

Determinants of Willingness to Donate Data from Social Media Platforms

Keywords: data donation, DDP, vignette experiments, digital trace data, linking survey data

Extended Abstract

Social media platforms provide a wealth of information for researchers across a variety of fields, offering reliable and error-free behavioural data that is often more accurate than self-reported survey responses (Araujo et al., 2017; Ohme et al., 2021; Parry et al., 2021). However, in the wake of privacy concerns, many platforms have restricted the use of APIs, forcing researchers to explore alternative methods like data donation. Data donation involves users voluntarily providing their own social media data, typically through data download packages (DDPs). While this approach offers greater transparency and helps to overcome privacy concerns associated with platform-side data collection, it requires active participation from users and can be more burdensome compared to passive modes of data sharing (Boeschoten et al., 2020; Breuer et al., 2022). This study aims to understand the underlying mechanisms of willingness to participate in academic research as a social media data donor, with a focus on examining factors that influence users' willingness to share DDPs. By analysing previously uncovered factors, such as the type of social media data, in a cross-national comparison, the study aims to provide insights into nonresponse bias and improve recruitment strategies and research design. Ultimately, this research can help unlock the potential of social media data for scientific inquiry, while also ensuring the privacy and consent of users.

To study the determinants of data-sharing and help future data donation studies with detecting the conditions under which the willingness is the highest, we pre-registered two vignette experiments and embedded them in two online surveys conducted in Hungary and the US. In hypothetical requests for donating social media data via DDPs, we manipulated the amount of the monetary incentives (1), the presence or lack of non-monetary incentives (2), the number of requested platforms (3), the estimated upload/download time (4), and the type of requested data (5). In addition to this, personal characteristics were obtained from the survey including gender, education level, age, subjective wealth, frequency of Facebook usage, number of active social media platforms, Internet Users' Information Privacy Concerns, trust in businesses' data handling, affinity for technology, and the big five personality traits.

We used the following methodology to carry out the analysis. First, we used a variance component model to determine how much variation in willingness to donate digital footprint data was due to respondent and vignette level characteristics. Multilevel regressions were then performed using two levels in the data (one for the vignettes and another for the respondents) allowing for random intercepts and slopes by the individual. Then we tested for cross-level interactions between vignette and respondent level variables to determine how the willingness to donate was affected differently among social groups.

Results from the Hungarian study found that 30% of vignette respondents were very unlikely to donate their data, while 14% were always willing. In the US study these figures were 59% and 5%, respectively. On the respondent level, 18% refused any data sharing in the Hungarian sample, 52% in the US sample. Overall, the willingness rate was much lower in the US sample, than in the Hungarian one.

The multilevel regression analyses showed that regarding study parameters, monetary incentives increased willingness to participate in both countries, while other effects were not

consistent between the two countries. Non-monetary incentives (providing a report) and time to download/upload data influenced willingness in the US sample but not in the Hungarian one, whereas the type of data affected the willingness to participate only in the case of the Hungarian respondents.

Regarding respondent characteristics, older and highly educated respondents were more likely to share their data in the Hungarian sample, while in the US sample results reinforced previous findings that privacy and security concerns lower the willingness to participate. Consistently in the two samples, more frequent social media users were more likely to donate their data, but for instance, the affinity for technology or personality traits (e.g., openness) of respondents did not influence willingness significantly.

To the extent that these self-reports overlap with actual sharing behaviour, our findings altogether suggest that data donation studies should expect strong and systematic selection bias. Our findings help design more effective future data donation requests and provide insights into the potential patterns of selection bias in data donation studies.

References

- Araujo, T., Wonneberger, A., Neijens, P., & de Vreese, C. (2017). How Much Time Do You Spend Online? Understanding and Improving the Accuracy of Self-Reported Measures of Internet Use. *Communication Methods and Measures*, 11(3), 173–190.
<https://doi.org/10.1080/19312458.2017.1317337>
- Boeschoten, L., Ausloos, J., Moeller, J., Araujo, T., & Oberski, D. L. (2020). Digital trace data collection through data donation. *ArXiv:2011.09851 [Cs, Stat]*.
<http://arxiv.org/abs/2011.09851>
- Breuer, J., Kmetty, Z., Haim, M., & Stier, S. (2022). User-centric approaches for collecting Facebook data in the ‘post-API age’: Experiences from two studies and recommendations for future research. *Information, Communication & Society*, 1–20.
<https://doi.org/10.1080/1369118X.2022.2097015>
- Ohme, J., Araujo, T., de Vreese, C. H., & Piotrowski, J. T. (2021). Mobile data donations: Assessing self-report accuracy and sample biases with the iOS Screen Time function. *Mobile Media & Communication*, 9(2), 293–313. <https://doi.org/10.1177/2050157920959106>
- Parry, D. A., Davidson, B. I., Sewall, C. J. R., Fisher, J. T., Mieczkowski, H., & Quintana, D. S. (2021). A systematic review and meta-analysis of discrepancies between logged and self-reported digital media use. *Nature Human Behaviour*, 5(11), Article 11.
<https://doi.org/10.1038/s41562-021-01117-5>

