

BIG DATA IN DER AMTLICHEN STATISTIK - MÖGLICHKEITEN UND GRENZEN -

Dr. Susanne Schnorr-Bäcker
Statistisches Bundesamt, Wiesbaden

DAGStat-Symposium am 24. April 2015 in Berlin

Überblick

1. Bundesstatistik – Aufgaben und Arbeitsweise
2. Big Data – Begriff und Wesen
3. Big Data in der Bundesstatistik
4. Ausblick



Foto: S. Schnorr-Bäcker

1. Bundesstatistik – Aufgaben und Arbeitsweise

Aufgaben

- Hauptquelle für statistische Informationen über Massenphänomene
- Für Politik, Unternehmen, Wissenschaft, Bürger
- Neutral, objektiv und wissenschaftlich unabhängig
- Hohe Qualität

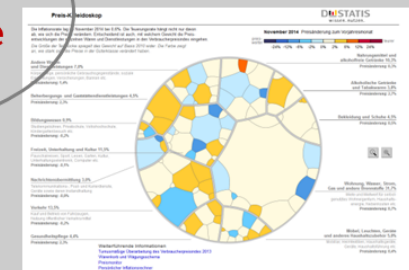
Prinzipien

- Gesetzgebung
- Funktionale Zentralisierung
- Regionale Dezentralisierung
- Internationale Zusammenarbeit

für die Diskussion,
Evaluation,
Entscheidungs-
findung



Berechnungsgrund-
lage für finanzielle
Leistungen, Vorteile
usw.



Quellen

https://www.destatis.de/DE/Publikationen/Thematisch/UmweltoekonomischeGesamtrechnungen/Umweltindikatoren/IndikatorenPDF_0230001.pdf?__blob=publicationFile
<https://www.destatis.de/Voronoj/PreisKaleidoskop.svg>

Amtliche Statistik Deutschland – Einige Fakten

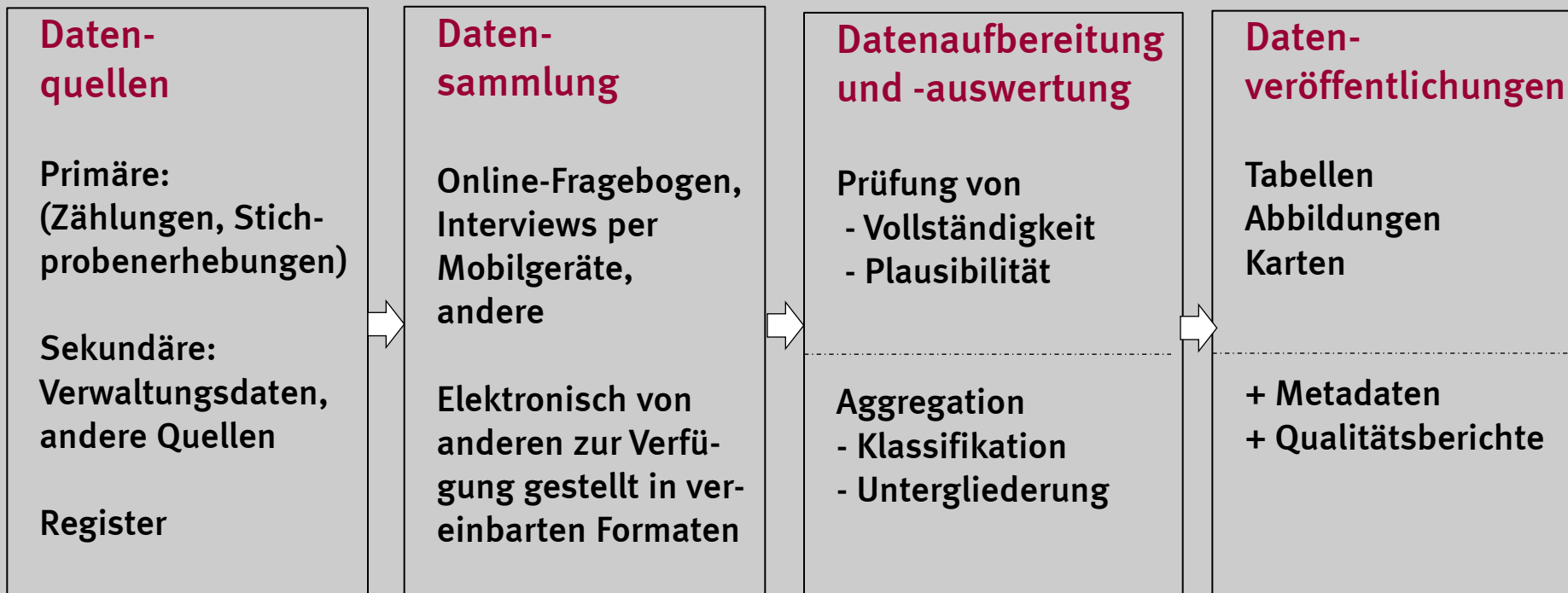
Rund 400 Statistiken, davon 2/3 EU-weit harmonisiert,
mit 4,5 Mio. monatlichen Besuchen und ca. 760.000 Downloads

