

# DATA MANAGEMENT: GEOCODED UNIT RECORD DATA – CONFIDENTIALITY MANAGEMENT AND GEOCODED UNIT RECORD DATA: LESSONS FROM GERMANY

Dr. Susanne Schnorr-Bäcker

Federal Statistical Office of Germany

UN Global Forum of Statistical and Geospatial Information

4-5 August 2014, New York

# Outline

1. About Official Statistics Germany – basics
2. Regional statistics – types of presentation
3. Grid based mapping – possibilities and limitations
4. Comparison of administrative and grid based data
5. Political impact
6. Conclusions



# 1. About Official Statistics Germany – basics

- (1) Every statistical value is related to a point in time or timespan and a specific territory (the nation itself or a locality)**
- (2) Official Statistics provides a lot of data for heterogenous user groups and political and other purposes**
- (3) Official Statistics uses various data sources**
- (4) Surveys for Official Statistics require always a legal basis**
- (5) Confidentiality of statistical data has always to be guaranteed**
- (6) Statistical data are produced jointly by the statistical offices of Länder and Federal Statistical Office according to UN's GSBPM**

# Sources of Official Statistics

## Primary

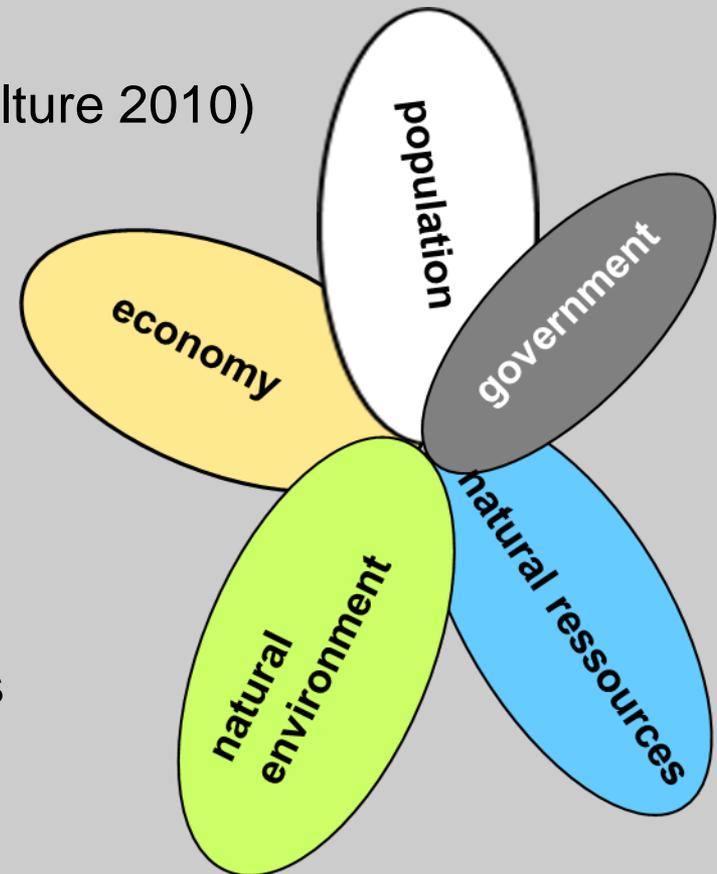
- Census (population 2011, agriculture 2010)
- Sample surveys (LFS, SBS)

## Secondary

- Federal Employment Agency
- Fiscal authorities
- Central Bank
- .....

## Registers

- Hospitals and rehabilitation units
- Schools
- Enterprises
- .....



# Regional statistical units in official statistics

- **Administrative districts/area**  
different states' levels  
e.g. municipalities
- **Non-administrative areas**  
e.g. employment office districts,  
water catchment areas
- **Points**  
e.g. street car accidents
- **Lines**  
e.g. motorways, rivers
- **Grids**  
e.g. agricultural census

Image: s.schnorr-baecker

## 2. Regional Statistics – types of presentation

### Federal Statistical Office

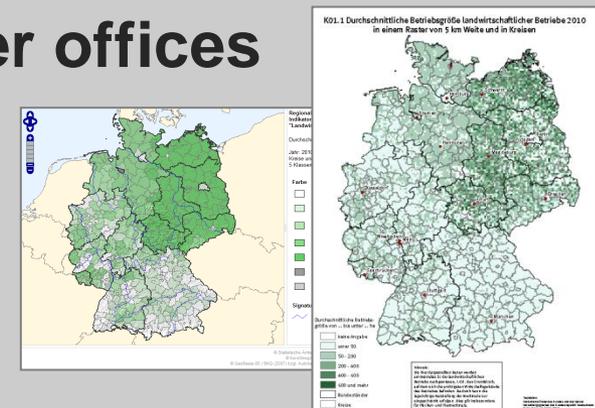
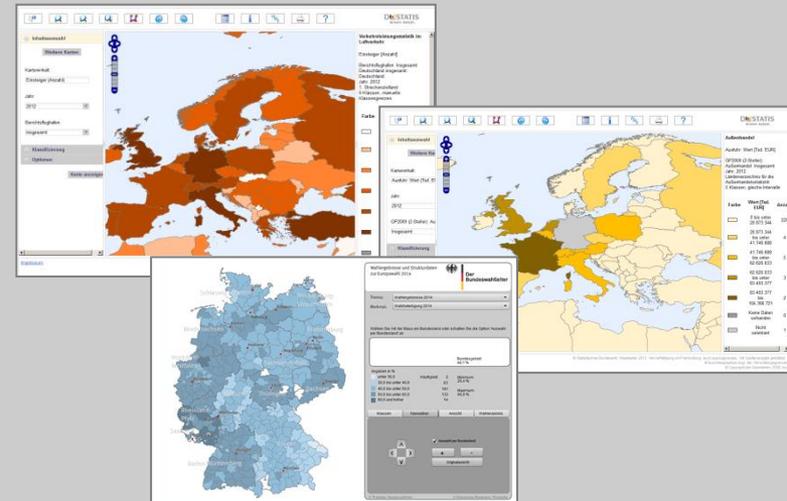
- air traffic
- foreign trade

