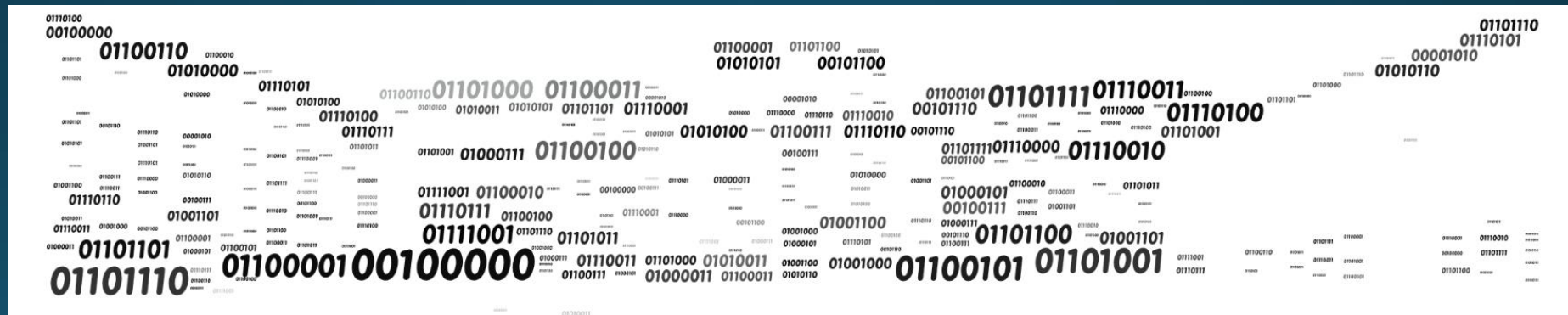


# Survey and Data Science



Prof. Dr. Frauke Kreuter

# Survey and Data Science



Prof. Dr. Frauke Kreuter

# Was hat sich geändert?

# Datenquellen bisher

## Administrative Prozesse

201300303201

2

3 **Steuernummer** [redacted]

eTIN lt. Lohnsteuerbescheinigung(en), sofern vorhanden [redacted] eTIN lt. weiterer Lohnsteuerbeschei [redacted]

