

# Digitale Datenspuren nutzbar machen

## Datenspenden als Methode der Kommunikationswissenschaft

---

Sitzung **3**: Einführung: Datenspenden

Valerie Hase (Ludwig-Maximilians-Universität München)

👉 [github.com/valeriehase](https://github.com/valeriehase) & [valerie-hase.com](https://valerie-hase.com)

# 1) Was sind Datenspende-Studien?



Quelle: Foto von Markus Winkler auf Unsplash

*Wer von euch hat - im Vorfeld dieses Workshops - schon einmal die eigenen Daten von einer digitalen Plattform angefragt? 🤔*

# Was sind Datenspende-Studien?

**Definition** 💡: Datenspende-Studien sind eine nutzerzentrierte Methode zur Sammlung von Daten:

- Nutzende haben das Recht, Daten, die Plattformen über sie sammeln, anzufragen und herunterladen.
- Sie können ihre Datenpakete (DDPs) via Datenspende-Tools (DDTs), oft im Rahmen von Umfragen, der Wissenschaft zur Verfügung stellen.
- Forschende nutzen CSS-Methoden zur Filterung, Anonymisierung und Analyse dieser Daten.

# Was für Daten enthalten “DDPs”?

Für klassische soziale Medien wie Facebook, Instagram oder Youtube z. B.

- digitale Nutzerprofile (soziodemografische Angaben, Privatsphäre-Einstellungen)
- digitale Aktivitäten (Freunde, Likes, Suchen, Privatnachrichten, passive Nutzung)
- digitales Targeting (Werbung, algorithmische inferierte Interessen)
- analoge Aktivitäten (Teilnahme an Events, Lebensereignisse)

# Beispiel I

Beispiel-Studie mit Smartphone-Datenspenden (Wu-Ouyang & Chan, 2022, S. 10)

**Table 3.** Regression models of well-being.

	Model 1		Model 2		Model 3	
	Loneliness		Depression		Negative outcomes	
	B	CIs	B	CIs	B	CIs
Sex	.08*	(.01, .16)	.07	(.01, .15)	.03	(−.05, .10)
Income	−.13**	(−.21, −.05)	−.06	(−.15, .03)	.03	(−.05, .11)
Age	−.09*	(−.17, .00)	−.05	(−.14, .05)	.01	(−.07, .09)
Education	−.03	(−.10, .03)	−.02	(−.10, .06)	.09*	(.02, .16)
Self-reported time	.09*	(.00, .19)	.10*	(.00, .21)	.14**	(.04, .23)
Log time	−.02	(−.11, .07)	−.04	(−.13, .05)	−.02	(−.10, .08)
R <sup>2</sup> (%)	6		2.6		2.8	
F statistic	7.66***		3.28**		3.49**	

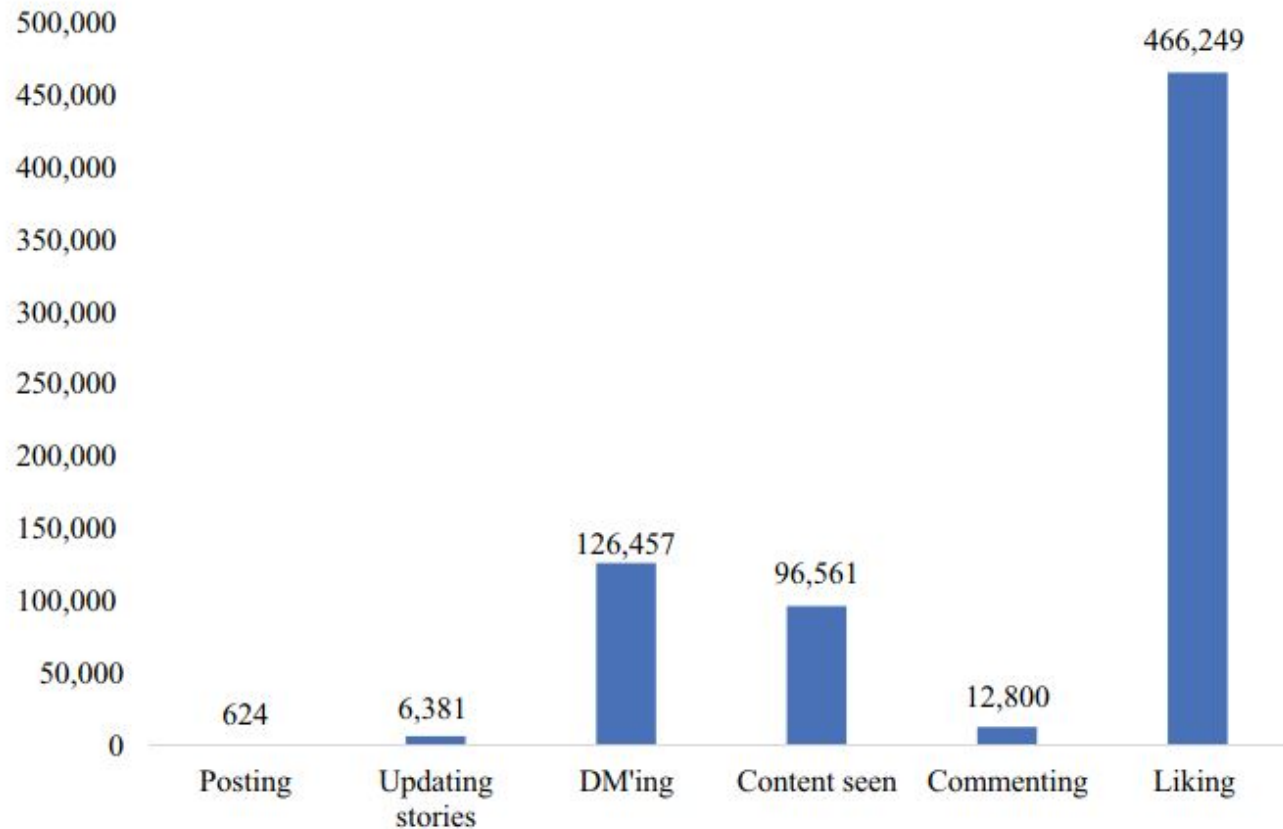
Note.  $n = 728$ . Beta are standardized coefficients with the final entry results. Each data point was transformed to z-score and the 95% confidence intervals (CIs) using 5,000 bootstrapped samples were calculated.

