

Die Bedeutung von Daten für die Sozialplanung

VSOP Jahrestagung 2019, Darmstadt

Barbara Engels

Institut der deutschen Wirtschaft

engels@iwkvoeln.de



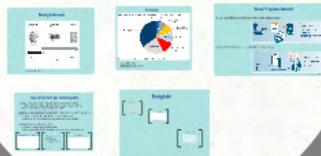
Perspektiven

- öffentliche Verwaltungen können proaktiv handeln
- Gleichheit und Ungleichheit kann besser gemessen werden
- Ressourcen können besser genutzt werden
- ! Daten können immer auch fehlerleiten

Welche Unterstützung würden Sie sich wünschen?

Chancen

- Mix aus Datenquellen >>
- das Internet als Datenquelle >>
- schnellere, präzisere Daten -> Smart Government >>



Herausforderungen

- **garbage in, garbage out:** Datenaufbereitung aufwendig - aber zentral (noisy data)
- hohe Bias-Wahrscheinlichkeit, Endogenität, **Diskriminierungsgefahr**
- **Streetlights- und Beobachtungseffekte:** Selektionsproblem, digital divide (Daten leicht zugänglich, aber nicht repräsentativ)
- Korrelation vs. Kausalität: **Nathaway-Effekt**
- technische Probleme: Zugriff, Bottleneck Computerressourcen
- relevant vs. readily available data - **big is not smart** >>
- Wer macht's? Integration in die Fachverfahren der Verwaltung >>
- **Datenschutz:** Big Data gefährdet Privatheit, persönliche Autonomie, Persönlichkeitsrechte >>



Ausgangslage

- Immer mehr Daten >>
- **Digitalisierung** macht Daten leicht erhaltbar, messbar, teilbar
- Menge, Vielfalt, Geschwindigkeit, Kombinierbarkeit sind neu >>
- **neue Möglichkeiten für Wirtschaft und Gesellschaft:** verbesserte Forschung, neue Produkte/Dienstleistungen, Optimierung der Produktivität- & Logistikprozesse >>
- auch für **Sozialplanung?**
- Sozialberichte: v.a. Aggregatdaten, die Kontextinformationen über Teilhabebedingungen bieten; Teilhabekonzept verlangt nach **Individualdaten** (Langzeitstudien) - für Kommunen so nicht machbar; wertvolle Daten aus Verwaltungsverfahren & Leistungsprozessen müssen mit anderen Mikrodaten verknüpft werden



Praxistipp

Planen, was Sie überhaupt messen wollen, bevor Sie überhaupt nach Daten schauen!

Praxistipp

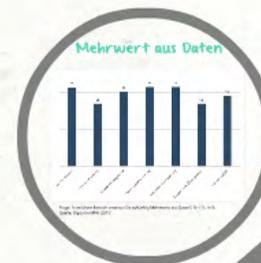
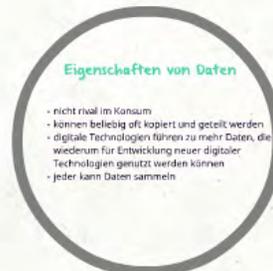
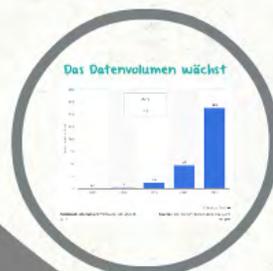
Welchen Prozess oder Vorgang wollen Sie messen?

Welche Daten haben Sie schon?

Welche Möglichkeiten haben Sie, um Daten zu bekommen?

Ausgangslage

- immer **mehr** Daten >>
- **Digitalisierung** macht Daten leicht erhebbar, messbar, teilbar: Menge, Vielfalt, Geschwindigkeit, Kombinierbarkeit sind neu >>
- **neue Möglichkeiten für Wirtschaft und Gesellschaft:** verbesserte Forschung, neue Produkte&Dienstleistungen, Optimierung der Produktions- & Logistikprozesse >>
- auch für **Sozialplanung?**
- Sozialberichte: v.a. Aggregatdaten, die Kontextinformationen über Teilhabebedingungen bieten; Teilhabekonzept verlangt nach **Individualdaten** (Längsschnitt) - für Kommunen so nicht machbar; wertvolle Daten aus Verwaltungsverfahren & Leistungsprozessen müssen mit anderen Mikrodaten verknüpft werden