4

**Einkünfte aus nichtselbständiger Arbeit**

**Angaben zum Arbeitslohn** Lohnsteuerbescheinigung(en) Steuerklasse 1 - 5

5 **Steuerklasse** 168 [redacted]  
EUR Ct

6 Bruttoarbeitslohn 110 [redacted]

7 Lohnsteuer 140 [redacted]

8 Solidaritätszuschlag 150 [redacted]

## Befragungen



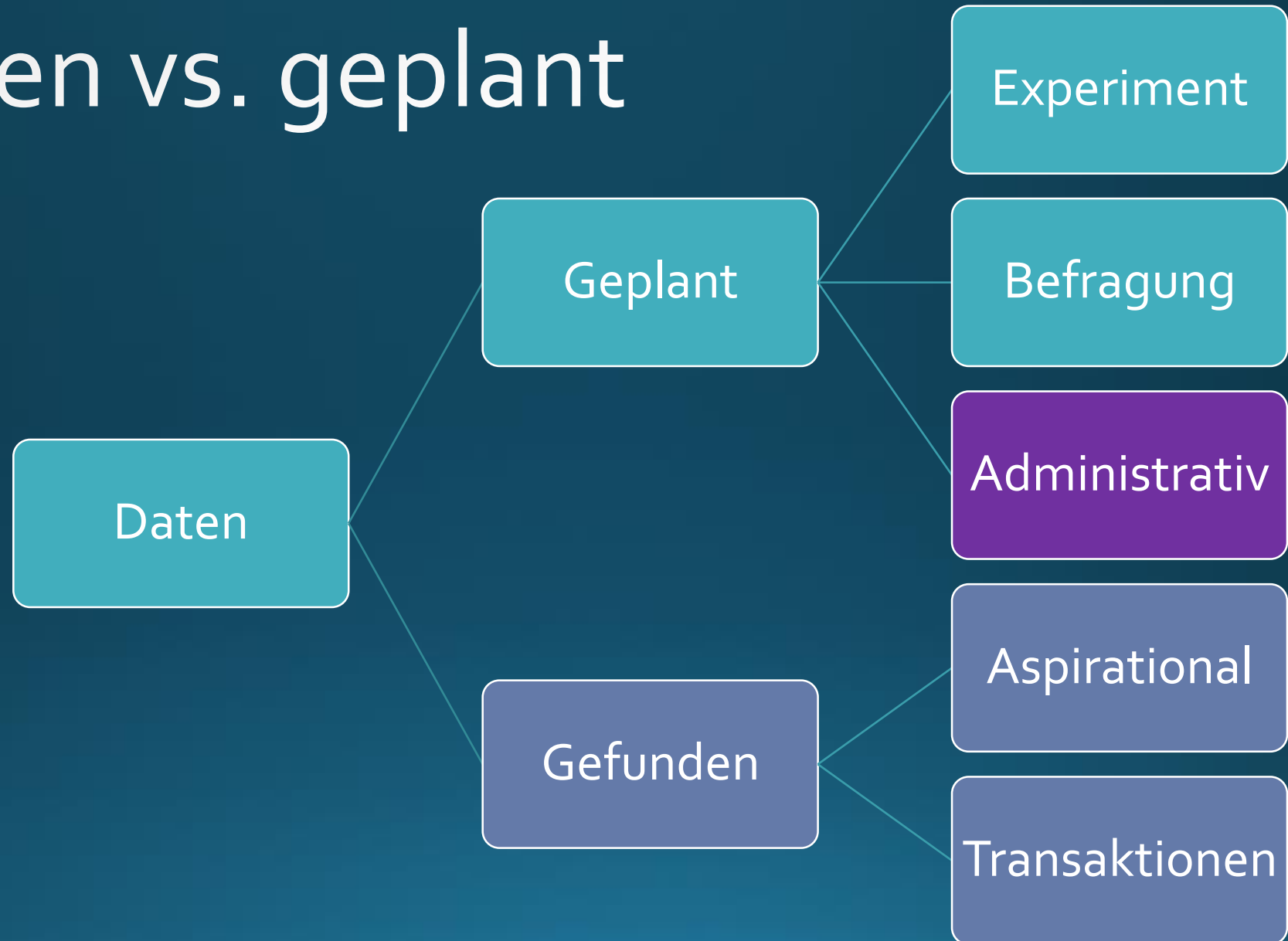
# Drei Paradigmenwechsel

D Digital erhoben

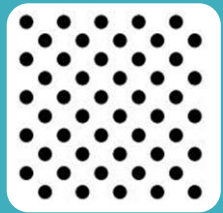
A Allen verfügbar

S Selbst entstanden

# Gefunden vs. geplant

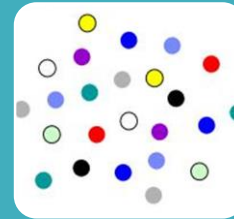


# Eigenschaften der neuen Daten



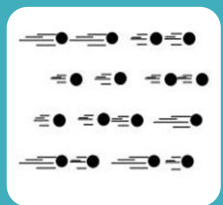
## “Volume”

