

**Linking surveys and digital trace data: Insights from two studies on determinants of  
data sharing behavior**

Henning Silber\*†

Florian Keusch\*\*

Johannes Breuer\*

Pascal Siegers\*

Christoph Beuthner\*

Sebastian Stier\*

Tobias Gummer\*

Bernd Weiß\*

\*GESIS – Leibniz Institute for the Social Sciences

\*\*University of Mannheim

† corresponding author, contact: [henning.silber@gesis.org](mailto:henning.silber@gesis.org)

**March 2021**

**Keywords:** data linkage, consent, social network sides, data sharing rates, incentives, data  
donation

Recommended citation of this preprint:

Silber, H., Breuer, J., Beuthner, C., Gummer, T., Keusch, F., Siegers, P., Stier, S., and Weiß, B. (2021), “Linking surveys and digital trace data: Insights from two studies on determinants of data sharing behavior”, SocArXiv, Available at <https://doi.org/10.31235/osf.io/dz93u>.

**Abstract**

Combining surveys and digital trace data can enhance the analytic potential of both data types. We present two studies on factors influencing data sharing behavior for different types of digital trace data: Facebook, Twitter, Spotify, and health app data. Across those data types, we compare the relative impact of five factors on data sharing: data type, data sharing method, respondent characteristics, sample composition, and incentives. The results show large differences between the data types and sharing methods, especially related to task difficulty and respondent burden. Higher incentives generally increase data sharing rates, whereas the influence of respondent characteristics depends on the respective data types.

## 1. Introduction

The widespread use of digital devices and online platforms produces vast amounts of data. These data are often subsumed under the category of digital trace data and can be a rich source of information for the social and behavioral sciences (King, 2011; Ledford, 2020; Shlomo & Goldstein, 2015). For social scientists, digital traces are especially interesting when they are available on the individual level and can be linked with data on person-level variables, such as attitudes, values, personality, and personal characteristics which are typically collected via surveys (Stier et al., 2020).

A fruitful approach for increasing the analytic potential of digital trace data is to combine them with data from surveys so that the two data sources can complement and enrich one another (Al Baghal et al., 2020; Amaya et al., 2019; DiGrazia et al., 2013; Harari et al., 2017; Stier et al., 2020). While surveys usually rely on reported or self-reported behavior, digital trace data allow to record and track many different types of behavior over time and with high granularity. On the other hand, surveys allow getting more profound insights into personal belief systems, which helps to understand behavior and investigate causal relationships (Stier et al., 2020). While the combination of those two data types holds great promise for the social sciences, the practicalities of linking surveys and digital trace data and their implications have not yet been systematically studied. Understanding the different ways in which surveys and digital trace data can be linked is important to assess the nature and quality of the resulting combined data. The options for linking surveys and digital trace data and their outcomes depend on a variety of factors, including technological developments and privacy considerations (Boeschoten et al., 2020; Nissenbaum, 2009, 2018; Oberski & Kreuter, 2020).

Against this background, this paper presents results from two studies that explored different ways of combining surveys and digital trace data. The main objective of both studies was to gain insights into the data sharing process. In Study 1, we explored the combination of

surveys with different types of social media and web tracking data. In Study 2, we experimentally explored the combination of a survey with social media and health app data. By identifying which factors determine participants' data sharing behavior in studies that link surveys and digital trace data, our aim is to derive recommendations and best practice advice for future research.

## **2. Linking Survey Data and Digital Trace Data**

A straightforward way to link survey and digital trace data is to ask respondents within a survey whether they are willing to share additional data (Kreuter et al., 2020; Sakshaug, 2020; Sloan et al., 2020). However, such questions can constitute an additional response burden (Eckman & Haas, 2017), leading to high item nonresponse or even to break-off.

Since a data sharing request introduces a new demand into the survey interview, respondents must evaluate whether they approve the request. As for any behavioral choice, cost and benefit considerations are likely to guide the decision-making process (Biner & Kidd, 1994; Dutwin et al., 2015; Esser, 1986; Leeper, 2019; Porter & Whitcomb, 2003). Benefits of an affirmative answer can include congruence with conversational norms, time savings, and possible incentives. First, conversation norms of human interactions suggest that it is perceived as impolite to say no and decline a request. Second, time savings may include a shorter questionnaire because information that would have otherwise been collected through self-reports can be derived from digital traces. Third, incentives are often monetary stimuli that will be received if a person agrees to the request for data. Costs of an affirmative answer can include the effort necessary to share the data and the data's potential sensitivity.

Additional considerations that may guide the decision process are attitudes and norms toward privacy, data sharing, and scientific research (Keusch et al., 2020; Oberski & Kreuter, 2020; Sloan et al., 2020). In a situation where costs are high, for example, if a person is asked to go through multiple complicated steps to share their data or if the requested data is especially sensitive, attitudes and norms are unlikely to strongly affect the decision-making

process (Best & Kroneberg, 2012; Riker & Ordeshook, 1968; Stern, 1992). However, in the context of the data-sharing request where the costs are low, attitudes regarding privacy or conversational norms are more likely to influence the decision-making process for data sharing.

Typically, three conditions must be met before respondents share their digital trace data (Elevelt et al., 2019; Keusch et al., 2019; Revilla et al., 2018, 2019). First, only users of the platforms, services, or devices that generate the data can be asked to share it (*usage*). For instance, for social media data, those rates can vary dramatically between age groups, occupations, platforms, and countries (e.g., among US adults: 68% use Facebook, 40% use Instagram, and use 25% for Twitter, see Shearer and Mitchell, 2021). Second, researchers have to obtain *informed consent* from participants to prove that they are willing to share their digital traces. Third, users who have given informed consent have to successfully complete the data sharing procedure to share their data (*data sharing behavior*). This is, for instance, necessary when respondents have to download and install an app or a browser plug-in or are asked to export and share data from the platforms or devices under study.

### **2.1. Factors influencing data sharing**

While there are some previous studies that compared willingness to share additional data in surveys across different *data types* (Jenkins et al., 2006; Revilla et al., 2016, 2019; Wenz et al., 2019), it remains largely an open question how different digital trace data sources directly compare to each other and to which extent cost-benefit structures as well as attitudes and norms affect the willingness to share these data.