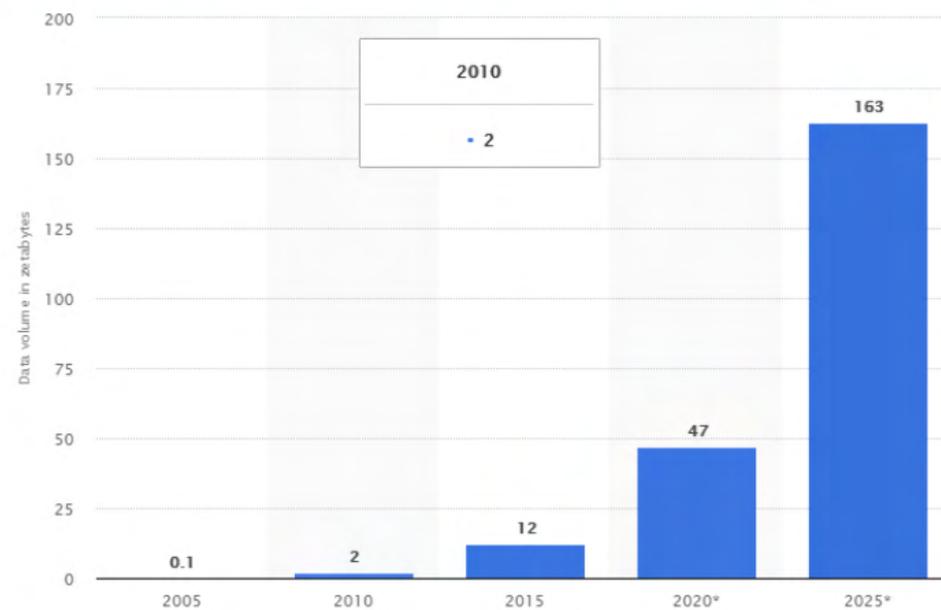
Worum geht's?

Welche Möglichkeiten gibt es, **neue/andere** Fürsorge, Bildung, Integration, Wohnungs-, Gesundheitsförderung, Armut, Alten- und Behindertenhilfe zu erheben/analysieren?

Ist **Sozialberichterstattung** auf Basis der Daten anders möglich?

Inwiefern schaffen neue/andere Zahlen ein **Gerechtere**? Einen besseren **Lebensstandard** können kommunale **Ressourcen** dadurch werden?

Das Datenvolumen wächst



© Statista 2019

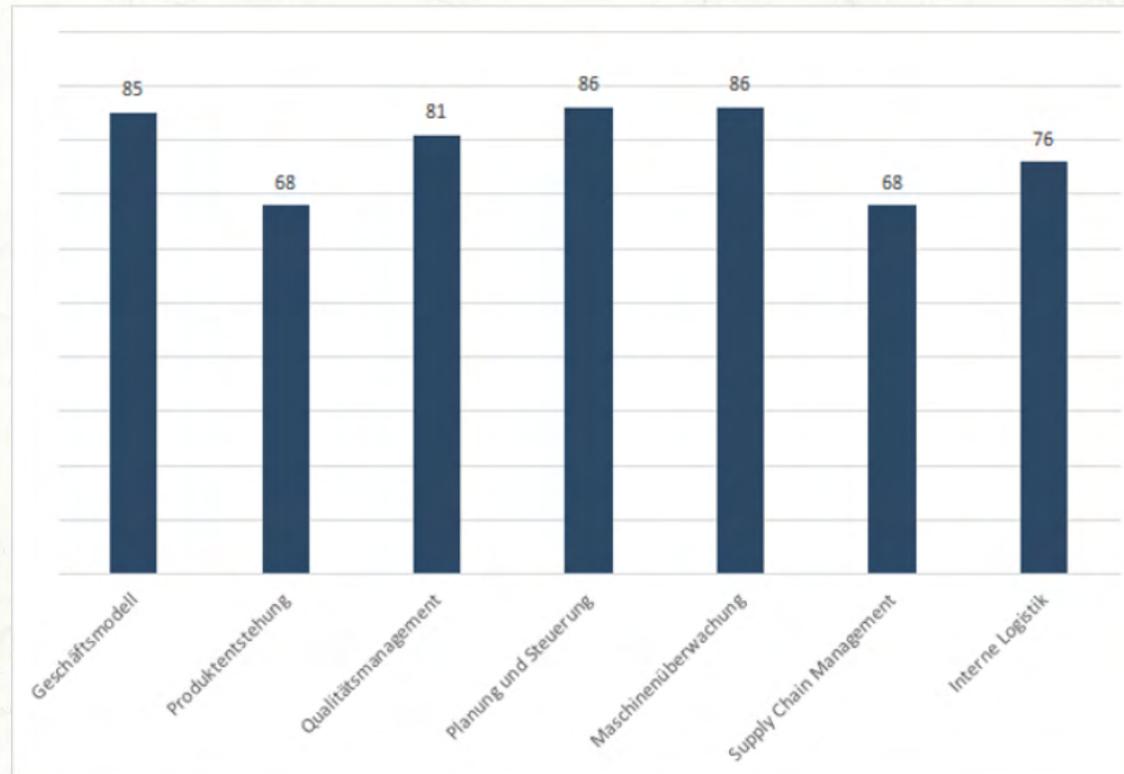
Additional Information: Worldwide; IDC; 2005 to 2017

Sources: IDC; Kleiner Perkins Caufield & Byers; Seagate

Eigenschaften von Daten

- nicht rival im Konsum
- können beliebig oft kopiert und geteilt werden
- digitale Technologien führen zu mehr Daten, die wiederum für Entwicklung neuer digitaler Technologien genutzt werden können
- jeder kann Daten sammeln

Mehrwert aus Daten



Frage: In welchem Bereich erwarten Sie zukünftig Mehrwerte aus Daten?, N=111; in %
Quelle: Digital in NRW (2019)

Worum geht's?

