



Digitale Verhaltensdaten bei GESIS

KSWD 2020 Berlin, 03.03.2020

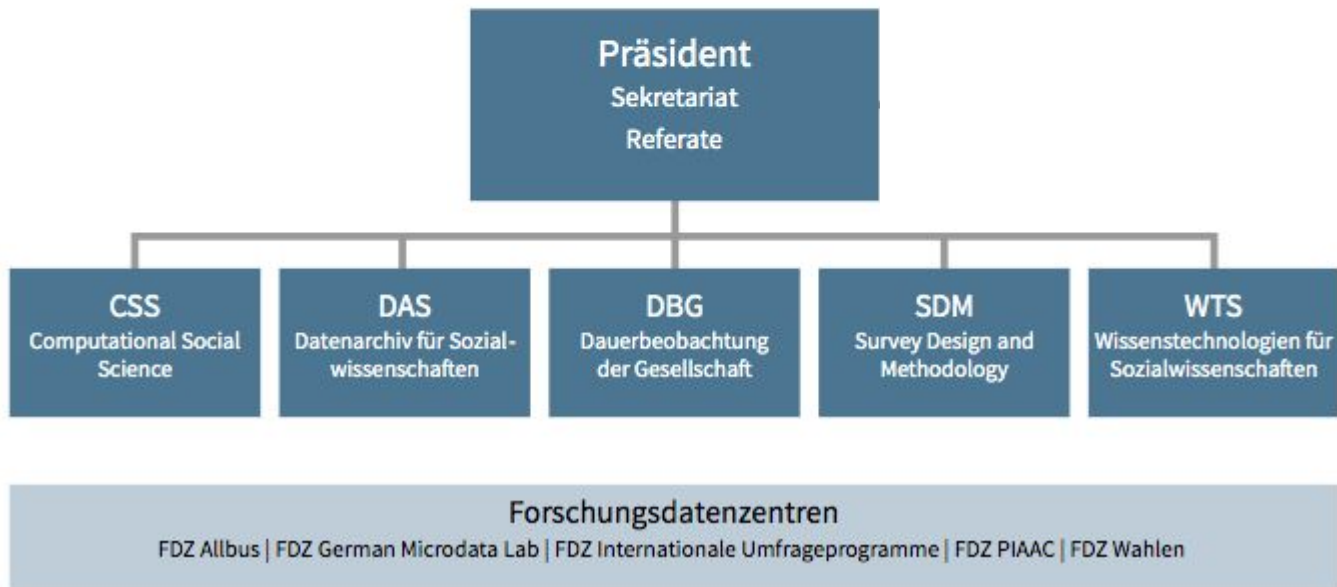
Dr. Fabian Flöck, Team Data Science, Abteilung CSS, GESIS Köln