First of all, the way in which people are asked to share their digital trace data (*data sharing method*) is likely to influence the sharing decision (Boeschoten et al., 2020; Settanni et al., 2018). There are various ways of collecting digital trace data that differ, among other things, in the type of data they generate as well as in the amount of effort the data sharing requires (Breuer et al., 2020). Many of these data collection options can also be used in

studies where researchers partner with users to access digital trace data (Halavais, 2019). For example, participants can be asked to simply share their Twitter handle, allowing researchers to collect their Twitter data through the platform's API. Or researchers can ask participants to use an app or browser-plugin that records (parts of) their digital traces (de Haan & Hendriks, 2013; Haim & Nienierza, 2019; Kosinski et al., 2013). Another option is to ask participants to export (parts of) their digital trace data themselves, which is a functionality that most platforms and services offer, and then share these "data download packages" (Boeschoten et al., 2020) with the researchers, e.g., by uploading them through a web tool. This relatively new approach is often called "data donation" due to the active role of the respondents within the data sharing process. Notably, those different data sharing procedures vary with regard to task difficulty and respondent burden, which translates into perceived costs. Specifically, more active data sharing procedures, such as data donation, are usually more burdensome for respondents than passive ones, such as providing consent and a username (Keusch et al., 2019).

It is uncertain in which situations attitudes and norms (*respondent characteristics*) regarding privacy, data sharing, or science guide data sharing behavior (Keusch et al., 2019). Using the distinction between low-cost and high-cost situations introduced above (Best & Kroneberg, 2012), it is an open question under which circumstances a data sharing situation is considered a low-cost or high-cost situation, and which attitudes and norms are more likely to affect decision making in those scenarios. In addition, sociodemographic attributes of respondents may also be relevant for data sharing decisions. For example, older people are typically less knowledgeable when it comes to digital media (Kuru et al., 2018; Smith & Page, 2015) which might be especially important for data sharing methods that require a substantial amount of effort from the participants.

While it is evident that *monetary incentives* can increase the likelihood of data sharing, it is not clear which amount should be offered (Jäckle et al., 2019; Keusch et al., 2019) and

whether it should be offered as a prepaid or postpaid incentive. This is not only relevant from a cost and efficiency perspective, since researchers typically do not have unlimited funds and would like to implement an optimal incentive strategy, but also from another practical perspective. Specifically, if an unusually high incentive is offered, respondents might even be less likely to share their data because they might regard the data as very valuable – potentially because of its perceived sensibility – or suspect a possibly harmful use.

Data sharing behavior may also vary depending on the *sampling method and sample composition* (Brosnan et al., 2017; Elevelt et al., 2019; Jäckle et al., 2019; Keusch et al., 2019). For example, respondents from a cross-sectional, general population sample might be less likely to share additional data than respondents from special populations, such as participants of commercial online access panels, who might be more familiar with requests to share digital content.

### 3. Study 1

#### 3.1. Methods

**Data.** The data for Study 1 comes from a non-probability panel of German Internet users who agreed to use software that tracks their web browsing behavior on desktop computers and/or smartphones. The panel is managed by a German market research company and contains around 2,000 participants<sup>1</sup>. For our study, we acquired access to the web tracking data from June 2018 to May 2019. During that period, participants of the tracking panel were invited to complete two web surveys which included the data sharing requests. All 2,042 individuals, who participated in the web tracking panel in July 2018, were invited to participate in the first online survey, of which 1,411 followed the invitation and started the survey (participation rate = 69.1%). 1,355 panelists completed the first survey (completion rate = 96.0%). For the second survey, all 1,931 individuals, who participated in the web tracking in March 2019,

---

<sup>1</sup> The sample size of the sample fluctuated due to dropout and consecutive sample refreshing but was always around 2,000 respondents for each month.

were invited and 1,325 took part in the survey (participation rate = 68.6%), of which 1,240 completed the survey (completion rate = 93.6%).

In the first survey, we asked respondents whether they are willing to share their Twitter data, and in the second survey, whether they are willing to share Facebook and Spotify data.

Survey 1 also included an incentive experiment, in which respondents were randomly assigned to receive a 5€ data sharing incentive after (postpaid) or before (prepaid) the completion of the data sharing procedure for their Twitter data. In Survey 2, for the sharing of their Facebook data, respondents received a 5€ postpaid incentive, and for the sharing of their Spotify data, respondents received a 2.50€ postpaid incentive<sup>2</sup>; there was no experimental variation of incentive condition in the second survey. The median response time for those who completed the first survey was 15 minutes and 2 seconds. For the second survey, it was 13 minutes and 34 seconds.

**Data Sharing Procedure.** Different methods were used to collect the social media data in Study 1. First, respondents were asked to provide informed consent for each of the data requests in the two surveys (see Section 2 of the Online Appendix). In addition to the information provided in the survey questions, participants had the opportunity to read further information about privacy and data handling via a website URL prominently placed in the survey text. However, only a small number of respondents used this opportunity<sup>3</sup>. Twitter data was collected using the platform's public APIs and included profile information and up to 3200 past tweets (collected via the REST API) as well as new tweets by the participants from the end of field phase of the second online survey in August 2018) until the end of the project's overall data collection phase in May 2019 (via the STREAM API). The Twitter data could be linked with the survey data via the username/handle. Respondents who indicated that

---

<sup>2</sup> These incentives were higher than the amount participations typically get paid for only answering a survey. This was meant to reflect the assumed response burden as well as the perceived sensitivity of the data.