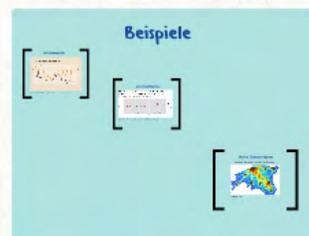
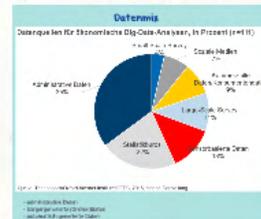
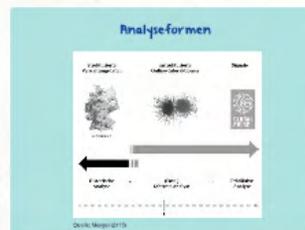
Welche Möglichkeiten gibt es, **neue/andere Zahlen** für Arbeitsmarkt, Fürsorge, Bildung, Integration, Wohnungsmarkt, Gesundheitsförderung, Armut, Alten- und Seniorenpolitik und Behindertenhilfe zu erheben/analysieren?

Ist **Sozialberichterstattung** auf Basis der neuen Datenlage auch anders möglich?

Inwiefern schaffen neue/andere Zahlen eine höhere **Gerechtigkeit**? Einen besseren **Lebensweltbezug**? Inwiefern können kommunale **Ressourcen** dadurch besser genutzt werden?

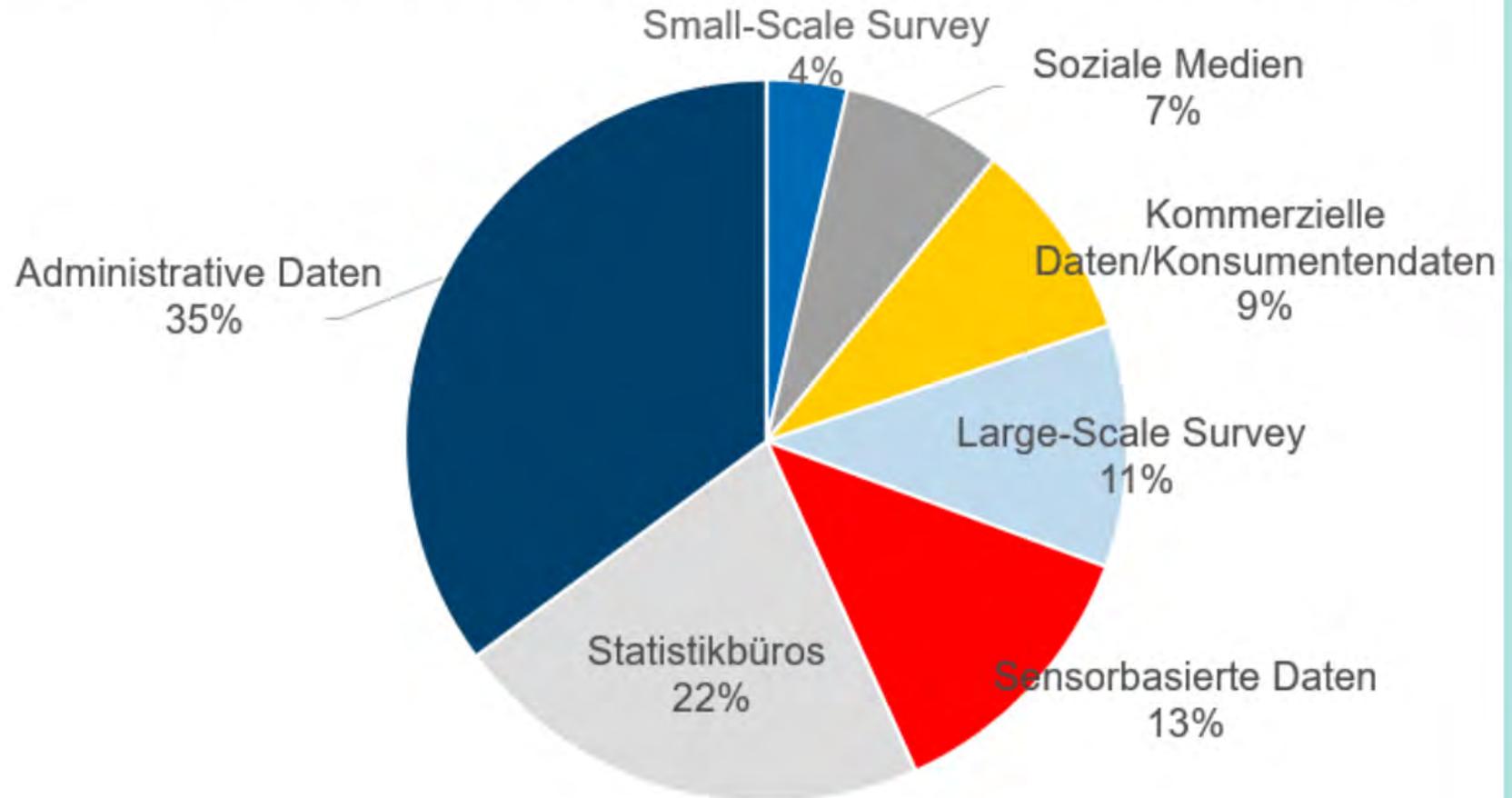
Chancen

- Mix aus Datenquellen >>
- das Internet als Datenquelle >>
- schnellere, präzisere Daten -> Smart Government >>



