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

Sebastian Prechsl
Institute for Employment Research & LMU Munich

October 22, 2025

Our Agenda

1️⃣ What is digital trace data?

2️⃣ What is data donation? - The participant’s perspective.

3️⃣ 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

About me: Sebastian Prechsl

...

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

Dr. Valerie Hase

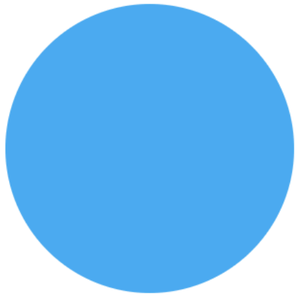
PI, LMU Munich

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

A huge thanks to Valerie Hase, for conceptualizing a previous data donation workshop at CompText in Vienna.

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)

1️⃣ 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.*

This might include:

- Tweets, likes, shares on social media
- Geo data (locations, movements)
- Digital payments
- 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 (latent) constructs can we measure?

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

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)
- **Availabillity**
  - 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, 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)

🤔 Questions?

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.

Ohme, Jakob, Theo Araujo, Laura Boeschoten, Deen Freelon, Nilam Ram, Byron B. Reeves, and Thomas N. Robinson. 2024. “Digital Trace Data Collection for Social Media Effects Research: APIs, Data Donation, and (Screen) Tracking.” *Communication Methods and Measures* 18 (2): 124–41. https://doi.org/10.1080/19312458.2023.2181319.

Parry, Douglas A., Brittany I. Davidson, Craig J. R. Sewall, Jacob T. Fisher, Hannah Mieczkowski, and Daniel S. Quintana. 2021. “A Systematic Review and Meta-Analysis of Discrepancies Between Logged and Self-Reported Digital Media Use.” *Nature Human Behaviour* 5 (11): 1535–47. https://doi.org/10.1038/s41562-021-01117-5.

Reiss, Michael V. 2023. “Dissecting Non-Use of Online News – Systematic Evidence from Combining Tracking and Automated Text Classification.” *Digital Journalism* 11 (2): 363–83. https://doi.org/10.1080/21670811.2022.2105243.

Scharkow, Michael. 2016. “The Accuracy of Self-Reported Internet Use—A Validation Study Using Client Log Data.” *Communication Methods and Measures* 10 (1): 13–27. https://doi.org/10.1080/19312458.2015.1118446.

Sepulvado, Brandon, Michael Lee Wood, Ethan Fridmanski, Cheng Wang, Matthew J. Chandler, Omar Lizardo, and David Hachen. 2022. “Predicting Homophily and Social Network Connectivity From Dyadic Behavioral Similarity Trajectory Clusters.” *Social Science Computer Review* 40 (1): 195–211. https://doi.org/10.1177/0894439320923123.

Sloan, Luke, Curtis Jessop, Tarek Al Baghal, and Matthew Williams. 2020. “Linking Survey and Twitter Data: Informed Consent, Disclosure, Security, and Archiving.” *Journal of Empirical Research on Human Research Ethics* 15 (1-2): 63–76. https://doi.org/10.1177/1556264619853447.

Struminskaya, Bella, Peter Lugtig, Vera Toepoel, Barry Schouten, Deirdre Giesen, and Ralph Dolmans. 2021. “Sharing Data Collected with Smartphone Sensors.” *Public Opinion Quarterly* 85 (S1): 423–62. https://doi.org/10.1093/poq/nfab025.

Wagner, Michael W. 2023. “Independence by Permission.” *Science* 381 (6656): 388–91. https://doi.org/10.1126/science.adi2430.

Yan, Pu, Ralph Schroeder, and Sebastian Stier. 2022. “Is There a Link Between Climate Change Scepticism and Populism? An Analysis of Web Tracking and Survey Data from Europe and the US.” *Information, Communication & Society* 25 (10): 1400–1439. https://doi.org/10.1080/1369118X.2020.1864005.