### Bundeswahlleiter

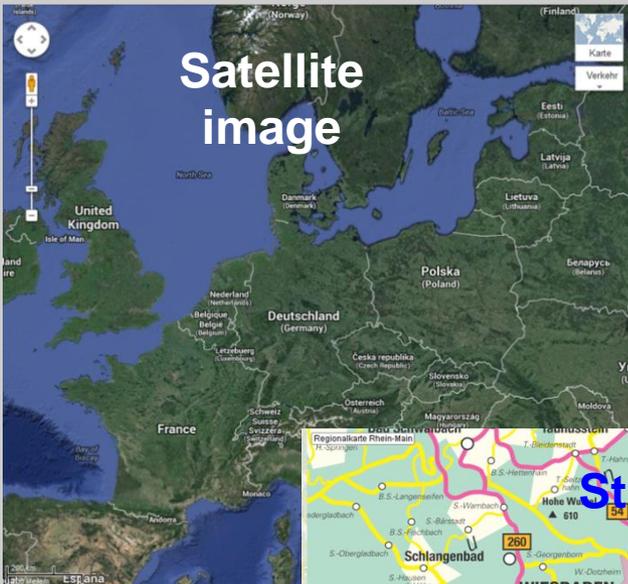
- federal elections
- elections for European Parliament

### Joint publications together with Länder offices

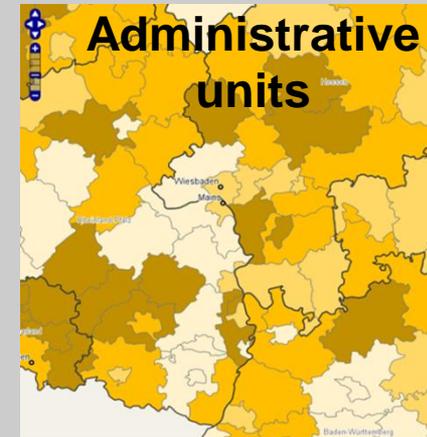
- regional atlas
- *agricultural atlas (grids)*
- *population census 2011- Living in Germany*
- *street car accidents (planned)*



# Real world from different perspectives



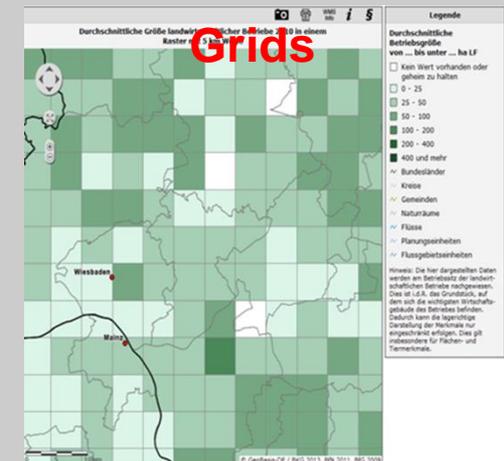
[https://maps.google.de/maps?f=q&source=s\\_q&hl=de&geocode=&authuser=0&q=google+street+view&aq=&vps=2&jsv=485b&vpsrc=0&t=h&ie=UTF8&hq=http://maps.google.com/intl/de/help/maps/streetview/mapleft.kml&ct=onebox&cd=3&cad=onebox\\_streetview](https://maps.google.de/maps?f=q&source=s_q&hl=de&geocode=&authuser=0&q=google+street+view&aq=&vps=2&jsv=485b&vpsrc=0&t=h&ie=UTF8&hq=http://maps.google.com/intl/de/help/maps/streetview/mapleft.kml&ct=onebox&cd=3&cad=onebox_streetview)



Source: <https://www-genesis.destatis.de/gis/genView?GenMLURL=http://www-genesis.destatis.de/regatlas/ai002-1.xml&CONTEXT=REGATLAS01>



See: Kartenwerke der Landeshauptstadt Wiesbaden;  
<http://maps.wiesbaden.de/cityguide/html/de/1280x1024.html>