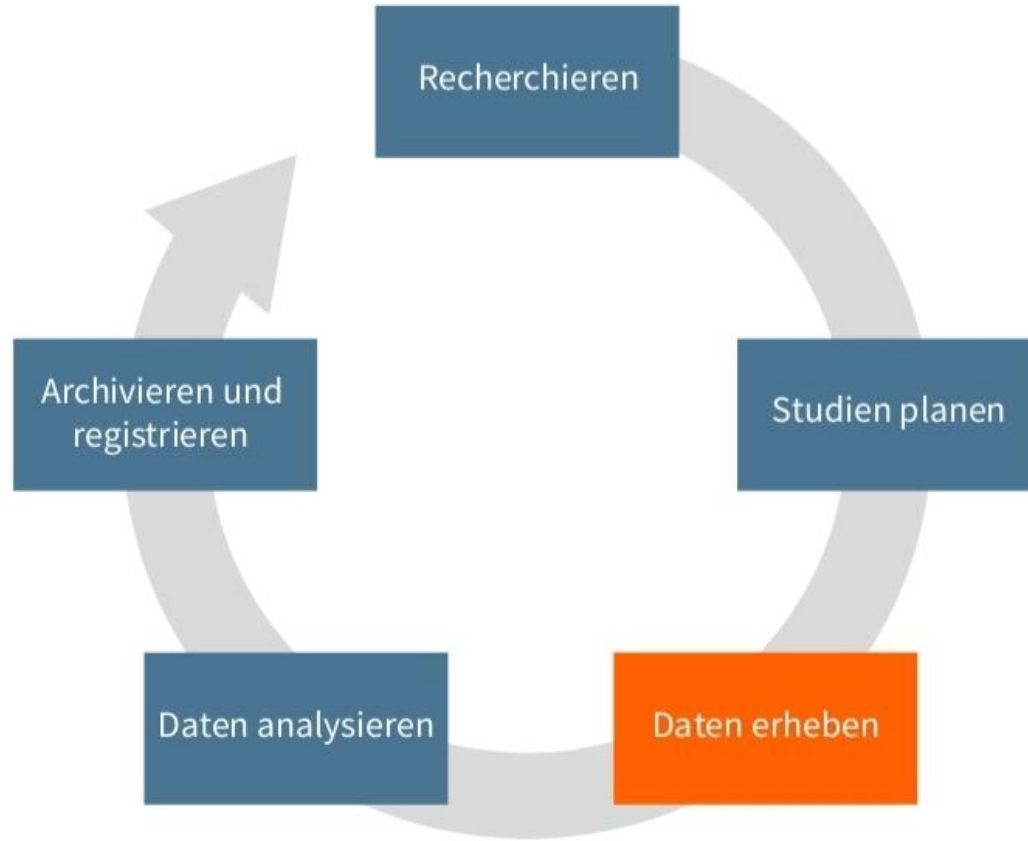


Digitale Verhaltensdaten (Spurdaten) bei GESIS

KSWD 2020 Berlin, 03.03.2020

Dr. Fabian Flöck, Team Data Science, Abteilung CSS, GESIS Köln

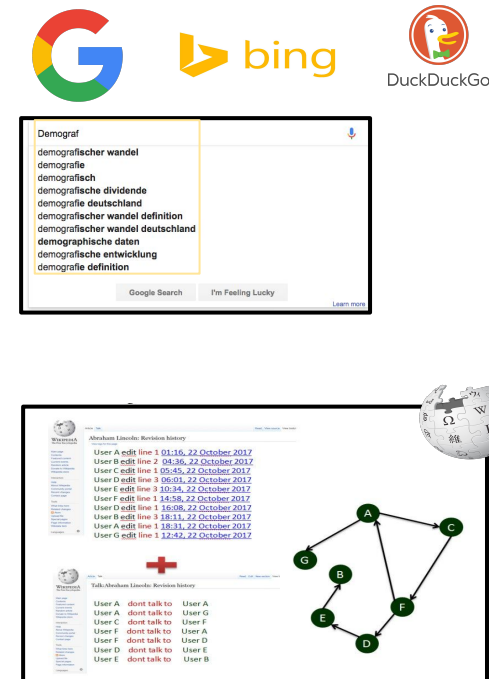




Forschungsthemen

Offline-Pendants

Online-spezifisch



Forschungsthemen

Offline-Pendants



Online-spezifisch



fiverr

Online- = Offline-
Verhalten?
Komplement vs.
Ersatz

Verbreitung von
Information/Trends/Fake
News/... ?

What does this mean, where a little grey
icon appears beside a user's profile pic on
some tweets but not others?

Diskursdynamiken?
Eskalation/Hate Speech?

Replying to @john_self
Who started the thread?

08:34 · 19/07/2019 from Dublin City, Ireland · Twitter
for iPhone

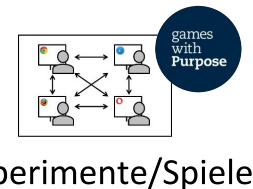
Interaktion von
Mensch und
System/KI?

Online-
Kollaboration/
Rollen?