- Große Mengen
- Groß ist relativ



## “Variety”

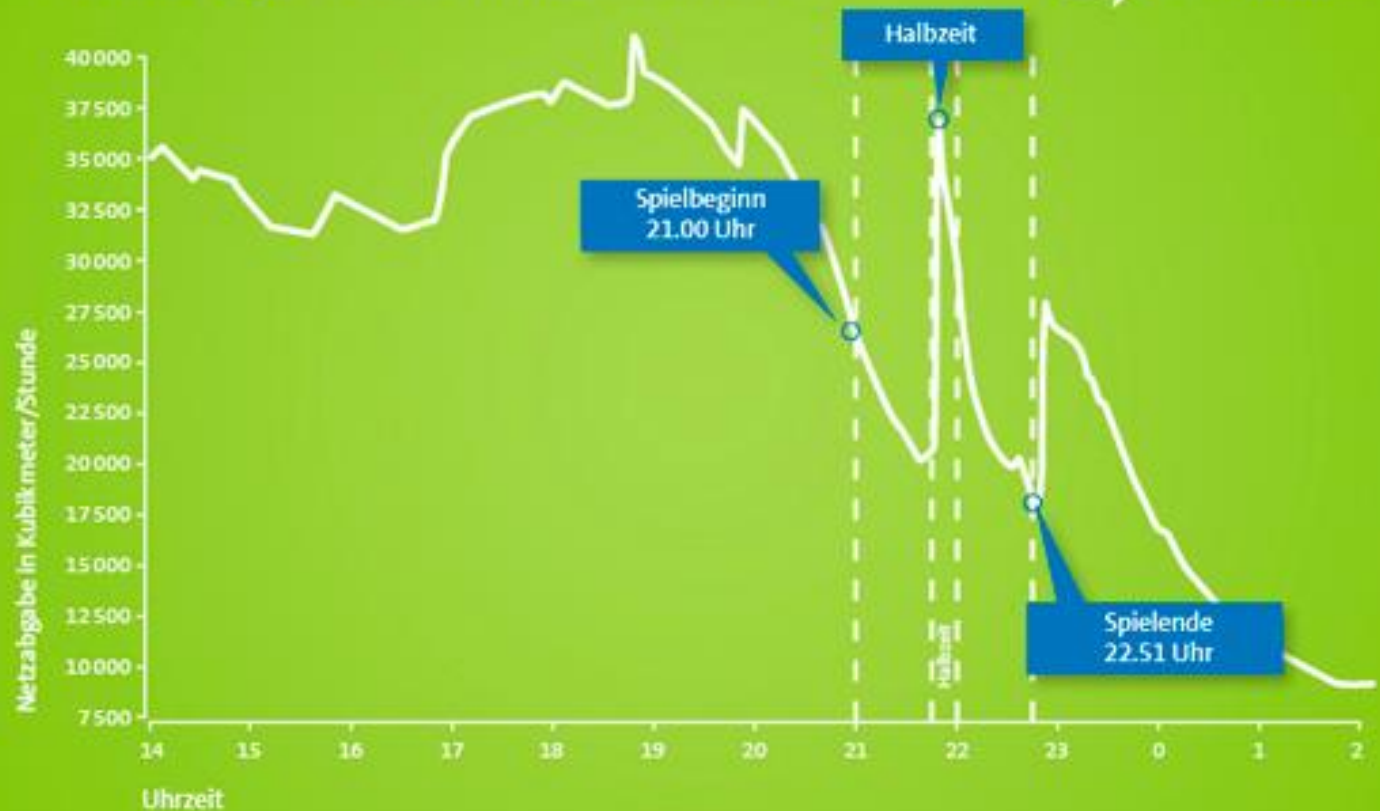
- Vielzahl an Formen
- Text, Audio, Bilder