<sup>3</sup> The web tracking data show that a maximum of 4 to 7 respondents accessed this extended information for at least one of the three data types.

they have a personal Facebook account were asked whether they are willing to install a browser plugin that collects public posts from their Facebook news feed whenever users login to their Facebook account and see or scroll through their feed (for details see Haim & Nienierza, 2019). The plugin was available for the desktop versions of the Firefox and Chrome browsers and could be downloaded and installed through the respective official plugin stores (Haim & Nienierza, 2019). Participants were able to deactivate the plugin and could also delete the data that it collected. Notably, none of the participants made use of the latter option. The Facebook data could be linked with the survey data through an anonymized ID code that participants were asked to generate following a specific pattern in the survey and during the plugin installation process. Respondents who reported that they have a Spotify account were asked to provide data on the last 50 songs played, their playlists, and music preferences as defined by Spotify through a web app accessible via a link in the survey. To use the web app, participants were asked to log in with their Spotify account. The app then collected the data via the Spotify API. Participants could review the data and decide whether they want to share it or not. The Spotify data and the survey data could be linked via a numeric participant ID that was passed on as a URL parameter from the panel management site of the market research company to the online survey site and then to the Spotify web app.

**Measures.** Regarding data sharing behavior of Twitter, Facebook, and Spotify data, we generated a dichotomous measure for each data type, indicating whether a respondent shared the respective data. Through the web and app tracking data, platform usage for Twitter, Facebook, and Spotify was tracked. Specifically, we used the number of website visits and app usage for the tracked period before the surveys. Survey 1 included a survey evaluation measure, and Survey 2 a measurement of privacy concerns. In addition, the device

respondents used to answer the surveys (smartphone/tablet vs. PC)<sup>4</sup> and demographic information about respondents' age, gender, education, and income were collected. See Table A3 in the Online Appendix Section 1 for the description and coding of all measures used in Study 1, and Tables A5 and A6 for the descriptive statistics of those variables.

**Analysis.** The results for Study 1 are based on descriptive analyses and logistic regression models predicting data sharing behavior. For the descriptive analyses, our main goal was to learn how many participants drop out during each of the three steps of the data sharing process (Step 1: platform usage; Step 2: informed consent; Step 3: data sharing). In the regression models, the dependent variable is dichotomous with “shared the respective data” coded as 1 and “did not share data” coded as 0. For those models, the category “did not share data” includes respondents who agreed to share the data but eventually did not share it either because they decided against it during the data sharing process or because of technical difficulties. Additional models predicting “informed consent” vs. “no informed consent” are displayed in the Online Appendix Section 1 (Table A1 for Study 1 and Table A2 for Study 2). All models were estimated using R 4.0.3 (R Core Team, 2020).

### 3.2. Descriptive Results

The descriptive results from Study 1 (see Figure 1 for a flowchart illustrating the entire data sharing procedure) show that the sample included 79.4% Facebook users, 22.8% Twitter users, and 22.7% Spotify users. Of those users, 31.2% shared their Facebook data, 41.4% their Twitter data, and 59.1% their Spotify data. Considering the full samples, 24.3% of the respondents shared their Facebook data, 13.4% their Spotify data, and 9.4% their Twitter data. Notably, only 785 of the 985 Facebook users (79.7%) received the consent request for the Facebook plugin. The reason for this was a technical issue during the data collection.

---

<sup>4</sup> We assumed that this distinction between device types (smartphone and tablet vs. desktop computer) would be important especially for the Facebook data as the plugin was only available for desktop browsers, which may lead to lower data sharing rates on smartphones and tablets compared to desktop computers.

Specifically, Facebook changed the newsfeed during our data collection phase, which caused the plugin to freeze the browser tab when users clicked on a picture from their feed to enlarge it. Of the respondents who agreed to the data sharing request, 83.0% shared their Spotify data, 65.0% their Facebook data, and 63.2% their Twitter data.

[Figure 1 about here]

### 3.3. Predicting Data Sharing

Higher platform usage, as measured by the tracking data, increased the likelihood of data sharing for all three data types (see Table 1). Also, some of the sociodemographic variables were significant predictors of the data sharing probability for all three data types. Specifically, for Twitter and Spotify, younger respondents were more likely to share their digital data. For Facebook and Spotify, male respondents were more likely to share their data, and for Facebook, respondents with higher levels of education were more likely to share their data. The device used to answer the survey only affected the likelihood of data sharing for Facebook. Respondents who answered the survey on a mobile device (tablet or phone) were less likely to share their data than respondents who answered the survey on their desktop computer. Respondents' survey evaluation and privacy concerns did not affect the likelihood of data sharing. Likewise, prepaid and postpaid incentives led to the same sharing rates for Twitter data.

[Table 1 about here]

## 4. Study 2

### 4.1. Methods

**Data.** The data for Study 2 comes from a German non-probability online panel ( $n = 3,136$ ; invitations = 26,339; participation rate = 39.8%; completion rate = 86.2%). The field period of the web survey in Study 2 was from October 2019 to December 2019. To increase the number of eligible respondents for the data sharing requests, Twitter usage and smartphone type (iPhone or Samsung) were used as screening variables (screen-out rate = 65.5%).

Respondents were asked whether they are willing to share their Twitter or health app data (the introductory texts are reported in the Online Appendix Section 2). Each respondent received only one of the two data sharing requests (Twitter or health app data) within the survey. One third of the respondents received the Twitter data sharing request and two thirds the health app data sharing request. For both data sharing requests, an incentive experiment with four groups was implemented: Respondents were randomly assigned to receive 0€, 2.50€, 5€, or 10€ for sharing their digital traces. The median response time of the survey was 12.4 minutes.

**Data Sharing Procedure.** For Twitter data, respondents were randomly assigned to share their data either by providing their user handle (passive procedure) or through data donation (active procedure). Health app (only for iPhone or Samsung devices) data was shared via data donation. To donate their data, respondents were first asked to export their data from the respective application or via the website (only Twitter). In a second step, they were asked to upload the data through a secure web tool, which is regularly used for sharing research data. To lower the burden, we provided instructions for each service, describing each step of the downloading and uploading process.