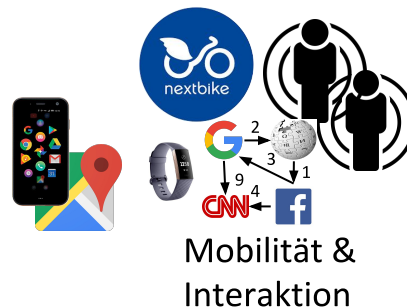


amazon

Digitale Spurdaten: Thematische Unterscheidung



Umfragen



Registerdaten

Digitale Spurdaten: Thematische Unterscheidung



Digitale Spurdaten: Erhebungsmodus



Digitale Spurdaten: Erhebungsmodus



Digitale Spurdaten: Erhebungsmodus



“Found”



Nachrichten/
Medien



Soziale Medien/
Netzwerke



mechanical turk

“Arbeitsmärkte”



games
with
Purpose

Experimente/Spiele



Umfragen



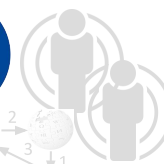
Organisations
-intern



Kulturelle Inhalte
(Essen, Musik, ..)



Persönliche
Kommunikation

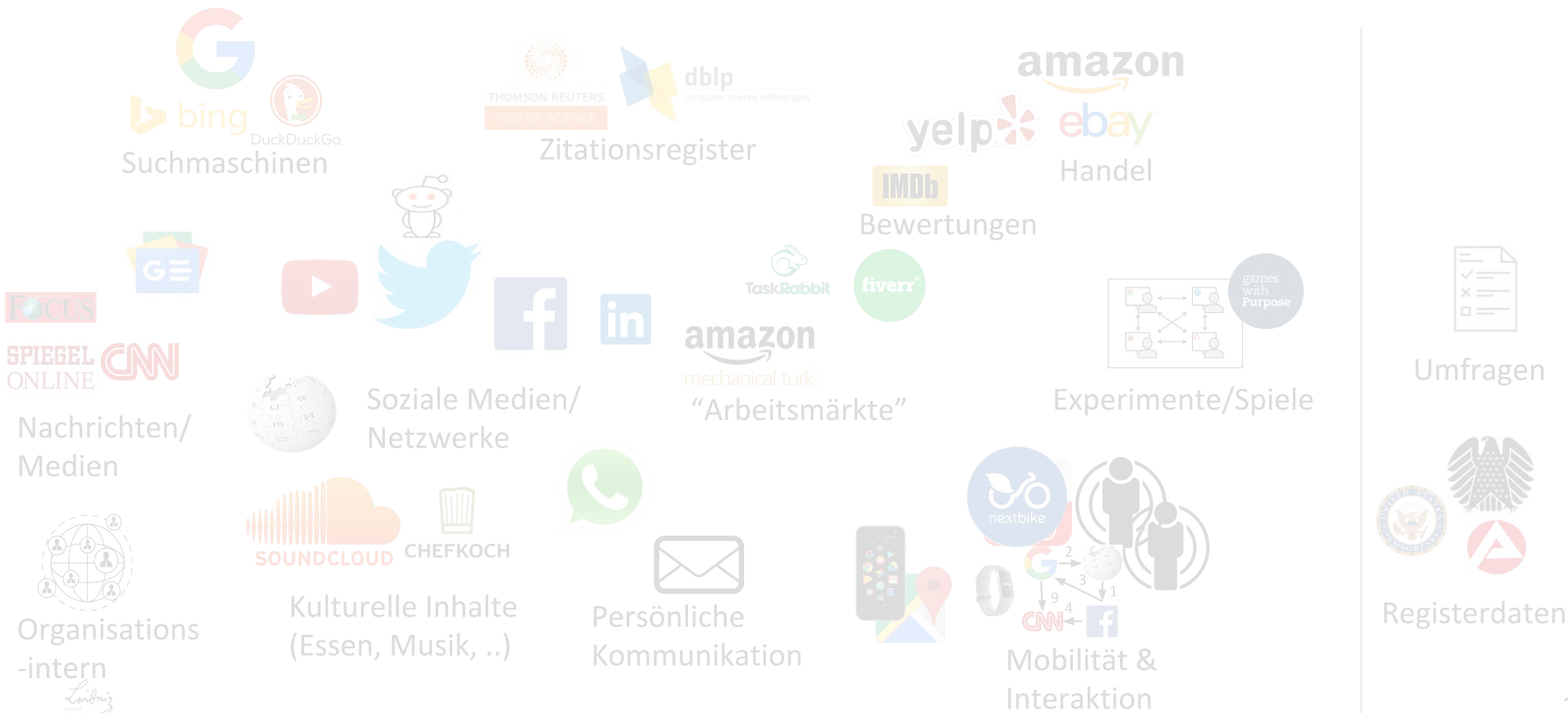


Mobilität &
Interaktion

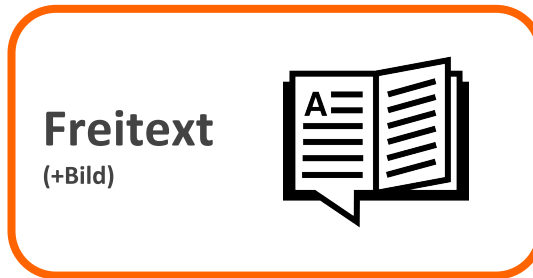


Registerdaten

