



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2024

Comparing Online Recruitment Strategies for Data Donation Studies

Pfiffner, Nico ; Reiss, Michael V ; Mitova, Eliza ; Blassnig, Sina

DOI: <https://doi.org/10.5117/CCR2024.2.8.PFIF>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-262555>

Journal Article

Published Version



The following work is licensed under a Creative Commons: Attribution 4.0 International (CC BY 4.0) License.

Originally published at:

Pfiffner, Nico; Reiss, Michael V; Mitova, Eliza; Blassnig, Sina (2024). Comparing Online Recruitment Strategies for Data Donation Studies. *Computational Communication Research*, 6(2):1-19.

DOI: <https://doi.org/10.5117/CCR2024.2.8.PFIF>

Comparing Online Recruitment Strategies for Data Donation Studies

Nico Pfiffner

Department of Communication and Media Research, University of Zurich, Switzerland

Michael V. Reiss

Leibniz-Institute for Media Research | Hans-Bredow-Institut, Germany

Eliza Mitova

Department of Communication and Media Research, University of Zurich, Switzerland

Sina Blassnig

Institute of Digital Communication and Media Innovation, University of Fribourg, Switzerland; University of Applied Sciences of the Grisons, Switzerland

Abstract

The collection of data donations holds great potential for making digital trace data accessible to communication research. However, a major challenge is the recruitment of participants who are willing to donate their data for research. Hence, this research report presents our experiences with different strategies used in one study to recruit citizens to donate their personal media usage data. Based on our experiences during a data donation study in which Google Search and YouTube usage histories were collected, we compare three different online recruitment channels—(a) an access panel, (b) social media advertisements, and (c) a data marketplace—in terms of five dimensions: (1) administrative effort, (2) setup effort, (3) participation rate, (4) costs, and (5) analytical considerations. Overall, participation rates were rather low across all recruitment channels. Nevertheless, access panels remain a viable option for recruiting data donors even if significantly lower response rates must be expected compared to survey studies. Furthermore, we see potential in recruiting data donors through online advertisements, especially if they can be targeted through mobile ads. By contrast, our experience with the data marketplace was discouraging.

Keywords: data donation, participant recruitment, online recruitment strategies, digital trace data

Introduction

Digital trace data that unobtrusively capture people's online behaviors and interactions with digital services are of great interest for studying a variety of digital phenomena. However, researchers still face significant hurdles when it comes to getting access to such data (Lazer et al., 2020). The collection of data donations from citizens is one way of overcoming these hurdles. This approach has gained importance following the introduction of the general data protection regulation (GDPR) in the European Union, which gives users the right to request and share the data relating to them that service providers have stored (Ausloos & Veale, 2021). One way to collect data donations in practice is to make use of the takeout functionality offered by digital media platforms, such as YouTube, Instagram, and Spotify. To do this, researchers instruct study participants on how they can request their personal data takeout from a platform of interest (often referred to as *data download packages* [DDPs]) and provide the participants with ways to then submit their data to the researchers (Boeschoten et al., 2022). Collecting data donations has the advantage that active and informed consent is obtained from individuals whose behaviors are captured in the data; moreover, the donated data can easily be enriched with survey data (Breuer et al., 2020; Halavais, 2019).

However, a major challenge is the recruitment of participants (Ohme & Araujo, 2022; Stier et al., 2020). Data donation collections deal with personal and thus sensitive data that were generated by a third party, and they constitute a new approach with which neither participants nor recruitment partners are familiar. It remains an open question whether researchers' experiences in recruiting participants for more traditional participant-centric approaches, such as survey studies, translate to the collection of data donations (Trappmann et al., 2022).

Previous research has thus begun to explore the factors that influence people's willingness to participate in data donation studies (Breuer et al., 2023; Ohme et al., 2020; Pfiffner & Friemel, 2023; Sen et al., 2021; Skatova & Goulding, 2019), but so far, a conclusive meta study integrating all these results is still missing. There are also multiple initiatives under way to develop and advance technical solutions to implement data donation collections (e.g., Araujo et al., 2022; Boeschoten et al., 2023; DataSkop, 2023; Haim et al., 2023; Pfiffner et al., 2022). Nevertheless, the research community still lacks experience regarding the best practices for collecting data donations (Ohme & Araujo, 2022) and how well suited different channels are for recruiting participants for data donation studies (in terms of administrative effort,

setup effort for implementing the data donation collection, participation rates, costs, epistemological consequences, etc.).