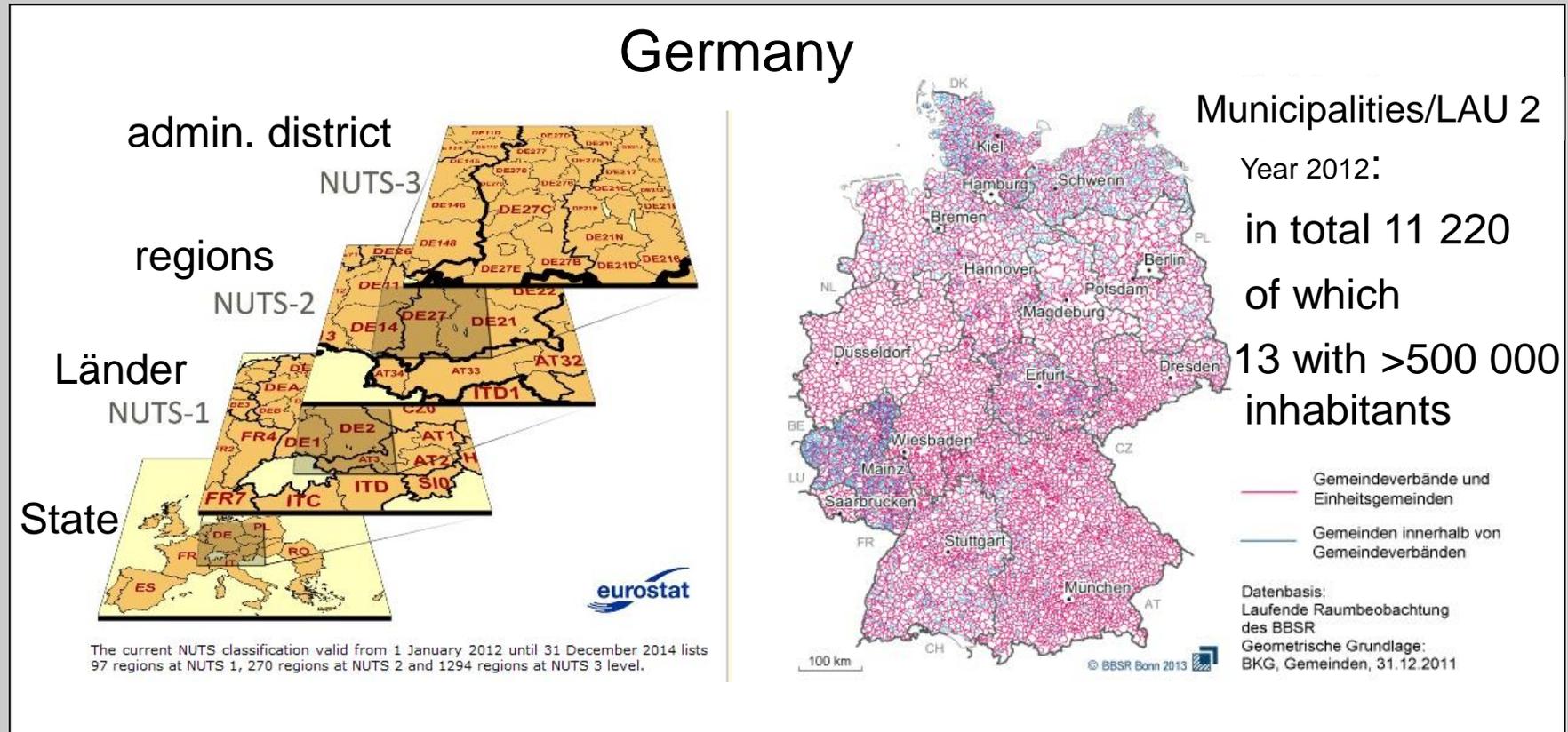


Source: <http://www.atlas-agrarstatistik.nrw.de/>

# Requirements for producing administrative regional statistical data

- thematic issue (person, company, others)
- regional identifiers such as
  - post code
  - borough code (Gemeindekennziffer)
  - city / part of a city
  - block side/census district
  - polling districts
- time stamp concerning
  - date of collection and corresponding statistical unit
  - period for which results are valid in line with the validity of the regional unit's boundaries

# Administrative Units & Regional Classifications



Sources: [http://epp.eurostat.ec.europa.eu/portal/page/portal/nuts\\_nomenclature/introduction](http://epp.eurostat.ec.europa.eu/portal/page/portal/nuts_nomenclature/introduction);  
[http://www.google.de/imgres?imgurl=http://www.bbsr.bund.de/nn\\_1112616/BBSR/DE/Raumbeobachtung/Raumabgrenzungen/Gemeinden/Karte\\_\\_GemVBG10,property%3Dposter.gif&imgrefurl=http://www.bbsr.bund.de/nn\\_1067638/BBSR/DE/Raumbeobachtung/Raumabgrenzungen/Gemeinden/gemeinden\\_node.html%3F\\_nnn%3Dtrue&usq=\\_\\_l8FD9BOBT3dqf4kSY5DUUDeBiQc=&h=681&w=800&sz=155&hl=de&start=15&zoom=1&tbnid=VxCzXYwW2QOwWM:&tbnh=122&tbnw=143&ei=k6K1UaHcNoKR4ATS4IDQAg&prev=/search%3Fq%3Dgemeindekarte%2Bstatistisches%2BBundesamt%26um%3D1%26hl%3Dde%26tbn%3Disch&um=1&itbs=1&sa=X&ved=0CEgQRQMwDg](http://www.google.de/imgres?imgurl=http://www.bbsr.bund.de/nn_1112616/BBSR/DE/Raumbeobachtung/Raumabgrenzungen/Gemeinden/Karte__GemVBG10,property%3Dposter.gif&imgrefurl=http://www.bbsr.bund.de/nn_1067638/BBSR/DE/Raumbeobachtung/Raumabgrenzungen/Gemeinden/gemeinden_node.html%3F_nnn%3Dtrue&usq=__l8FD9BOBT3dqf4kSY5DUUDeBiQc=&h=681&w=800&sz=155&hl=de&start=15&zoom=1&tbnid=VxCzXYwW2QOwWM:&tbnh=122&tbnw=143&ei=k6K1UaHcNoKR4ATS4IDQAg&prev=/search%3Fq%3Dgemeindekarte%2Bstatistisches%2BBundesamt%26um%3D1%26hl%3Dde%26tbn%3Disch&um=1&itbs=1&sa=X&ved=0CEgQRQMwDg)

# Cooperation with German Federal Agency for Cartography and Geodesy (BKG)

In particular:

- ✓ **Digital Basic Landscape Model**
- ✓ **Digital Orthophotos Ground Resolution 20 cm (DOP 20)**
- ✓ **WMS: DOP20**
- ✓ **Georeferenced Address Data**
- ✓ **Building Polygons**