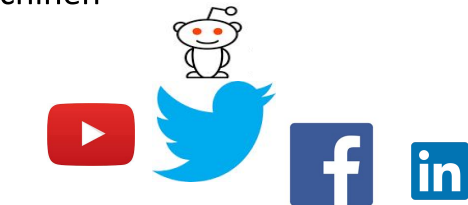
Digitale Spurdaten: Datenarten



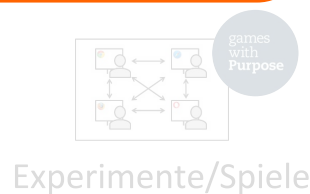
Digitale Spurdaten: Datenarten



Nachrichten/
Medien



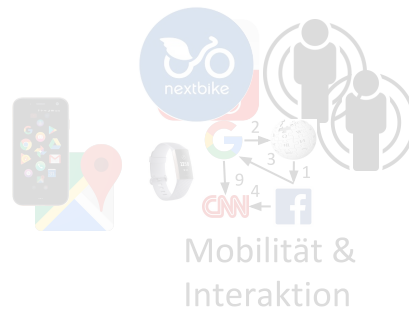
Soziale Medien/
Netzwerke



Kulturelle Inhalte
(Essen, Musik, ..)



Persönliche
Kommunikation



Digitale Spurdaten: Datenarten



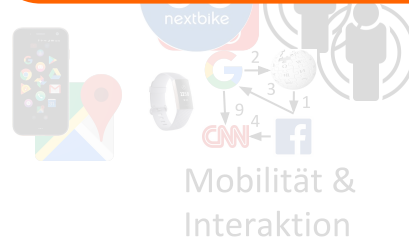
Freitext
(+Bild)

Aufb. via NLP

Wort-häufig-keiten
Stance: Against
dbr:Angela_Merkel



Persönliche
Kommunikation



Umfragen



Registerdaten

Digitale Spurdaten: Datenarten



Digitale Spurdaten: Datenarten

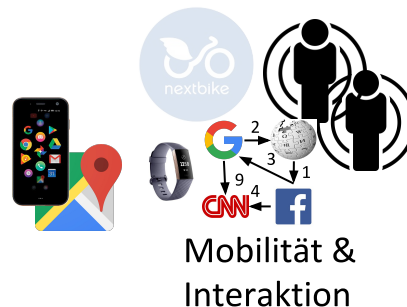
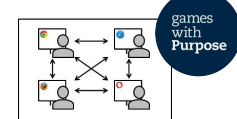
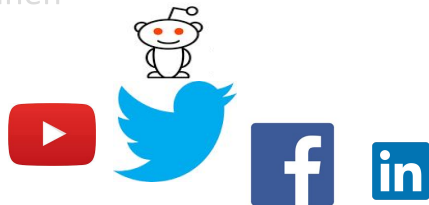