This research report addresses this shortcoming by sharing our experiences during a research project that collected Google Search and YouTube data donations to investigate how people used these platforms to inform themselves about upcoming democratic votes in the context of national referendum campaigns (Blassnig et al., 2023). We report our observations regarding three different strategies that we employed to recruit data donors: recruitment through an access panel, recruitment through social media advertisements, and recruitment through a commercial data marketplace.

Background

The purpose of the original study was to explore how Swiss citizens used search engines to find information about upcoming national democratic votes and how proponents and opponents of specific referendum proposals differed regarding their information search behaviors. To investigate this question, participants were asked to donate their Google Search and YouTube histories and to answer a survey containing self-reported measures on variables such as voting behaviors, political attitudes, and information behaviors. The study was conducted in Switzerland, and the data collection took place from 29 November to 22 December 2021 (for more details on the original study, see Blassnig et al., 2023).

Recruitment of Participants

During the study, we recruited participants through three different channels: (1) an online access panel of a market research company, (2) social media advertisements, and (3) a commercial data marketplace. The initial strategy was to recruit participants exclusively through the online access panel of the market research company *respondi*, essentially applying the same recruitment strategy that is frequently utilized for online surveys. The original sample size target for the recruitment through the access panel was 300 participants, which was significantly lower than a sample size that can realistically be achieved through the same approach in a standard online survey (without targeting specific demographic quotas). Participation was rewarded with monetary compensation (on average 23.6 CHF). Unfortunately, the response rate was much lower than anticipated. Therefore, we decided to invest the remaining financial recruitment resources in additional recruitment channels to increase our sample size: On the one hand, we

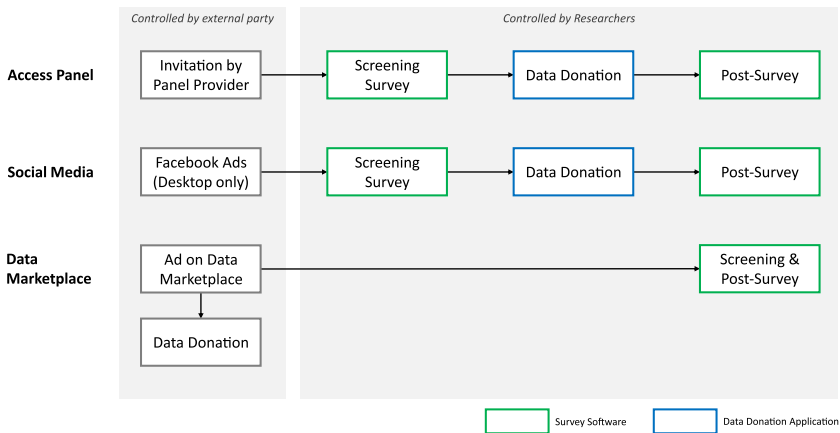


Figure 1: Participation Process Across Online Recruitment Channels

placed Facebook advertisements to recruit participants. On the other hand, we promoted the study on a commercial data marketplace (bitsabout.me).

The social media advertisement campaign was aimed at German-speaking Swiss users above the age of 18. It was limited to desktop users because the instruction materials and web application were optimized for desktop use. This was also the reason why the social media advertisements were restricted to Facebook, as advertisements placed on Instagram could not be targeted at desktop devices. On social media, participation was incentivized by having a high chance of winning a 100 CHF voucher for a big retail company.

The data marketplace *bitsabout.me* is a commercial platform through which private individuals can share their personal data with companies or research groups.¹ This marketplace provided an infrastructure for uploading and connecting personal data with a user account and subsequently selling the data to projects that are advertised on the platform. On the marketplace, participation was rewarded with a monetary compensation of 20 CHF.

Implementation of the Data Collection Across Online Recruitment Channels

The implementation of the data collection and participation steps varied across recruitment channels, as illustrated in Figure 1.

¹bitsabout.me no longer offers the possibility of sharing/collecting Google Takeout data (checked in May 2024).

Participants recruited through the access panel first received a study invitation from the market research company. When they clicked on the link in the invitation, they were redirected to a screening survey implemented with the software Unipark (Tivian, 2022). In the screening survey, participants indicated whether they were generally logged in with their user account when using YouTube and Google Search (a precondition for data donation) and received information on the upcoming data donation task. Participants who passed the screening and indicated their willingness to continue with the study were redirected to a web application developed and hosted by the researchers. This web application was an early prototype of the Data Donation Module (DDM; Pfiffner et al., 2022). It consisted of (a) a set of step-by-step instructions in the form of both text and GIFs that guided participants through the process of requesting their personal data through the Google Takeout platform and (b) an upload form through which they could upload their personal data. To request their data, participants visited the Google Takeout website (<https://takeout.google.com/settings/takeout>), where they could select the types of data they wanted to export: in this case, their Google Search and YouTube usage data stored under “My Activities” (see the online appendix² for a detailed overview of the data request steps). During the upload, the data were filtered (i.e., only activities registered after 14 June 2021 were kept in the donated data), and the reduced data were then presented back to participants in the form of tables containing the extracted information for each of the donated files. Based on the reduced datasets, participants could make a final decision regarding whether they wanted to donate. Afterwards, they were redirected to the post-survey, which included questions related to the core research question.

