

# gesis

Leibniz Institute  
for the Social Sciences



## The Past, Present and Future of Geocoded Survey Data at the GESIS Data Archive

Stefan Müller | Wolfgang Zenk-Moeltgen | Stefan Schweers

{stefan.mueller|wolfgang.zenk-moeltgen|stefan.schweers}@gesis.org

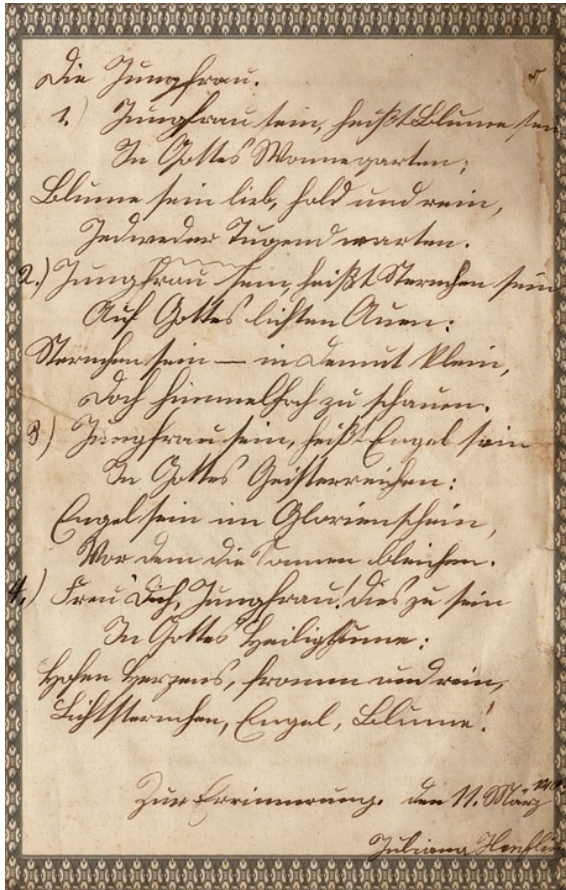


EDDI16 - 8<sup>th</sup> Annual European DDI User Conference, December 6-7, 2016



<http://creativecommons.org/licenses/by/4.0/>

# Disclaimer: What I won't present today



Once upon a time,  
 there were the Past,  
 Present and Future of  
 Geocoded Survey Data  
 at the GESIS Data  
 Archive.

The End.

## Because we are not alone

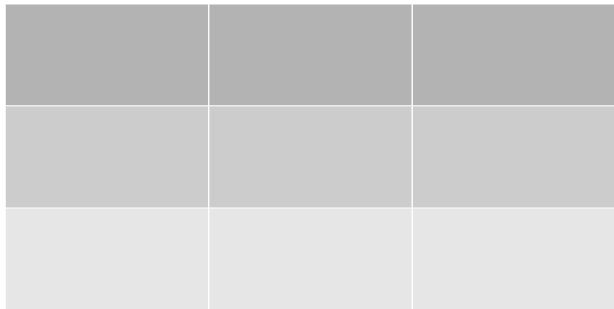
- Maybe Geocoded Data are a special case
- But it's always about adding observations or attributes



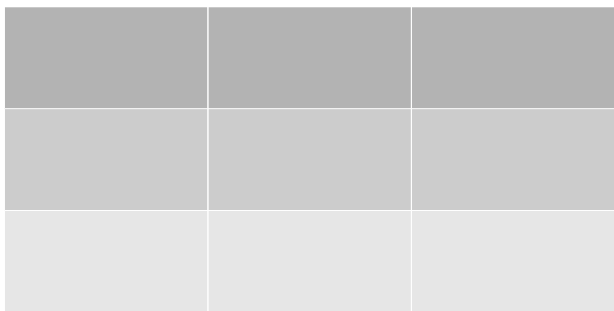
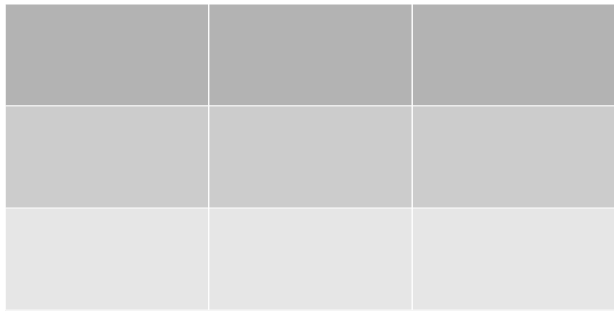
Source: [pixabay.com](https://pixabay.com)

