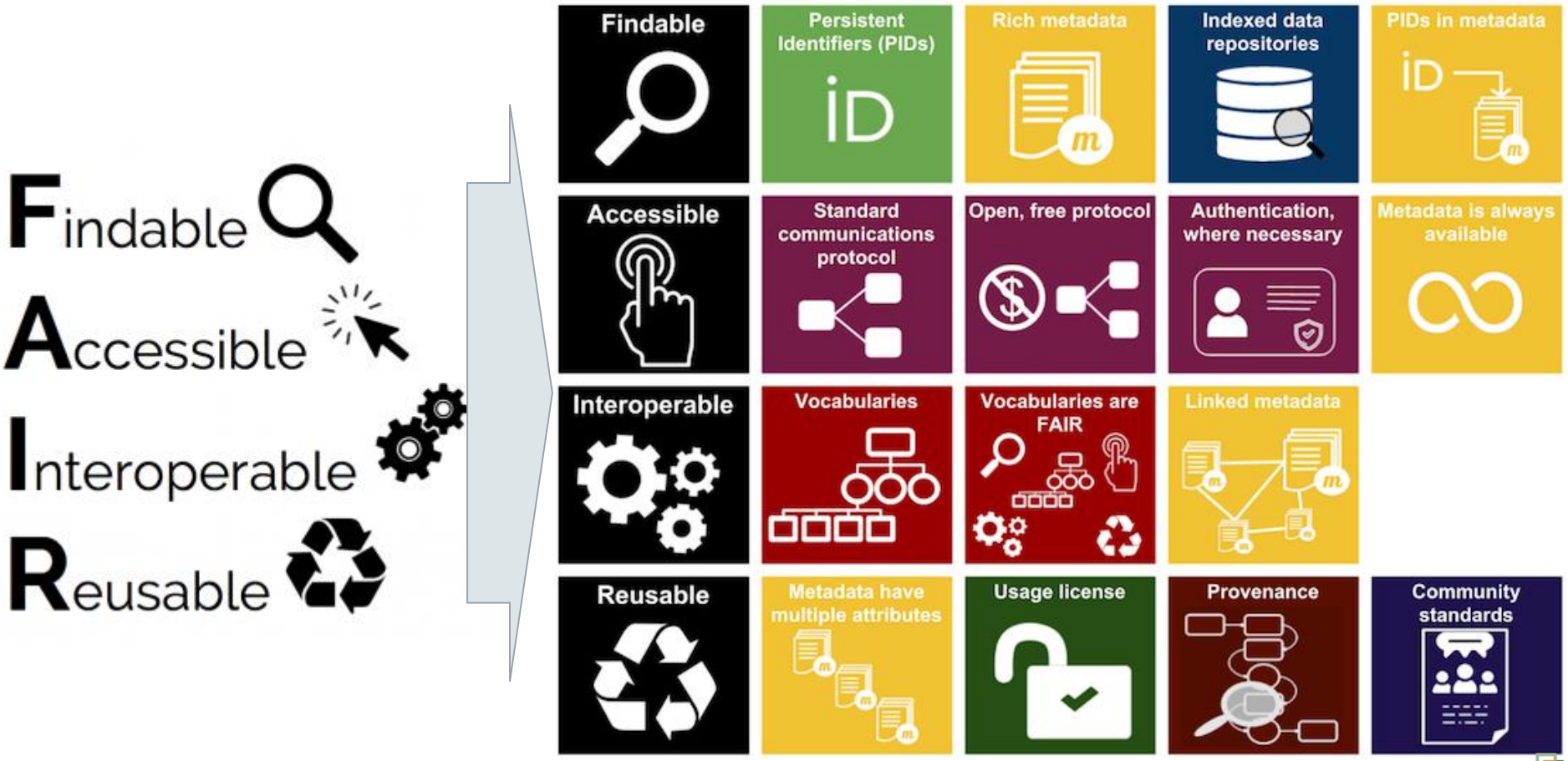
Peter Löwe (ploewe@diw.de) , Tobias Gebel, Hans Walter Steinhauer