The procedure for participants recruited through social media was comparable to the access panel, except that participants’ initial point of contact with the study was a social media advertisement displayed on Facebook. The advertisement campaign was designed and managed by the researchers.

Participants recruited through the data marketplace saw a promotion consisting of a description of the study on the data marketplace when they were logged in with their marketplace account. In addition, they received a promotional e-mail from the data marketplace. If participants decided to provide their data for the study, they first had to confirm that they wanted to share their data with the research project. In contrast to participants recruited through the other channels, these participants received instructions from the data marketplace on how to request and upload their data from

²<https://doi.org/10.17605/OSF.IO/G6SWX>

Google, and it was possible that they had already uploaded their media usage data to the marketplace beforehand. Hence, these participants were not instructed by the researchers on how to access their data, and they only decided whether they wanted to share the data that they had already uploaded to the data marketplace. Therefore, the data collection and storage were handled by the marketplace provider. Second, participants were asked to fill out the post-survey. Participants had to go through these steps separately, as the data donation and the survey part were not directly linked, and both had to be accessed one after another through the data marketplace interface. Participation was only successful if both steps were completed.

Experiences and Learnings Across Recruitment Channels

Across the three channels, a total of 155 participants successfully donated their data. In the following, we compare and discuss our experiences with these three recruitment approaches in terms of five dimensions that are of practical relevance for the planning and conducting of a data donation study: *administrative effort*, *setup effort for implementing the data collection* (i.e., the data donation), *participation rate*, *costs*, and *analytical considerations*.

Administrative Effort

Administrative effort refers to all administrative tasks needed to organize a study before the data collection can start. This includes communication and negotiation with recruitment partners, as well as the management of the participant recruitment.

In this regard, setting up the social media advertisement campaign was the quickest, as there was no communication and negotiation needed with another party. Regarding collaboration with the access panel, the first challenge was to find a company that was willing to open its pool of participants to a data donation study, as several panel providers were hesitant to invite their samples to take part in such a study. This was mostly due to the novelty of the approach, and one panel provider explicitly expressed the fear that such a study design could irritate panel members and harm the reputation of the panel. Once a suitable access panel was found, correspondence took place via e-mail and followed a relatively standardized and predictable process comparable to administering a traditional survey study. Communication and negotiation with the data marketplace were less straightforward in our case because the relatively young company had no standardized work-

flows in place. However, it can be expected that when collaborating with a more experienced and established data marketplace, the administrative effort will be similar to or less than that of an access panel.

As this dimension is often the first step of a research project, we recommend planning considerable time to find a suitable cooperation partner when aiming to work with an access panel or data marketplace. This may be more difficult compared to a normal survey project. Hence, if the country context is not important, it may even be worthwhile to broaden the scope to other national contexts, as there may be a larger availability of suitable cooperation partners. This is particularly applicable to smaller countries or when looking for a commercial data marketplace, as such companies are not yet common or established in many countries.

For the social media context, the administrative burden is limited, and setup routines are relatively simple and well documented, but certain skills and knowledge are required for setting up the campaign, especially to optimize the cost per participant. Besides being able to navigate the social media platform's interface, no negotiation with a third party is necessary. However, this means that researchers—depending on the type of platform and campaign—must prepare advertising materials (e.g., different invitation texts and banners to catch users' attention), which can impact the recruitment success. While all of this can be implemented independently and very quickly, unforeseen incidents make regular monitoring by the researcher necessary. In our case, the advertisement campaign was flagged as fraud by a user and automatically deactivated. This required active intervention by us and a review by the social media platform before the campaign was reactivated.

Compared to an access panel, recruiting participants through a data marketplace has the clear advantage that all parties involved already have a shared understanding of the data transaction that is about to take place. Consequently, less energy needs to be invested in the explanation of the data donation procedure and the setup of the briefing and guidance materials. Furthermore, if the information on participants' demographic characteristics is sufficient (i.e., no extra survey constructs are measured), the effort for researchers and participants can be even lower, as this information is often already collected by the data marketplace and available as meta data, which means that no additional survey must be implemented. Moreover, if participants have already uploaded the data needed for a given study to the data marketplace, the effort for participants is minimal.