**Measures.** With respect to data sharing behavior for Twitter and health app data, we included a dichotomous measure for each data type, which indicated whether a respondent shared the respective data. For the health app data, we included a variable indicating a respondent's smartphone type (iPhone or Samsung). Platform usage was measured via self-reports for both Twitter and health app usage. Moreover, the survey included measures of privacy concerns, earlier experiences with privacy intrusion, technical affinity, perceptions of surveys in general (value, enjoyment, and burden; see De Leeuw et al., 2019), and the evaluation of the survey (see Gummer & Daikeler, 2020). In addition, data on the device (smartphone vs. PC/tablet)<sup>5</sup>

---

<sup>5</sup> We coded device as smartphone (1) versus PC/tablet (0) in Study 2. This was because donating data from the Apple and Samsung health apps was only possible using a smartphone. Thus, responding to the questionnaire on a non-smartphone required using an additional device to complete the data sharing task.

used for answering the survey and demographic information about respondents' age, gender, education, income, and personality characteristics (conscientiousness and openness) were collected. See Table A4 in the Online Appendix Section 1 for the description and coding of all measures of Study 2, and Table A7 for the descriptive statistics of those variables.

**Analysis.** As for Study 1, the results for Study 2 are based on descriptive analyses and logistic regression models predicting data sharing behavior.

#### **4.2. Descriptive Results**

The descriptive results from Study 2 (see Figure 2) show that the survey included 33.8% Twitter users and 66.2% health app users (32.6% iPhone users and 33.6% Samsung phone users). Of those Twitter users, 24.0% shared their data. In addition, 9.4% of the iPhone users and 2.7% of the Samsung phone users shared their health app data. Considering the full sample, 7.0% of the respondents shared their Twitter data through providing their user handle and 1.1% via data donation; 3.1% shared their iPhone health app data and 0.9% shared their Samsung health app data. Of the respondents who agreed to the respective data sharing requests, 95.6% shared their Twitter data by providing a user handle, and 17.3% via data upload; 22.2% shared their iPhone health app data and 5.6% shared their Samsung health app data.

[Figure 2 about here]

#### **4.3. Predicting Data Sharing**

Higher incentives increased the likelihood of data sharing for both data types (see Table 2). However, while for Twitter data an incentive of 2.50€ already increased the likelihood compared to not providing an incentive, for health app data, only the 10€ incentive condition led to a significant increase in the data sharing rate. Higher health app usage led to a decrease in the likelihood of data sharing, while self-reported platform usage did not show a significant effect for Twitter data. A positive survey evaluation increased the probability of data sharing for both data types. In addition, earlier privacy intrusion experiences decreased the likelihood

of data sharing for health app data, and the personality trait of openness decreased the likelihood of data sharing for Twitter data. In contrast, an affinity for technology increased the likelihood of data sharing for Twitter data. With respect to demographics, younger respondents were more likely to share their health app data and respondents with higher education were more likely to share their Twitter data than respondents with lower levels of education. Respondents who answered the survey on their smartphones were less likely to share health app data, and respondents who owned a Samsung smartphone were less likely to share health app data than iPhone owners. Perceptions of surveys and privacy concerns did not affect the likelihood of data sharing for the two data types in Study 2.

[Table 2 about here]

## **5. Discussion**

### **5.1. Summary of Results**

Taken together, our studies show that data sharing rates can vary dramatically between samples, data types, data sharing methods, devices, respondents, and incentive conditions. Since the crucial factor during the data sharing process appears to be respondents' burden, data sharing requests generally seem to represent a high-cost situation (Best & Kroneberg, 2012). As a consequence, attitudes and values only affect sharing behavior to a small degree. Hence, factors that directly relate to the data sharing difficulty, such as the data sharing method (Boeschoten et al., 2020) should be tested and optimized for research that seeks to combine surveys and digital trace data. Providing survey respondents with additional incentives (Keusch et al., 2019) also seems beneficial for increasing the willingness to share data from digital platforms.

For *data type*, our research shows that the users were most willing to share Spotify data (59.1%). Health app data (6.1%), in contrast, had the lowest overall data sharing percentages. However, Facebook (79.4%) and health apps (67.2%) had the largest share of

users in the full sample<sup>6</sup>. Considering the various steps of the data sharing process, researchers should try to anticipate the percentage of users as well as the likelihood of data sharing when planning their survey. Our results confirm findings from previous research comparing different data types (Revilla et al., 2019; Wenz et al., 2019), which also showed that differences in sharing rates can be very large (e.g., between 73.7% for receiving a product at home and 5.5% for letting respondents' children wear a small device that delivers real-time information about stress levels, see Revilla et al., 2019).

With respect to the *data sharing method*, our experimental design in Study 2 showed that respondents were more likely to share their data when they were asked to provide their user handle compared to exporting and uploading their Twitter data themselves. Since respondents only had to type in their Twitter handle for the API data collection, this sharing procedure was considerably less effortful. A similar effect was found for health app data sharing, which was especially effortful on Samsung devices (due to the need to identify and upload only certain files of those exported from the app), which resulted in a data sharing rate below 1%. These results indicate that, while the active data sharing method of data donation (Boeschoten et al., 2020) is certainly a promising tool for social research, researchers need to improve automatization and reduce respondents' burden to increase the likelihood of a donation.

Looking at *respondent characteristics* and *device usage*, we found that demographics and devices used were more likely to influence data sharing behavior than respondents' attitudes and values. Specifically, we found that younger respondents as well as male respondents were more likely to share their digital trace data. Potential reasons for these demographic differences are technical affinity and usage behavior. In fact, when not including technical affinity in Study 2 (Table 2), male respondents were more likely than female

---

<sup>6</sup> In case of the health app this is due to the screening process employed for Study 2.

respondents to share both Twitter and health app data. This effect, however, disappeared when a measure for technical affinity was included. With respect to user behavior, previous studies have shown gender differences in the use of social media platforms. For example, a study by Muscanell and Guadagno (2012) found that men use Facebook more often for forming new relationships, whereas women use it more often to maintain existing relationships. Another study found that women are more concerned about privacy and engage more often in privacy-protecting behavior on Facebook than men (Hoy & Milne, 2010). We also found mean differences in privacy concerns ( $t(1230.5) = 3.01, p = 0.003, d = 0.17$ ) between women ( $M = 3.75, SD = 0.93$ ) and men ( $M = 3.59, SD = 0.92$ ) in Study 1. Notably, while privacy concerns were included in the regression model for sharing Facebook data, gender was a significant predictor, with men being more likely to share their Facebook data than women. An explanation for this finding might be that higher privacy concerns among women are associated with different usage behaviors, which may have caused the gender differences in willingness to share the data.

