

# “I Really Thought I Would Use More Than Just Google” – Developing and Testing a Browsing History Donation Approach With Journalists

*Keywords: data donations, journalism research, platform centralization, information systems, mixed methods*

## Extended Abstract

In our current digital environment it is difficult for any user to recall their media use (Guess, 2015, Haenschen, 2019). Social desirability and professional socialization might render the task of reliably surveying journalists on their professional online habits challenging. These dynamics of personalization have been linked to a lack of diversity or the polarization of political opinions (e.g. Bail et al., 2018; Bakshy et al., 2015). Journalists use information intermediaries as an essential tool in their professional work (Humayun & Ferrucci 2022) and information distribution and the amplification of content via social media by journalists have been extensively studied through surveys (e.g Tandoc & Vos, 2016; Vázquez-Herrero et al., 2019). However, while journalistic reception practices in these algorithmic environments significantly inform the production of journalistic content (Humanyun & Ferruci 2022) studies have been limited to self-reports.

**In this paper, we explore if and in which ways we can gather information on the online information repertoires of journalists through data donations while taking specific characteristics of these professionals into account.** By building on previous work in journalistic social media use (Hedman & Djerf-Pierre, 2013), data donations (Menchen-Trevino, 2016), and trace interviews (Dubois & Ford 2015) we gathered data in a pilot study with a multi-method approach that includes three steps: a survey, a data donation of their browsing history, and a semi-standardized interview.

**Online Survey:** Journalists were asked to provide information on their demography, their technical setup, and their online media use in an online questionnaire. Information on demographics and work background allowed us to recruit a diverse sample. We asked participants a) to estimate how many hours of their workday they spend researching issues online and b) to rank relevant online sources for their work. We compared this information on their online media use with the log data from their browser donations.

**Data Donation:** The journalists were asked to install [an application][redacted for peer review], a cross-platform, open-source application developed by the authors. [The software][redacted for peer review] allows the participants to collect and visualize their browsing histories (main URL and timestamp) that is stored locally in their browser. The software allowed to include only selected browsers, filter by working hours, and generate an overview of all the site domains that they visited more than ten times over ninety days. Each participant could obscure sensitive data (marked as "verborgen" in Fig. 2) before uploading the information to encrypted online storage. While WebHistorian (Menchen-Trevino, 2016) represented a more established data-gathering tool that includes more visualization features for general users, we decided to develop a bespoke iteration for specific needs and ethical requirements; it didn't collect the full URLs as they were not needed for our research question, allowed for the collection across different browsers and gathered the participants' browsing data during their working hours only. As the tool is Python-based we hope for further advancement for other features and populations within the CCS community.

**Interview:** In a post-donation interview, participants reflected on the data donation in terms of the organizational, technical, and privacy challenges they faced. In dialogue with the participants, the researchers compared the browser history data to information from the individual survey. The stimulus of experiencing a visualization of their browsing habits and potential mismatches between self-reported and trace data led to reflections on the perceived dominance of certain websites, work-related habits, and values and the reliability of our data.

**Presented Results:** We collected data from twelve journalists representing a broad spectrum of age, gender, and professional characteristics. Therefore, we will present four key results:

- Willingness to donate: Which journalists were (not) willing to donate their browsing data and under which conditions? How was the request perceived by their employers?
- Donation experience: Which technical and literacy-related challenges did the journalists face when donating? Which problems and problematizations of data protection arose? How do journalists reflect on the experience of not receiving but handing over such personal data?
- Comparison to self-reports: How did the browsing data and the participants' self-assessments differ? How do journalists perceive their use of personalized intermediaries?
- Data quality: Based on the information on their willingness to engage, their experiences of donating data, and the comparisons with their self-reporting we will reflect on the quality of the gathered exposure data, the limitation of our approach, and the feasibility of [the software][redacted], as a research tool, a data collection method and interview stimulus.

## References

- Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. B. F., Lee, J., Mann, M., Merhout, F., & Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, 115(37), 9216–9221. <https://doi.org/10.1073/pnas.1804840115>
- Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348(6239), 1130–1132. <https://doi.org/10.1126/science.aaa1160>
- Dubois, E., & Ford, H. (2015). Qualitative Political Communication| Trace Interviews: An Actor-Centered Approach. *International Journal of Communication*, 9(0), Article 0. <https://ijoc.org/index.php/ijoc/article/view/3378>
- Guess, A. M. (2015). Measure for measure: An experimental test of online political media exposure. *Political Analysis*, 23(1), 59–75. <https://doi.org/10.1093/pan/mpu010>
- Haenschen, K. (2019). Self-reported versus digitally recorded: Measuring political activity on Facebook. *Social Science Computer Review*, 38(5), 567–583. <https://doi.org/10.1177/0894439318813586>
- Hedman, U., & Djerf-Pierre, M. (2013). The social journalist. *Digital Journalism*, 1(3), 368–385. <https://doi.org/10.1080/21670811.2013.776804>
- Humayun, M. F., & Ferrucci, P. (2022). Understanding Social Media in Journalism Practice: A Typology. *Digital Journalism*. <https://www.tandfonline.com/doi/abs/10.1080/21670811.2022.2086594>
- Menchen-Trevino, E. (2016, March 15). Web historian: Enabling multi-method and independent research with real-world web browsing history data. *ICConference 2016 Proceedings*. <https://doi.org/10.9776/16611>
- Tandoc, E. C., & Vos, T. P. (2016). The Journalist Is marketing the news. *Journalism Practice*, 10(8), 950–966. <https://doi.org/10.1080/17512786.2015.1087811>
- Vázquez-Herrero, J., Direito-Rebollal, S., & López-García, X. (2019). Ephemeral journalism: News distribution through Instagram stories. *Social Media + Society*, 5(4), 2056305119888657. <https://doi.org/10.1177/2056305119888657>