## “Velocity”

- Daten in Bewegung
- Ständige Updates

## Netzabgabe für den 12. Juni 2016



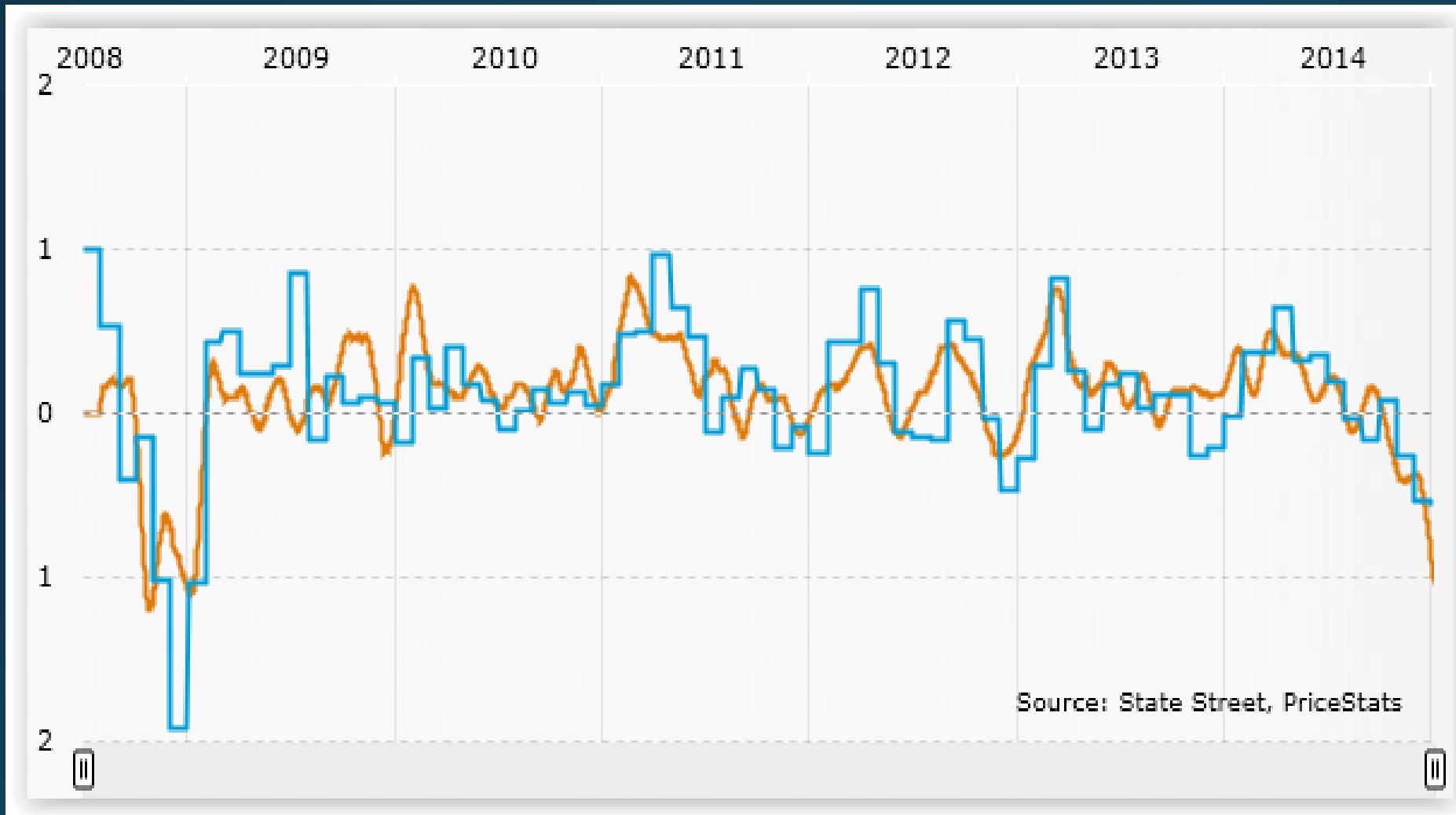
Beispiel Wasser

Volume  
Netzabgabe  
Kubikmeter pro  
Stunde

Velocity  
Kontinuierliche  
Messung

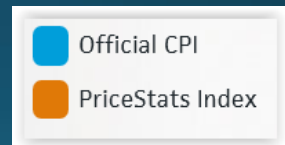
Variety  
Vielzahl an Formen  
(Abwasser)





USA  
 Aggregierte  
 Inflationsraten  
 (monatlich)

PriceStats  
 Index und  
 Daten des  
 Bureau of  
 Labor Statistics



# Hoffnung:

Digitalisierung ersetzt teure Datenerhebung,  
bietet mehr (bessere) Daten als Entscheidungsgrundlage und  
erlaubt Zugriff auf Informationen in Echtzeit