Datenmix

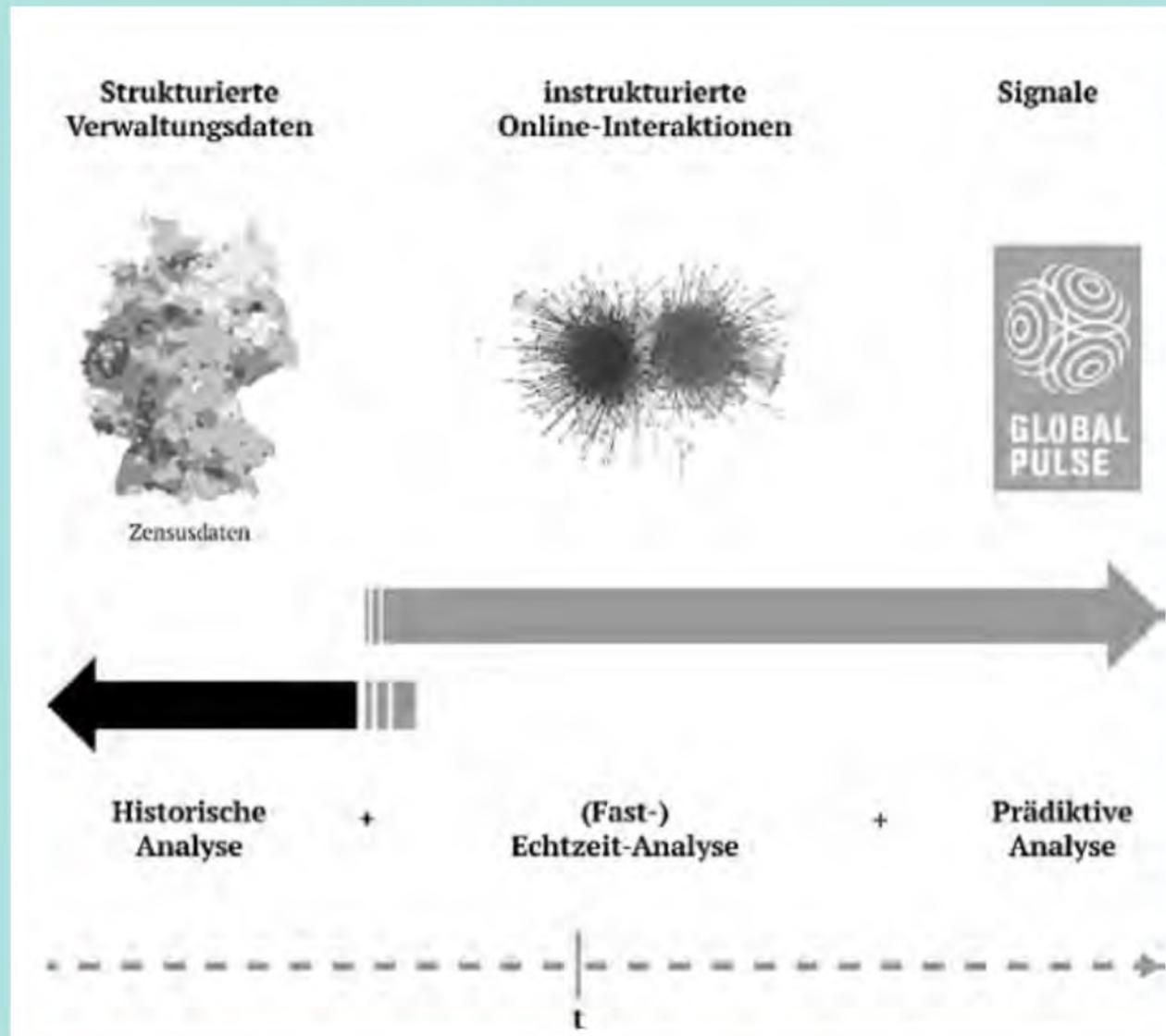
Datenquellen für ökonomische Big-Data-Analysen, in Prozent (n=111)



Quelle: Technopolis/Oxford Internet Institute/CEPS, 2015; eigene Darstellung

- administrative Daten
- bürgergenerierte (Online-)Daten
- automatisch generierte Daten

Analyseformen



Quelle: Mergel (2018)

Das Internet als Datenquelle

Dumps Daten, die öffentlich und gebündelt bereitgestellt werden
Analogie: Kopie der öffentlichen Buchhaltung nach Hause nehmen
Problem: u.U. fehlerhaft, nicht aktuell, selektiv

Application Programming Interface (API) Daten über speziellen Zugang laden
Analogie: Schlüssel zum öffentlichen Teil der Buchhaltung
Problem: jede API ist eigenes Programmieruniversum, selektiv

Scrapen Daten von Webseite auslesen
Analogie: Schaufenster abfotografieren
Problem: Snapshot; läuft bei Umstrukturierung ins Leere; Botstopper