Setup Effort for Implementing the Data Collection

The setup effort dimension refers to the setup, maintenance, and monitoring of the data collection process. For participants recruited through the access panel and social media advertisements, a data donation application was set up and hosted by the researchers themselves. This setup was time-consuming and demanded technical skills in which many communication scholars may not be trained. In the case of the data marketplace, the data collection was handled by the provider of the marketplace. However, this meant that we were not able to preprocess and filter the data, as only a “take all or no data of a participant” option was offered by the marketplace.

In our case, setting up the data donation application was a time-consuming task because at the time, no suitable data donation collection application existed off the shelf. Therefore, we used an application that we developed ourselves which was an early version of the Data Donation Module. Today, some applications are available that are potentially faster to set up and that enable the easy, secure, and transparent collection of data donations (e.g., Open Source Data Donation Framework [OS2DF; Araujo et al., 2022], Data Donation Module [DDM; Pffnner et al., 2022], and Port [Boeschoten et al., 2023]). These existing applications have different foci and may still require some technical skills for implementation, depending on the application and the researchers’ access to shared infrastructures. Therefore, we recommend comparing the applications in detail regarding their availability to the researchers as well as their fit not only to the research question and design but also to the technical skills of the research team.

Another relevant aspect of the data donation collection to consider is the monitoring of the ongoing donation process. In the case of a self-hosted donation application, the responsibility for this lies with the researchers. This can be an additional burden, but it also provides flexibility in the case of unforeseen issues. For example, in our study, participants were instructed to request their Google takeout in JSON format, but a change or bug introduced by Google in its takeout function during our study led to the data being consistently exported in HTML rather than JSON format. Consequentially, we had to promptly update the data extraction script on our application server to accommodate the new data format.

Considering all these aspects, using a data marketplace is the simplest choice in terms of setup effort because the data marketplace implements the whole process. However, looking at the other dimensions, we only recommend this recruitment channel when the pool of participants is very large and when the data quality and processing are transparent and fit the

researchers' needs.

Participation Rate

Overall, all three strategies resulted in relatively low participation rates. In the case of the access panel, of 4,235 persons contacted, 1,182 followed the link to the study, but only 134 (11.3%) successfully donated their data. This number was significantly lower than the estimation of 300 participants made by the access panel provider prior to the data collection. For the social media advertisements, out of 16,212 persons who saw the advertisement at least once on their desktop device, 296 clicked on the link to the data donation platform, and 15 (5.1%) completed the donation (see Table 1). In the case of the data marketplace, only eight people donated their data, so the yield of this approach was also low. This was particularly unexpected given that this platform is specifically targeted at users who are willing to share their data. The marketplace did not disclose the exact number of active users, but it made a confident prediction of 50 participants.

Predicting and optimizing the participation rate for data donation studies are challenging tasks and should have a high priority in the planning process. In our experience, the recruitment partners were overly optimistic in their estimations of the number of expected participants. We recommend interpreting such predictions cautiously. Therefore, choosing a recruitment channel with a large base of potential participants and experience with similar research designs is recommended. Furthermore, carefully designing communication materials that address factors that may increase people's motivation to participate, such as highlighting the relevance of individuals' data to the research project or fostering participants' beliefs that the donation of their data is an achievable task (Pffiffner & Friemel, 2023), can also affect the willingness to participate.

In the case of social media advertisement campaigns, the participation rate is essentially a function of the advertisement budget, as the pool of potential participants is extremely large for most countries. Investing time and administrative effort in experimenting with different ways of addressing and incentivizing social media users can increase the participation rate, as other studies have shown (e.g., Vogler et al., 2023).

Table 1: Overview of costs per recruitment channel

| Recruitment channel | Recruitment period | Overhead costs | Incentive costs | Total | N Participants | | | Cost per donation |
|---|-----------------------------|----------------|-----------------|-----------|----------------|------------------------|-------------|-------------------|
| | | | | | Started | Reached donation stage | Donated | |
| Access panel | 29 Nov 2021– 13 Dec 2021 | 1,002 CHF | 3,112 CHF | 4,114 CHF | 1,182 | 573 (41.5%) | 134 (11.3%) | 30.7 CHF |
| Social media | 15 Dec 2021– 23 Dec 2021 | 370 CHF | 1,000 CHF | 1,370 CHF | 296 | 59 (12.4%) | 15 (5.1%) | 91.3 CHF |
| Data marketplace | 22 Dec 2021– 03 Jan 2022 | 232 CHF | 160 CHF | 392 CHF | - | - | 8 | 49.0 CHF |
| Gross Costs | Total | 1,604 CHF | 4,272 CHF | 5,876 CHF | - | - | 157 | 37.4 CHF |
| Net cost per donation included in the final sample after data cleansing: ^a | | | | | | | | |
| | | | | | | | 128 | 45.9 CHF |

Note. ^aCases where the Google Search history only consisted of very few entries or was empty were excluded from the sample analyzed in our empirical study.