Positive survey evaluation increased the likelihood of data sharing in Study 2. Interestingly, we found that the likelihood of sharing digital trace data increased with more regular platform usage for Facebook, Twitter, and Spotify data (Study 1), but decreased the probability of data sharing for health app data. Possibly, health information might be considered more sensitive compared to social media data that is often freely available for other users or at least directly visible to the users' contacts on these networks. The type of device used for answering the survey affected data sharing for Facebook and health app data. Both were particularly effortful (health app) or impossible (Facebook desktop application) to engage in on a smartphone. Similar to Elevelt et al. (2019), we did not find substantial effects of survey attitudes and privacy considerations on data sharing behavior in our studies.

Our *incentive* experiments within the surveys showed that respondents were more willing to share additional data when they received higher incentives (Study 2). However,

whether the incentives were prepaid or postpaid did not influence data sharing behavior (Study 1). An incentive of 2.50€ increased the likelihood that respondents agreed to share their Twitter data but for the health app data only an incentive of 10€ made a difference with respect to data sharing behavior. This difference between Twitter and health app data suggests that if the sharing process is more effortful, only high incentives can possibly compensate for the increased burden. However, even with high incentives, the data sharing rates remained relatively low for health app data, especially for respondents with Samsung devices where the data sharing process was most difficult. This finding is in line with previous research showing that higher incentives do not always increase data sharing rates (Jäckle et al., 2019) and that the time when incentives are given does not influence the likelihood of data sharing (Keusch et al., 2019).

With respect to the *sample composition*, we obtained higher linkage rates when we recruited respondents from the web tracking sample (Study 1), compared to the regular online access panel sample (Study 2). Participants in the tracking study had already agreed to provide more information than regular panel respondents so that it appears logical that they were also more willing to provide additional digital trace data. In Study 2, we successfully implemented a screening procedure to increase the number of Twitter users as well as iPhone and Samsung device users. If researchers aim for a high number of affirmative responses to the data sharing request, this practice is recommendable for future studies, especially when prepaid incentives are used to reduce survey costs.

## **5.2. Recommendations**

Requests for sharing digital trace data are likely to represent a high-cost situation in which respondents need to see their benefits clearly to be willing to share their data. Based on our two studies, we can derive four recommendations regarding the sample composition, data sharing method, incentives, and the devices used to answer a survey.

**Recommendation 1** *When designing studies, researchers should consider the percentage of users of their targeted platforms or devices in the sample. A possibility to increase the percentage of users is to implement a screening procedure as we have done in our second study.*

**Recommendation 2** *It is important to minimize the effort to share the data. This is especially important for data sharing methods in which respondents are asked to actively share their digital trace data (“data donation”) since these are often more burdensome than, for example, merely asking for a username and informed consent.*

**Recommendation 3** *Small incentives can increase the likelihood of data sharing if the data sharing task is not too burdensome. While large incentives can help to increase the likelihood of data sharing for very burdensome tasks, we recommend simplifying the data sharing process wherever possible so that small incentives are sufficient.*

**Recommendation 4** *We recommend pretesting the feasibility of the data sharing process on multiple devices, especially on smartphones and tablets. An increasing number of respondents prefer filling out questionnaires on mobile devices, so that researchers need to consider this during the study planning and design phase.*

### **5.3. Limitations**

While our research included several experimental and quasi-experimental (iPhone vs. Samsung device) designs, further experiments are required for gaining additional insight into respondents’ data sharing behavior. Given that linking surveys and digital trace data is a relatively new approach, there may well be other factors influencing data sharing behavior that we did not consider in our studies. Hence, the findings from our study need to be tested further for other samples, sharing methods, and types of data in replication studies.

The data for both of our studies came from non-probability samples, which is adequate for explorative and experimental studies. However, additional evidence is needed from samples drawn with probability-based methods to test whether the results are generalizable.

Given that our respondents were part of commercial access panels, we expect data sharing rates to be lower in general population samples. For such samples, following our Recommendations 2 to 4 will likely be even more important for increasing consent and data sharing rates. If researchers are interested in one or more groups of users (of specific platforms or devices), a screening procedure (Recommendation 1) may be a suitable way of increasing the likelihood of data sharing.

## 6. Literature

- Al Baghal, T., Sloan, L., Jessop, C., Williams, M. L., & Burnap, P. (2020). Linking Twitter and survey data: The impact of survey mode and demographics on consent rates across three UK studies. *Social Science Computer Review*, 38(5), 517–532.
- Amaya, A., Bach, R., Keusch, F., & Kreuter, F. (2019). New data sources in social science research: Things to know before working with Reddit data. *Social Science Computer Review*, 0894439319893305.
- Best, H., & Kroneberg, C. (2012). Die low-cost-hypothese. *KZfSS Kölner Zeitschrift Für Soziologie Und Sozialpsychologie*, 64(3), 535–561.
- Biner, P. M., & Kidd, H. J. (1994). The interactive effects of monetary incentive justification and questionnaire length on mail survey response rates. *Psychology & Marketing*, 11(5), 483–492.
- Boeschoten, L., Ausloos, J., Moeller, J., Araujo, T., & Oberski, D. L. (2020). Digital trace data collection through data donation. *ArXiv Preprint ArXiv:2011.09851*.
- 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>
- Brosnan, K., Grün, B., & Dolnicar, S. (2017). Use, preference and completion rates for web surveys. *International Journal of Market Research*, 59(1), 35–55.
- de Haan, J., & Hendriks, R. (2013). Online data, fixed effects and the construction of high-frequency price indexes. *Economic Measurement Group Workshop*, 28–29.
- de Leeuw, E., Hox, J., Silber, H., Struminskaya, B., & Vis, C. (2019). Development of an international survey attitude scale: measurement equivalence, reliability, and predictive validity. *Measurement Instruments for the Social Sciences*, 1(1), 1-10.
- DiGrazia, J., McKelvey, K., Bollen, J., & Rojas, F. (2013). More tweets, more votes: Social media as a quantitative indicator of political behavior. *PloS One*, 8(11), e79449.

