

gesis

Leibniz-Institut
für Sozialwissenschaften



The German Federal Election: Social Media Data for Scientific (Re-)Use

Meet the Experts! – GESIS online talks

Marius Sältzer & Sebastian Stier · October 14, 2021

Speakers



Dr. Marius Sältzer

- Postdoctoral Researcher in the team Designed Digital Data, Department Computational Social Science
- Ph.D. in political science, University of Mannheim
- Political parties, party competition, intra-party politics
- Contact: marius.saeltzer@gesis.org



Dr. Sebastian Stier

- Senior Researcher in the team Designed Digital Data, Department Computational Social Science
- Ph.D. in political science, University of Heidelberg
- Political communication, political behavior, linking surveys and digital behavioral data
- Contact: sebastian.stier@gesis.org

Social media and elections

1. Why?

2. How to?



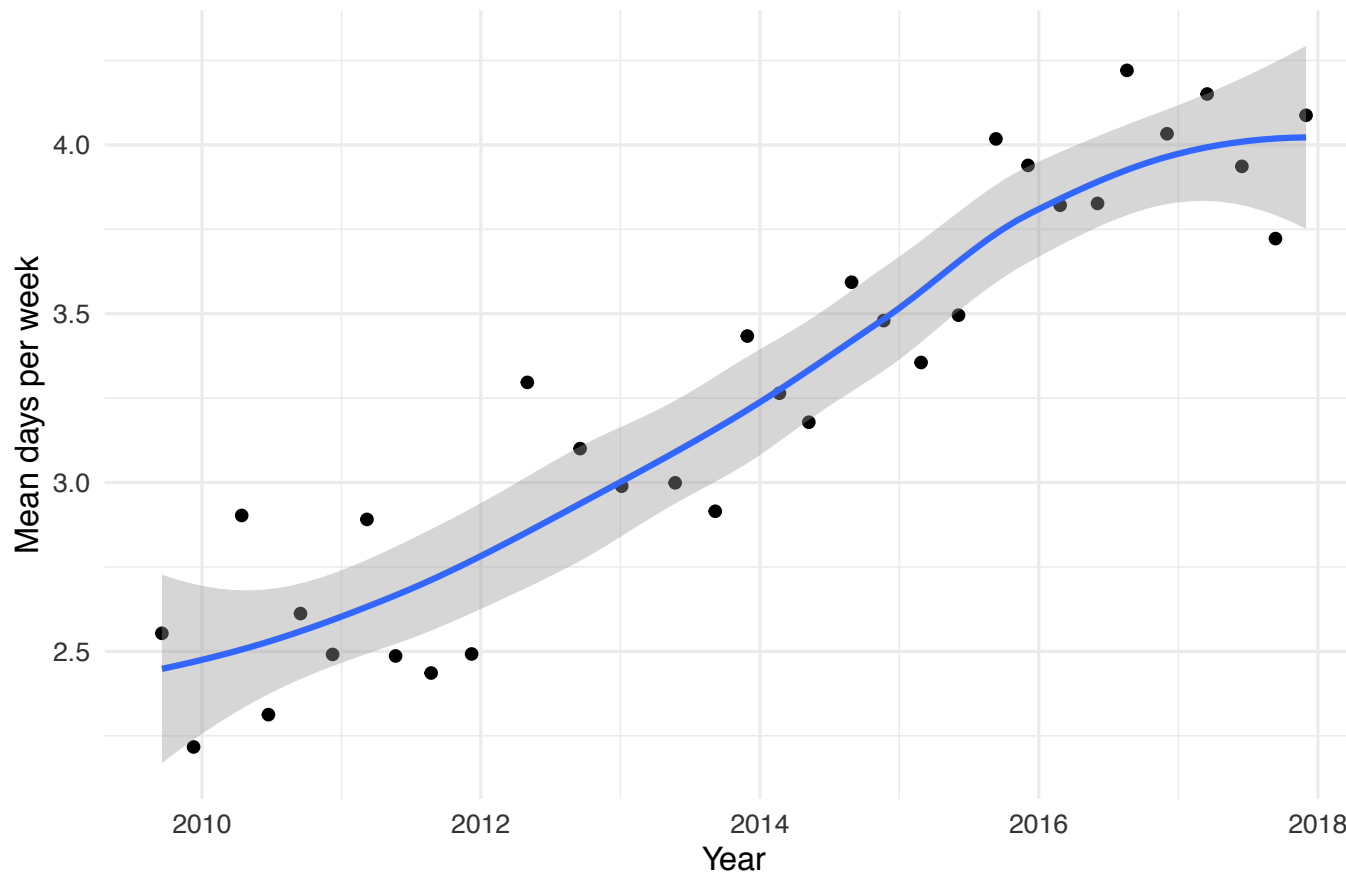
3. Applications

4. Data sharing

1. Why collect social media data on (German) elections?

Digital media as political information sources

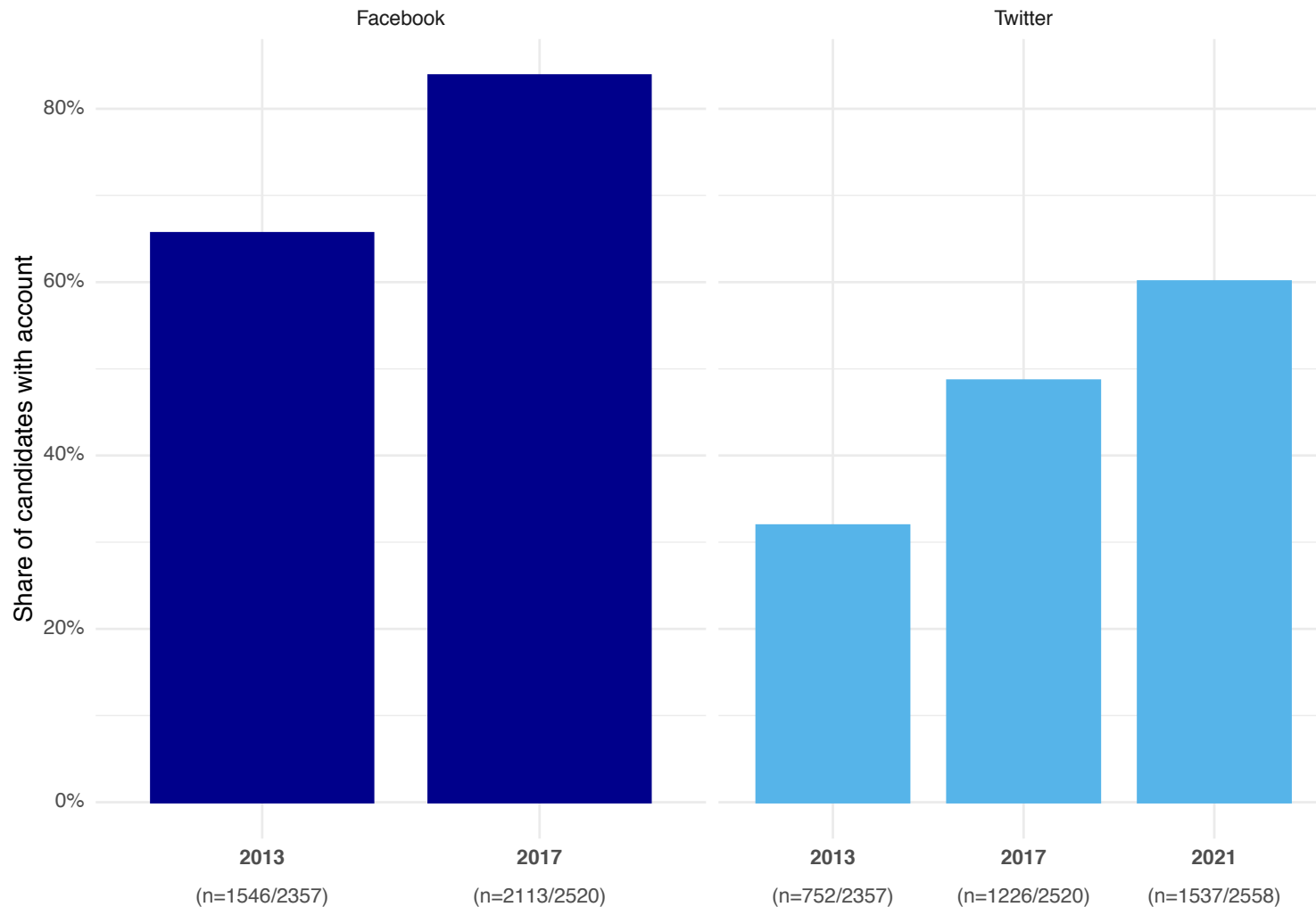
On how many days in the last week did you read about politics or political parties on the Internet?



German Longitudinal Election Study (GLES), Longterm-Online-Tracking, Cumulation 2009-2017.
GESIS Data Archive. ZA6832 Version 1.1.0. DOI: 10.4232/1.13416

The rise of social media campaigning

Social media accounts of candidates in German federal elections, 2013–2021



The rise of social media campaigning

Demand

- Increasing diffusion of social media in society
- Use as information medium
- Interactive
- Corona pandemic

Supply

- Internet-savvy generation of politicians
- Hybrid media system
- Targeted advertising
- Corona pandemic

Scientific use

Substantive

- New type of communication
- SM as a social network
- New phenomena: mobilization, polarization, hate speech, ...

Methodological

- Easy access to large data sets
- Fine grained data (text, images, metadata)
- Complete time series
- All relevant actors in one arena: politicians, parties, voters, media orgs, journalists

2. How to collect social media data?

Conceptualization: What to collect?