EGU 2020 - ESSI 3.5 - EGU2020-18956

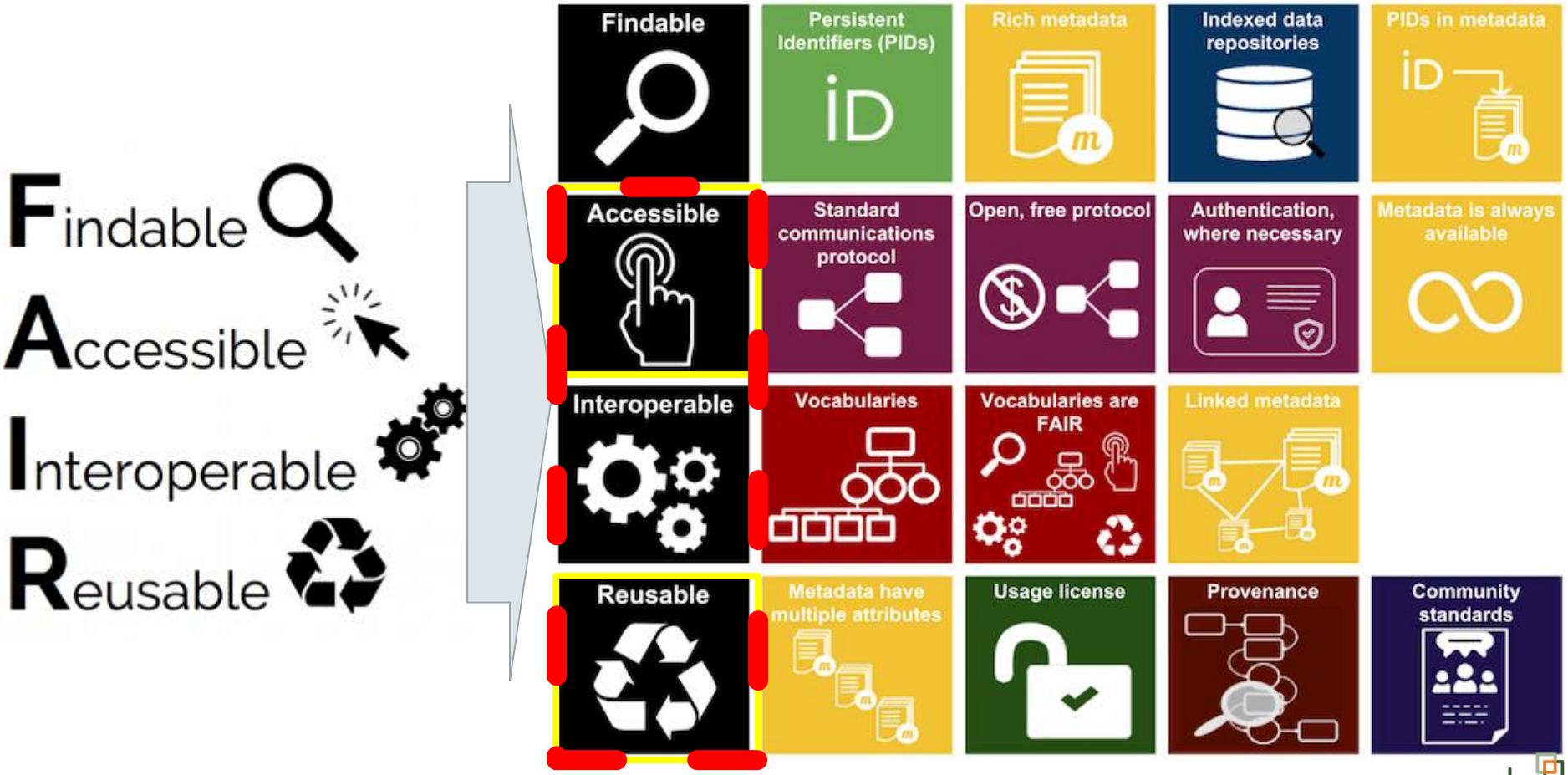


Setting the scope: FAIR

Setting the scope: FAIR



Setting the scope: FAIR

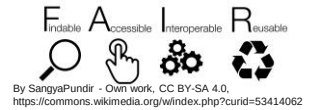


The Challenge

- **Goal**: Application of Building Model Data for Germany provided by the Federal Agency for Cartography and Geodesy (BKG) .