→ **It's a Data Linking issue!**

# Adding observations and attributes

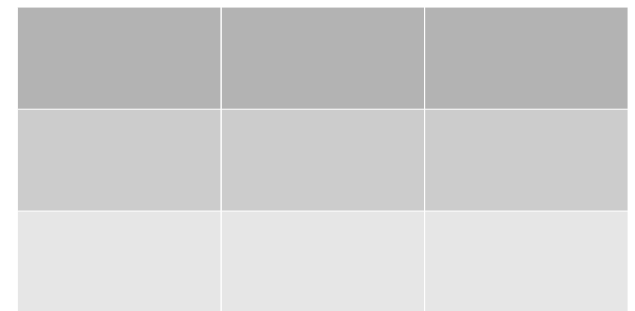
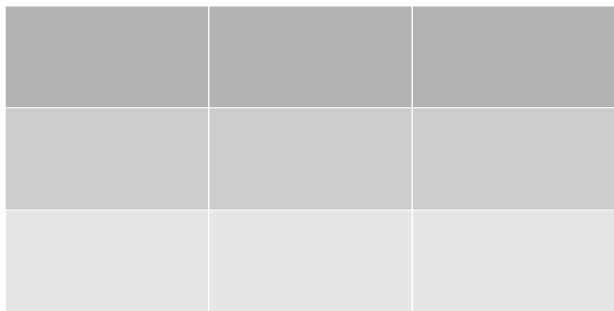