## Figures

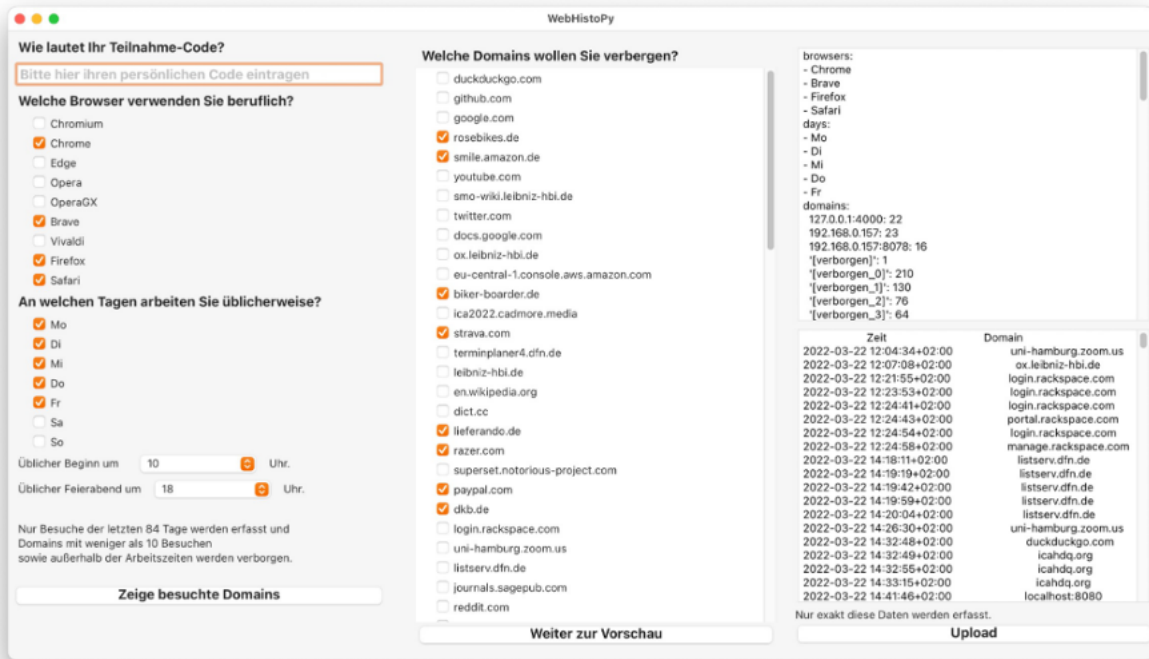


Figure 1: Screenshots of the [application][redacted for peer review] frontend in macOS

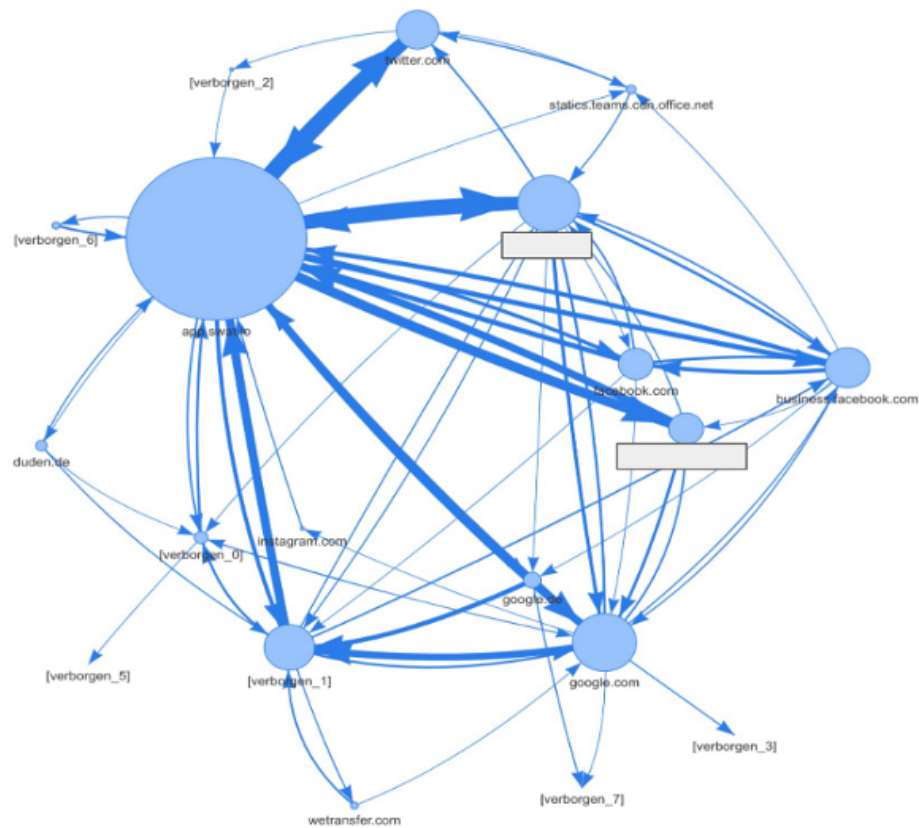


Figure 2: Exemplary visualization of browsing patterns of a participant based on the data gathered, made with pyvis (node = domains visited, edge = visits of the nodes within 5 minutes of each other, node size by betweenness centrality, edge width by number of visits, edge direction chronologically)