

“I Really Thought I Would Use More Than Just Google” - Investigating Professional Journalistic Online Use With Browser History Donations

This work has not been published or presented for a similar audience before

Journalists use information intermediaries as an essential tool in their work (Humayun & Ferrucci, 2022). The role of social media and search engines in the production, distribution, and consumption of news has been extensively studied through surveys and content analysis (ibid.). However, practices of media reception in these algorithmically personalized environments inform the first step of the datafied journalistic production cycle, the observation of society and the selection of newsworthy content (Loosen, 2018). Understanding the relevance of information intermediaries for journalistic routines is crucial, as these sort information based on users' behavior. Although current research is inconclusive these personalizations have been linked to a lack of diversity and to a polarization of political opinions (Bail et al., 2018; Fletcher & Jenkins, 2019) and may affect news selection, production, distribution, and consumption. In a digital environment characterized by frequent short episodes of online use and personalization, it is difficult to recall media use validly (Parry et al., 2021). For the study of journalistic media use, social desirability and professional socialization makes reliably surveying journalists about their online habits more challenging. In this presentation, we explore how we can collect information about journalists' online information repertoires through data donations, taking into account their specific professional characteristics.

Research Design

Building on previous work on journalistic social media use (Hedman & Djerf-Pierre, 2013), data donations (Menchen-Trevino, 2016) and trace interviews (Dubois & Ford, 2015), we collected data in an explorative pilot study using a mixed-methods approach involving three steps for each participant (n = 11): a survey about their media use, technical setup, and professional biography; a data donation of their browsing history; and a semi-standardized interview. See Figure 1 for a display of the frontend of our self-developed data donation tool. In the Github repository of the tool further screenshots, and other documentation as well as the

open source code can be found: [anonymized link (without deanonymizing screenshots) for peer review: <https://anonymous.4open.science/r/browser-history-donation-tool/>]

Figure 1

Screenshot of [Tool Name] Frontend in MacOS, [some Websites also Redacted for Peer Review]

The screenshot shows the following sections:

- Wie lautet Ihr Teilnahme-Code?**: A text input field with the placeholder "Bitte hier ihren persönlichen Code eintragen".
- Welche Browser verwenden Sie beruflich?**: A list of browser options with checkboxes: Chromium, Chrome, Edge, Opera, OperaGX, Brave (checked), Vivaldi, Firefox (checked), and Safari (checked).
- An welchen Tagen arbeiten Sie üblicherweise?**: A list of days with checkboxes: Mo (checked), Di (checked), Mi (checked), Do (checked), Fr (checked), Sa, and So.
- Üblicher Beginn um**: A time input field set to "9 Uhr".
- Üblicher Feierabend um**: A time input field set to "17 Uhr".
- Nur Besuche der letzten 84 Tage werden erfasst und Domains mit weniger als 10 Besuchen sowie außerhalb der Arbeitszeiten werden verborgen.**
- Zeige besuchte Domains**: A button.
- Welche Domains wollen Sie verbergen?**: A list of domains with checkboxes, including docs.google.com, github.com, duckduckgo.com, google.com, reddit.com, twitter.com, strava.com (checked), [redacted] (checked), linkedin.com, dhl.de (checked), [redacted] (checked), mas.to, osf.io, chat.openai.com, komoot.de (checked), localhost:3000 (checked), [redacted] (checked), lichess.org (checked), app.datacamp.com, mszturc.github.io, [redacted] (checked), statics.teams.cdn.office.net, pypi.org, and soscिसurvey.de.
- Weiter zur Vorschau**: A button.
- browsers:**: A list of browser types: Safari, Firefox, Brave, days: Mo, Di, Mi, Do, Fr.
- domains:**: A list of domain counts: N/A: 29, '[verborgen]': 1, '[verborgen_0]': 359, '[verborgen_1]': 292, '[verborgen_2]': 107, '[verborgen_3]': 97, '[verborgen_4]': 88, '[verborgen_5]': 88.
- Table of browser history:**

Zeit	Domain
2023-01-11 14:01:27+02:00	[verborgen]
2023-01-11 14:04:19+02:00	[verborgen]
2023-01-11 14:05:24+02:00	[verborgen]
2023-01-11 14:05:34+02:00	[verborgen]
2023-01-11 14:05:44+02:00	[verborgen]
2023-01-11 15:49:38+02:00	duckduckgo.com
2023-01-11 15:49:45+02:00	duckduckgo.com
2023-01-11 15:50:18+02:00	google.com
2023-01-11 15:50:21+02:00	google.com
2023-01-11 15:50:22+02:00	google.com
2023-01-11 15:51:27+02:00	google.com
2023-01-11 15:51:27+02:00	google.com
2023-01-11 15:51:29+02:00	google.com
2023-01-11 15:51:30+02:00	google.com
2023-01-11 15:52:06+02:00	google.com
2023-01-11 15:52:24+02:00	google.com
2023-01-11 16:02:41+02:00	duckduckgo.com
2023-01-11 16:03:51+02:00	duckduckgo.com
2023-01-11 16:03:56+02:00	duckduckgo.com
- Nur exakt diese Daten werden erfasst.**
- Upload**: A button.