### 3. Grid based mapping possibilities and limitations

#### Georeferencing

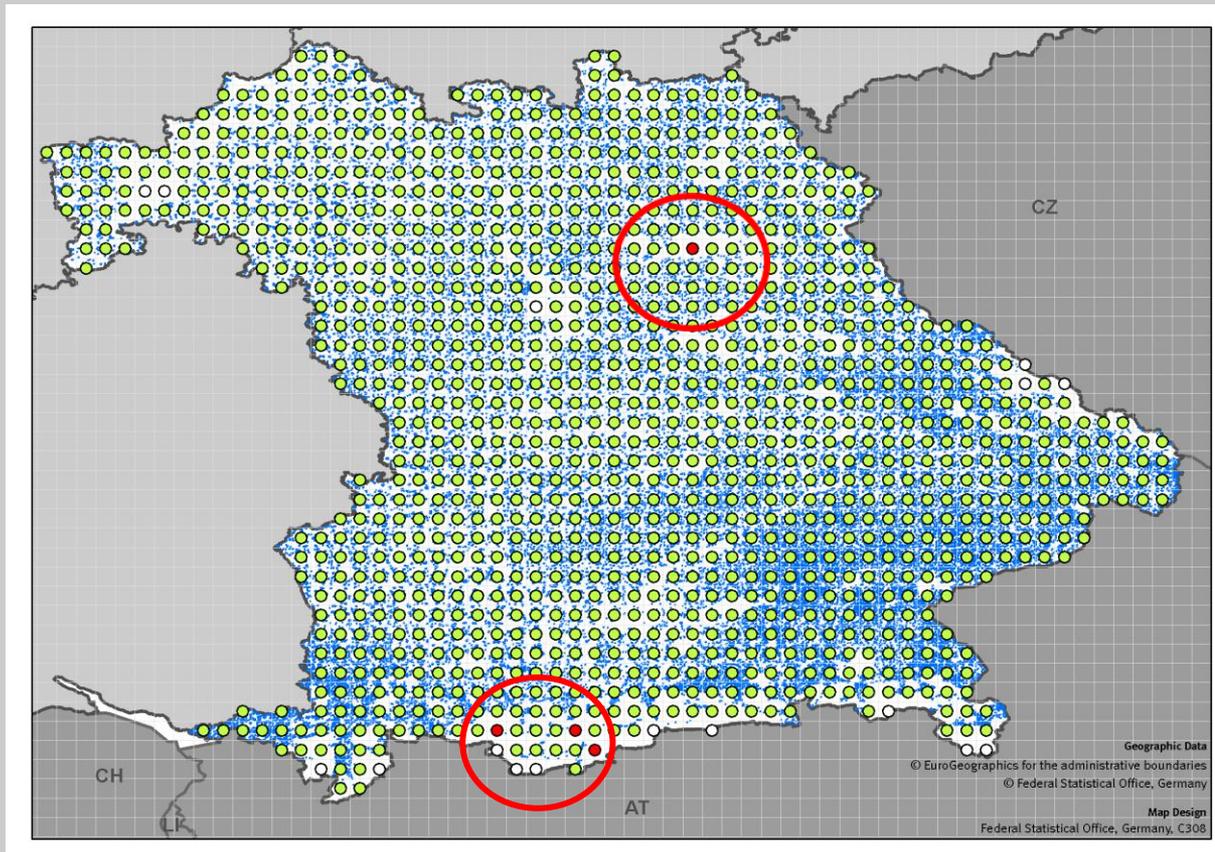
differs according to thematic issues  
(for privacy/confidentiality reasons)

- individual information about people can be saved in 100m\*100m grids (1 ha)
- economic and environmental facilities can be geocoded in registers

for data releases and publications appropriate measures are necessary, i.e.

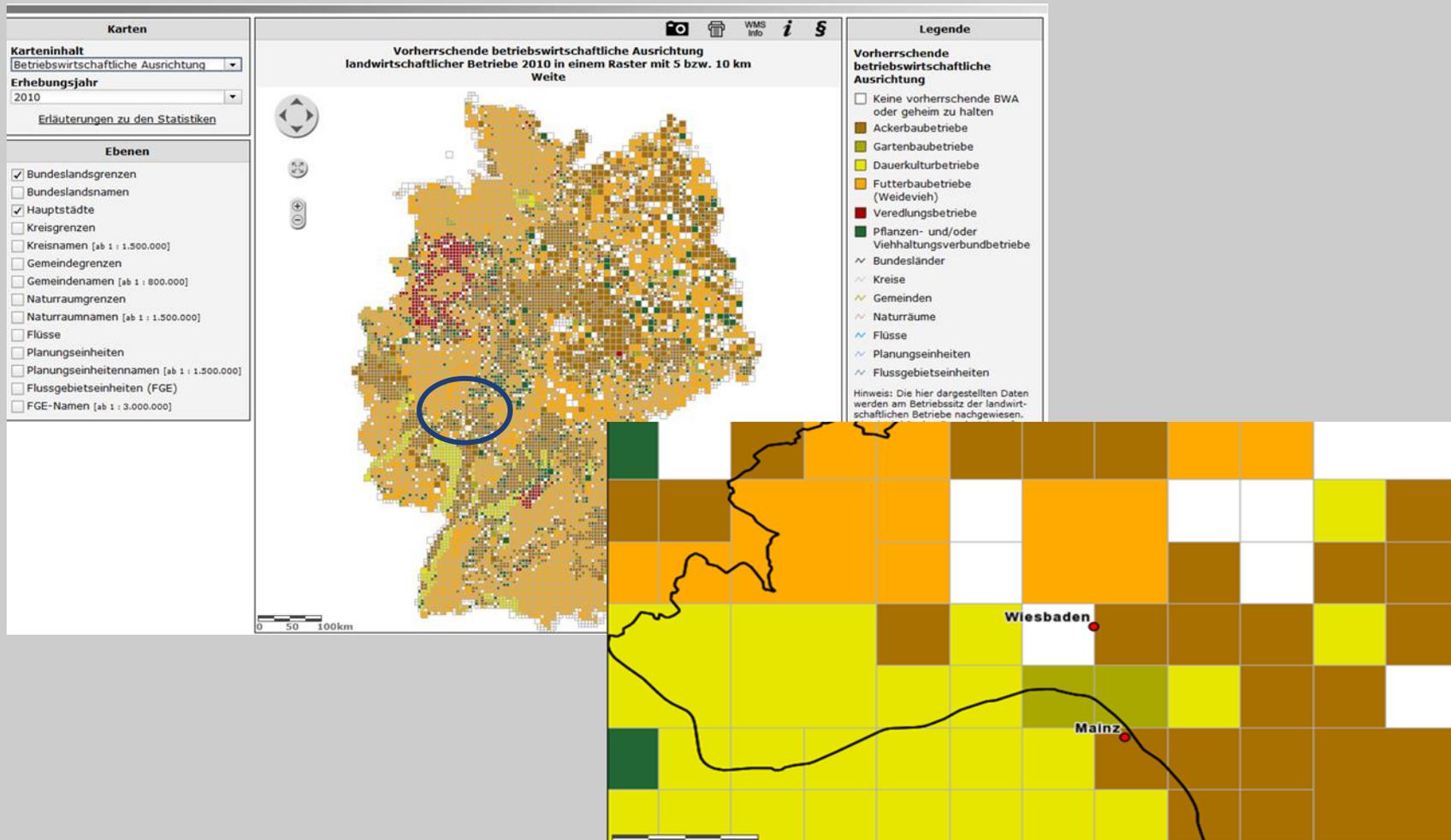
- for choosing an appropriate frequency distribution, in particular the size classes
- data protection is required for confidentiality reasons