# Adding observations and attributes

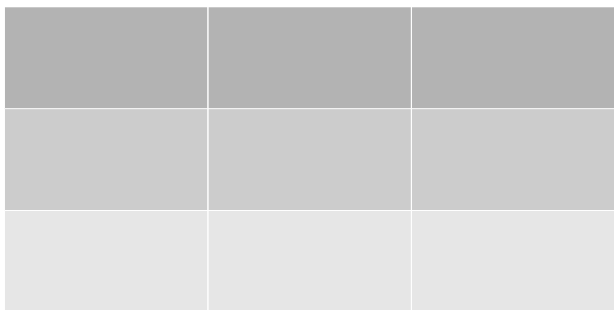


row-wise

# Adding observations and attributes

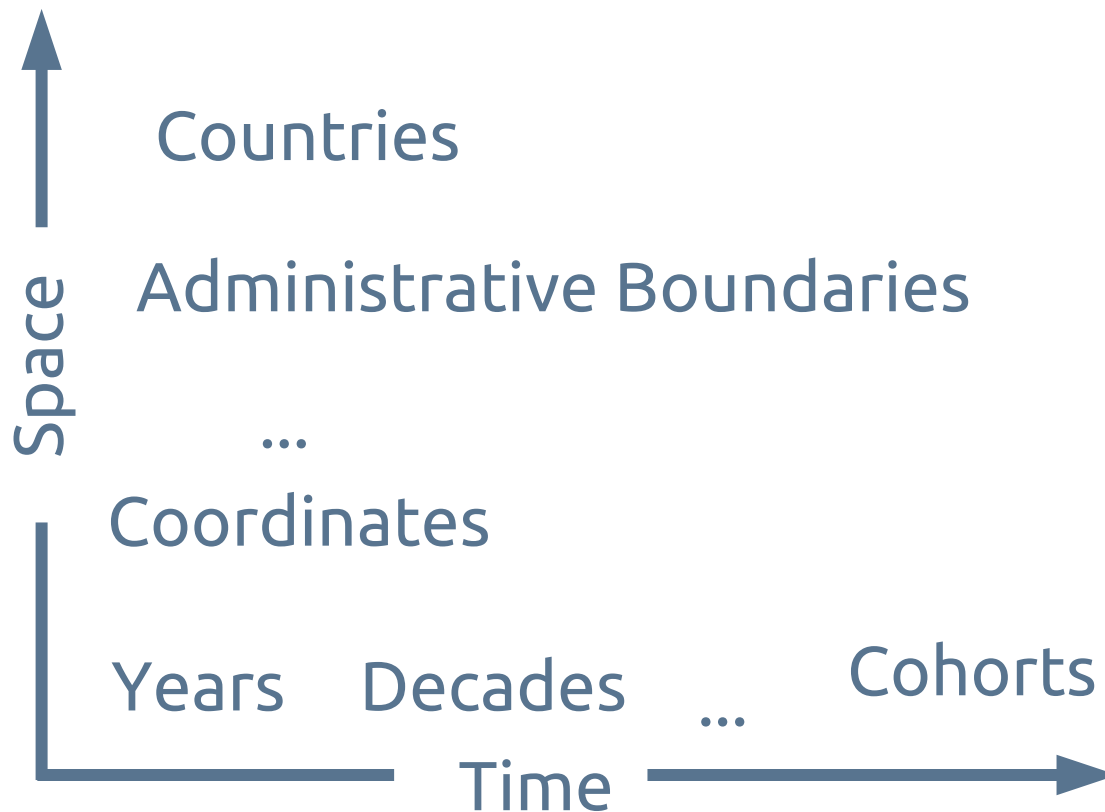


column-wise

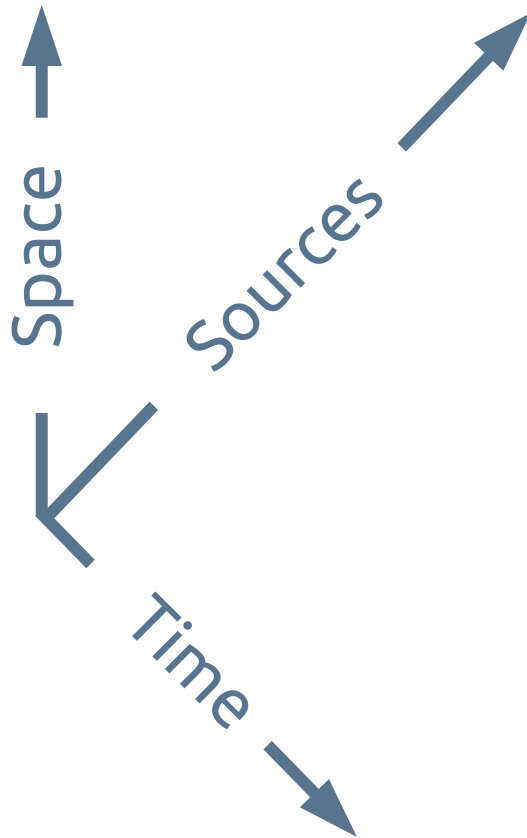


row-wise

# Data Linking across Space and Time



# Data Linking across Space and Time and different sources



Social Media Data

Historical Data

Experimental Data

**Geospatial Data**

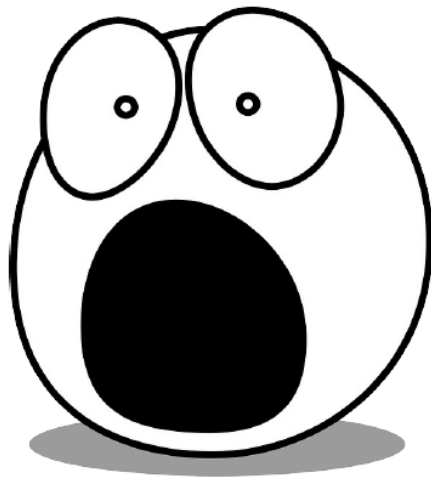
...