# Was sollte man können?

**Datenzugang/  
Info-Weitergabe**

**Datenanalyse**

**Datenaufbereitung/  
Speichern**

**Daten generierender  
Prozess**

**Fragestellung**



Visualisierung, Kommunikation, Datenschutz

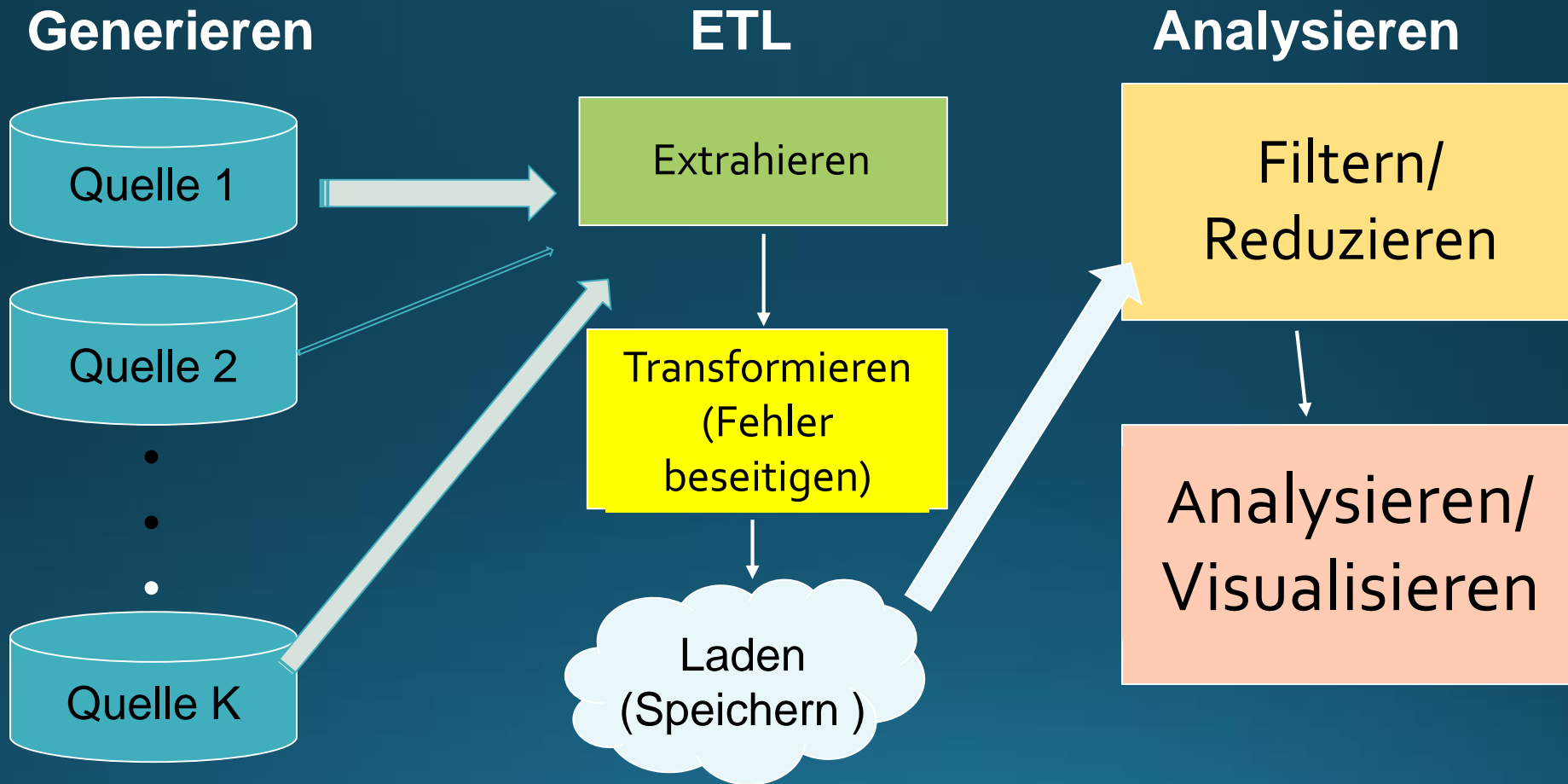
Machine Learning, Regression, Inferenz

Datenbank Management, Record Linkage etc.