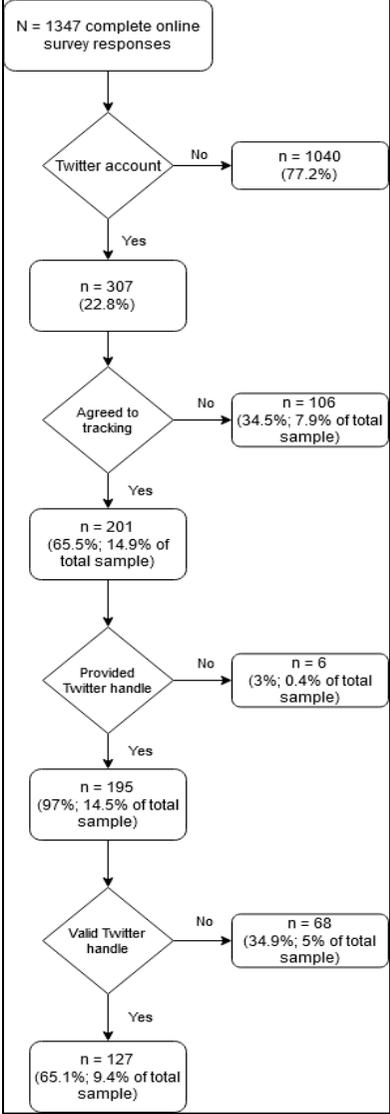
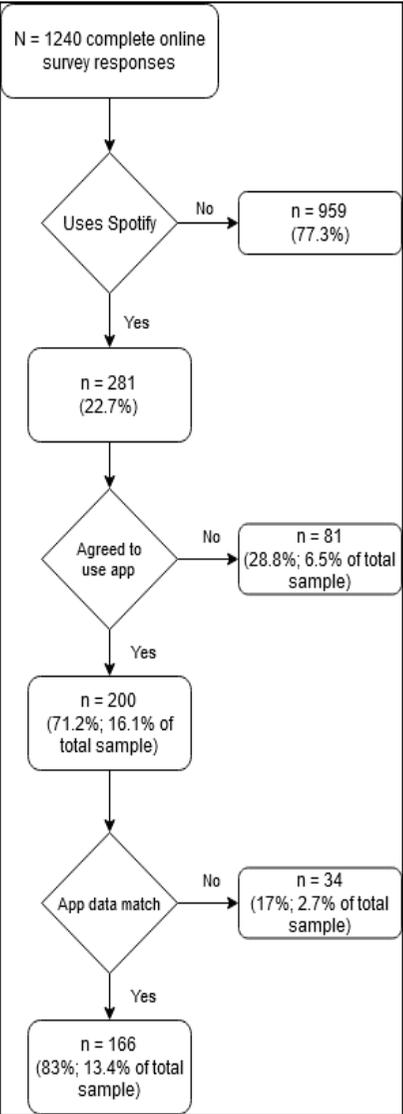
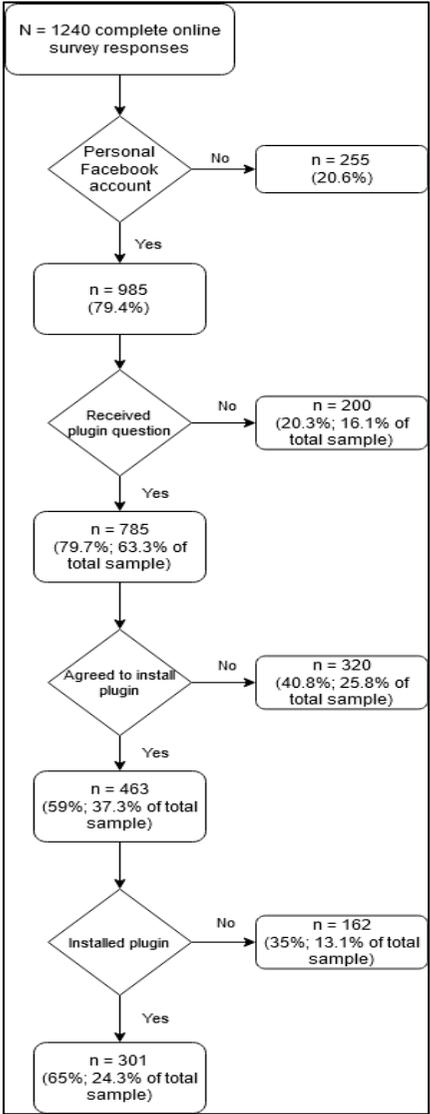
- Dutwin, D., Loft, J. D., Darling, J. E., Holbrook, A. L., Johnson, T. P., Langley, R. E., Lavrakas, P. J., Olson, K., Peytcheva, E., & Stec, J. A. (2015). Current knowledge and considerations regarding survey refusals: Executive summary of the AAPOR task force report on survey refusals. *Public Opinion Quarterly*, 79(2), 411–419.
- Eckman, S., & Haas, G.-C. (2017). Does granting linkage consent in the beginning of the questionnaire affect data quality? *Journal of Survey Statistics and Methodology*, 5(4), 535–551.
- Elevelt, A., Lugtig, P., & Toepoel, V. (2019). Doing a Time Use Survey on Smartphones Only: What Factors Predict Nonresponse at Different Stages of the Survey Process? *Survey Research Methods*, 13(2), 195–213.
- Esser, H. (1986). Über die Teilnahme an Befragungen. *Zuma Nachrichten*, 10(18), 38–47.
- Haim, M., & Nienierza, A. (2019). Computational observation: Challenges and opportunities of automated observation within algorithmically curated media environments using a browser plug-in. *Computational Communication Research*, 1(1), 79–102.
- Halavais, A. (2019). Overcoming terms of service: A proposal for ethical distributed research. *Information, Communication & Society*, 22(11), 1567–1581.
- Harari, G. M., Müller, S. R., Aung, M. S., & Rentfrow, P. J. (2017). Smartphone sensing methods for studying behavior in everyday life. *Current Opinion in Behavioral Sciences*, 18, 83–90.
- Hoy, M. G., & Milne, G. (2010). Gender Differences in Privacy-Related Measures for Young Adult Facebook Users. *Journal of Interactive Advertising*, 10(2), 28–45.  
<https://doi.org/10.1080/15252019.2010.10722168>
- Jäckle, A., Burton, J., Couper, M. P., & Lessof, C. (2019). Participation in a mobile app survey to collect expenditure data as part of a large-scale probability household panel: Coverage and participation rates and biases. *Survey Research Methods*, 13(1), 23–44.