Costs

The recruitment costs were the highest and least predictable in the case of the social media advertisement campaign. On the one hand, the incentives (10 vouchers worth 100 CHF each) were raffled among only 15 participants. On the other hand, the costs for the campaign (370 CHF) were mainly associated with displaying the advertisement and were not directly linked to the number of recruited participants. Regarding recruitment through the panel provider and the data marketplace, most of the costs were incurred per actual participant, and the remaining parts of the costs were attributable to overheads known in advance. Table 1 provides an overview of the cost composition.

Based on our experiences, the recruitment of participants for a data donation study is considerably more expensive than for survey studies. On average, we paid 45.9 CHF for every data donation that was used in our study. This cost per donation comprised all the overhead costs paid to recruitment partners and the incentives paid directly to participants. Due to the low participation rate, the incentive for participants recruited through the access panel was gradually increased over the course of the study. Participants were first offered 7.40 CHF for completing the survey as well as 5 CHF for a successful data donation, which was increased to 7.40 CHF plus 22.20 CHF a week into the study. Unfortunately, this increase of the financial incentive had no noteworthy impact on the participation rate.³

While we would generally prefer channels where costs are directly tied to the number of participants, which makes expenses predictable, these channels may result in a situation in which there is a remaining budget, but the pool of participants is depleted. In such cases, a combination with another recruitment channel, such as social media, can be promising. For Facebook, the maximum amount to be spent on advertising the study could be planned, but this amount was not directly linked to the number of participants; rather, it was linked to the number of times that the advertisement was displayed.

Analytical Considerations

This section addresses our observations regarding the characteristics of the samples that were accessible through the different channels and the challenges associated with the combination of data donations collected

³In interpreting these amounts, please take into consideration that the relative price level of Switzerland is comparatively high.

from them.

The participant pools for the three approaches, and hence the characteristics of the obtainable samples, differed in terms of their size, the transparency of the sample characteristics (e.g., demographics), whether they were self-recruited or actively recruited, and the control over the recruitment process.

Unfortunately, due to the imbalance in the number of participants recruited through the three channels, we cannot provide a meaningful statistical comparison of the sample characteristics. However, theoretically, recruiting via social media provides by far the largest pool of potential participants, but it comes with deficits regarding both the transparency of and the control over the recruitment process and the characteristics of the sample. Although the tools for managing a social media campaign provide researchers with a multitude of options for targeting their advertisement to specific groups of users, whom the advertisement will be displayed to is ultimately opaque. In the case of the access panel, the characteristics of the pool of potential participants are made transparent, and it is possible to target specific audiences based on the sociodemographic characteristics that the panel provider has stored regarding its panel members. In theory, the same is true for a data marketplace, although in our case, the marketplace provider did not disclose to us detailed information about its user base. Furthermore, in our case, the panel provider was the only channel that actively recruited participants for its access panel, leading to a diverse sample. Therefore, for general purposes or stratified samples, we recommend using an access panel with an actively recruited panel. This also allows for controlling whether the recruitment process is biased, at least in relation to the panel characteristics. However, for fringe groups, social media with its very large user pool can still be an attractive option, as a panel may not contain enough participants with very specific or niche characteristics. Social media even allows for more tailor-made approaches, such as directly contacting specific individuals through personal messages or placing invitations in certain Facebook groups.

Furthermore, although our study was specifically targeted at desktop users, it is worth considering that many participants now use their mobile devices to take part in online studies. Therefore, instructions for data donations should ideally be addressed at both mobile and desktop users. Otherwise, a considerable share of participants will effectively be excluded from participating in the study, structurally biasing the data donation sample. Additionally, depending on the platform, data takeouts can be more or

less cumbersome to acquire on mobile or desktop devices, and not all participants are equally skilled at following instructions to down- and upload the data takeout. While these considerations pertain to all the approaches, researchers have control over this when implementing the data donation themselves, whereas they have no control over how it is executed in a data marketplace.