Federal Agency for Cartography and Geodesy

- .. in the Social Sciences to optimize annual longitudinal surveys.



By SangyaPundir - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=53414062

- **Challenge:**
  - **Data is provided as state-level file-based silos (no web service)**
  - **Data reduction necessary to focus on areas of interest.**
  - **No reference implementation based on Free and Open Source (FOSS) tools.**
- **Approach:**
  - Enable FAIR data use by introducing a standard-based FOSS-based service layer

https://www.bkg.bund.de/SharedDocs/Produktinformationen/BKG/DE/P-2019/190211_LoD1.html

DIW BERLIN

Application Scenario: Residential buildings for household-based surveys

- **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 (15.000 households) across Germany.**

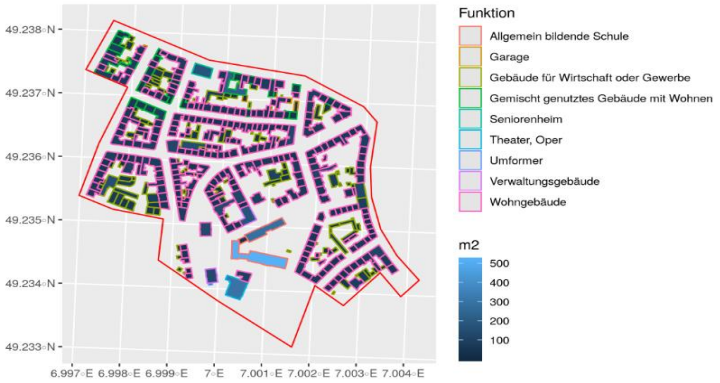
- **SOEP samples households in certain neighborhoods within cities, areas of the so called „Soziale Stadt“ (Social Integrative City)**

- **Regional clustering based on census tile grid**

- **Random sampling based on official address data.**

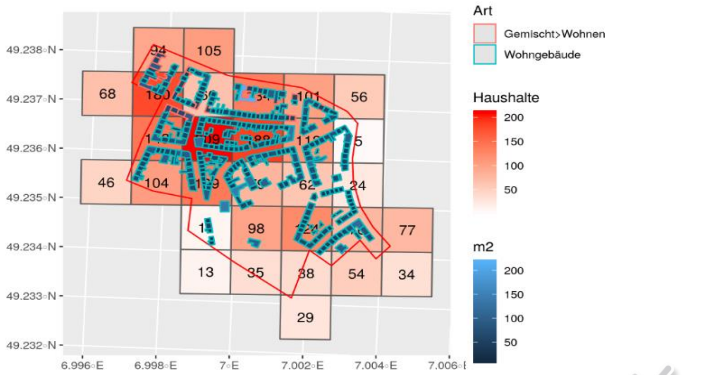
- **New: <u>Filtering by building model attributes for building function to seperate residential buildings from commercial buildings.</u>**

Layout, Lot size, building function



Source: Steinhauer (2020)

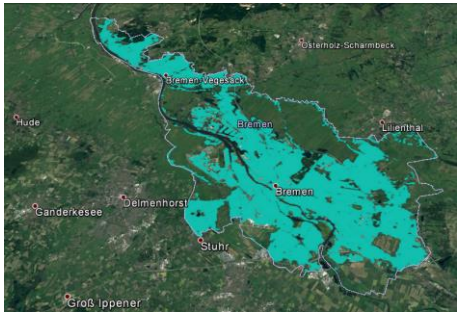
Census grid clipping



Source: Steinhauer (2020)

Hands-on Challenge: Working with Data Silos

- Data is provided as silos on State-level as compressed file archives.
- Decompressing thousands of XML-files (CityGML) on a desktop computer leaves a large storage footprint.
- Need to convert building polygons into other geospatial formats – on State level (KML, ESRI-Shape are bad ideas) ?
- The German CityGML-dialect extends the standard OGC definition.
- **Reality Check**: Social Scientists use off-the-shelf statistical software (R, Stata) on desktop computers and are not by default versed in geoinformatics- and scripted file-processing.