2 Datenbanken nach „*open Government*“-Standards
für Deutschland mit rund 200 Statistiken und 1/2 Mrd. Werten
und über 700.000 Tabellenabrufen pro Quartal

für Deutschland auf regionaler Ebene für Verwaltungseinheiten,
d.h. Kreise (NUTS 3), 80 Statistiken und 20 Mio. Werte,
darunter 20 Statistiken auf Gemeindeebene

Interaktive Kartenangebote wie
Regionalatlas für Verwaltungs- und nichtadministrative Einheiten
interaktive rasterbasierte Atlanten (Agrarstatistik, Zensus)
andere (z.B. Außenhandel, Luftverkehr, Wahlen)

Wichtige Schritte in der Statistikproduktion*



Rechtsgrundlage

*gemäß Geschäftsprozessmodell Amtliche Statistik (GMAS), aufbauend auf dem Generic Statistical Business Process Model (GSBPM) der UN ECE

**Standards
Klassifikationen
Qualitätskriterien**

**Vertraulichkeit und
andere Rechte
garantiert**

2. Big Data (BD) – Begriff und Wesen

Big Data [ˈbɪɡ ˈdeɪtə] (von engl. *big* = groß, *data* = Daten) bezeichnet Daten-Mengen, die zu groß, oder zu komplex sind, oder sich zu schnell ändern, um sie mit händischen und klassischen Methoden der Datenverarbeitung auszuwerten. Der Begriff "Big Data" unterliegt als Schlagwort derzeit einem kontinuierlichen Wandel; so wird mit Big Data ergänzend auch oft der Komplex der Technologien beschrieben, die zum Sammeln und Auswerten dieser Datenmengen verwendet werden.[1][2] Die gesammelten Daten können aus nahezu allen Quellen stammen: Angefangen bei jeglicher elektronischer Kommunikation, über von Behörden und Firmen gesammelte Daten, bis hin zu den Aufzeichnungen verschiedenster Überwachungssysteme.[3] Big Data können so auch Bereiche abdecken, die bisher als privat galten.

Quelle: http://de.wikipedia.org/wiki/Big_Data

Digitalisierte Daten (z.B. Maschinen-, Sensordaten),

- überwiegend bei Dritten,
- in unstrukturierter Form,
- in großer Menge, Komplexität und/oder Variabilität,

die besondere Methoden und Verfahren zur Aufbereitung und Auswertung mittels Computern erfordern.

Aktuelle Entwicklungen und Big Data



Alle Fotos: S. Schnorr-Bäcker

Big Data-Quellen

1 Internet, besonders

- Suchmaschinen
- Angebotsplattformen (Güter, Jobs, öffentliche Dienstleistungen)
- Unterhaltung (Spiele, Fotos, Bücher, Filme, Musik)
- Crowd- Sourcing (Funding, Umfragen)

