DESTATIS
wissen.nutzen.

COMMON ISSUES ON BENEFITS AND CHALLENGES OF BIG DATA SOURCES

Dr. Susanne Schnorr-Baecker
Federal Statistical Office of Germany

International Conference on Big Data for Official Statistics
28-30 October 2014 Bejing, China

Official Statistics Germany – Some Basics

Aims

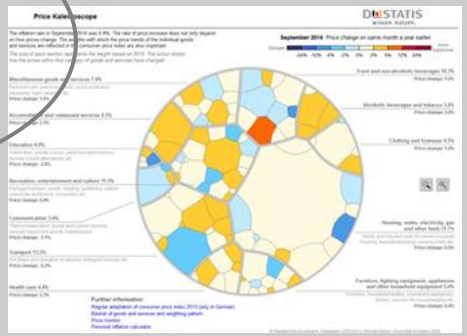
- **Main Source of statistical information on mass phenomena**
- **For government, companies, academia, citizens**
- **Neutral, objective and scientific independent**
- **High quality**

Principles

- **Legislation**
- **Functional centralisation**
- **Regional decentralisation**

for discussion, evaluation, decision making

calculation basis for financial payments, benefits etc.



Sources
https://www.destatis.de/EN/Publications/Specialized/EnvironmentalEconomicAccounting/Indicators2014.pdf?__blob=publicationFile
https://www.destatis.de/Voronoi/PriceKaleidoscope.svg

Official Statistics Germany – Some Facts

about 400 statistics of which 2/3 are EU-wide harmonised

with 4,5 million monthly visits and 760.000 downloads

2 databases according to „*open Government*“ standards

for Germany with about 200 statistics and ½ billion values and about 600.000 quarterly table retrievals

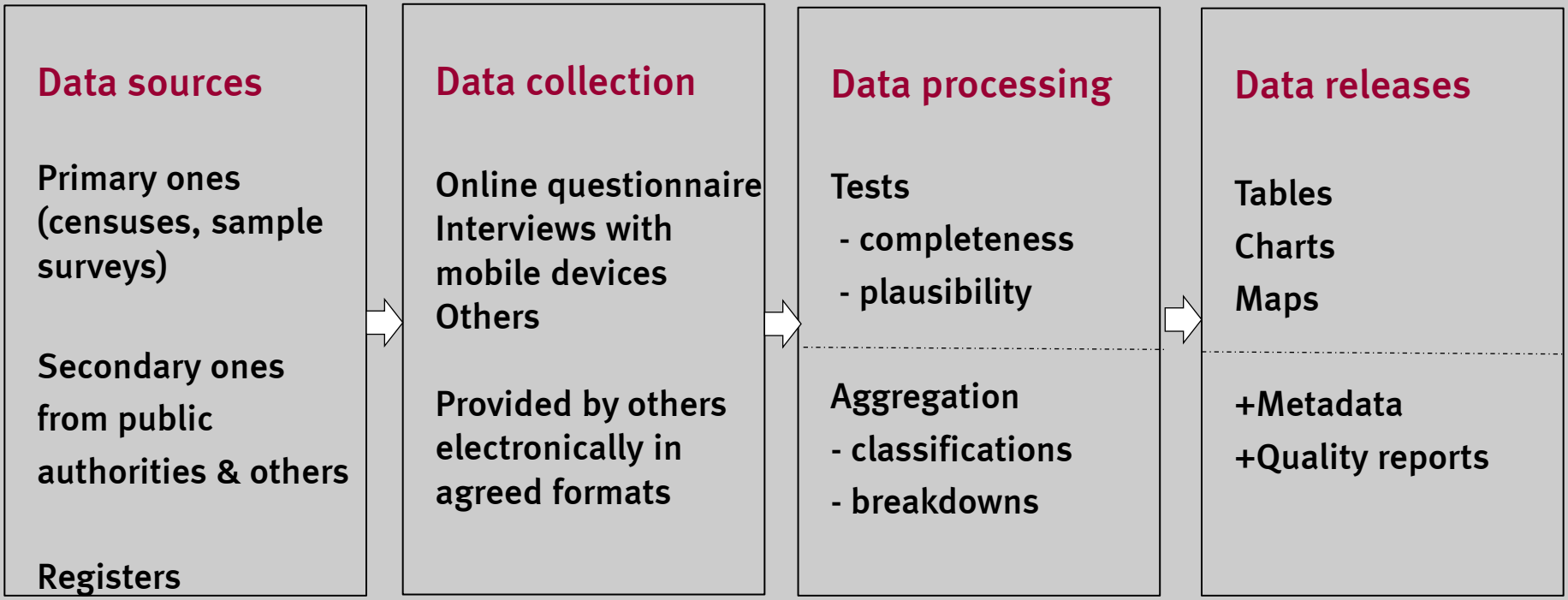
for Germany at regional level (administrative units (NUTS 3)) and municipalities with 80 statistics of which 20 statistics and 20 million values at LAU 2