- Jenkins, S. P., Cappellari, L., Lynn, P., Jäckle, A., & Sala, E. (2006). Patterns of consent: Evidence from a general household survey. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 169(4), 701–722.
- Gummer, T., & Daikeler, J. (2020). A Note on How Prior Survey Experience With Self-Administered Panel Surveys Affects Attrition in Different Modes. *Social Science Computer Review*, 38(4), 490-498.
- Keusch, F., Struminskaya, B., Antoun, C., Couper, M. P., & Kreuter, F. (2019). Willingness to participate in passive mobile data collection. *Public Opinion Quarterly*, 83(S1), 210–235.
- Keusch, F., Struminskaya, B., Kreuter, F., & Weichbold, M. (2020). Combining active and passive mobile data collection: A survey of concerns. *Big Data Meets Survey Science: A Collection of Innovative Methods*, 657–682.
- King, G. (2011). Ensuring the data-rich future of the social sciences. *Science*, 331(6018), 719–721.
- Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, 110(15), 5802–5805.
- Kreuter, F., Haas, G.-C., Keusch, F., Bähr, S., & Trappmann, M. (2020). Collecting survey and smartphone sensor data with an app: Opportunities and challenges around privacy and informed consent. *Social Science Computer Review*, 38(5), 533–549.
- Kuru, O., Bayer, J., Pasek, J., & Campbell, S. W. (2017). Understanding and measuring mobile Facebook use: Who, why, and how? *Mobile Media & Communication*, 5(1), 102-120.
- Ledford, H. (2020). How Facebook, Twitter and other data troves are revolutionizing social science. *Nature*, 582(7812), 328–330.

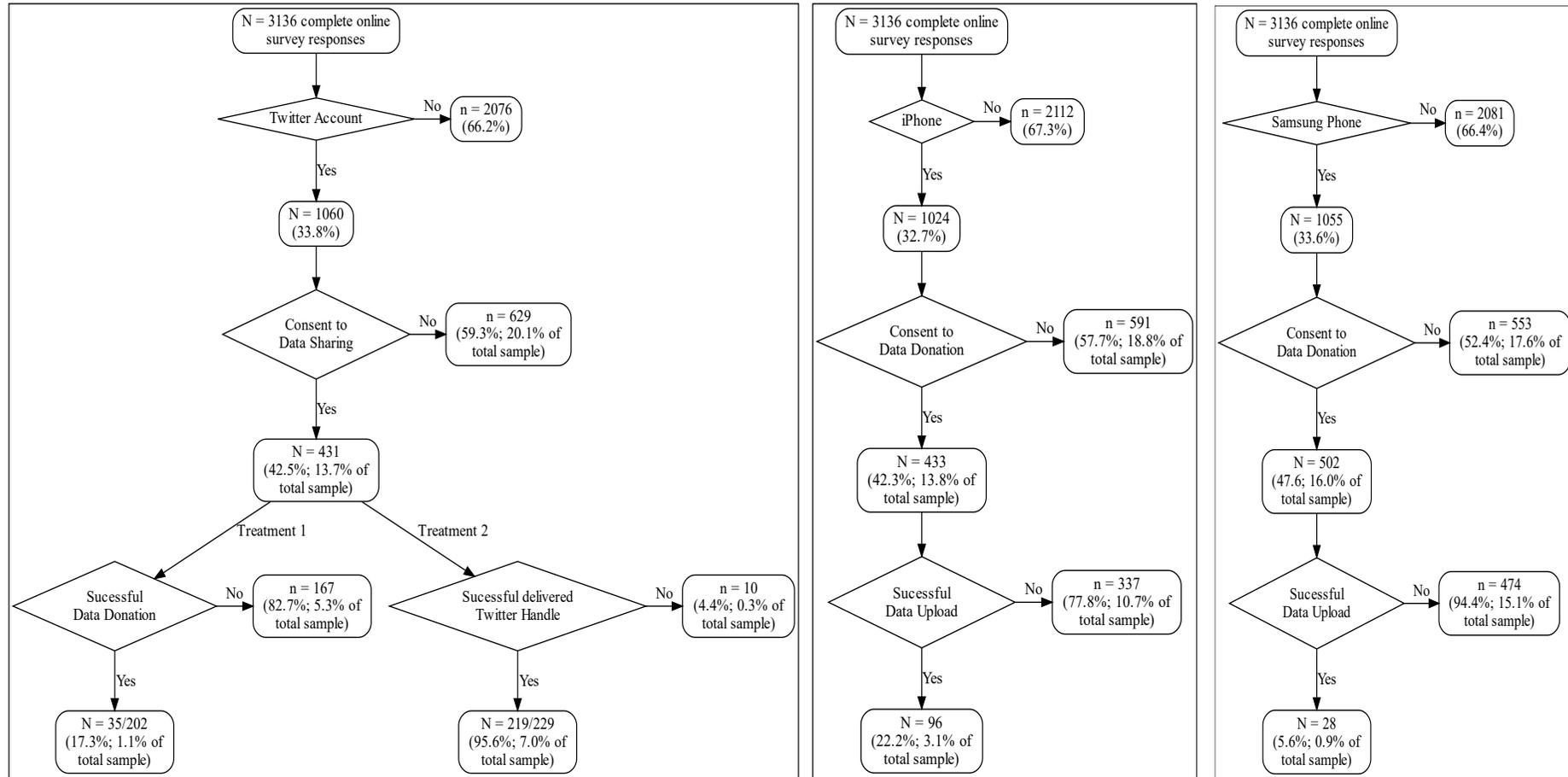
- Leeper, T. J. (2019). Where have the respondents gone? Perhaps we ate them all. *Public Opinion Quarterly*, 83(S1), 280–288.
- Muscanell, N. L., & Guadagno, R. E. (2012). Make new friends or keep the old: Gender and personality differences in social networking use. *Computers in Human Behavior*, 28(1), 107–112.
- Nissenbaum, H. (2009). *Privacy in context: Technology, policy, and the integrity of social life*. Stanford University Press.
- Nissenbaum, H. (2018). Respecting context to protect privacy: Why meaning matters. *Science and Engineering Ethics*, 24(3), 831–852.
- Oberski, D. L., & Kreuter, F. (2020). Differential privacy and social science: An urgent puzzle. *Harvard Data Science Review*, 2(1).
- Smith, A., & Page, D. (2015). US smartphone use in 2015. *Pew Research Center*, available at: <https://www.pewresearch.org/internet/2015/04/01/us-smartphone-use-in-2015>.
- Porter, S. R., & Whitcomb, M. E. (2003). The impact of lottery incentives on student survey response rates. *Research in Higher Education*, 44(4), 389–407.
- R Core Team. (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- Revilla, M., Couper, M. P., & Ochoa, C. (2018). Giving respondents voice? The feasibility of voice input for mobile web surveys. *Survey Practice*, 11(2), 2713.
- Revilla, M., Couper, M. P., & Ochoa, C. (2019). Willingness of online panelists to perform additional tasks. *Methods, Data, Analyses*, 13(2), 29.
- Revilla, M., Toninelli, D., Ochoa, C., & Loewe, G. (2016). Do online access panels need to adapt surveys for mobile devices? *Internet Research*, 26(5), 1209–1227.
- Riker, W. H., & Ordeshook, P. C. (1968). A Theory of the Calculus of Voting. *The American Political Science Review*, 62(1), 25–42.