Audience-centered

- Keywords (e.g., #btw17, #triell, klimakrise)
- Comments/ @mentions on a specific account
- Specific “issue public” and all participants
- Limited information about users

Elite-centered

- User accounts (parties, politicians)
- Full timelines
- All issues from a specific set of users
- Information about users

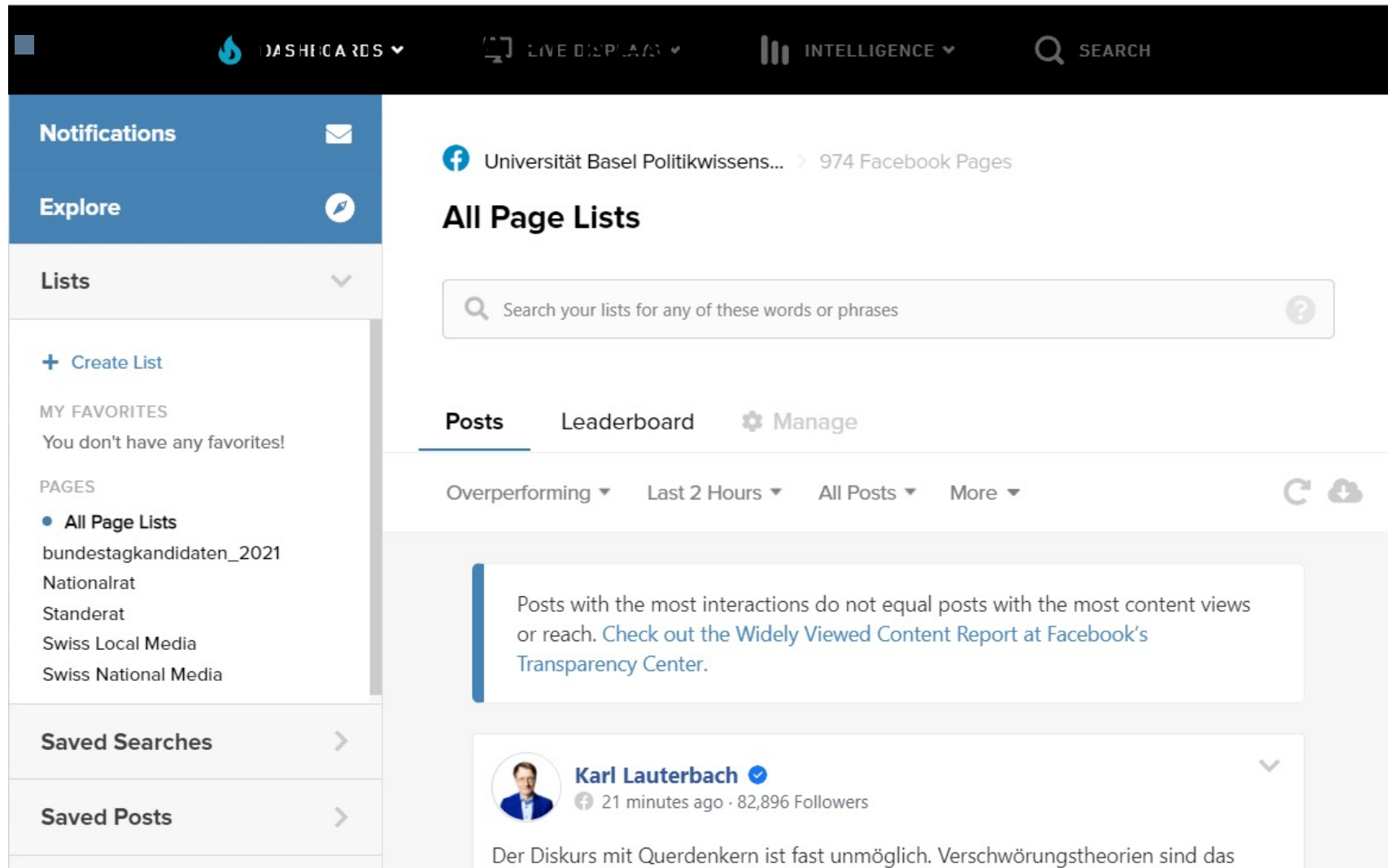
Facebook & Instagram



CrowdTangle

- Provided via Facebook
- Register as an academic
- Access
 - ▶ Interface (very user friendly)
 - ▶ Experimental Application Programming Interface (API)

