DESTATIS
Statistisches Bundesamt

INNOVATIVE DATA COLLECTION USING BIG DATA SOURCES FOR OFFICIAL STATISTICS

Natalie Rosenski, Maren Köhlmann (Destatis)

UNECE Workshop on Statistical Data Collection

14 October 2019, Geneva

Content

Background

Improvement of timeliness, accuracy and relevance

Reduction of response burden

→ Use of new digital data for official statistics

→ Diverse data sources

- Mobile Network Data
- Remote Sensing Data
- Data from the Internet of Things (IoT)

© Gerd Altmann, Pixabay

ESSnet Big Data II – WP L

12 partner from 11 countries: AT, BG, FI, FR, IT, NL, NO, PL, PT, UK& DE

Germany (Destatis) in lead for WP L “Preparing Smart Statistics”

Duration: November 2018 – October 2019

Goals:

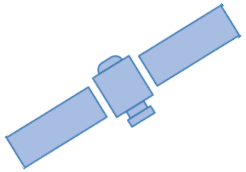
- Exploration of the IoT in order to produce trusted smart statistics
- Overview of relevant topics for official statistics
- Recommendations for possible follow-up studies

ESSnet Big Data II – WP L

4 Tasks

- Smart Farming
- Smart Cities
- Smart Devices
- Smart Traffic

DIGITALSTADT DARMSTADT



Remote Sensing Data

DLR

Federal Agency for Cartography and Geodesy

Smart Business Cycle Statistics

- Flash estimates of economic indicators

Makswell

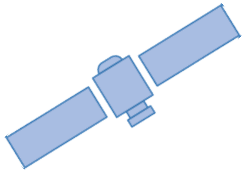
- MAKing Sustainable development and WELL-being frameworks work for policy analysis

Deep Solaris

- (Semi-)automatic analysis of satellite and aerial images for energy system transformation and sustainability indicators

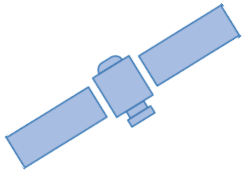
ESSnet Big Data II – WP H "Earth Observation"

- Linking remote sensing data to survey data



Remote Sensing Data - Challenges

- Frequency of images and weather-conditions
- Radar images vs. optical satellite images
- High resolution data only commercially available and costly
- Low coverage of high resolution images



Remote Sensing Data – Example: ESSnet Big Data II

ESSnet Big Data II - WP H “Earth observation”

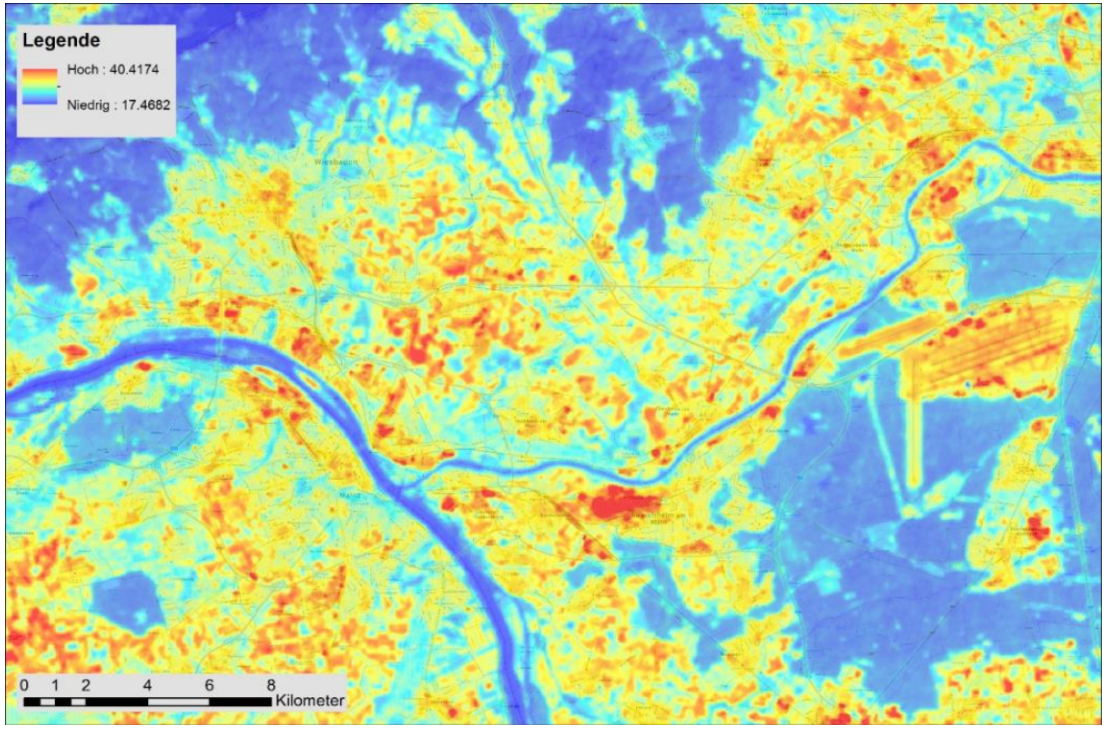
→ Analysis of quality of life indicators