Microdata Linkage



# Yet: Back to EDDI 2015

gesis

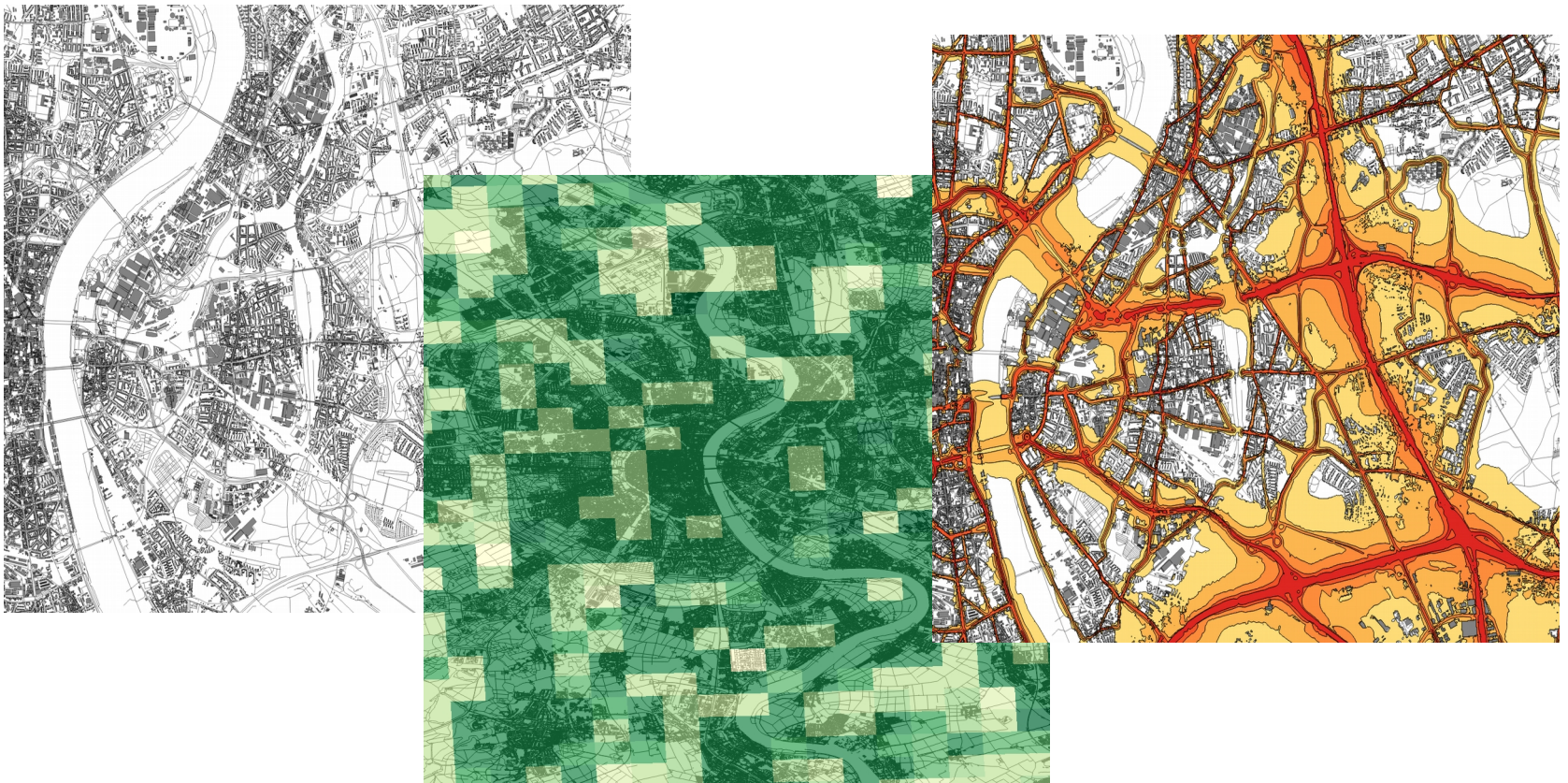


Source: pixabay.com

**New data types,  
formats, contents  
etc.**

**...new documentary  
demands!**

# Indeed, geospatial data are complicated



## In need of documentary workarounds

- There is no easy way to document geographic structures (compatible with ISO 19115) on the variable level in DDI Lifecycle
- But the data are distributed via secure access anyway
  - It therefore supports our distribution process to logically split the data

# Again: Back to EDDI 2015

gesis

## '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

# Again: Back to EDDI 2015

gesis

## 'Whole' story II

- Create one study unit for each geographic structure

**WE WILL PUBLISH SOON**

**SU I:**

'Ordinary'  
survey  
variables

**SU II:**

Variables  
extracted  
from rasters

**SU III:**

Variables  
extracted  
from polygons

## That's it?

- No.
  - Metadata such as ISO 19115 are way too integrated in spatial data infrastructures (SDI) to not automate as much as possible to make our lives easier
- We want to build our own SDI!

gesis

# Georeferencing of Survey Data

gesis

Leibniz Institute  
for the Social Sciences

ALBUS



Pilot project for geocoding social science survey data and linking to small scale spatial data addressing technical, documentary and data protection challenges