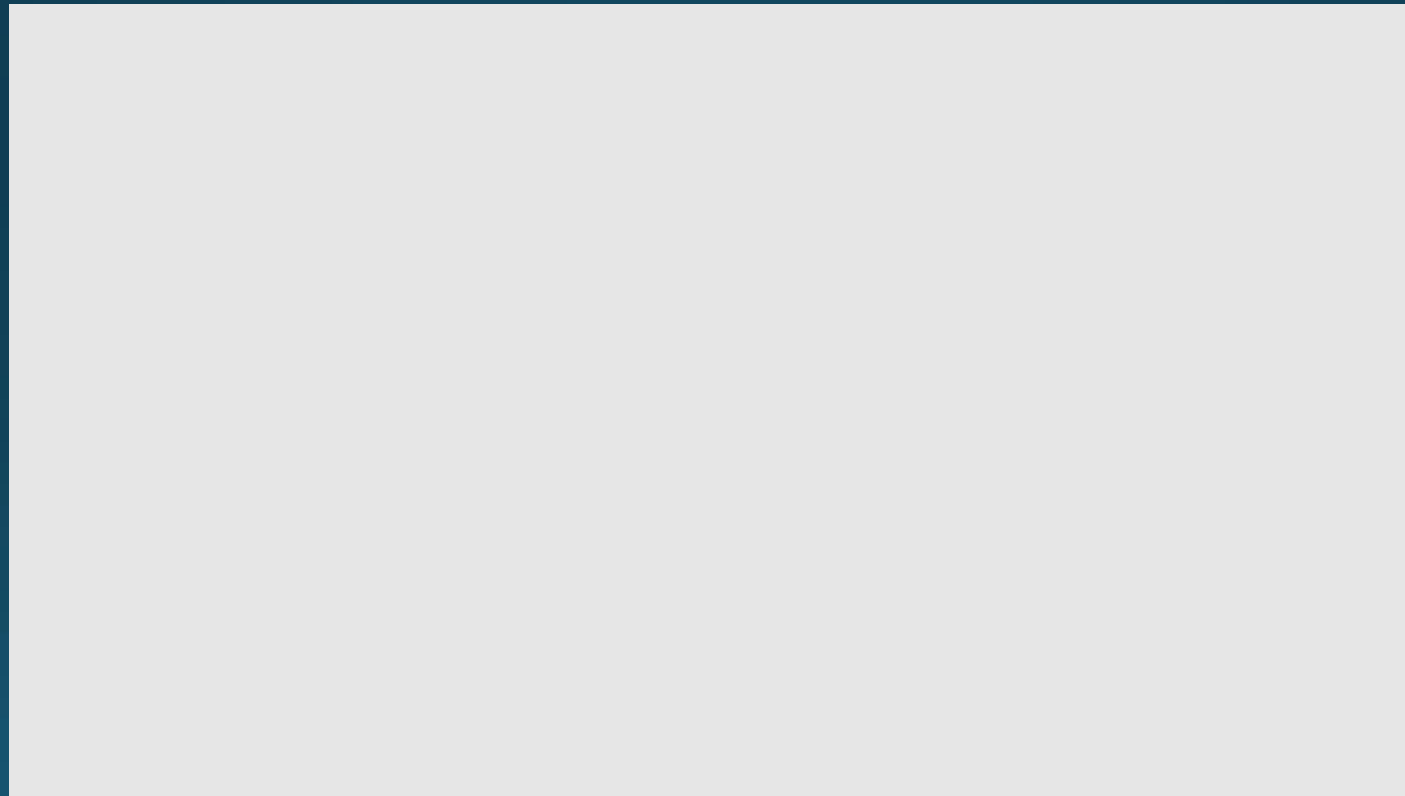
Erhebungen, Scraping, Linkage, Monitoring