CrowdTangle Interface

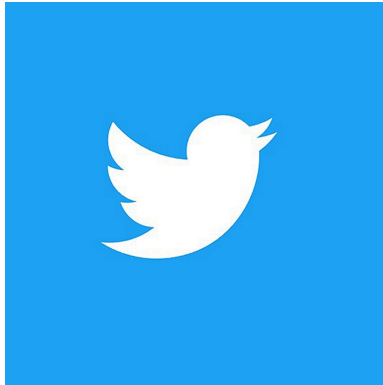


The screenshot displays the CrowdTangle interface. At the top, there is a navigation bar with icons for DASH CARDS, LIVE DISPLAYS, INTELLIGENCE, and a search function. On the left side, a sidebar menu includes 'Notifications', 'Explore', 'Lists' (with a dropdown arrow), '+ Create List', 'MY FAVORITES' (with a message 'You don't have any favorites!'), 'PAGES' (listing 'All Page Lists', 'bundestagkandidaten_2021', 'Nationalrat', 'Standerat', 'Swiss Local Media', and 'Swiss National Media'), 'Saved Searches', and 'Saved Posts'. The main content area shows a Facebook page list for 'Universität Basel Politikwissens...' with 974 pages. Below this is a search bar for lists. The 'Posts' tab is selected, showing a filter for 'Overperforming' and a time range of 'Last 2 Hours'. A notification banner states: 'Posts with the most interactions do not equal posts with the most content views or reach. Check out the Widely Viewed Content Report at Facebook's Transparency Center.' Below the notification, a post by Karl Lauterbach is visible, posted 21 minutes ago with 82,896 followers. The post text begins with 'Der Diskurs mit Querdenkern ist fast unmöglich. Verschwörungstheorien sind das'.

CrowdTangle

- Limitations
 - ▶ Only Public Pages
 - ▶ No comments
 - ▶ No reliable API access
 - ▶ No reliable access to images
- Dependency on Facebook's goodwill

Twitter API



- Collect hashtags (up to 1 week backwards)
- Most recent 3,200 tweets on a timeline
- Problematic:
 - ▶ large follower networks
 - ▶ rate limited: number of accesses

Twitter API

Academic Research product track

Advance your research objectives with public data on nearly any topic.

<https://developer.twitter.com/en/products/twitter-api/academic-research>

- Create Twitter Account
- Fill out a form
- Wait for Twitter: 1 to 7 days

Academic API

- Going back any time
- Both # and timelines
- Problems:
 - ▶ Requires a research affiliation
 - ▶ Heavy rate limits
 - ▶ Limited to 10mil. tweets/month

Best practice: Data storage

Small

- Autoparser included
- Save as CSV

Medium

- Store JSON
- Store in local SQL database

Large

- Run on Server / Cloud
- Automation via Cronjobs

3. Applications: German federal election campaigns on social media

The German Federal Election 2021 Twitter Dataset

- All candidates of the 7 major parties
- >1,200 Twitter accounts
- Here: all Tweets since 2021



https://search.gesis.org/research_data/ZA7721?doi=10.4232/1.13789

Use case: Individual behavior on SM

- Parties are not monolithic actors
- Important for party positioning
- Individuals differ!
 - ▶ Candidates versus MPs
 - ▶ Leaders versus Backbenchers
 - ▶ Regional differentiation
 - ▶ Wings inside parties

Use case: Individual behavior on SM

← **Hans-Georg Maaßen** Folgen
1.865 Tweets



Hans-Georg Maaßen @HGMaassen · 11. Sep. ⋮

„Die Migrationdpolitik seit 2015 war ein politischer Fehler, deshalb kämpfe ich für einen politischen Kurswechsel.“