\*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$ .

Quelle: Wu-Ouyang & Chan, 2022, S. 10

# Beispiel II

Beispiel-Studie mit Instagram-Datenspenden (Driel et al., 2022, S. 273)



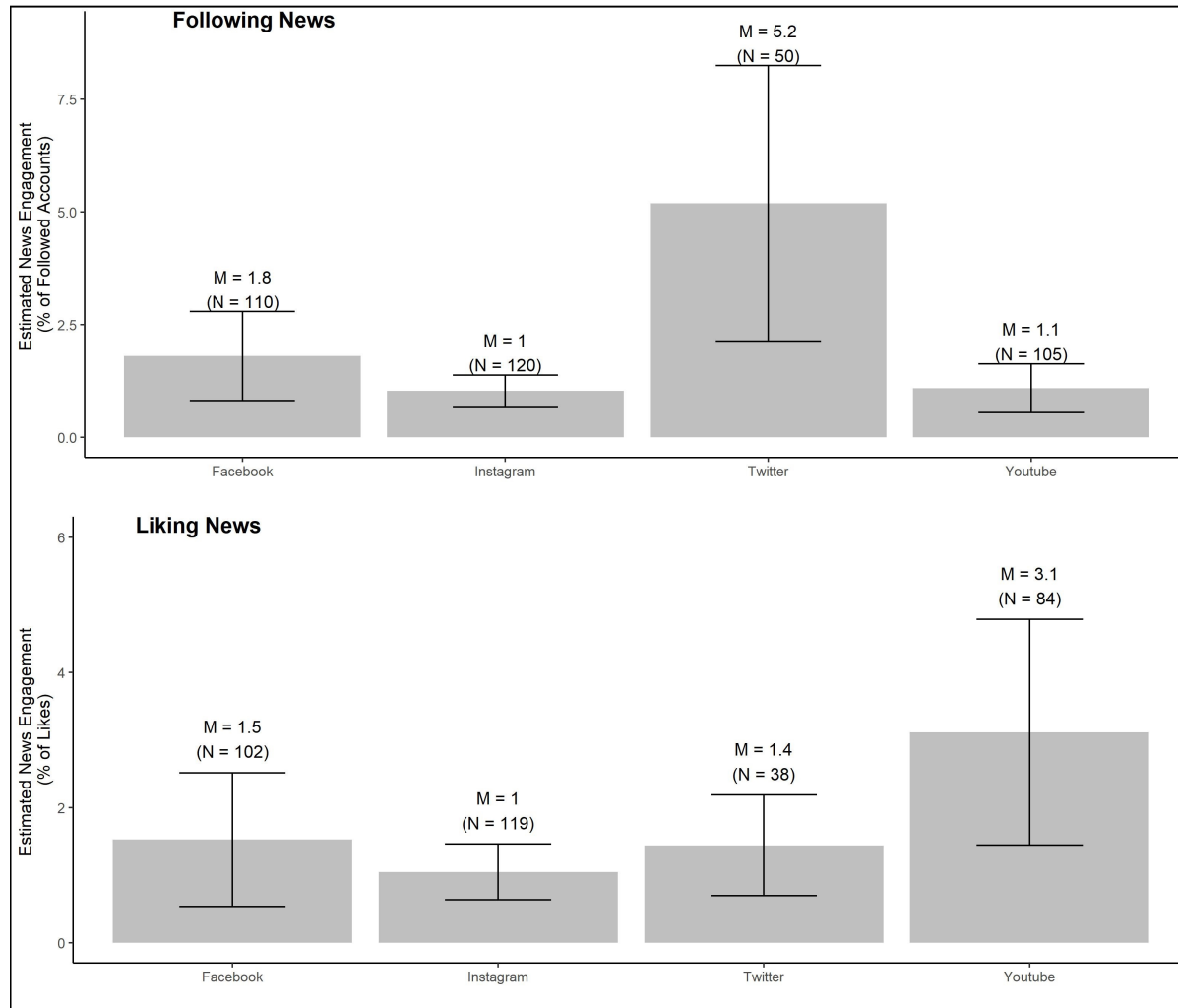
**Figure 2.** Frequencies of different Instagram activities.