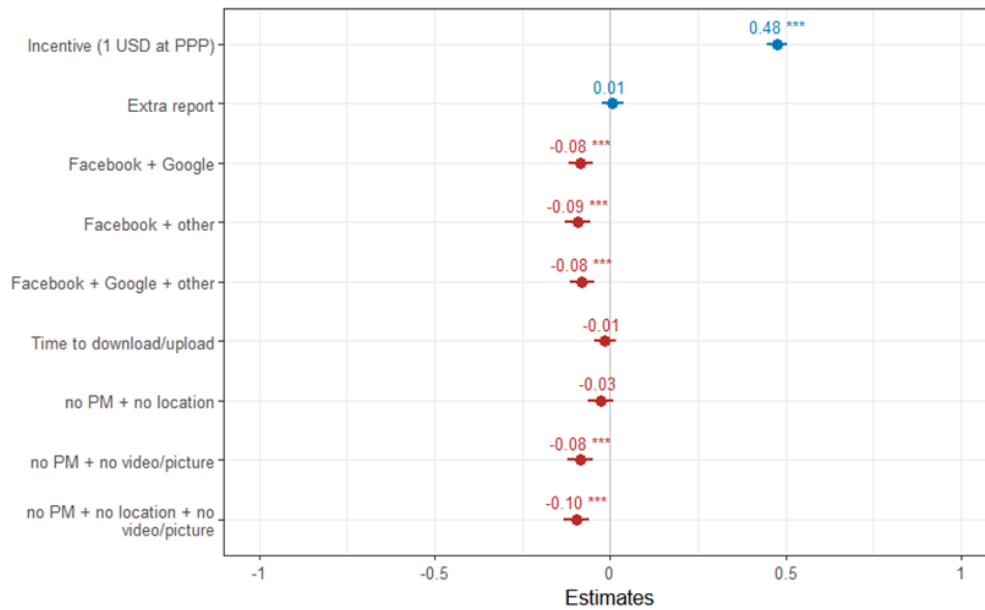


Figure 1. Relative effects of vignette level characteristics on willingness to donate data; only vignette level independent variables; allowing for random intercept (Model 2). Standardized regression coefficients. Hungary.

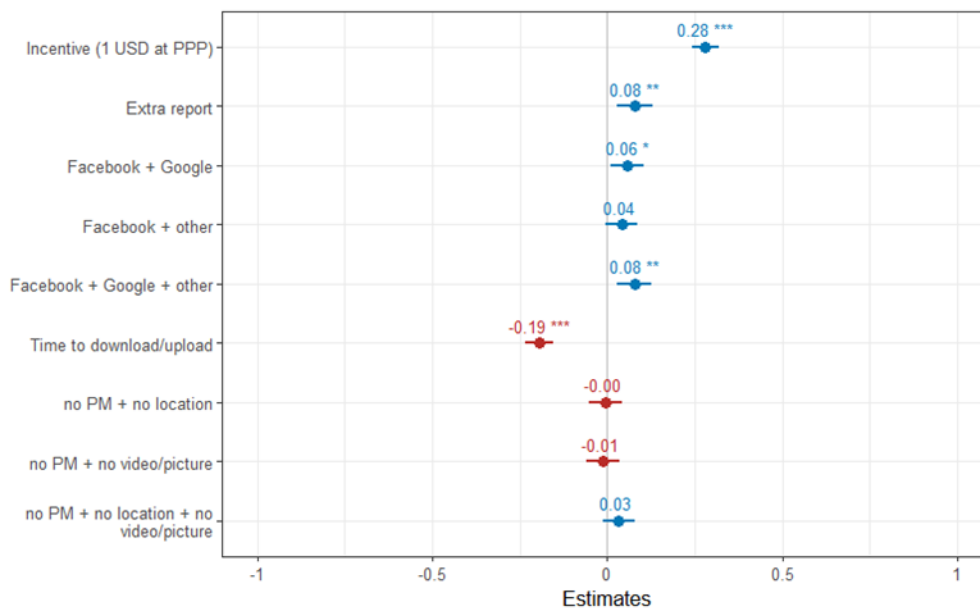


Figure 2. Relative effects of vignette level characteristics on willingness to donate data; only vignette level independent variables; allowing for random intercept (Model 2). Standardized regression coefficients. US.