# Confidentiality problems\*, here single farms



Agricultural Census:  
data delivery for  
Eurofarm for Bavaria,  
\*Grid size based on 5  
arc minutes

# Grid size and confidentiality



# Grid based data - Pros and Cons -

## Pros

- more detailed regional information
- stable over time (changes of boundaries do not matter)
- possibilities of matching with other layers
- easy to collect

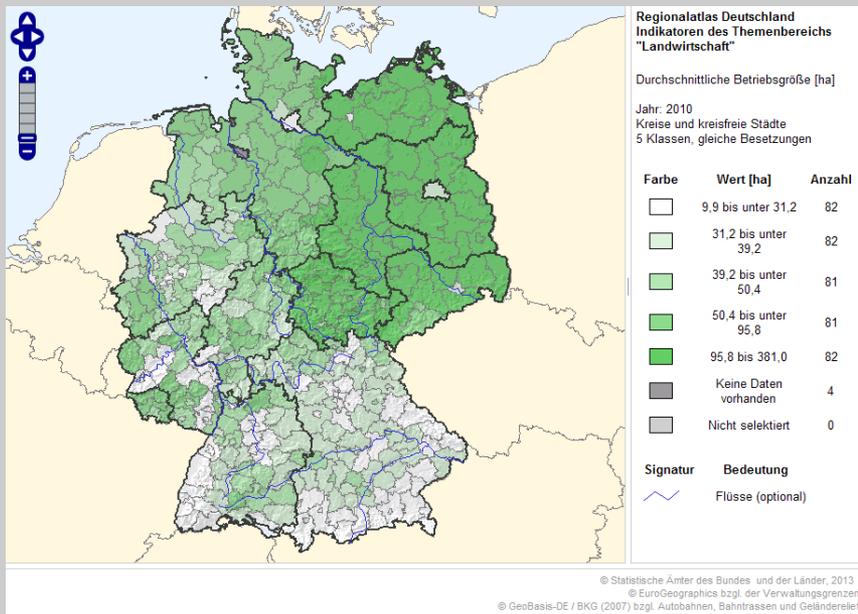
## Cons

- higher initial effort (reviewing of addresses)
- limited regional accuracy (1 pixel  $\neq$  one object)
- high data volumes
- not suited for regional analyses at exact parcel level

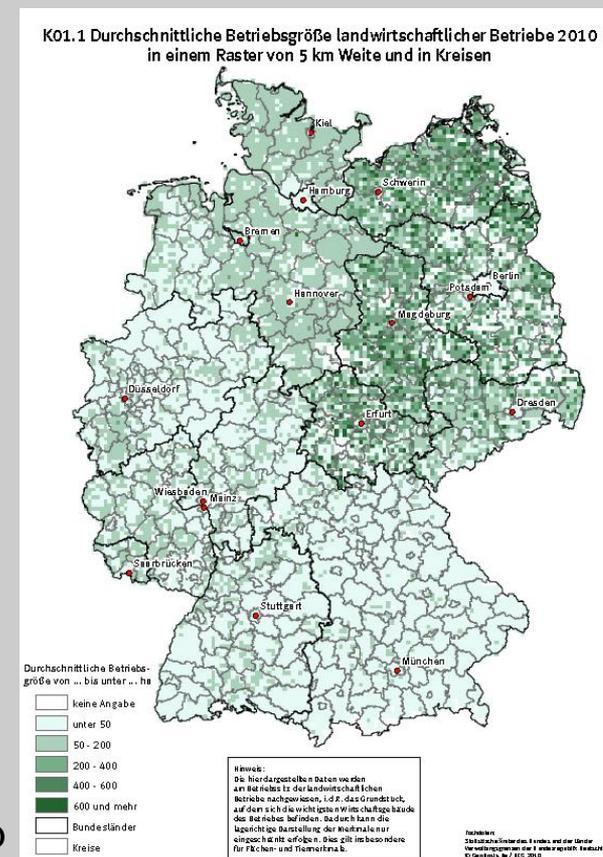
# 4. Comparison of administrative and grid based data