interactive mapping tools such as

regional atlas for administrative and non-administrative units
online atlas (grid-based) for agricultural statistics

several others (foreign trade, air transports, elections)

Major steps in statistical data processing



Legal basis

Standards
Classifications
Quality criteria

Confidentiality and other rights guaranteed

Big Data as potential data sources – Examples

Commercial data and transactional data

- credit card transactions, electronic ticketing, property purchases,
- purchases of consumer goods and services

Sensor data or other tracking devices

- RFID for tracking animals, merchandise, etc.
- Satellite images
- Electronic toll collection

Behavioural data, opinions and perceptions

- Social Media (Facebook, Twitter, etc)
- Mobile phones

Images 1, 2: Susanne Schnorr-Bäcker; no. 3: https://www.destatis.de/DE/Startseite.html

Big Data and Official Statistics – Germany

1. Identification of big data as potential sources for Official Statistics
2. Identification of statistical areas benefitting from Big Data
3. Feasibility study on web-scraping for specific areas of price statistics (consumer price index, purchasing power parities)
4. Collaboration in a European Task Force on Big Data
5. Other Participation in activities on inter- and supranational level

Image: Susanne Schnorr-Bäcker

Common Benefits and Challenges

Traditional sources

- Information provided by a person
- Refers to a statistical unit (person, household, company etc.)
- With some background information on statistical unit
- Information required refers to a point in time or a period

- Limited number of respondents
- Can be structured and classified
- User oriented
- Is available even after a long time
- Can be documented in detail

Big Data

- Information provided by a technical device/identifier
- Refers mostly to an activity, transaction, opinion
- Often with no background information on statistical unit
- Information is often collected continuously

- Mostly unlimited number of information
- Difficult to structure and classify
- User orientation has to be checked
- Availability is not always guaranteed
- Difficult to document in detail

Quality challenges of Big Data

*Dimensions of quality**	*Challenges*
Mandate for data collection	▪ Could be solvable (access to public authorities easier than to private ones) ▪ Infringements of other rights have to checked (e.g. property rights)
Statistical confidentiality	▪ Could be solvable (in case of private Big Data owners probably more difficult)
Appropriate statistical procedures	▪ Could be solvable, if it is clearly defined what to look for and to show
Cost effectiveness	▪ Must be determined in terms of cost-benefit (including long term prospects)
Accuracy and reliability	▪ Difficult to decide because of lack of knowledge of tracked entities; estimation of bias and errors could be difficult
Timeliness and punctuality	▪ Seem to be a clear benefit because of real time data ▪ For time series and comparisons over time at least a time span of observation should be fixed

*) according to European Code of Practice, see also UN Fundamental Principles of Official Statistics

Impact on Official Statistics & Conclusions

- Increasing needs of a limited set of reliable indicators about economic, societal and ecological developments for planning and decision-making
- Impact on methodologies, data representation, data quality & documentation
- Increasing cooperation between various data producers and users (also globally)


Image: S. Schnorr-Baecker

Conclusions

- Big data as a data source for Official Statistics possible
- Relevance of critical fundamental principles has to be checked for each source

THANK YOU FOR LISTENING!

Dr. Susanne Schnorr-Baecker
Telefon: +49/(0) 611 / 75 2082
Susanne.schnorr-baecker@destatis.de
www.destatis.de