Finally, from a more technical perspective, when using different channels to collect data donations, it is of great importance to verify in advance that the structure of the collected data is compatible across donations. In the case at hand, the combination of donations from participants recruited through the access panel and the social media advertisement campaign was unproblematic because the data were collected through the same application within the same time span, and participants received identical instructions. However, combining these data with the cases recruited through the data marketplace proved problematic. The data uploaded by participants to the marketplace were structured differently (probably due to data preprocessing steps implemented by the marketplace itself) and combined different activities (e.g., likes and watched videos) in the same structure, making it difficult to reliably distinguish between activities. In the end, we did not include any of the cases collected through the data marketplace in our data analysis because of this issue. Therefore, based on this experience, we recommend always double checking in advance whether datasets gathered through different channels are compatible. Furthermore, it is advisable to collect data donations for one study in a specific and relatively tight time frame, as platforms can sometimes change the structure and content of their users' data takeouts, which can complicate or prevent the combination of data donations.

Limitations

Before summarizing and discussing our experiences, some limitations must be noted that have implications for the interpretation of this report. First, the primary goal of the study on which this report is based was not to systematically or experimentally compare different recruitment channels but to explore how citizens used Google Search to acquire information regarding upcoming political referendums. Therefore, participants were not evenly distributed across recruitment channels, and we cannot report statistical evidence but only share our observations and experiences. Second, the study was conducted in Switzerland, and some of our observations might not transfer to other contexts (e.g., due to differences between populations regarding

digital literacy, privacy concerns, or general digitalization advancements). Third, it is important to consider potential sampling biases of different recruitment channels when planning a data donation study (Boeschoten et al., 2022; Sen et al., 2021). It is also noteworthy that maximizing participation rates is not always the main objective, as it may introduce additional bias in the sample. For example, measures to maximize participation rates, such as increasing financial incentives, can have heterogenous effects across a population and only increase participation rates in certain groups (Singer & Ye, 2013), which may also differ across recruitment channels.

Unfortunately, the report at hand does not allow us to derive definitive conclusions in this regard, and further research explicitly designed to compare the biases of different recruitment channels is needed.

Discussion & Conclusion

This contribution provides important insights into our experiences with three different channels in recruiting participants for data donation studies. The comparison of our experiences along the dimensions of *administrative effort*, *setup effort*, *participation rate*, *costs*, and *analytical considerations* is of direct practical value for researchers planning to conduct data donation studies. Our observations are summarized in Tables 2 and 3. Overall, we find that participation rates were rather low across all recruitment channels, which must be considered when planning a data donation study. Based on our experiences, we derive the following recommendations for researchers planning a data donation study:

1. Access panel providers may lack experience with data donation studies and may not be willing to open their panels to a data donation study. Therefore, we recommend planning considerable time for finding a suitable, and ideally experienced, cooperation partner. Nonetheless, access panels are still a viable option for recruiting data donors, although significantly lower response rates and higher costs must be expected compared to survey studies.
2. Because data donations involve several steps for study participants, the process is prone to errors. For example, participants may request the wrong data format, upload unexpectedly large files, or encounter unforeseen technical errors. Therefore, we recommend using a data donation application that offers an easy-to-use monitoring solution for the ongoing data donation collection, and an application that can be directly controlled and troubleshot by the researchers. In this regard, we also

recommend comparing the available data donation applications (e.g., Araujo et al., 2022; Boeschoten et al., 2023; Haim et al., 2023; Pfiffner et al., 2022) in detail regarding their features and fit to the research question and research design.

3. We only recommend using a data marketplace if it is an established platform, the size of the pool of active users is substantial, and the data collection and preprocessing steps are transparently reported. Furthermore, to avoid data quality and compatibility issues, we recommend always double checking in advance whether datasets collected by external partners are complete and contain the information one would expect (e.g., does the data contain all the information that one would obtain by using the official data takeout from a platform?).
4. Recruitment partners often lack experience with data donation studies. As a result, they may be overly optimistic in their estimation of the number of expected participants. In our case, this also applied to the number of expected clicks predicted in the Meta ad manager. We recommend interpreting such predictions cautiously.
5. We see potential in recruiting participants for data donation studies through social media advertisements, especially if the study setting allows for running the ad campaign over a longer period of time and participation on mobile devices is possible. If participation on mobile devices is not possible, many social media users cannot be reached, as these platforms are mostly accessed through mobile devices. Some platforms, such as Instagram, run their ads on both mobile and desktop devices.

This report is a response to calls for more exchange surrounding the opportunities and challenges of data donations. We want to emphasize that the recommendations stated in this report are based on our experiences during one research project and should thus be generalized cautiously. However, only through the accumulation of practical experiences will the research community be able to leverage the potential of data donations. Future research involving data donations can profit from the experiences reported here, as they inform the planning phase and safeguard against avoidable mistakes.