2 Informationsaustausch (Mobiltelefone, soziale Netze)

3 Satellitenbilder

4 Sensoren zur Prozesssteuerung (z.B. Verkehr, Logistik)

3. Big Data in der Bundesstatistik

1. Bereichsübergreifendes strategisches Projekt
2. Identifikation von Statistikbereichen, die für BD geeignet sind
3. Machbarkeitsstudie zu Web-Scraping für bestimmte Bereiche der Preisstatistik (Verbraucherpreisindex, Kaufkraftparitäten)
4. Mitarbeit in einer europäischen Task Force für Big Data
5. Teilnahme an anderen Aktivitäten auf inter- und supranationaler Ebene



Big Data im Europäischen Statistiksistem (ESS) - Meilensteine und Zeitplan -

Scheveningen Memorandum (DGINS 2013)

**Einrichtung einer Task Force mit Vertretern aus NSI (auch D),
Wissenschaft und inter- und supranationalen Einrichtungen
(2014)**

Ziele

- **bis 2016: Vorbereitende Arbeiten und Pilotstudien**
- **bis 2020: volle Integration von BD in nationale BD-Strategien**
- **ab 2020: Integration von BD in Produktionsprozesse**

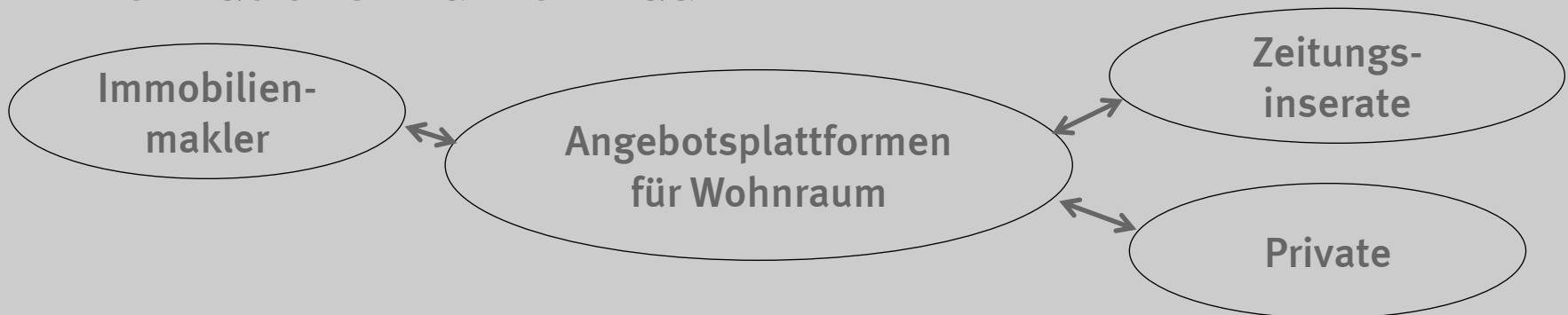
Anwendungsbeispiele im ESS*

	Datenquellen	Datentyp	Statistikbereiche
1	Mobile Kommunikation	Mobilfunkdaten	Tourismus-, Bevölkerungsstatistiken
2	www	Websuche	Arbeitsmarkt-, Wanderungsstatistiken
		Spezielle Webseiten	Statistiken zu Verbraucher-, Immobilienpreisen , IKT, Unternehmensregister
		Soziale Netze	Statistiken zum Konsumenten-Vertrauen, GDP and Beyond, IKT
3	Sensordaten	Verkehrsüberwachung (wie Induktionsschleifen, Lichtschranken)	Verkehrs-, Transportstatistiken
		Smart Meter (z.B. intelligente Stromzähler)	Energiestatistiken
		Satellitenbilder	Landnutzungs-, Landwirtschafts-, Umweltstatistiken
		Automatische Schiffserkennung	Statistiken über Transporte und Emissionen
4	Transaktionen mittels Prozess generierten Daten	Flugbewegungen	Statistiken über Transporte und Schadstoffeinträge in der Luft
		Supermarkt Scanner und Verkaufsdaten	Statistiken über Preise und Haushaltskonsumtion

Quelle: Eurostat Big Data Task Force ESSC 2014/22/8/EN – 26/09/2014

Big Data versus Bundesstatistik

- Informationen zu Wohnraum -



Bundesstatistische Daten (Auswahl)

Erreichbarkeit mit ÖPNV

?