Crawlen: Logisch, aber nicht trivial



1. Liste mit zu durchsuchenden Webseiten erstellen
2. Im Seitenquelltext nach gewünschter Daten suchen
3. Coden (R, Java): Ziehe mir von jede Seite der Liste diese Daten raus und lege sie ab
4. Datenaufbereitung (!)
5. Business as usual: Ökonometrie

Legende

Wort	Suchanfragen	Suchergebnisse
Wort	Suchanfragen	Suchergebnisse

1970 1980 1990 2000 2010 2014

Text Mining & Sentiment Analysis

- oft für **Nowcasting** verwendet
- Text Mining entdeckt z.B. Worthäufigkeiten
- Sentiment Mining kategorisiert Wörter
- Datenquellen: Social Media, Newsletter, Blogs, ...
- Problem: hohe Fehleranfälligkeit (Sarkasmus, Typos), Bias

Open Data



Das Datenportal für Deutschland
Suchen Sie nach offenen Daten für Deutschland

Das Datenportal für Deutschland ist ein zentraler Anlaufpunkt für alle öffentlichen Daten der Bundesländer und Kommunen. Hier können Sie nach offenen Daten suchen, diese herunterladen und für Ihre eigenen Zwecke nutzen.

Suchen Sie in diesen Kategorien:

- Energie
- Umwelt
- Wirtschaft
- Verkehr
- Kultur
- Bildung
- Gesundheit
- Soziales
- Verwaltung
- Öffentliche Einrichtungen
- Öffentliche Dienstleistungen
- Öffentliche Gebäude
- Öffentliche Räume
- Öffentliche Dienstleistungen
- Öffentliche Gebäude
- Öffentliche Räume

Crawlen: Logisch, aber nicht trivial

SPIEGEL ONLINE DER SPIEGEL SPIEGEL TV Suchen Anmelden

SUCHE Schlagzeilen | Wetter | DAX 10.577,16 | TV-Programm | Abos

Nachrichten > Suche

Verfassungsschutz Suchen

Zeitraum Volltext

SPIEGEL ONLINE DER SPIEGEL manager-magazin.de

Verfassungsschutz (6959 Suchergebnisse)

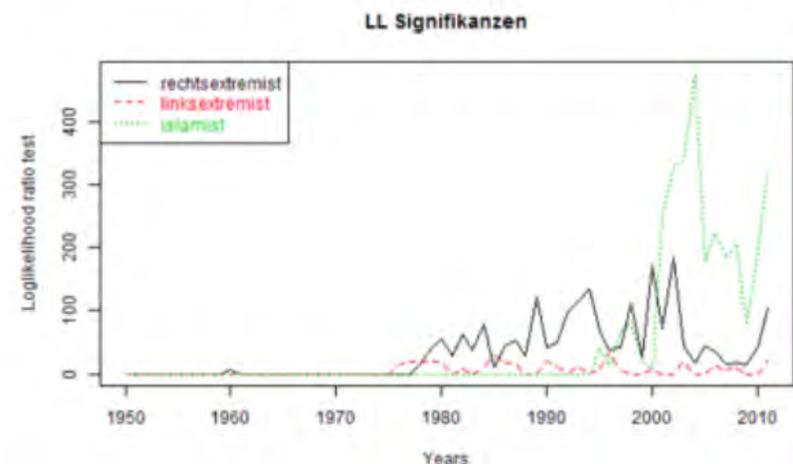
Kostenfreies Archiv: Die SPIEGEL-ONLINE-Suche bietet Ihnen in dieser Trefferliste auch Artikel aus dem kompletten **SPIEGEL-Archiv seit 1947** (bis auf die jeweils letzten zwölf Monate kostenlos) und Themenseiten zum Stichwort. Wollen Sie mit **Verknüpfungen, Einschränkungen, Platzhaltern** suchen? [zur Hilfe...](#)

Themen

Bundesamt für Verfassungsschutz **Verfassungsschutz**
Landesämter für Verfassungsschutz

Festnahme von Jaber Albakr: Wachsame Geheimdienste, mutige Syrer, zögerliche Polizei
SPIEGEL ONLINE - Politik - 11.10.2016
Die Pläne des Terrorverdächtigen Albakr sind in letzter Minute vereitelt worden. Der Fall zeigt: Der Informationsaustausch der Geheimdienste funktioniert - aber

1. Liste mit zu durchsuchenden Webseiten erstellen
2. Im Seitenquelltext nach gewünschter Daten suchen
3. Coden (R, Java): Ziehe mir von jeder Seite der Liste diese Daten raus und lege sie ab
4. Datenaufbereitung (!)
5. Business as usual: Ökonometrie



Text Mining & Sentiment Analysis

- oft für **Nowcasting** verwendet
- Text Mining entdeckt z.B. Worthäufigkeiten
- Sentiment Mining kategorisiert Wörter
- Datenquellen: Social Media, Newsletter, Blogs, ...
- Problem: hohe Fehleranfälligkeit (Sarkasmus, Typos), Bias

Open Data

Das Datenportal für Deutschland

Open Government: Verwaltungsdaten transparent, offen und frei nutzbar

Nach Datensätzen suchen

Suchen

Erweiterte Suche

Kartensuche

DCAT-AP.de als formaler Metadatenstandard für offene Verwaltungsdaten bestätigt

Der IT-Planungsrat hat DCAT-AP.de als formalen Austauschstandard für allgemeine offene Verwaltungsdaten festgelegt.