🗨 216

↻ 454

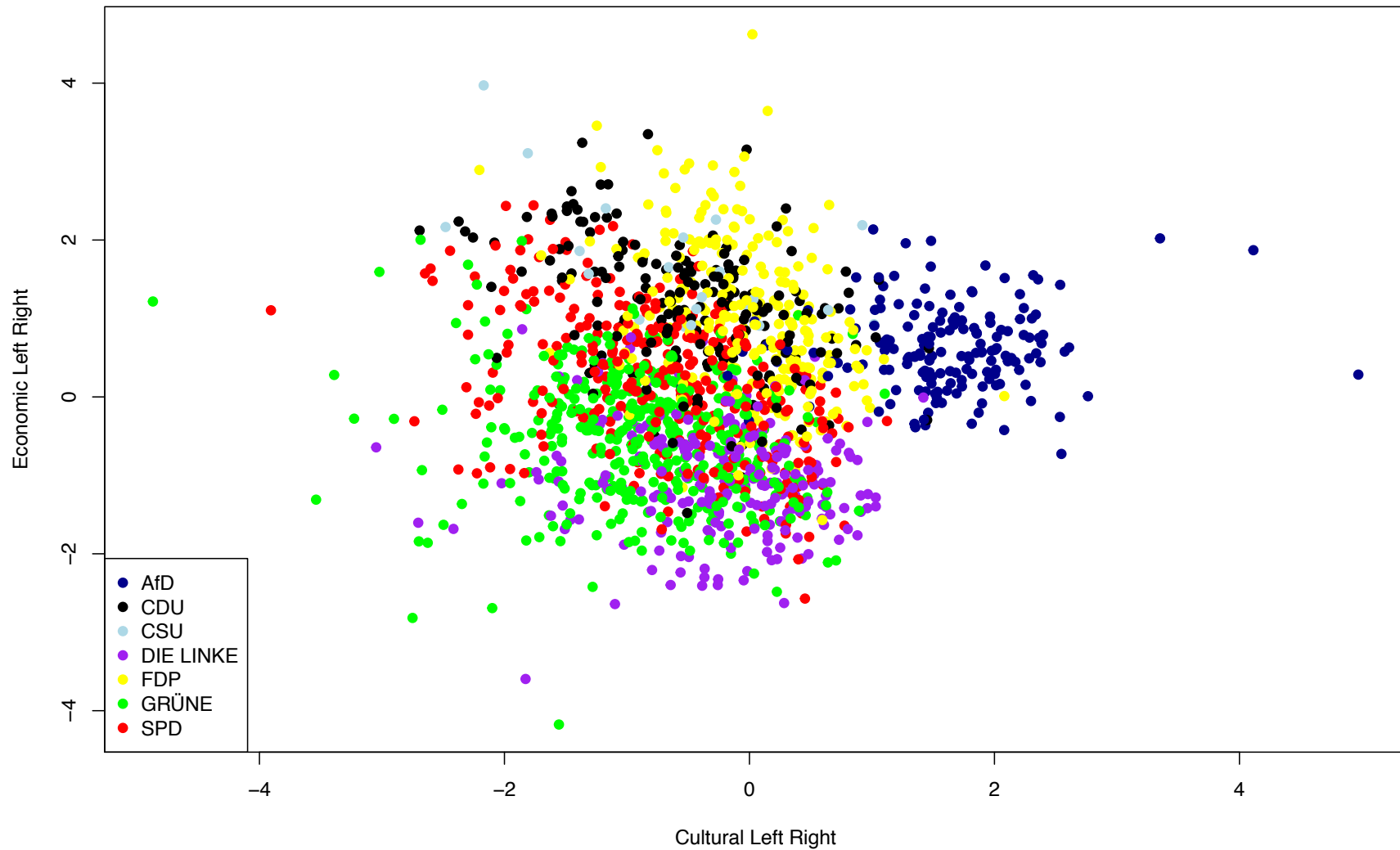
❤ 3.021

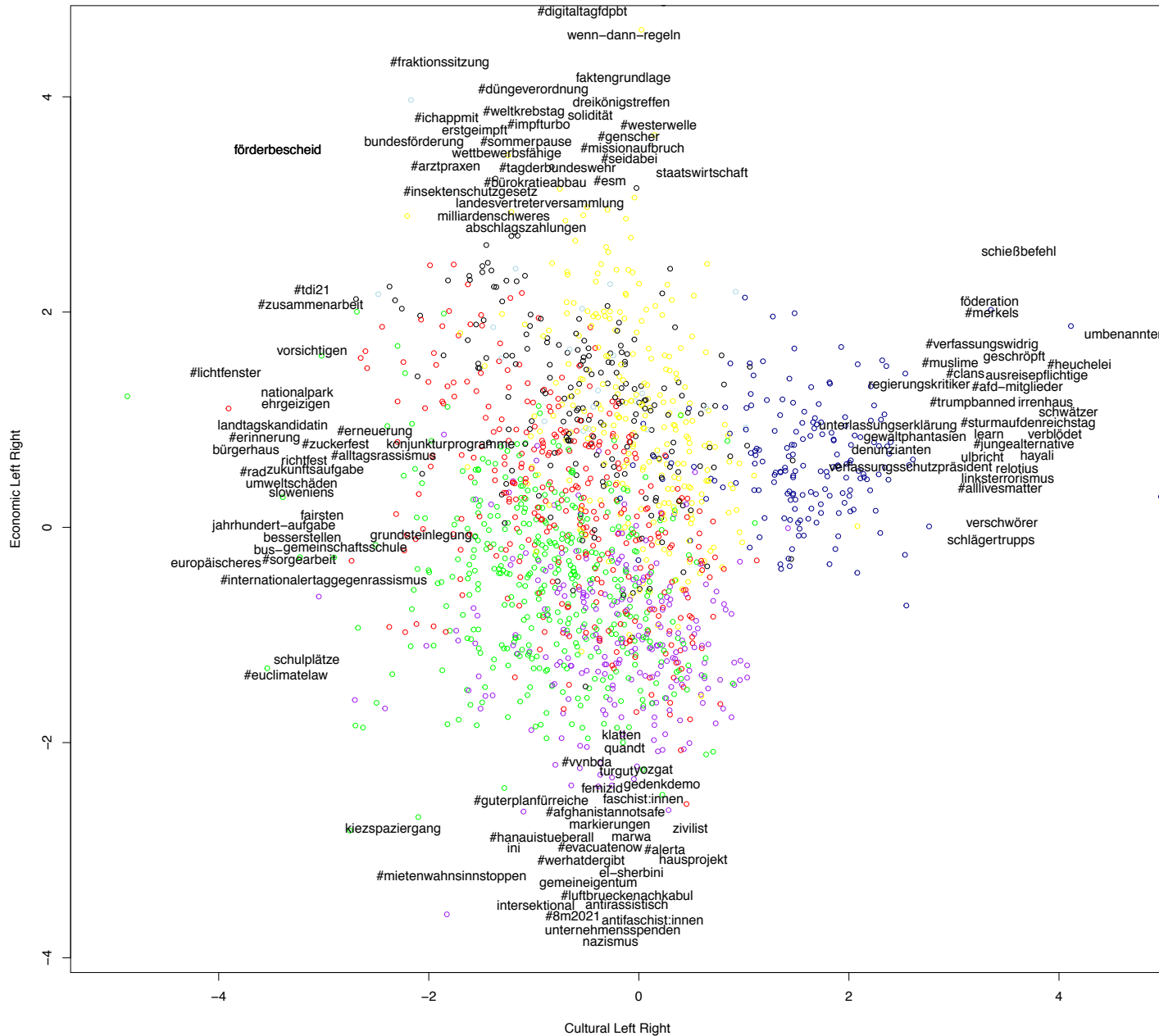


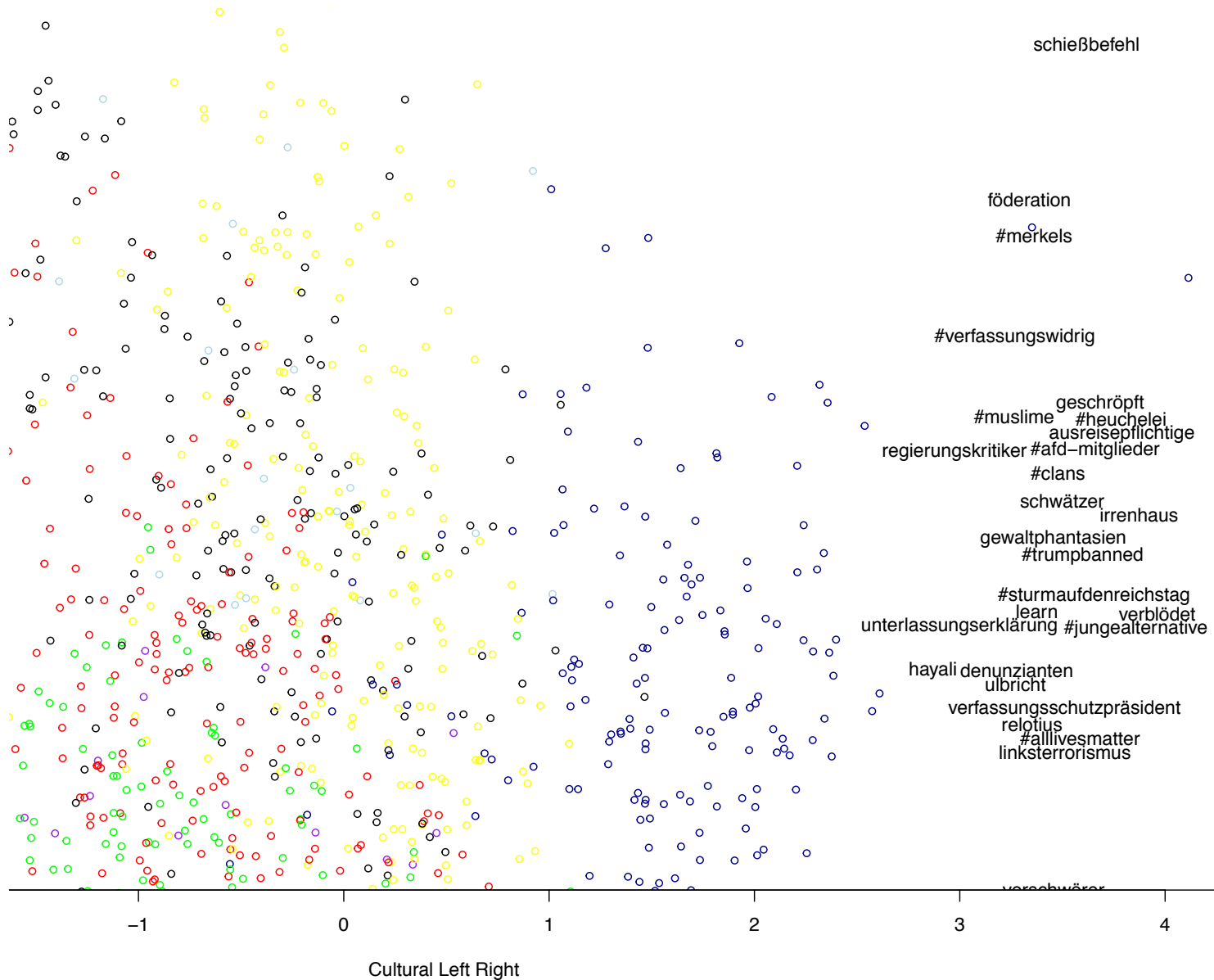
Approach: Text analysis

- What are candidates talking about and can we derive a political space from the terms they use?
- Analysis of all tweets
- Dimensional Reduction (Latent Semantic Analysis)
- Deriving the main dimensions of political conflict
- Interpretable as a left-right dimension