Umfragen



Registerdaten

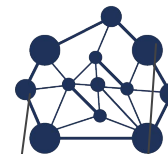
Digitale Spurdaten: Datenarten



Digitale Spurdaten: Datenarten



Netzwerke



**Anno-
tation**

“Nepotismus-
Cluster”
besonders
defensiv



Digitale Spurdaten: Datenarten



Digitale Spurdaten: Datenarten



Umfragen

Registerdaten

Konkrete Ansätze

Archivierung/Nachnutzung

Social Media

- Vorgehen je nach TOS, Datenstruktur, Aufbereitung und Personenkreis
- Bislang v.a. Twitter (u. FB/ Wikipedia)
- Entwicklung Datenschutzprüfungsprozess
- Arten Archivierung/Zugang:
 - Komplett
 - IDs+“Rehydrierung” (v.a. Twitter)
 - Indexe (ü. Abfragen) bzw. nur abstrahierte Annotationen von Inhalten

mediamonitoring.gesis.org

gesis

Leibniz-Institut
für Sozialwissenschaften

Home Politicians API Organizations API Documentation

Social Media Monitor (beta)

Track activity of politicians and
organizations related to the *German
General Election 2017*®

Explore:
Politicians

Select:

Peter Altmaier

01-09-2017

31-12-2018

☐ Daily ☒ Weekly ☐ Monthly

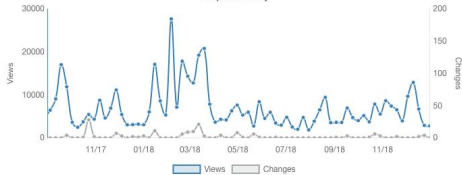
Hint

Use the [Politicians API](#) and the [Organizations API](#) to create your own visualizations. Examples of how to **reproduce these visualizations** are provided by clicking on the 📌 and 📄 icons next to the plots. Or, simply download the csv data of the plots by clicking on the 📄 icon.

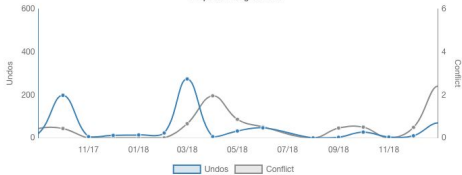
Twitter (Tweets and Replies Activity)



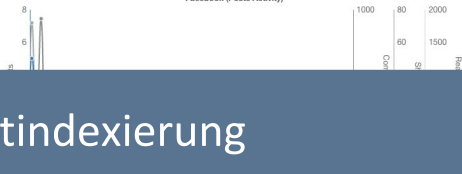
Wikipedia Activity



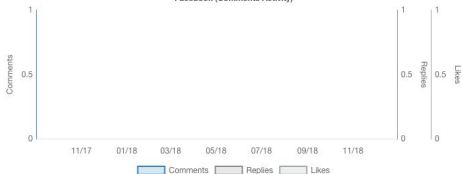
Wikipedia Disagreement



Facebook (Posts Activity)