## Grid based map for average farm size in ha

### Average farm size in ha at NUTS 3

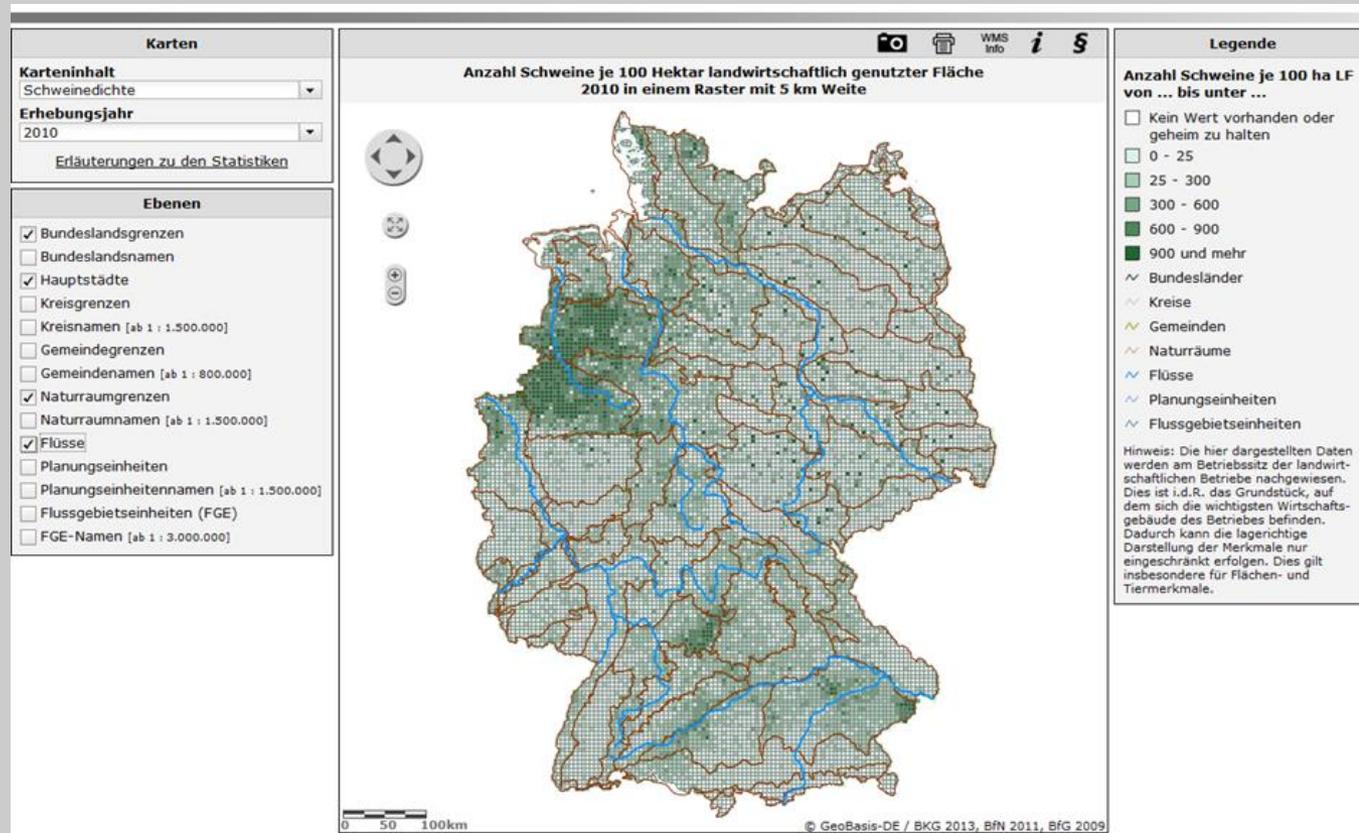


administrative units



grid based map

# Grid based maps with different layers

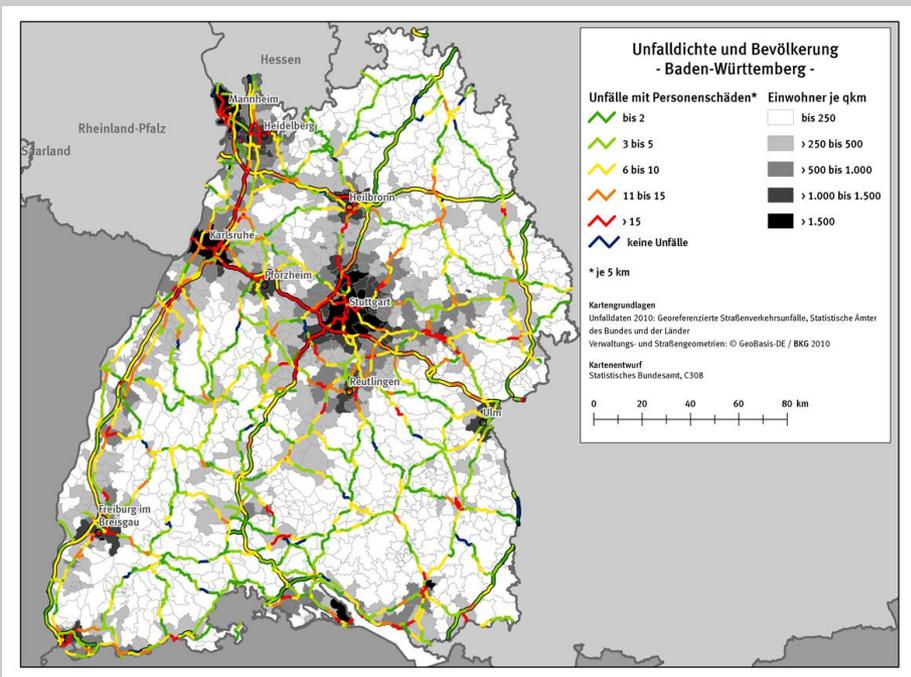
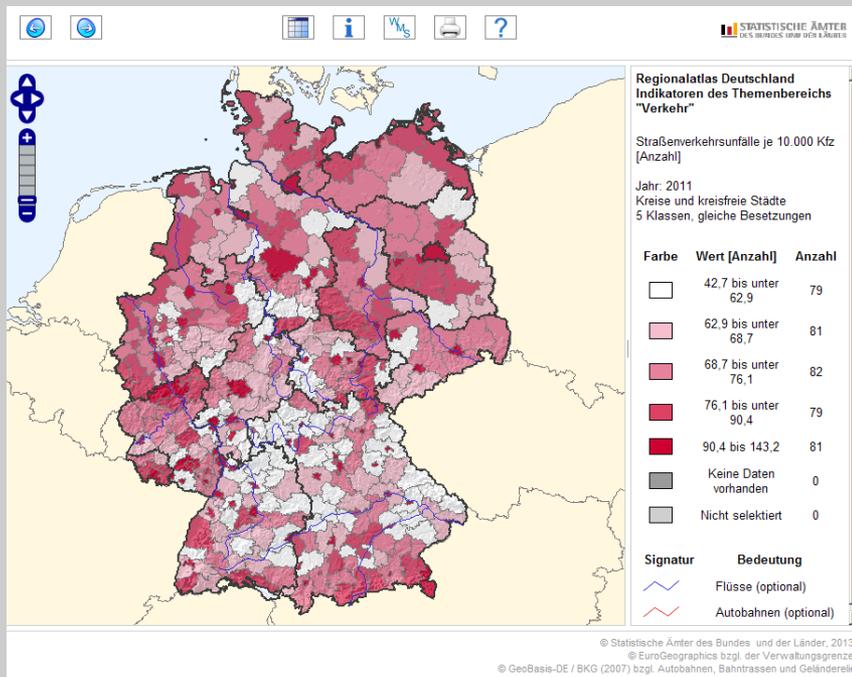