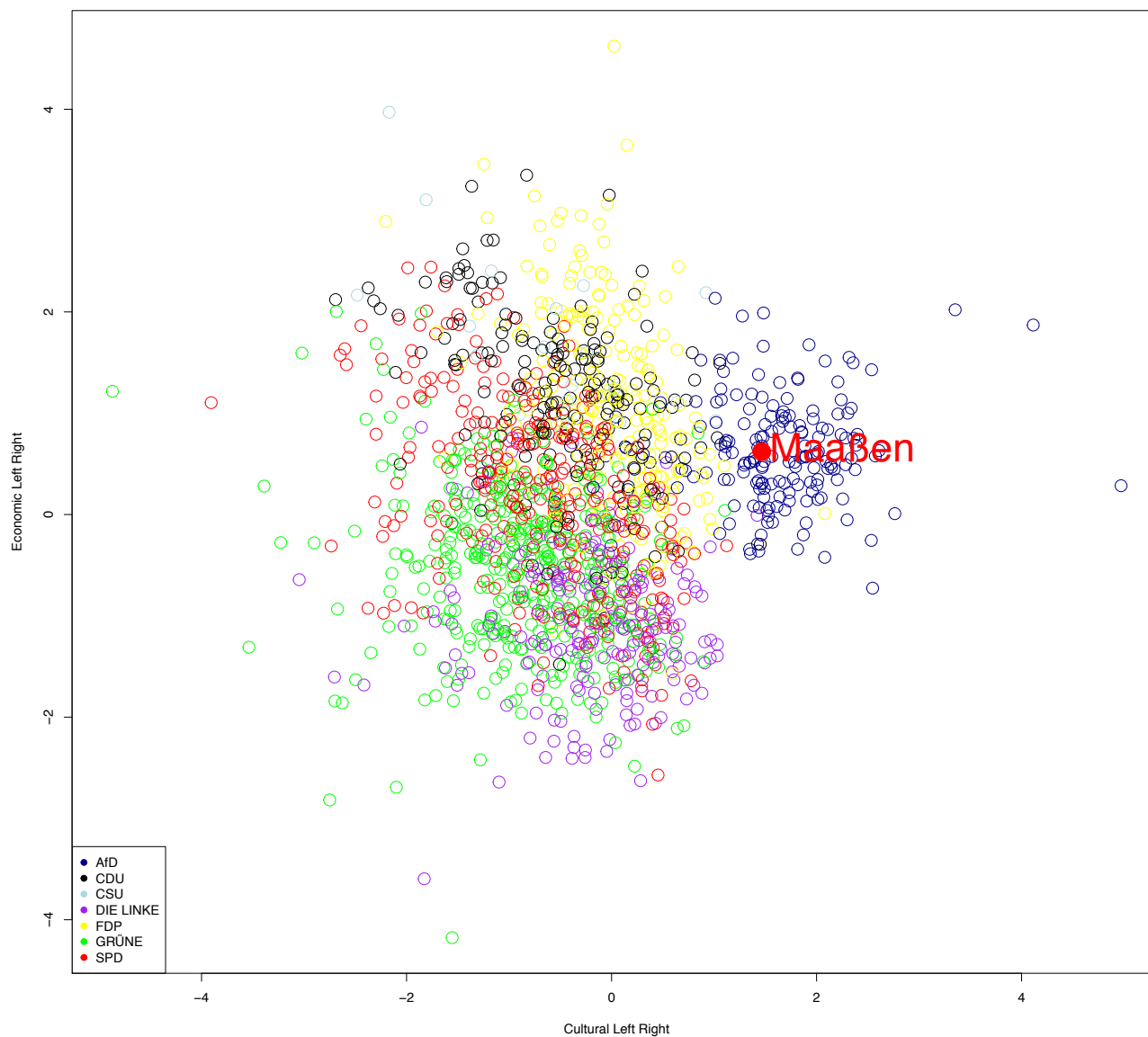
The German Twittersphere







Where is Maaßen?