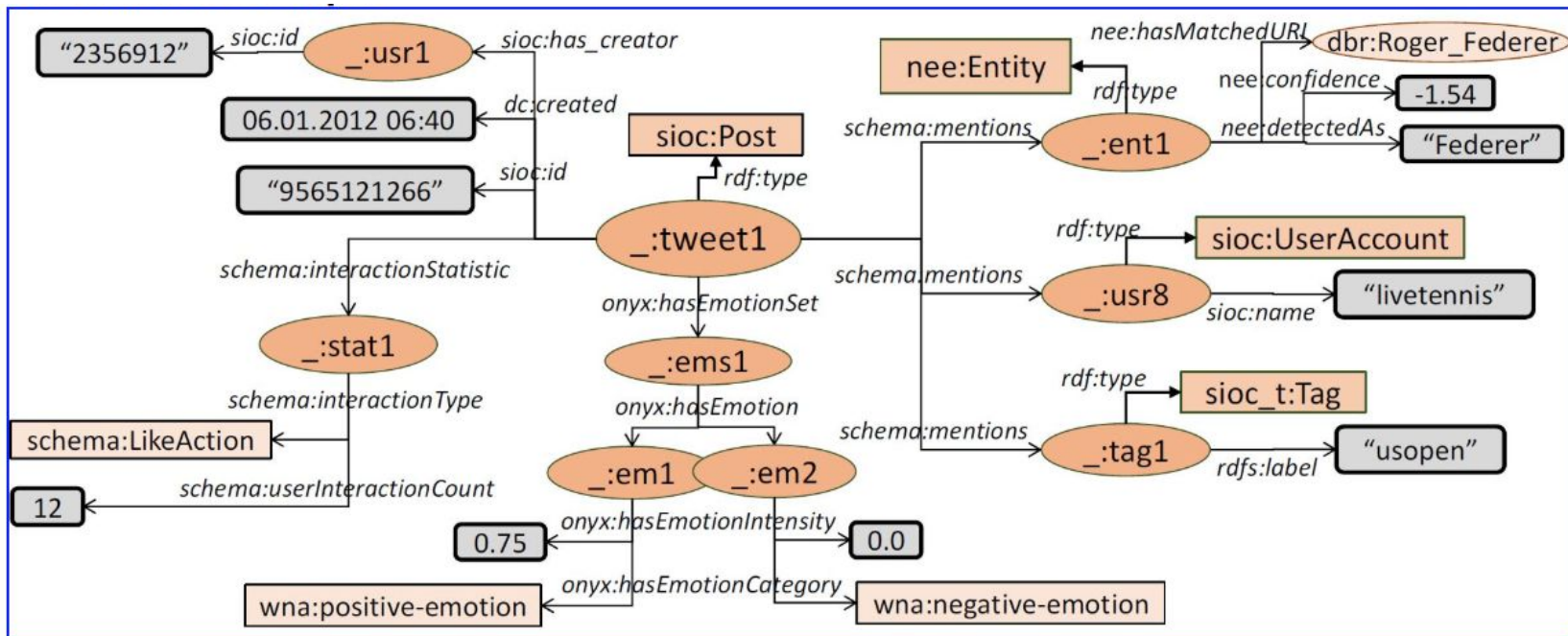
Facebook (Comments Activity)



Suche durch Textindexierung

TweetsKB

1,5 Milliarden Tweets mit Entitäten, Sentiments



Social Media

- Eruierung Datenfernverarbeitung (geeignet für Python/R und NLP-Libraries) über GESIS Notebooks
- Eruierung Zugang über GESIS Secure Data Center

Webtracking

- Bisläng Kauf von kommerziellen Anbietern, Nutzung nur für Koautoren, nur im Haus
 - Möglichkeiten der Weitergabe von Aggregaten wird eruiert (z.B. Exploration v. Trends per Strata in Interface)
- In Zukunft mglw. Angebot eines Webtracking-Tools (Software)
- Webinhalte: Weitergabe unklar

SocioPatterns

- Public Use Files wenn voll anonymisierbar
 - Nur sehr aggregierte Fragebogendaten
- Verknüpfung mit vollen Fragebogendaten über Secure Data Center / Fernverarbeitung angedacht

Generell

- Starker Verlass auf verfügbare APIs/ Webinhalt und konstanten Zugang
- Standards für Metadaten nicht einheitlich
- Plattform-Kontexte nicht erfasst (v.a. soziotechnisches System)
- Wenig Standards zur Dokumentation von Sammlungs- und Aufbereitungsmethoden und deren Fehlerraten