Table 2: Summary of practical experiences for each recruitment channel: Administrative and setup effort

| | Access panel | Social media | Data marketplace |
|--|--|---|---|
| Administrative effort | <ul style="list-style-type: none"> + There is a familiar administrative process for the panel provider and participants. It is comparable to survey studies. - Not all panel providers are willing to open their panels for a data donation study. Therefore, more time must be planned for finding a suitable panel provider. | <ul style="list-style-type: none"> + No communication with a third party is required, as all administrative work is done via the social media platform's ad management interface. + It can be implemented very fast, and the administrative effort is comparatively low. - Preparation of ad materials is necessary (e.g., invitation text and visual banners to catch users' attention). - Constant monitoring of the running ad campaign by the researchers is necessary. | <ul style="list-style-type: none"> + In theory, if a marketplace is established, has an active userbase, and has experience working with academic partners, the administrative effort can be minimal. + No donation instructions must be designed for participants. - Data marketplaces may not be available in many countries. - It is a comparatively novel and often not very established recruitment channel. Thus, administrative processes and routines may not be optimized. |
| Setup effort for implementing the data collection | <ul style="list-style-type: none"> Researchers usually have to set up and maintain the data collection process themselves: - There is a high setup effort, usually requiring technical knowledge. | | <ul style="list-style-type: none"> + There is no setup effort for the data donation, as data collection is handled by the data marketplace. + Data can be accessible instantly after a data donor has agreed to participate. o If specific information on participants is required (basic socio-demographics are usually available), an additional survey may have to be set up by the researcher. - There is no control over the data collection process. |

Note. + advantage; o neutral observation; - disadvantage.

Table 3: Summary of practical experiences for each recruitment channel: Participation turnout, costs, and analytical considerations

| | Access panel | Social media | Data marketplace |
|----------------------------------|--|---|--|
| Participation rate | <ul style="list-style-type: none"> - A large participant pool is required, as participation rates are low. - Participation rates are considerably lower compared to a normal survey project. - Participation rates may be overestimated by the panel provider. | <ul style="list-style-type: none"> + The number of participants is dependent on the budget, as the pool size of potential participants is usually large. Depending on the budget, many participants can be recruited. - Participation rates (i.e., conversion rates) may be overestimated by the application to manage the ads. | <ul style="list-style-type: none"> - Participation rates may be overestimated by the panel provider. |
| Costs | <ul style="list-style-type: none"> + Most costs are directly connected to the participation turnout. + Overhead costs are known in advance. + The cost per participant is known in advance. - It is more expensive than survey studies. | <ul style="list-style-type: none"> + It is easy to monitor the running costs in the ad manager, and the desired amount to be spent can be predefined. - Costs are only partially linked to participation turnout. - Cost per participant is difficult to predict in advance. | <ul style="list-style-type: none"> + Most costs are directly connected to participation turnout. + Overhead costs are known in advance. + The cost per participant is known in advance. |
| Analytical considerations | <ul style="list-style-type: none"> + The panel is often actively recruited. + High transparency regarding population characteristics + High control over the recruitment process + Data collection is controlled by researchers. - It has the same drawbacks that all access panels have (e.g., the panel population presumably already contains biases). | <ul style="list-style-type: none"> + Potential access to fringe groups or selected individuals + Data collection is controlled by researchers. - Limited control over the recruitment / advertisement process | <ul style="list-style-type: none"> + High transparency regarding population characteristics + High control over the recruitment process o Control over and transparency of the data collection process depend on the communication policy of the data marketplace. - Presumably strong self-selection bias - Potential data compatibility issues when combining with data donations of other recruitment channels |

Note. + advantage; o neutral observation; - disadvantage.