- e.g.
- rivers
  - administrative districts
  - natural spaces

# Others sorts of mapping for points and lines

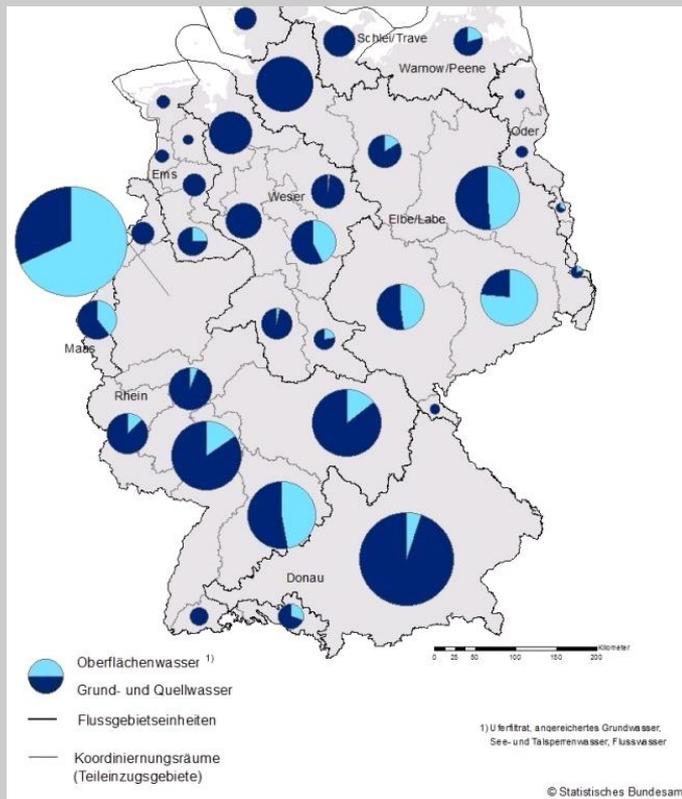
Street car accidents per 10,000 street car vehicles

Street car accidents based on geocodes

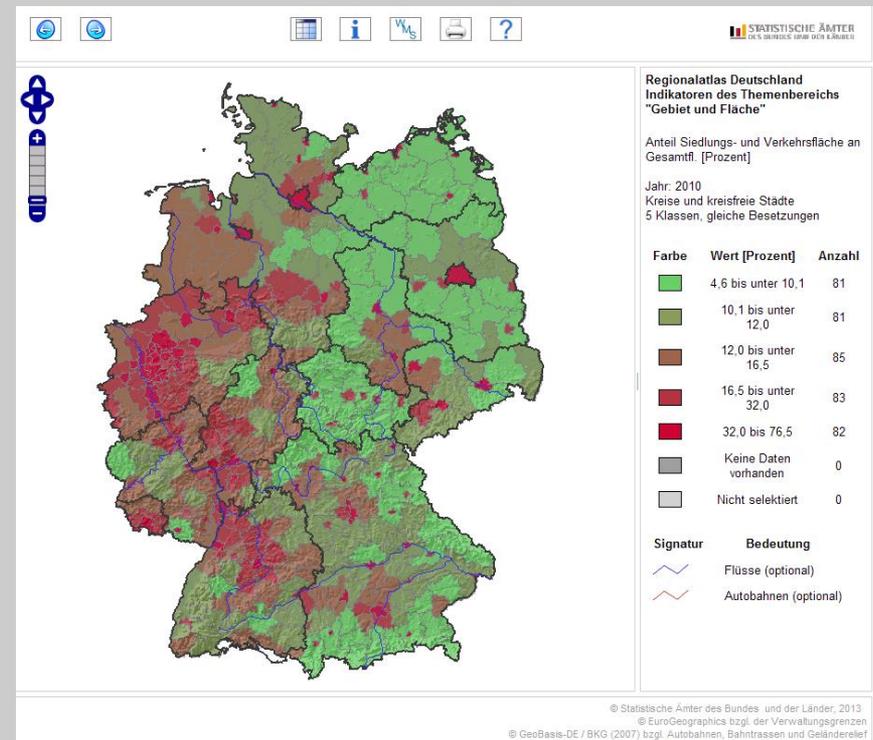


## Catchment areas

Production of ground and surface water 2010 according to coordination areas (sub basin-districts of river basin districts of water frame directive 2000/60/EC)



## Settlement and traffic areas



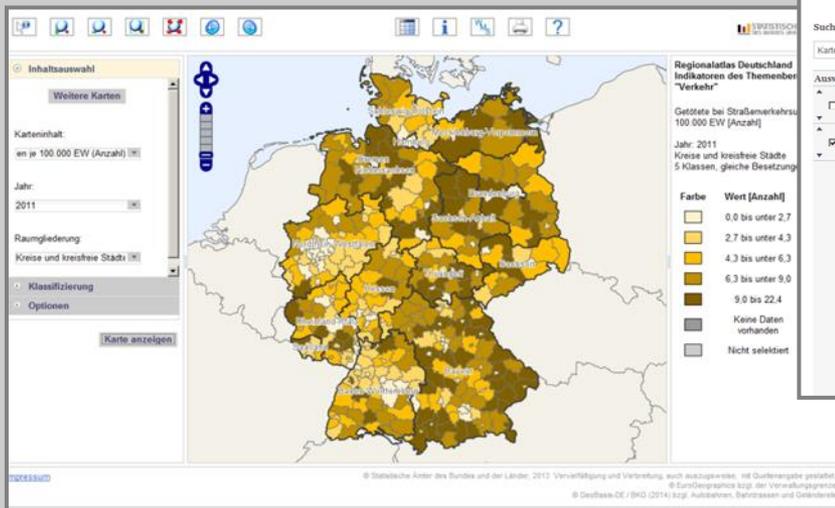
Source: Federal Statistical Office, Environmental Statistics/ G 202; mapping: Bundesanstalt für Gewässerkunde (BfG)