Vor knapp drei Jahren hat die Geschäfts- und Koordinierungsstelle GovData die Verantwortung für das Standardisierungsvorhaben "Metadatenstruktur für offene Verwaltungsdaten" des IT-Planungsrates übernommen. Am 28. Juni konnte dieses Vorhaben erfolgreich mit dem Beschluss des IT-Planungsrates zum Abschluss gebracht werden, mit dem DCAT-AP.de als formaler Austauschstandard für allgemeine offene Verwaltungsdaten festgelegt wurde. Weitere Informationen finden Sie [hier](#).

Stöbern Sie in diesen Kategorien



Bevölkerung und
Gesellschaft
(5837)



Bildung, Kultur und
Sport
(2096)



Energie
(374)



Gesundheit
(1111)



Internationale Themen
(91)



Justiz, Rechtssystem
und öffentliche
Sicherheit
(3231)



Landwirtschaft,
Forstwesen,
Forstwirtschaft und
Nahrungsmittel
(1901)



Regierung und
öffentlicher Sektor
(8971)



Regionen und Städte
(246)



Umwelt
(6636)



Verkehr
(2579)



Wirtschaft und Finanzen
(5026)



Wissenschaft und
Technologie
(800)

Smart Government

nutzerfreundliche und effiziente Verwaltungsleistungen



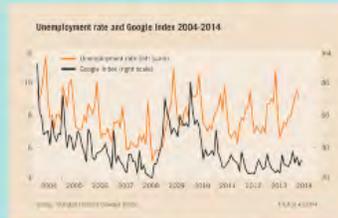
datengestützte Entscheidungshilfen für die Verwaltung



Quelle: McKinsey/Bitkom (2018)

Beispiele

Arbeitslosenrate



Quelle: Hui Akin, J., 2014, Big data: Google Analytics Predict Unemployment in Ireland

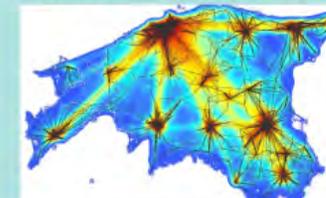
Lebensmittelpreise



Quelle: U4 Global Price, 2014, Monitoring food prices in Indonesia using Social Media Sentiment