Beschreiben, Verstehen, Erklären

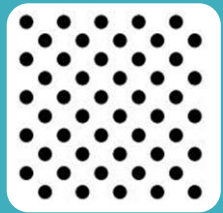
# Big Data Prozess



# Beispiel: Daten generierender Prozess

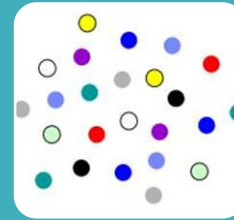


# Die vierte Eigenschaft!



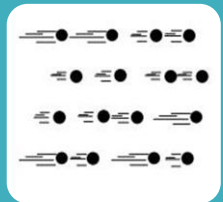
## Volumen

- Große Mengen
- Groß ist relativ



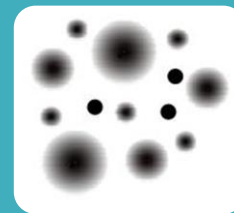
## Variety

- Vielzahl an Formen
- Text, Audio, Bilder



## Velocity

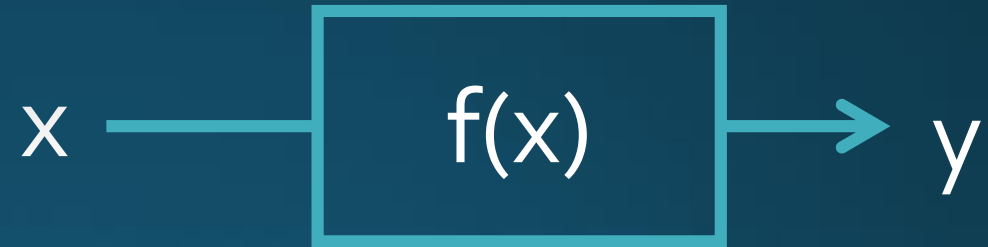
- Daten in Bewegung
- Ständige Updates



## Veracity

- Fehlerhafte oder
- Fehlende Daten

# Fragestellung: Vorhersage vs. Verstehen?





# Wo findet man Leute?

# Expertise im Team

Inhaltliche  
Expertise

Statistik



Informatik

System  
Administration

# Personal(entwicklung)

Weiterbildung


Kooperationen

Unkonventionelle Fächer

# Berufsbegleitende Weiterbildung


- Arts and Humanities
- Business
- Computer Science
- Data Science
- Life Sciences
- Math and Logic
- Personal Development
- Physical Science and Engineering
- Social Sciences
- Language Learning

## Data Analysis




Data  
10-co

## Data Science




Data  
6-cou  
Univer

## Excel




Exce  
5-cou

## Statistik




Statis  
5-cou


## Business



Busir  
5-cou



- Home
- Program
- Project
- Contact
- FAQs
- Imprint
- Search
- [APPLY NOW](#)



## INTERNATIONAL PROGRAM IN SURVEY AND DATA SCIENCE

offered through the University of Mannheim and the Joint Program in Survey Methodology (Universities of Maryland and Michigan, Westat)

[BE PART OF IT](#)

● ○ ○

We are pleased to announce the launch of the International Program in Survey and Data Science (IPSDS). Fundamental changes in the nature of data, their availability, the way in which they are collected, integrated, and disseminated are a big challenge for all those working with designed data from surveys as well as organic data. IPSDS was developed in response to the increasing demand from researchers and practitioners for the appropriate methods and right tools to face these changes. We offer a multidisciplinary curriculum, world-class faculty, and a web-based learning environment that allows you to take courses from anywhere in the world.