Preise

Verbraucher-, Erzeugerpreise (z.B: Baupreise, Immobilienerwerb), Kaufkraftparitäten

unterjährig

Wohnungsausstattung

Zensus, Mikrozensus (MZ), Einkommens- und Verbrauchsstatistik (EVS), Laufende Wirtschaftsrechnung (LWR)

mehrfährig

Räumliche Lage

Zensus/Anschriften- und Gebäuderegister (AGR)

Vollerhebung, selten

Neubauten und Baufertigstellungen

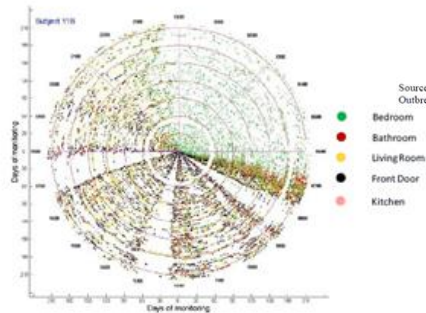
Statistiken im Baugewerbe, Mikrozensus (MZ)

jährlich

Anwendungsbeispiele (1)

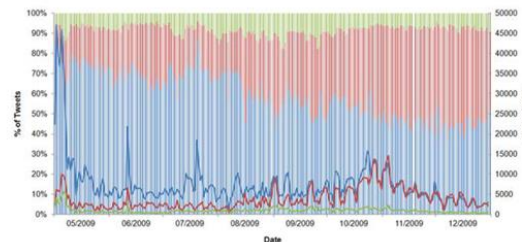
Gesundheit

Figure 5. Sensor events in a residential facility



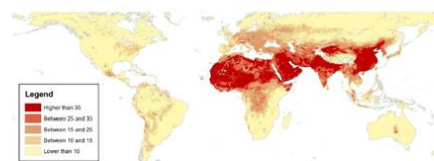
Source: Misha Pavel. OECD-NSF Workshop, 2011.

Figure 6. Tweets containing H1N1, swine flu, or both from May to December 2009



Source: Chen C. G. Eysenbach (2010), "Pandemics in the Age of Twitter: Content Analysis of Tweets during the 2009 H1N1 Outbreak." PLoS ONE 5(11): e14118. doi:10.1371/journal.pone.0014118

Figure 2. PM_{2.5} concentrations based on satellite-based data, 2011



Note: The data refer to three-year average (2010-2012). The 50% relative humidity standard has been adopted for consistency with the ground-level measurements. The map displays PM_{2.5} concentrations according to five levels based on the WHO guidelines.

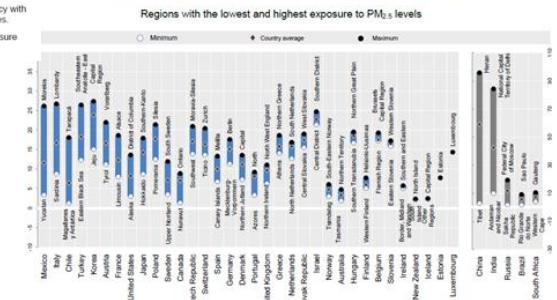
Source: van Donkelaar, A., R. V. Martin, M. Brauer and B. L. Boys (2014) "Use of Satellite Observations for Long-Term Exposure Assessment of Global Concentrations of Fine Particulate Matter." Environmental Health Perspectives, in press.

Quelle: OECD (2014) „Unleashing the Power of Big Data for Alzheimer’s Disease and Dementia Research: Main Points of the OECD Expert Consultation of Unlocking Global Collaboration to Accelerate Innovation for Alzheimer’s Disease and Dementia,“ OECD Digital Economy Papers, No. 233, OECD Publishing;
<http://www.oecd-ilibrary.org/docserver/download/5jz73kvmvbw.pdf?expires=1421405824&id=id&acname=guest&checksum=C75245D9FD783877403D524076C52D62>, S. 17 und 19

Quelle: BREATHING THE SAME AIR? MEASURING AIR POLLUTION IN CITIES AND REGIONS
Monica Brezzi and Daniel Sanchez-Serra
OECD, Regional Development Policy Division, vorgestellt auf der WPTI-Sitzung am 1. Dezember 2014 bei der OECD in Paris, S. 6 und S. 9

Feinstaub-Belastung

Figure 5. Regional disparities in average exposure to air pollution, 2011



Note: Data refer to three-year average measures (2010-2012).
Source: OECD calculations based on van Donkelaar et al. (2014).