Building models for the State of Bremen rendered in GoogleEarth

Building models for the State of Bremen rendered in QGIS (via a WMS)

Solution: WMS-Service Reference Implementation

New: FOSS-based service layer

- **Ingest:** gdal/ogr (GMLAS - Geography Markup Language (GML) driven by application schemas)
- **Backend:** PostGIS
- **Service-Frontend**: MapServer (OGC-WxS)

Enabling FAIR reuse in Social Science desktop computing environents :

- R mit sf (Pebesma, 2018), tidyverse (Wickham, 2017)
- STATA
- QGIS

Quelle: Biljecki, Ledoux, and Stoter (2016)

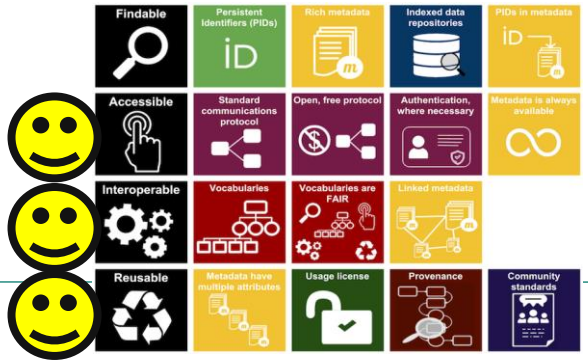
Results & Outlook

A reference web-service implementation built on open source software was implemented.

Survey studies at DIW/SOEP can now harness official geospatial building models to optimise annual survey planning and execution.

Todo:

- Public documentation of know-how about gdal/ogr-based wrangling for german-style CityGML data through OSGeo (umbrella foundation for the gdal/ogr-project).
- Extend FOSS-based reference implementation at DIW for WFS and WCS.
- Discuss with BKG options for an official web-based access to the data.



Sources

- **Census**
- https://www.zensus2011.de/DE/Home/Aktuelles/DemografischeGrunddaten.html

- **Georeferenced Adress Data**
- http://gdz.bkg.bund.de/index.php/default/georeferenzierte-adressdaten-ga.html

- **3D Building Models**
- https://gdz.bkg.bund.de/index.php/default/digitale-geodaten/sonstige-geodaten/3d-gebaudemodelle-lod1-deutschland-lod1-de.html

- Biljecki, Filip, Hugo Ledoux, and Jantien Stoter. 2016. “An Improved Lod Specification for 3D Building Models.” *Computers, Environment and Urban Systems* 59. Elsevier: 25–37. https://doi.org/10.1016/j.compenvurbsys.2016.04.005.

- Pebesma, Edzer. 2018. “Simple Features for R: Standardized Support for Spatial Vector Data.” *The R Journal* 10 (1): 439–46. https://doi.org/10.32614/RJ-2018-009.

- R Core Team. 2019. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

- Steinhauer, H. W. (2020). Nutzung geokodierter Zensus- und Gebäudedaten zur Ziehung und Gewichtung von Haushaltsstichproben. Vortrag auf dem 14. Workshop der Panelsurveys im deutschsprachigen Raum, Mannheim.

- Wickham, H. (2017). tidyverse: Easily Install and Load the 'Tidyverse'. R package version 1.2.1. https://CRAN.R-project.org/package=tidyverse

Vielen Dank für Ihre Aufmerksamkeit.

DIW Berlin — Deutsches Institut für Wirtschaftsforschung e.V.
Mohrenstraße 58, 10117 Berlin
www.diw.de

Redaktion
Peter Löwe (ploewe AT diw.de)

Tobias Gebel, Steinhauer