- Indicators with spatial dimension
  - Urban heat islands, urban greens, noise pollution, air quality
- Linking neighbourhood characteristics derived from EO data to socio-demographic characteristics from the (micro) census
  - Inhabitants, age structure, civil status, income, education

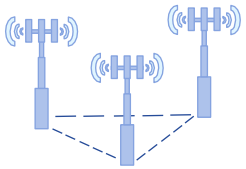
Remote Sensing Data – Example: ESSnet Big Data II

Surface temperature – heat islands

© Google Maps, Rhine-Main region



© Calculations by BKG using Landsat 8 data



Mobile Network Data

ESSnet Big Data II: WP I “Mobile Network Data”

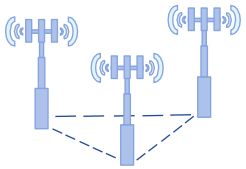
- Dynamic representation of the population via mobile network data

Comparison of static mobile network data with census data

- To represent day and resident population

Exploration of dynamic mobile network data for commuter statistics

- To illustrate commuter flows
- Compare data with census 2011 and existing commuter atlas

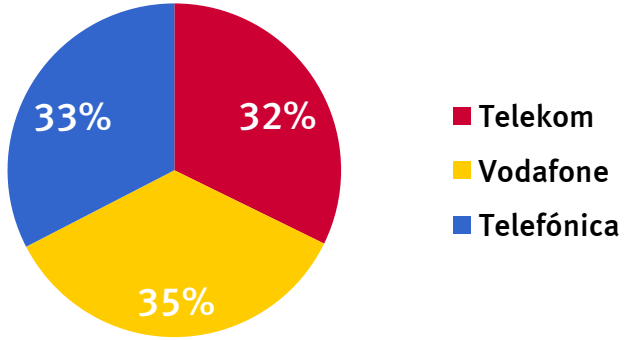


Mobile Network Data

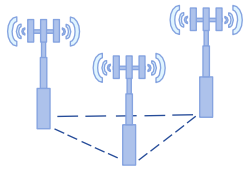
Anonymised and aggregated mobile network data

- Minimum activity per grid≥ 30 (lately ≥ 5)
- Test data of one German federal state for a statistical week
- Dwell time › 2 hours
- Specific geometry
- Demographic characteristics: age group, gender, mobile country code

Market share of mobile network operators (Germany, 2nd quarter 2019)