Big Data (2)

Tourismusstatistiken (Estland)

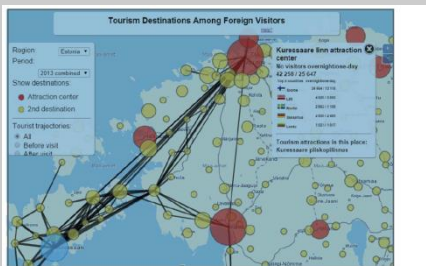


Figure 7. Application for monitoring tourism destinations for foreign tourists within Estonia (attraction centre = more than 10 000 yearly overnight stays) (Positium tourism demo, 2013).

<http://unstats.un.org/unsd/trade/events/2014/Beijing/Margus%20Tiru%20%20Mobile%20Positioning%20Data%20Paper.pdf>

<http://unstats.un.org/unsd/trade/events/2014/Beijing/documents/globalpulse/UN%20Global%20Pulse%20-%20Mining%20Indonesian%20Tweets%20to%20Understand%20Food%20Price%20Crises%20-%20Methodology.pdf>

Preise (Indonesien)

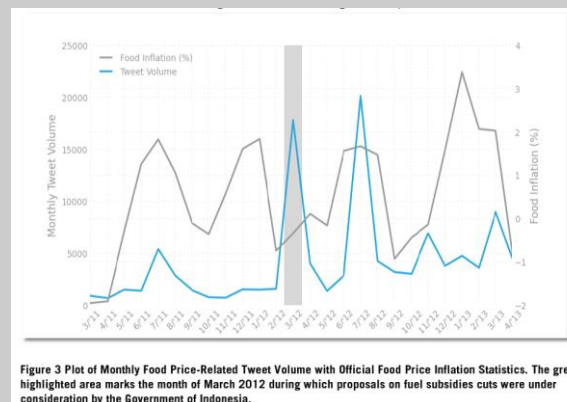
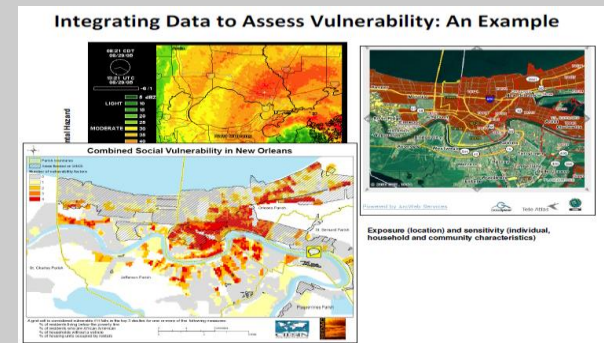


Figure 3 Plot of Monthly Food Price-Related Tweet Volume with Official Food Price Inflation Statistics. The grey highlighted area marks the month of March 2012 during which proposals on fuel subsidies cuts were under consideration by the Government of Indonesia.

Mehrdimensionale sozialökonomische Analysen (NASA)



Alex de Sherbinin, Deputy Manager, NASA Socioeconomic Data and Applications Center Center for International Earth Science Information Network, The Earth Institute, Columbia University Palisades, New York, USA; http://www.efgs.info/workshops/efgs-2010-tallinn-estonia/efgs2010/5_Keynote%20Speech%20Application%20of%20gridded%20population%20data%20sets%20for%20elucidating%20human%20environment%20interactions%20Alex%20de%20Sherbinin.pdf

4. Ausblick

Digitale Datenquellen für die Digitale Gesellschaft

Besonderheiten von Big Data

- Daten überwiegend bei Privaten
- Data Mining
- Data Design
- Kontinuität und Replizierbarkeit

Bundesstatistische Qualitätsanforderungen und BD

- gesetzl. Auftrag und Geheimhaltung
- wissenschaftliche Verfahren und Wirtschaftlichkeit
- Genauigkeit und Zuverlässigkeit
- Aktualität und Pünktlichkeit