Results

In the proposed presentation we present what shaped the journalists' recruitment and willingness to donate data, describe their donation experience, compare the donated data to their self-reports, and elaborate on the data quality. In Figure 2 we compare the survey data on professional online use with the individual browser history data regarding the ranking of the most visited websites (by count of page visits) from the individual browser history.

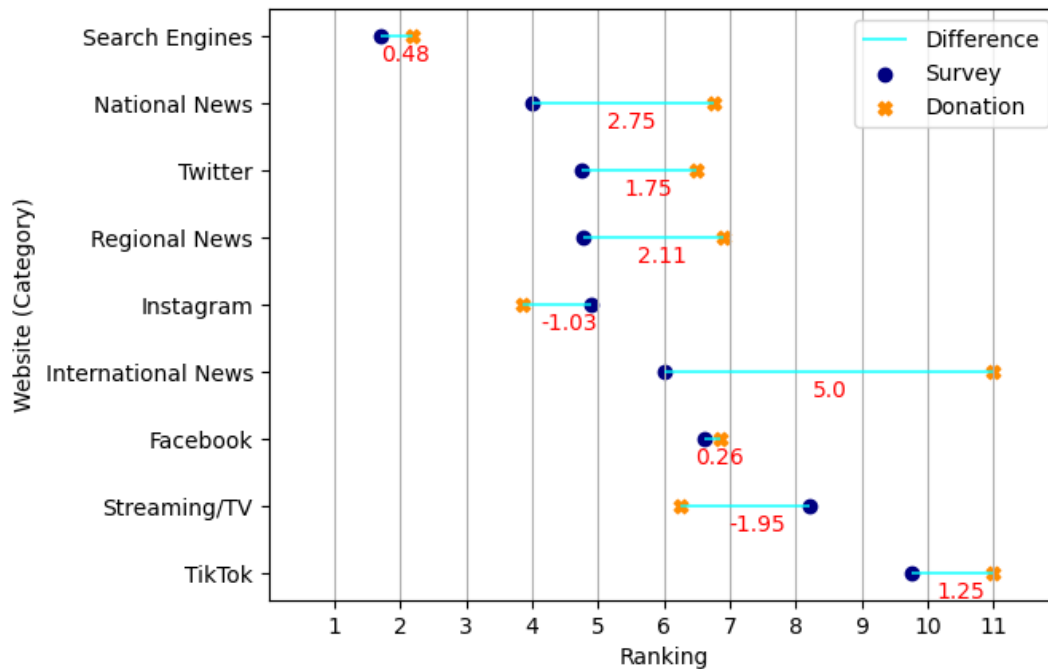
The most apparent differences between self-reported and measured data can be seen in how the participants ranked their estimated visits of news websites (regional/national/international news), which were ranked more relevant in the survey than in the web browsing data. October was one of the users that showed this pattern but also used the digital behavior data as a stimulus to reflect on their online use:

I would have thought that more local news sites would have appeared. Although I have actually just thought about it again, that I have really written articles in the last two months, where I simply didn't access them. [...] And I would have thought that there

would have been even more websites like this [a travel blog] that I actually use quite a lot, to get some information about stores or cafés in [destination]. But I probably also use more Instagram then actually. Or a lot more, because a lot of [travel] blogs can also be found on Instagram. (October, Freelance Journalist, City Magazine, 25–34)

Figure 2

Comparison of Mean Survey and Data Donation Rankings for Website Categories



Note: Differences (cyan/red) between the mean ranking of (estimated) visits of websites (categories, sorted by mean ranking in survey) according to survey (blue dots) and donation data (orange crosses) across all participants (n = 11)

Search engines were ranked as the most used websites in the survey and web browsing data where Google was an especially prevalent website. The dominance of Google was by far the most often mentioned result leading to surprise or irritation for the participants when discussing the details of their donations. The journalists often defended or rationalized their frequent Google use as they perceived its dominance as normatively not ideal. A participant, who is actively trying to avoid Google, was even hesitant to participate in the data donation as it would remind them of their failure to use other search engines:

I was a bit annoyed with myself and thought: Gosh! I'm somehow really still predominantly using Google as a search engine. And actually really often take it upon myself to say, no, just use a different search engine for once. Then I thought to myself, okay, now I have to look at my own browsing history, then I'm going to get angry about

this very point again, and I also have to share that with everyone, so to speak.(October, Freelance Journalist City Magazine, 25–34)

The data donation can be used as a stimulus to explain practices of online use in more detail. The accompanying interview fills gaps in digital behavior data that cannot be gathered reliably across platforms and devices, especially for professional media use in closed company systems.

Conclusion

Collaborating with individual users for the collection of information exposure data “turns the researcher-platform-user relationship around: Instead of relying on platforms to study users (often without their knowledge), researchers work directly with users willing to donate their data to academic research” (Araujo et al., 2022: 373-374).

In our study we chose to employ this method to study the professional online use of journalists, relying on meaningful informed consent, desire for self-awareness, and journalistic curiosity to link digital traces to self-reports. In comparing the donation and survey data it becomes apparent that differences in self-reported and measured media use data seem to be as common among the sample of journalists in our pilot study as they are in other study populations (Parry et al., 2021). Overall, the stimulus of seeing their browser history data and comparing this to self-reports lends itself to studies on practices around journalistic research and content production while also revealing assumptions about journalistic norms and role conceptions.