

What do people do online?

Using data donation to understand digital behavior.

a workshop at the SPP Junior Researcher Meeting

Frieder Rodewald 

University of Mannheim & Institute for Employment Research

October 22, 2025

Our Agenda

- 1 What is digital trace data?
- 2 What is data donation? – The participant's perspective.
- 3 What is data donation? – The researcher's perspective.

Who are you?

Please raise your hand 🙋 if you ...

- are familiar with the term digital trace data
- have worked with APIs
- have worked with data donation
- have worked with automated content analysis
- regularly use programming languages (e.g., R, Python)

About me: Frieder Rodewald

🎓 PhD, University of Mannheim & Institute for Employment Research

🤔 Research interests: “I study what people do online.”

More info: github.com/frodew & frieder-rodewald.de

A huge thanks to [Valerie Hase](#), for conceptualizing a previous data donation workshop at CompText in Vienna.

👉 part of the SPP project Integrating Data Donations in Survey Infrastructure



Dr. Valerie Hase

PI, LMU Munich



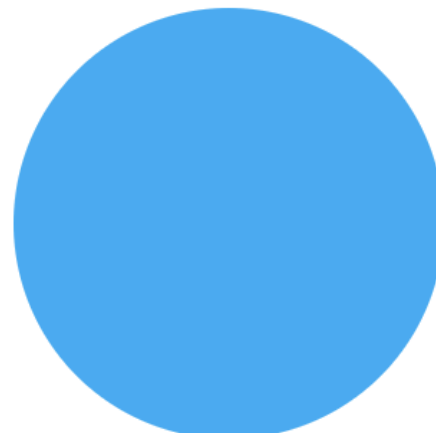
Prof. Dr. Florian Keusch

PI, University of Mannheim



Prof. Dr. Frauke Kreuter

PI, LMU Munich



Prof. Dr. Mark Trappmann

PI, IAB/University of Bamberg



Dr. Sebastian Prechsl

Postdoctoral Scholar, IAB/LMU Munich








Frieder Rodewald

PhD student, University of Mannheim



Our Team

What is the goal of this workshop?

-  Understanding digital data traces as a *type* of data
-  Understanding data donation as a *method* of data access
-  Working through key steps of data donation methods (participant & researcher view)
-  Discussing when (not) to use data donation studies
-  Detailed implementation (e.g., server set-up, coding data extraction scripts)

What is digital trace data?

🤔 Which examples for digital trace data you know?

What is digital trace data?

💡 **Definition:** *The recording and storing of activities on digital platforms to draw conclusions about digital and analog phenomena*

- e.g., tweets, likes, shares on social media
- e.g., geo data (locations, movements)
- e.g., digital payments
- e.g., Spotify playlists

Where can we find/collect digital trace data?

- Apps (e.g., running apps)
- Social media platforms (e.g., Instagram)
- Payment systems (e.g., Paypal)
- Wearable devices (e.g., smart watch)

Which types of data does this include?

Depending on the data collection method... ([Ohme et al. 2024](#)):

- often fine-grained (e.g., time-stamped)
- often longitudinal (e.g., over years, within-individual change)
- often less reactive (e.g., less concerns about social desirability)

Which (latent) constructs can we measure?

- **Internet use** ([Parry et al. 2021](#)) related to ...
 - well-being ([Ohme et al. 2024](#))
 - voting ([Bach et al. 2021](#))
- **News engagement** ([Reiss 2023](#)) related to ...
 - news diversity ([Jürgens and Stark 2022](#))
 - public opinion formation ([Yan, Schroeder, and Stier 2022](#))
- **Movements** related to ...
 - Mobility during pandemics ([Li et al. 2021](#))
 - Social networks ([Sepulvado et al. 2022](#))

Why are digital traces becoming more popular?

🤔 How many minutes a day do you use the internet to consume news?

- „internet“?

- „internet“?
- „news“?

- „internet“?
- „news“?
- „how many minutes“?






Why are digital traces becoming more popular?

- **Problems with self-reported data** (e.g., via survey)
 - Self-reported data subject to specific bias ([Scharkow 2016](#); [Parry et al. 2021](#))
 - Response rates in surveys are declining ([Luiten, Hox, and de Leeuw 2020](#))
- **Availability**
 - cheap (e.g., via APIs)
 - large data sets (“big data”)

⚠️ **Be careful:** These “advantages” are often claimed, but **not** empirically proven.

👉 Digital traces are neither necessarily less biased, nor cheaper, or larger.

(Dis-)advantages of digital trace data

-  More fine-grained, often longitudinal measures due to timestamps
-  Partly measurement of new variables (e.g., algorithmic inference)
-  Bias due to errors in representation and measurement
-  Implementation can be expensive and cumbersome
-  More data does not mean better data!

How can we collect digital traces?

Platform- and user-centric methods

- **Platform-centric** (based on platform cooperation)
 - API ([Jünger 2021](#))
 - Cooperation with platforms ([Wagner 2023](#))
- **User-centric** (based on user cooperation and informed consent) or “follow the user” approaches ([Caliandro 2024](#))
 - Data donation ([Carrière et al. 2025](#))
 - Linkage ([Sloan et al. 2020](#))
 - Sensors ([Struminskaya et al. 2021](#))
 - Tracking ([Christner et al. 2022](#))

References

- Bach, Ruben L., Christoph Kern, Ashley Amaya, Florian Keusch, Frauke Kreuter, Jan Hecht, and Jonathan Heinemann. 2021. "Predicting Voting Behavior Using Digital Trace Data." *Social Science Computer Review* 39 (5): 862–83. <https://doi.org/10.1177/0894439319882896>.
- Caliandro, Alessandro. 2024. "Follow the User: Taking Advantage of Internet Users as Methodological Resources." *Convergence: The International Journal of Research into New Media Technologies*, December, 13548565241307569. <https://doi.org/10.1177/13548565241307569>.
- Carrière, Thijs C., Laura Boeschoten, Bella Struminskaya, Heleen L. Janssen, Niek C. de Schipper, and Theo Araujo. 2025. "Best Practices for Studies Using Digital Data Donation." *Quality & Quantity* 59 (1): 389–412. <https://doi.org/10.1007/s11135-024-01983-x>.
- Christner, Clara, Aleksandra Urman, Silke Adam, and Michaela Maier. 2022. "Automated Tracking Approaches for Studying Online Media Use: A Critical Review and Recommendations." *Communication Methods and Measures* 16 (2): 79–95. <https://doi.org/10.1080/19312458.2021.1907841>.
- Jünger, Jakob. 2021. "A Brief History of APIs." In *Handbook of Computational Social Science, Volume 2*, 1st ed., 17–32. London: Routledge.
- Jürgens, Pascal, and Birgit Stark. 2022. "Mapping Exposure Diversity: The Divergent Effects of Algorithmic Curation on News Consumption." *Journal of Communication*, March, jqac009. <https://doi.org/10.1093/joc/jqac009>.
- Li, Xiao, Haowen Xu, Xiao Huang, Chenxiao Guo, Yuhao Kang, and Xinyue Ye. 2021. "Emerging Geo-Data Sources to Reveal Human Mobility Dynamics During COVID-19 Pandemic: Opportunities and Challenges." *Computational Urban Science* 1 (1): 22. <https://doi.org/10.1007/s43762-021-00022-x>.
- Luiten, Annemieke, Joop Hox, and Edith de Leeuw. 2020. "Survey Nonresponse Trends and Fieldwork Effort in the 21st Century: Results of an International Study Across Countries and Surveys." *Journal of Official Statistics* 36 (3): 469–87. <https://doi.org/10.2478/jos-2020-0025>.