References

- Araujo, T., Ausloos, J., van Atteveldt, W., Loecherbach, F., Moeller, J., Ohme, J., Trilling, D., van de Velde, B., de Vreese, C., & Welbers, K. (2022). Osd2f: An open-source data donation framework. *Computational Communication Research*, 4(2), 372–387. <https://doi.org/10.5117/CCr2022.2.001.ArAU>
- Ausloos, J., & Veale, M. (2021). Researching with data rights. *Technology and Regulation*, 2020, 136–157. <https://doi.org/10.26116/techreg.2020.010>
- Blassnig, S., Mitova, E., Pfiffner, N., & Reiss, M. V. (2023). Googling referendum campaigns: Analyzing online search patterns regarding swiss direct-democratic votes. *Media and Communication*, 11(1), 19–30. <https://doi.org/10.17645/mac.v11i1.6030>
- Boeschoten, L., Ausloos, J., Möller, J. E., Araujo, T., & Oberski, D. L. (2022). A framework for privacy preserving digital trace data collection through data donation. *Computational Communication Research*, 4(2), 388–423. <https://doi.org/10.5117/CCr2022.2.002.BoEs>
- Boeschoten, L., de Schipper, N. C., Mendrik, A. M., van der Veen, E., Struminskaya, B., Janssen, H., & Araujo, T. (2023). Port: A software tool for digital data donation. *Journal of Open Source Software*, 8(90), 5596. <https://doi.org/10.21105/joss.05596>
- Breuer, J., Bishop, L., & Kinder-Kurlanda, K. (2020). The practical and ethical challenges in acquiring and sharing digital trace data: Negotiating public–private partnerships. *New Media Society*, 22(11), 2058–2080. <https://doi.org/10.1177/1461444820924622>
- Breuer, J., Kmetty, Z., Haim, M., & Stier, S. (2023). User-centric approaches for collecting facebook data in the “post-api age”: Experiences from two studies and recommendations for future research. *Information, Communication Society*, 26(14), 2649–2668. <https://doi.org/10.1080/1369118X.2022.2097015>
- DataSkop. (2023). *Technologie open source [technology open source]. algorithmwatch, europa-universität viadrina, fachhochschule potsdam, universität paderborn, and verein mediale pfade*. <https://dataskop.net/technologie-open-source/>
- Haim, M., Leiner, D., & Hase, V. (2023). Integrating data donations in online surveys. *Medien Kommunikationswissenschaft*, 71(1-2), 130–137. <https://doi.org/10.5771/1615-634X-2023-1-2-130>
- Halavais, A. (2019). Overcoming terms of service: A proposal for ethical distributed research. *Information, Communication Society*, 22(11), 1567–1581. <https://doi.org/10.1080/1369118X.2019.1627386>
- Lazer, D., Pentland, A., Watts, D. J., Aral, S., Athey, S., Contractor, N., Freelon, D., Gonzalez-Bailon, S., King, G., Margetts, H., Nelson, A., Salganik, M. J., Strohmaier, M., Vespignani, A., & Wagner, C. (2020). Computational social science: Obstacles and opportunities. *Science*, 369(6507), 1060–1062. <https://doi.org/10.1126/science.aaz8170>
- Ohme, J., & Araujo, T. (2022). Digital data donations: A quest for best practices. *Patterns*, 3(4), 100467. <https://doi.org/10.1016/j.patter.2022.100467>

- Ohme, J., Araujo, T., de Vreese, C. H., & Piotrowski, J. T. (2020). 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>
- Pfiffner, N., & Friemel, T. N. (2023). Leveraging data donations for communication research: Exploring drivers behind the willingness to donate. *Communication Methods and Measures*, 17(3), 227–249. <https://doi.org/10.1080/19312458.2023.2176474>
- Pfiffner, N., Witlox, P., & Friemel, T. N. (2022). *Data donation module [computer software]*. <https://github.com/uzh/ddm>
- Sen, I., Flöck, E., Weller, K., Weiß, B., & Wagner, C. (2021). A total error framework for digital traces of human behavior on online platforms. *Public Opinion Quarterly*, 85(S1), 399–422. <https://doi.org/10.1093/poq/nfab018>
- Singer, E., & Ye, C. (2013). The use and effects of incentives in surveys. *The ANNALS of the American Academy of Political and Social Science*, 645(1), 112–141. <https://doi.org/10.1177/0002716212458082>
- Skatova, A., & Goulding, J. (2019). Psychology of personal data donation. *PLOS ONE*, 14(11). <https://doi.org/10.1371/journal.pone.0224240>
- Stier, S., Breuer, J., Siegers, P., & Thorson, K. (2020). Integrating survey data and digital trace data: Key issues in developing an emerging field. *Social Science Computer Review*, 38(5), 503–516. <https://doi.org/10.1177/0894439319843669>
- Tivian. (2022). *Unipark: Survey-software*. <https://www.unipark.com/en/survey-software/>
- Trappmann, M., Haas, G.-C., Malich, S., Keusch, F., Bähr, S., Kreuter, F., & Schwarz, S. (2022). Augmenting survey data with digital trace data: Is there a threat to panel retention? *Journal of Survey Statistics and Methodology*, 11(3), 541–552. <https://doi.org/10.1093/jssam/smac023>
- Vogler, D., Weston, M., Ryffel, Q., Rauchfleisch, A., Eisenegger, M., Schwaiger, L., & Christen, U. (2023). *Mobile news consumption and its relation to young adults' knowledge about and participation in referendums*. <https://doi.org/10.5167/uzh-230226>