Note. Posting = posting of photos and videos; Updating stories = all updates to stories; DM'ing = all direct messages sent; Content seen = clicking on profiles, photos, videos, and ads; Commenting = comments on posts; Liking = liking of posts.

Quelle: van Driel et al., 2022, S. 273

# Beispiel III

## Beispiel-Studie mit Social-Media-Datenspenden



Hase & Haim, under review



# Was ist neu an Daten aus DDPs?

Im Vergleich zu Daten via APIs ([Haim & Hase, 2023](#); [Ohme et al., 2023](#)) z.B.

- Kontrolle & informierte Einwilligung der Nutzer:innen, welche Daten Forschung nutzen darf
- z.T. vollständige längsschnittliche Daten ohne “Rate Limits” (z.B. seit erstmaliger Nutzung von Accounts)
- z.T. neue Daten (z.B., wenn auch selten, passiver Nutzung; nicht-öffentliche Daten wie Nachrichten)

# Was ist neu an Daten aus DDPs?

Im Vergleich zu Daten via Tracking ([Haim & Hase, 2023](#); [Ohme et al., 2023](#)) z.B.

- retrospektive Erhebung, die längsschnittliche Daten und weniger Reaktivität ermöglicht
- z.T. neue Daten (z.B. algorithmische Inferenzen)
- klarere Verankerung in EU-Recht (im Vgl. zu Rechtsstreitigkeiten bei Tracking) ([Edelson & McCoy, 2021](#))
- aber: ähnlich aufwendig für Befragte!

# Zusammenfassung: Datenspenden

- **Zusammenfassung:** Methode zur Sammlung digitaler Verhaltensdaten, bei der Nutzende ihre Datenpakete (DDPs) von Plattformen herunterladen und via Datenspende-Tools (DDTs) der Wissenschaft zur Verfügung stellen.
- **Weiterführende Literatur:**
  - Driel et al. (2022)
  - Boeschoten et al. (2022)

# Fragen? 🤔

- Boeschoten, L., Mendrik, A., Van Der Veen, E., Vloothuis, J., Hu, H., Voorvaart, R., & Oberski, D. L. (2022). Privacy-preserving local analysis of digital trace data: A proof-of-concept. *Patterns*, 3(3), 100444.  
<https://doi.org/10.1016/j.patter.2022.100444>
- Driel, I. I. van, Giachanou, A., Pouwels, J. L., Boeschoten, L., Beyens, I., & Valkenburg, P. M. (2022). Promises and Pitfalls of Social Media Data Donations. *Communication Methods and Measures*, 1-17.  
<https://doi.org/10.1080/19312458.2022.2109608>
- Edelson, L., & McCoy, D. (2021). Facebook is obstructing our work on disinformation. Other researchers could be next. *The Guardian*. <https://www.theguardian.com/technology/2021/aug/14/facebook-research-disinformation-politics>
- Haim, M., & Hase, V. (2023). Computational Methods und Tools für die Erhebung und Auswertung von Social-Media-Daten. In S. Stollfuß, L. Niebling, & F. Raczkowski (Eds.), *Handbuch Digitale Medien und Methoden* (pp. 1-20). Springer Fachmedien Wiesbaden. [https://doi.org/10.1007/978-3-658-36629-2\\_41-1](https://doi.org/10.1007/978-3-658-36629-2_41-1)
- Ohme, J., Araujo, T., Boeschoten, L., Freelon, D., Ram, N., Reeves, B. B., & Robinson, T. N. (2023). Digital Trace Data Collection for Social Media Effects Research: APIs, Data Donation, and (Screen) Tracking. *Communication Methods and Measures*. <https://doi.org/10.1080/19312458.2023.2181319>
- Wu-Ouyang, B., & Chan, M. (2022). Overestimating or underestimating communication findings? Comparing self-reported with log mobile data by data donation method. *Mobile Media & Communication*, 205015792211371.  
<https://doi.org/10.1177/20501579221137162>