gesis

# SORA

## Social-Spatial Science Infrastructure



Leibniz Institute of  
Ecological Urban and  
Regional Development



Karlsruhe Institute of Technology

# SOEP

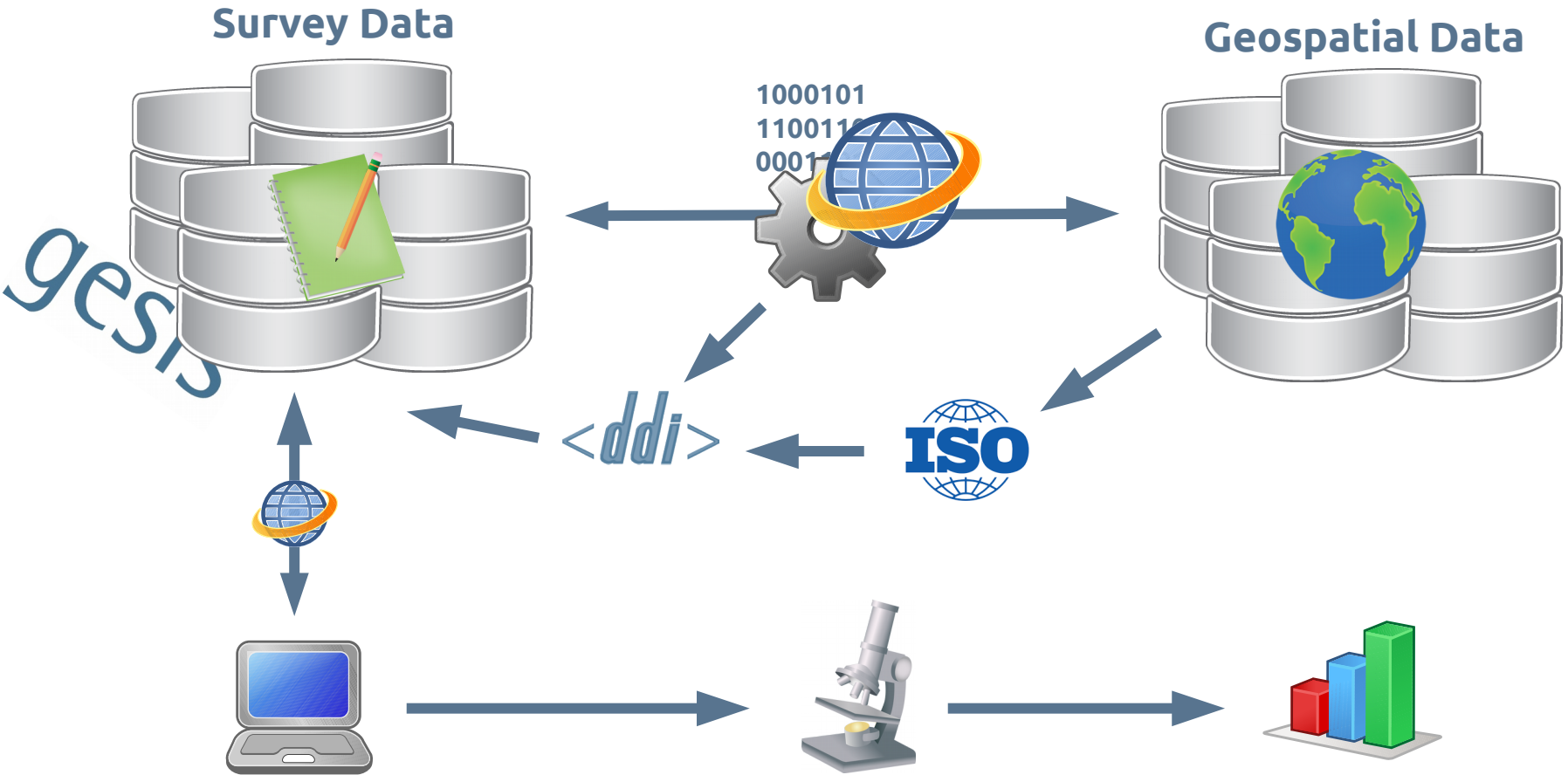
# gesis

Leibniz Institute  
for the Social Sciences

Support interdisciplinary research by making the research data infrastructures of spatial sciences and social sciences interoperable



# An SDI for Social Science Survey Data



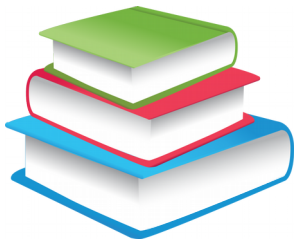
# Prospective and conclusive thoughts

- Documentation issues concern all projects that are engaged in data linking
- Workarounds are unsatisfying
  - Then again, it's working
- Let's automate as much as possible



Source: [pixabay.com](https://pixabay.com)

## Further information



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* 12 (1): 100-126. <http://dx.doi.org/10.1080/15420353.2015.1100152>

&

**GitHub**

<https://github.com/stefmue/georefum>

&

gesis

Leib  
for i

<http://www.gesis.org/forschung/drittmittelprojekte/projektuebersicht-drittmittel/georefum/>