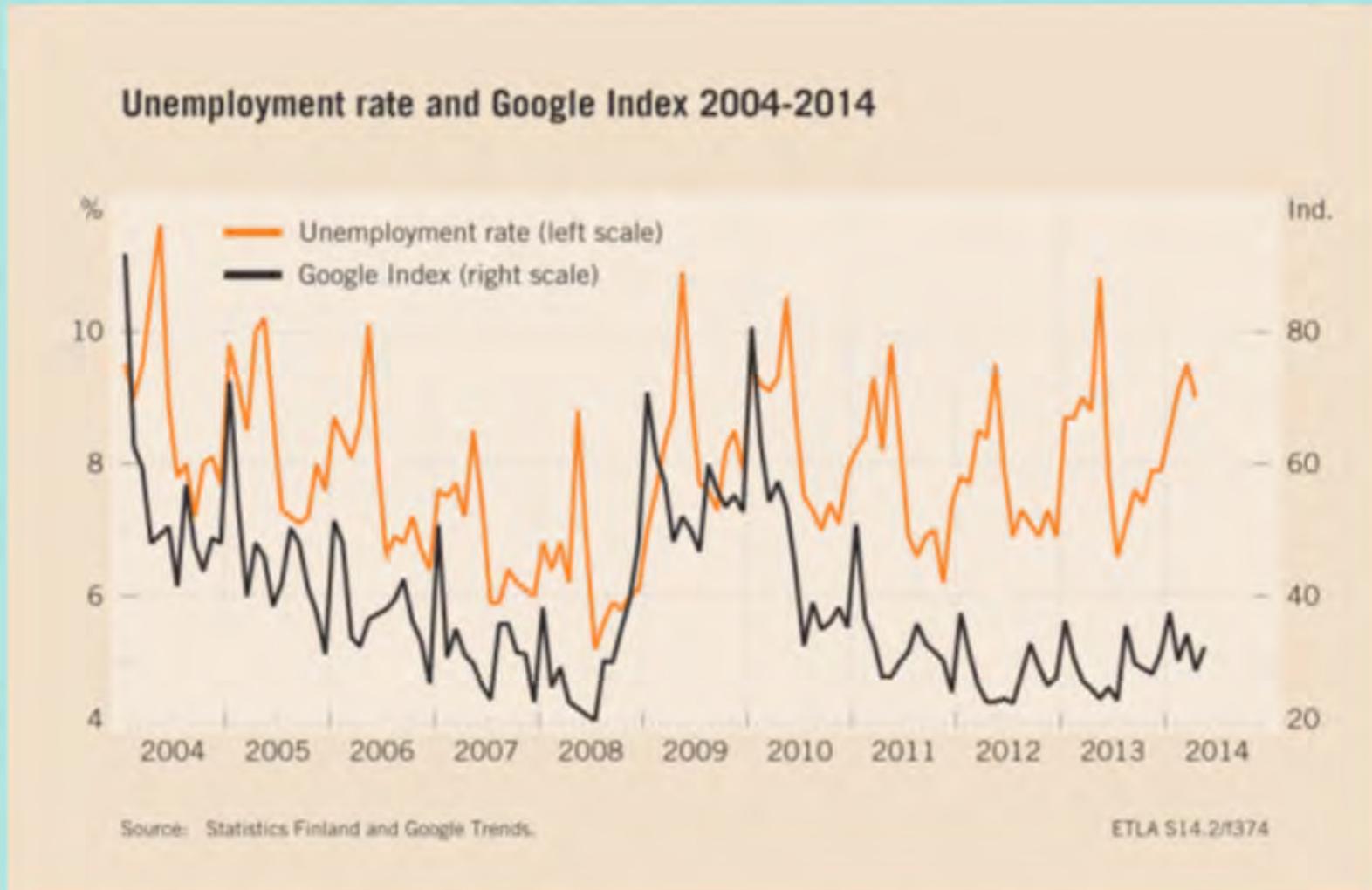
Mobile Standortdaten

Pendleraufkommen, De-Facto-Bevölkerung



Quelle: Masvio (2016)

Arbeitslosenrate

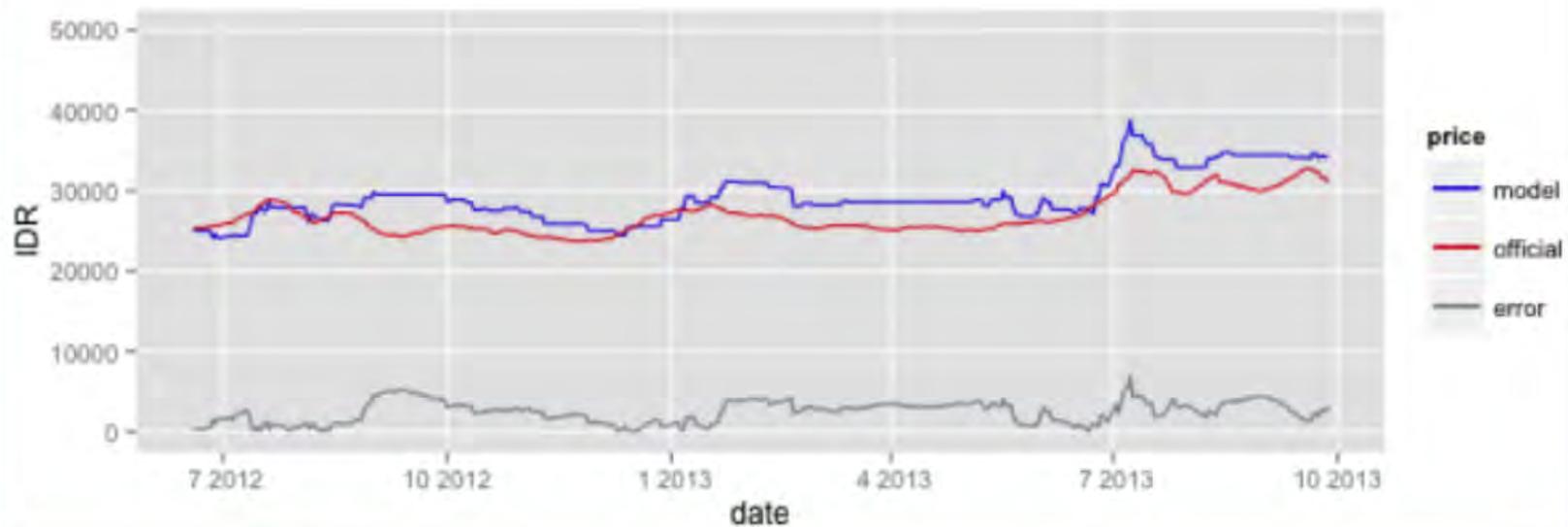


Quelle: Tuhkuri, J., 2014, Big data: Google Searches Predict Unemployment in Finland

Lebensmittelpreise

CHICKEN: The tool successfully modeled the price of chicken (“daging ayam”) with significant correlation for the duration.