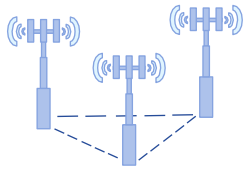


Data source: Federal Network Agency



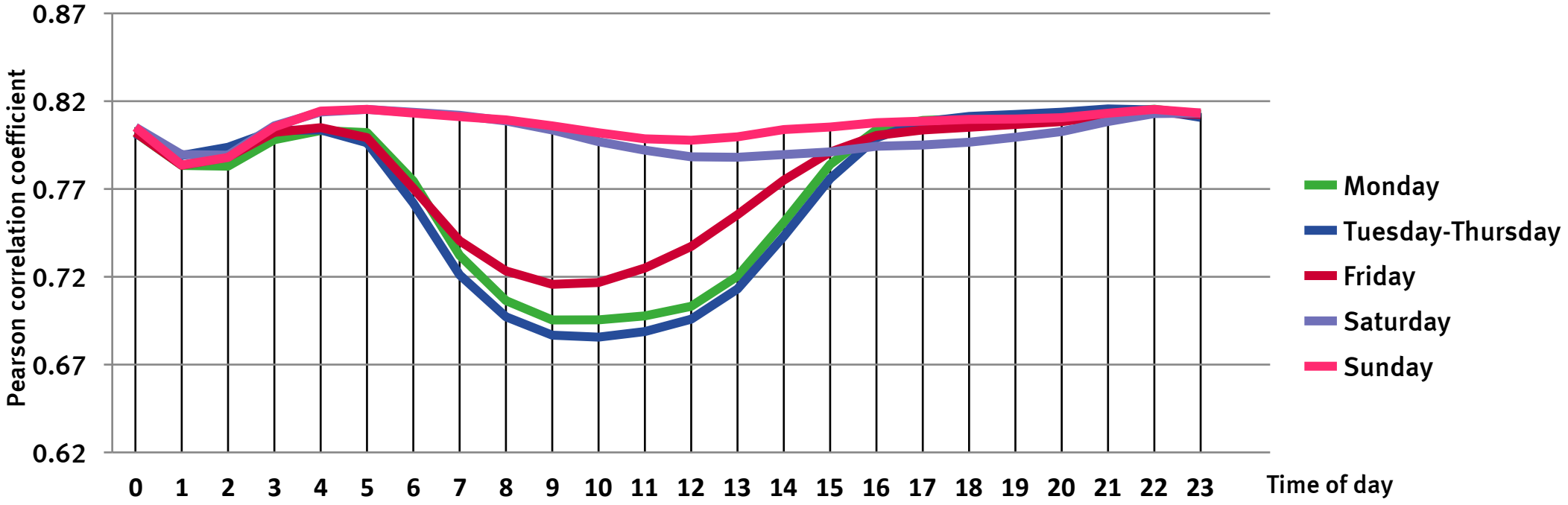
Mobile Network Data - Challenges

- No legally regulated permanent dataaccess
- No distinction between personal and device-related SIMcards
- Socio-demographic characteristics only for contractual customers
- Bias due to double counting of persons possessing several mobile phones and family tariffs
- Insufficient disclosure of the extrapolation methodology used by providers
- Available periods of mobile network data and census data do not always match well enough
- etc.



Mobile Network Data – Example

Illustration of the population via mobile network data



Graph: Pearson correlation coefficients for census values and mobile network data by statistical days and time periods

Outlook

- → Combination of survey, administrative and new digitaldata
- → Very promising to enrich the available official data
  - » Improve quality of data
    - » Accuracy: Provide more detailed data for policy decisions
    - » Timeliness: Improve punctuality of official statistics
    - » Relevance: Develop new research areas and questions
  - » Reduce response burden → reduction of costs ?

THANK YOU FOR YOUR ATTENTION!

Institute for Research and Development in Federal Statistics

newdigitaldata@destatis.de

Maren Köhlmann
+49 (0) 611 / 75 42 95
Maren.koehlmann@destatis.de

© Fancy by Veer/The World from Above/FAN2043621