# Kooperation: Projekte und Praktika

The screenshot shows the website for 'Elitenetzwerk Bayern'. At the top, there are navigation options for 'English', 'Gebärdensprache', and 'Leichte Sprache'. Below the header, there are four main program categories: 'ELITESTUDIENGÄNGE IM FREISTAAT BAYERN', 'MAX WEBER-PROGRAMM DES FREISTAATES BAYERN', 'INTERNATIONALE DOKTORANDENKOLLEGS', and 'NACHWUCHSFORSCHERGRUPPEN UND FORSCHUNGSSTIPENDIEN'. The 'Data Science' program is highlighted in red. The left sidebar contains a navigation menu with options like 'Überblick', 'Aktuelles', 'Im Detail', and 'Elitestudiengänge nach Wissenschaftsbereichen'. The main content area displays the 'Data Science' program details, including the participating universities (Ludwig-Maximilians-Universität München, Technische Universität München, Universität Augsburg), the degree type (Master of Science), duration (4 semesters), location (Munich), and language (English). Contact information for Prof. Dr. Göran Kauermann and Prof. Dr. Thomas Seidl is provided, along with a search bar and a 'Weitere Informationen' link.

Elitenetzwerk Bayern

English | Gebärdensprache | Leichte Sprache | Kontakt

ELITENETZWERK HOME

ELITESTUDIENGÄNGE IM FREISTAAT BAYERN

MAX WEBER-PROGRAMM DES FREISTAATES BAYERN

INTERNATIONALE DOKTORANDENKOLLEGS

NACHWUCHSFORSCHERGRUPPEN UND FORSCHUNGSSTIPENDIEN

Überblick

Aktuelles

Im Detail

Elitestudiengänge nach Wissenschaftsbereichen

Elitestudiengänge nach Studienorten

Bewerbung

ELITESTUDIENGÄNGE

Alle Elitestudiengänge

Data Science

Ludwig-Maximilians-Universität München (Sprecheruniversität), Technische Universität München, Universität Augsburg

DATA SCIENCE@LMU

STATISTICS + INFORMATICS

Abschlussart: Master of Science

Studiendauer: 4 Semester; Beginn jeweils zum Wintersemester

Studienort: München

Unterrichtssprache: Englisch

Zulassungsvoraussetzungen: Bachelor of Science (oder Gleichwertiges) in Statistik oder Informatik oder verwandten Disziplinen

Bewerbungsfrist: wird noch bekanntgegeben

Kontakt

Prof. Dr. Göran Kauermann

Tel.: 089-2180-2220

goeran.kauermann@statistik.uni-muenchen.de

Prof. Dr. Thomas Seidl

thomas.seidl@dbs.ifl.lmu.de

Dr. Constanze H. Schmalzing (Koordinatorin)

Tel.: 089-2180-2232

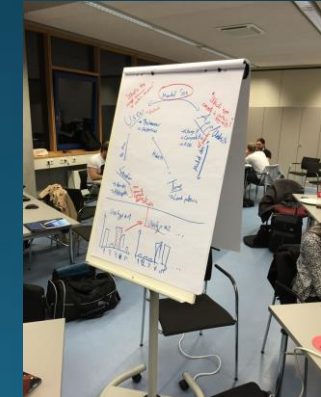
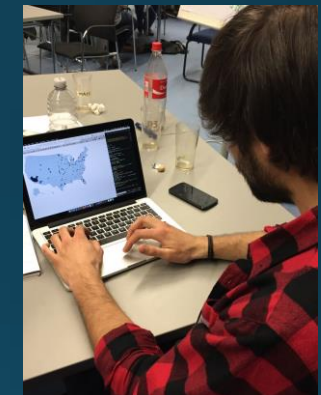
m-datascience@lmu.de

Weitere Informationen

datascience-munich.de

suchen

# Crowd Sourcing und Wettbewerbe



# In Kürze

1. Digitalisierung ist eine Chance für datengestützte Entscheidungsfindung
2. Dazu müssen Daten nicht nur verarbeitet sondern auch verstanden werden
3. Gezielte Weiterbildung und Kooperation kann vorhandene Lücken schließen

Freue mich auf eine gute Zusammenarbeit!