Big Data für ein Politik begleitendes statistisches Monitoring

„SDG/post 2015“ (UN)

„Better life initiative“ (OECD)

„Europa 2020“ (EU)



Nachhaltige Entwicklung
(seit 2002)

Enquêtékommision „Wachstum,
Wohlstand, Lebensqualität“
(2011-2013)

Bürgerdialoge „Gutes Leben“
(13.4.2015)

https://www.gut-leben-in-deutschland.de/DE/Home/home_node.html



https://www.destatis.de/DE/Publikationen/Thematisch/UmweltoekonomischeGesamtrechnungen/Umweltindikatoren/IndikatorenPDF_0230001.pdf?__blob=publicationFile



http://www.nachhaltigkeitsrat.de/uploads/media/Broschuere_Staedte_fuer_ein_nachhaltiges_Deutschland_texte_Nr_36_Juni_2011.pdf

Herausforderungen

- steigende Anforderungen an ein limitiertes Set von zuverlässigen Indikatoren zu wirtschaftlichen, gesellschaftlichen und ökologischen Entwicklungen für die Planung und Entscheidungsfindung
- Auswirkung auf die Methoden, Datendarstellung, -qualität und Dokumentation
- verstärkte Zusammenarbeit zwischen den verschiedenen Datenproduzenten und -nutzern (auch weltweit)

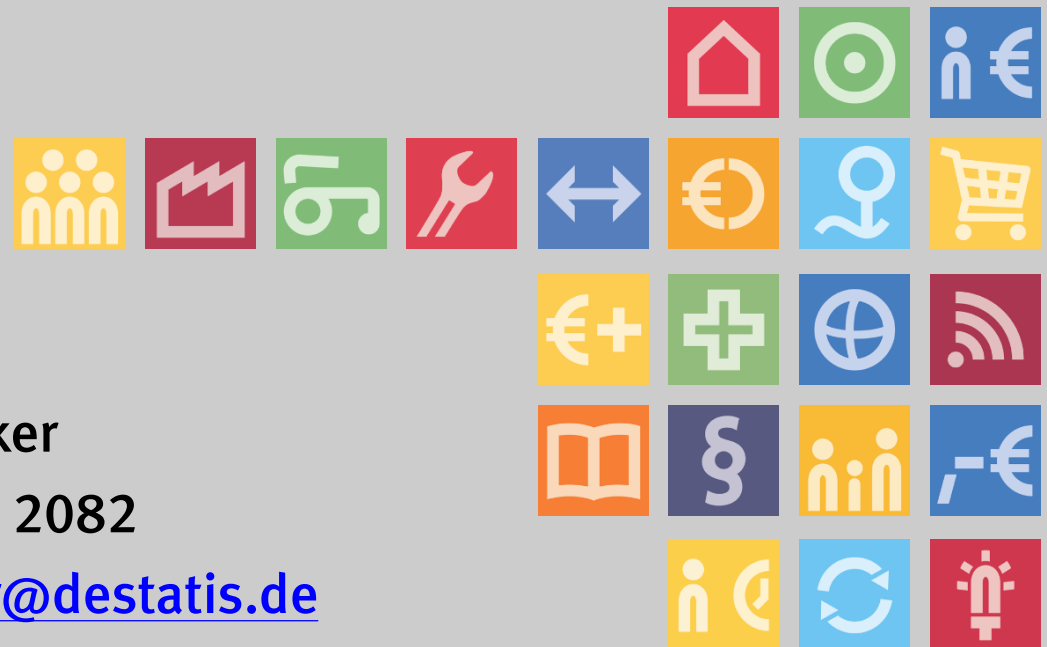


Foto: S. Schnorr-Baecker

Schlussfolgerungen

- Big Data als Datenquelle sollte schrittweise für die amtliche Statistik geprüft werden.
- Die Relevanz der Grundprinzipien muss für jede Quelle geprüft werden.

VIELEN DANK FÜR IHRE AUFMERKSAMKEIT!



Dr. Susanne Schnorr-Bäcker

Telefon: +49/(0) 611 / 75 2082

Susanne.schnorr-baecker@destatis.de

www.destatis.de