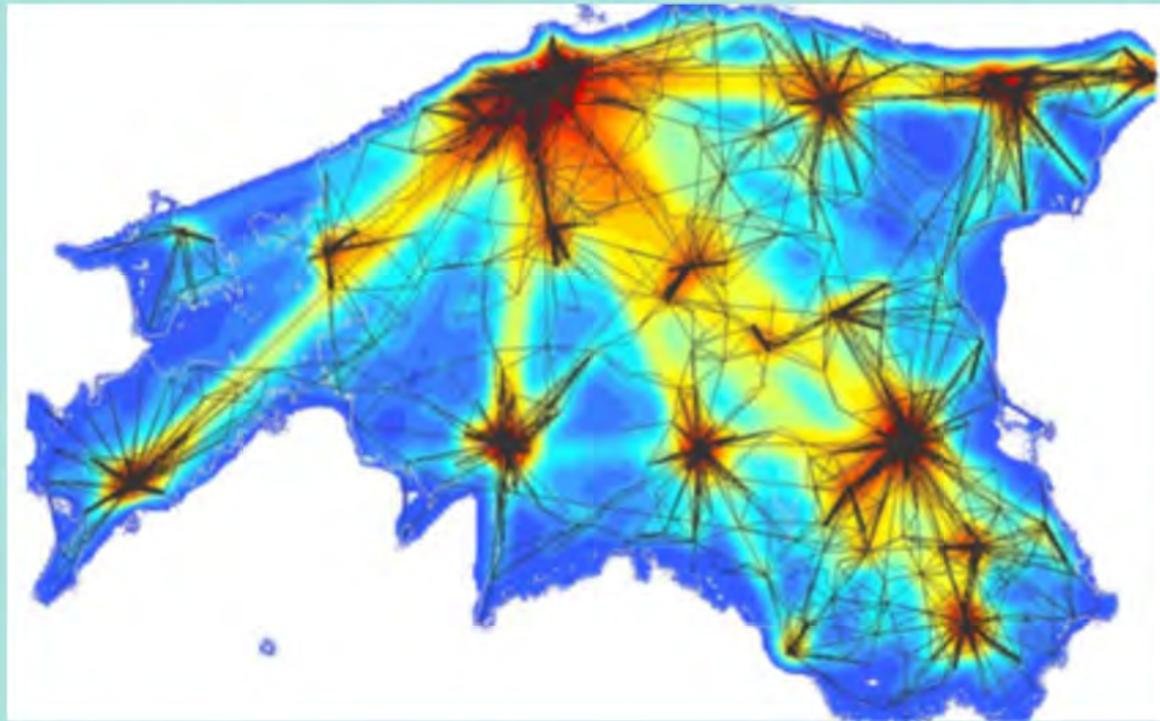
5,223 tweets contained price quotes on chicken.



Quelle: UN Global Pulse, 2014, Nowcasting Food prices in Indonesia using Social Media Signals

Mobile Standortdaten

Pendleraufkommen, De-Facto-Bevölkerung



Quelle: Masso (2016)

Praxischeck

Welche neuen Datenarten und Datenquellen nutzen Sie in Ihrem Arbeitsalltag?

Welche Datenarten und Datenquellen würden Sie gerne nutzen?

Welche Themenschwerpunkte sind für Sie besonders interessant?

Herausforderungen

- **garbage in, garbage out:** Datenaufbereitung aufwendig – aber zentral (noisy data)
- hohe Bias-Wahrscheinlichkeit, Endogenität, **Diskriminierungsgefahr**
- **Streetlights- und Beobachtungseffekte:** Selektionsproblem, digital divide (Daten leicht zugänglich, aber nicht repräsentativ)
- Korrelation vs. Kausalität; **Hathaway-Effekt**
- technische Probleme: Zugriff, Bottleneck Computerressourcen
- relevant vs. readily available data – **big is not smart** >>
- Wer macht's? Integration in die Fachverfahren der Verwaltung >>
- **Datenschutz:** Big Data gefährdet Privatheit, persönliche Autonomie, Persönlichkeitsrechte >>

Wer macht was?

	1	2	3	4	5
1					
2					
3					
4					
5					



Prax

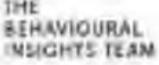
Mit welchen H...
sind Sie im An...
konfrontiert?

Welche Nutzu...
Daten gibt es...

Welche Mögli...
die Erhebung...
Daten zu vere...

Wer macht was?

Organisationseinheiten im öffentlichen Sektor, die Big Data verarbeiten

Thematische Data-Science-Teams	Social & Behavioral Insights Teams	Digital Operation Center	Bürger-Data-Scientists
Steuerbetrug vorhersagen	Verwaltungsperformance messen	Menschliche und physische Sensoren zusammenführen während und nach Katastrophen	Nicht-Experten oder Beamte in Freizeit
SNA	Öffentliche Angebote neu gestalten	Nowcasting	Hackathons
Risikoindikatoren	Data Analytics		Predictive modelling & analytics
Predictive Analytics			
			

Quelle: Mergel (2018)

Wer darf was?

Welche ethischen Grenzen der Ökonomisierung von Daten gibt es?



Quelle: BMI (2019)

Praxischeck

Mit welchen Herausforderungen sind Sie im Arbeitsalltag konfrontiert?

Welche Nutzungsbarrieren von Daten gibt es?

Welche Möglichkeiten sehen Sie, die Erhebung und Analyse von Daten zu vereinfachen?

Perspektiven

- öffentliche Verwaltungen können proaktiv handeln
- Gleichheit und Ungleichheit kann besser gemessen werden
- Ressourcen können besser genutzt werden
- ! Daten können immer auch fehlerhaft sein

Welche Unterstützung würden Sie sich wünschen?

Empfehlungen

- Zusammenarbeit in **interdisziplinären, agilen** Teams
- **zentrale Koordination** von Datenverknüpfung und Aufbereitung, gemeinsame Infrastruktur
- **Privacy by Design**
- **Stakeholder**-Management
- **Learning-by-Doing**-Mentalität entwickeln
- **Silos** aufbrechen
- Aufbau von **Partnerschaften mit Dritten**
- **Erfahrungsaustausch** mit anderen Behörden
- **Offenheit** für Veränderungen durch neue Technologien