# 5. Political Impact

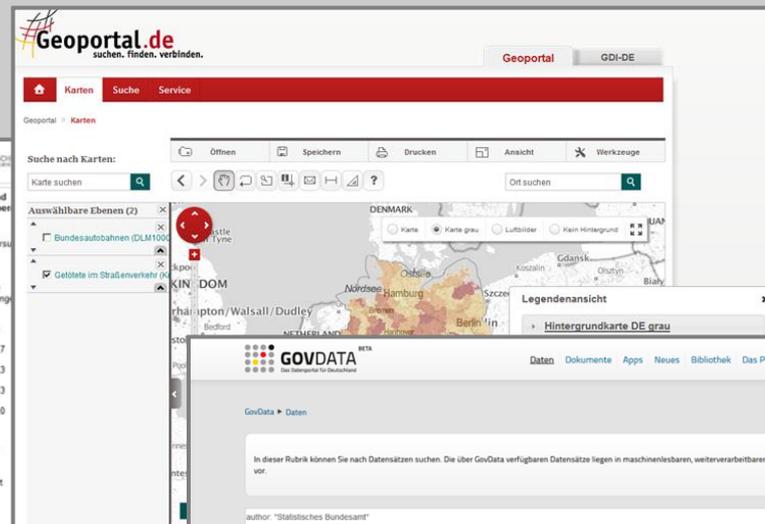
## Official Statistics' maps & other internet applications

persons killed in a road accident

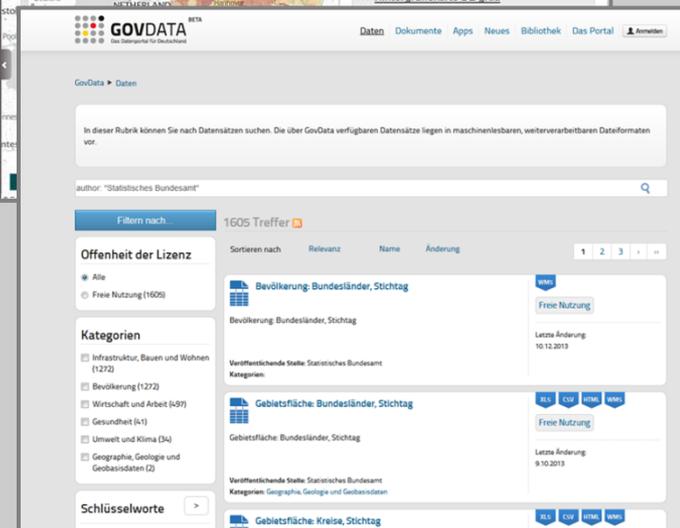
GDI-DE



regional atlas



GOVDATA



# Evidence based monitoring for politics, i.e. progress measurement resp. sustainable development

## Germany

### Sustainable Development in Germany (since 2002)

### Enquetekommission „Growth, Wealth and Quality of Life“ (2011-2013)

### Government commission „Gutes Leben“ (23.1.2014)

## Inter- and supranational level

### „Europa 2020“ (EU)

### „Better Life Initiative“ (OECD)

### „MDG post 2015“ (UN)



<https://www.destatis.de/EN/FactsFigures/Indicators/SustainableDevelopmentIndicators/SustainableDevelopmentIndicators.html>



[http://www.nachhaltigkeitsrat.de/uploads/media/Broschüre\\_Staedte\\_fuer\\_ein\\_nachhaltiges\\_Deutschland\\_texte\\_Nr\\_36\\_Juni\\_2011.pdf](http://www.nachhaltigkeitsrat.de/uploads/media/Broschüre_Staedte_fuer_ein_nachhaltiges_Deutschland_texte_Nr_36_Juni_2011.pdf)



# At-risk of poverty – different types of representation

Fig. 7a At-risk-of-poverty rates<sup>1</sup> by selected large towns 2011 ...  
... measured against the federal median

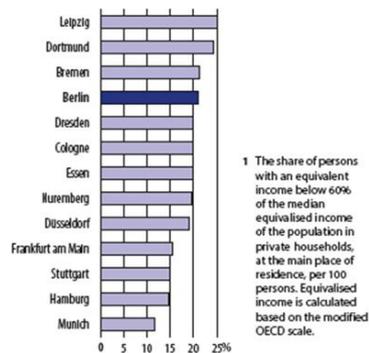
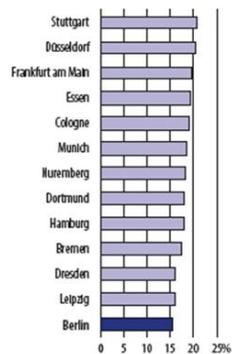


Fig. 7b ... measured against the median for each large town



Source: Social reporting 2012

Fig. 8 Regional atlas Germany Indicators relating to wealth and income



Disposable income per inhabitant 2009  
Administrative districts and towns not attached to an administrative district.  
5 classes, equal intervals

Income Class (EUR)	Number
13,895 EUR to under 17,320 EUR	119
17,320 EUR to under 20,745 EUR	228
20,745 EUR to under 24,170 EUR	57
24,170 EUR to under 27,595 EUR	5
27,595 EUR to under 31,020 EUR	3

Source: Schnorr-Bäcker, S.; Bömermann, H.: Regional statistical data on measuring progress The OECD approach “How’s Life? – Measuring Well-Being”, Journal of Official Statistics Berlin Brandenburg 2/2013, pp. 30

[http://www.statistik-hessen.de/fileadmin/media/files/infomaterial/e\\_reg\\_statistic\\_data\\_measuring\\_progress.pdf](http://www.statistik-hessen.de/fileadmin/media/files/infomaterial/e_reg_statistic_data_measuring_progress.pdf)

## 6. Conclusions

- more initial effort in data production and publication for grid based data is needed
- grid based statistical data provide more detailed insight in regional developments
- grid based data are much better suitable for political based monitoring strategies (such as comprehensive SDI strategies and / or specific policies)
- grid based data as a basis for new data sources such as big data (from internet, from sensors or even satellite images)

# Thinking global acting local!



Image: s.schnorr-baecker

# Thank you for listening!

## Questions?



Dr. Susanne Schnorr-Bäcker  
Telefon: +49/(0) 611 / 75 20 82  
susanne.schnorr-baecker@destatis.de  
[www.destatis.de](http://www.destatis.de)