4. Data Sharing

Datasets at GESIS

Monitoring sozialer Medien im Bundestagswahlkampf 2017

Stier, Sebastian; Bleier, Arnim; Bonart, Malte ■ Mörsheim, Fabian; Bohlouli, Mahdi; Nizhegorodov, Margarita; Posch, Lisa; Maier, Jürgen; Rothmund, Tobias; Staab, Steffen

GESIS Datenarchiv, Köln. ZA6926 Datenfile Version 1.0.0, <https://doi.org/10.4232/1.12992>

<https://doi.org/10.4232/1.12992>

Political Campaigning on Twitter During the 2019 European Parliament Election Campaign

Stier, Sebastian; Popa, Sebastian A.; Braun, Daniela

GESIS - Leibniz Institut für Sozialwissenschaften. Datenfile Version 1.0.0,
<https://doi.org/10.7802/1.1995>

<https://doi.org/10.7802/1.1995>



2013

2017

2019

2021

Deutsche Bundestagswahl 2013: Nutzung von Twitter durch Kandidaten

Kaczmirek, Lars; Mayr, Philipp

GESIS Datenarchiv, Köln. ZA5973 Datenfile Version 1.0.0, <https://doi.org/10.4232/1.12319>

<https://doi.org/10.4232/1.12319>

Twitter-Accounts der Kandidierenden zur Bundestagswahl 2021 (GLES)

Sältzer, Marius; Stier, Sebastian; Bäuerle, Joscha ■ Blumenberg, Manuela; Mechkova, Valeriya; Pemstein, Daniel; Seim, Brigitte; Wilson, Steven

GESIS Datenarchiv, Köln. ZA7721 Datenfile Version 2.0.0, <https://doi.org/10.4232/1.13790>

<https://doi.org/10.4232/1.13790>

Data sharing and scientific reuse

- Terms and conditions of social media platforms prohibit data sharing → replication crisis?
- Twitter
 - ▶ Sharing of tweet IDs → rehydrating tweets
 - ▶ Sharing of relevant accounts
- Facebook & Instagram
 - ▶ Sharing of relevant accounts → CrowdTangle
- Deleted posts will be lost
- *Meet the Experts* talk on research ethics by Katrin Weller and Oliver Watteler ([slides](#))

Conclusion

- Properly conceptualize your project before data collection: audience vs. elite-centered designs, do we need social media for this RQ?
- Ideally develop some expectations (*Open Science*)
- Think about the scope of your project before you start → Data management, server infrastructure, databases
- Social media holds huge – still untapped – potential for studying election campaigns!

Thank you for your attention!

gesis

Leibniz-Institut
für Sozialwissenschaften



The German Federal Election: Social Media Data for Scientific (Re-)Use

Meet the Experts! – GESIS online talks

Marius Sältzer & Sebastian Stier • October 14, 2021

Expert Contact & GESIS Consulting




Contact: you can reach the speaker/s via email:
marius.saeltzer@gesis.org & sebastian.stier@gesis.org

GESIS Consulting: GESIS offers individual consulting in a number of areas – including survey design & methodology, data archiving, digital behavioral data & computational social science – and across the research data cycle.

Please visit our website www.gesis.org for more [detailed information](#) on available services and terms.

More Services from GESIS

- Get materials for [capacity building in computational social science](#) and take advantage of our expanding expertise and resources in [digital behavioral data](#).
- Use GESIS data services for [finding data](#) for secondary analysis and [sharing your own data](#).
- Check out the [GESIS blog](#) "Growing Knowledge in the Social Sciences" for topics, methods and discussions from the GESIS cosmos – and beyond.
- Keep up with GESIS activities and subscribe to the monthly [newsletter](#).
-  for publications, tools & services.

*** Upcoming online workshop; Nov 2-5, 2021 ***

[Introduction to Social Media as Research Data: Potentials and Pitfalls](#)

More from CSS Experts in the Series

- June 24 Katrin Weller: **A Short Introduction to Computational Social Science and Digital Behavioral Data**
- July 01 Fabian Flöck, Indira Sen: **Digital Traces of Human Behavior from Online Platforms – Research Designs and Error Sources**
- July 08 Sebastian Stier, Johannes Breuer: **Combining Survey Data and Digital Behavioral Data**
- Sept 16 Oliver Watteler, Katrin Weller: **Research Ethics and Data Protection in Social Media Research**
- Sept 30 Roberto Ulloa: **Introduction to Online Data Acquisition**
- Oct 07 Roberto Ulloa: **Auditing Algorithms: How Platform Technologies Shape our Digital Environment**
- Oct 14 Marius Sältzer, Sebastian Stier: **The German Federal Election: Social Media Data for Scientific (Re-)Use**
- Nov 04 Arnim Bleier: **Introduction to Text Mining**
- Nov 11 Haiko Lietz: **Social Network Analysis with Digital Behavioral Data**
- Dec 2 Olga Zagovora, Katrin Weller: **Altmetrics: Analyzing Academic Communications from Social Media Data**
- Dec 16 Andreas Schmitz: **Online Dating: Data Types and Analytical Approaches**
- Jan 13 Gizem Bacaksizlar: **Political Behavior and Influence in Online Networks**
- Jan 27 David Brodesser: **SocioHub – A Collaboration Platform for the Social Sciences**