- Sakshaug, J. W. (2020). Linking Surveys with Big Data. In *Qualität bei zusammengeführten Daten* (pp. 163–173). Springer.
- Settanni, M., Azucar, D., & Marengo, D. (2018). Predicting individual characteristics from digital traces on social media: A meta-analysis. *Cyberpsychology, Behavior, and Social Networking*, *21*(4), 217–228.
- Shearer, E., & Mitchell, A. (2021). “News Use Across Social Media Platforms in 2020”, *Pew Research Center*, available at: <https://www.journalism.org/2021/01/12/news-use-across-social-media-platforms-in-2020/>
- Shlomo, N., & Goldstein, H. (2015). Big data in social research. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 787–790.
- Sloan, L., Jessop, C., Al Baghal, T., & Williams, M. (2020). Linking Survey and Twitter Data: Informed Consent, Disclosure, Security, and Archiving. *Journal of Empirical Research on Human Research Ethics*, *15*(1–2), 63–76.
- Stern, P. C. (1992). Psychological dimensions of global environmental change. *Annual Review of Psychology*, *43*(1), 269–302.
- 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>
- Wenz, A., Jackle, A., & Couper, M. P. (2019). Willingness to use mobile technologies for data collection in a probability household panel. *Survey Research Methods*, *13*(1), 1–22.

**Figure 1.** Data Sharing Flowcharts for Study 1: Facebook (first online survey), Spotify, and Twitter data (second online survey)



**Figure 2.** Data Sharing Flowcharts for Study 2: Twitter and health app data



**Table 1.** Predicting Data Sharing Behavior in Study 1

	<b>Facebook</b>		<b>Spotify</b>		<b>Twitter</b>	
	OR	CI	OR	CI	OR	CI
<b>Sociodemographics</b>						
Age	0.99	0.98, 1.00	0.96**	0.94, 0.99	0.97**	0.95, 0.99
Male	2.05***	1.50, 2.82	2.59***	1.49, 4.60	0.98	0.58, 1.68
Education (ref. high)						
Low	0.43***	0.27, 0.68	0.84	0.34, 2.10	0.71	0.33, 1.53
Medium	0.67*	0.47, 0.95	1.57	0.80, 3.16	0.89	0.50, 1.58
Income (ref. <1500€)						
1500€-3000€	1.03	0.71, 1.51	0.76	0.38, 1.51	1.00	0.54, 1.86
>3000€	0.80	0.53, 1.21	0.52	0.24, 1.11	1.17	0.62, 2.21
<b>Survey evaluation</b>						
Survey evaluation	<b>NA</b>		<b>NA</b>		1.44	0.95, 2.20
<b>Privacy</b>						
Privacy concerns	0.90	0.76, 1.06	0.86	0.62, 1.17	<b>NA</b>	
<b>Network usage</b>						
Network usage	1.09**	1.02, 1.16	1.34***	1.17, 1.55	1.26***	1.13, 1.42
<b>Incentive (ref. postpaid)</b>						
Prepaid	<b>NA</b>	<b>NA</b>	<b>NA</b>	<b>NA</b>	1.29	0.79, 2.12
<b>Device</b>						
Smartphone/tablet	0.43***	0.29, 0.64	0.79	0.43, 1.46	0.89	0.47, 1.64
<b>Nagelkerke R<sup>2</sup></b>		0.12		0.24		0.15
<b>N</b>		772		265		299

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

Note. OR = odds ratios; CI = confidence interval; NA = item not available

**Table 2.** Predicting Data Sharing Behavior in Study 2

	<b>Twitter</b>		<b>Health app</b>	
	OR	CI	OR	CI
<b>Sociodemographics</b>				
Age	1.01	1.00, 1.03	0.96***	0.94, 0.98
Male	1.45	1.00, 2.11	1.54	0.98, 2.41
Education (ref. high)				
Low	0.59	0.27, 1.24	1.11	0.37, 2.91
Medium	0.65*	0.45, 0.93	0.83	0.51, 1.35
Income (ref. <1500€)				
1500€-3000€	0.65	0.46, 0.93	0.88	0.56, 1.39
>3000€	0.72	0.45, 1.13	0.49	0.21, 1.01
<b>Personality</b>				
Openness	0.79*	0.66, 0.95	0.84	0.66, 1.08
Conscientiousness	0.86	0.73, 1.00	1.12	0.91, 1.38
<b>Survey evaluation</b>				
Survey evaluation	1.17*	1.03, 1.34	1.26*	1.02, 1.56
<b>Survey attitudes</b>				
Value	1.09	0.93, 1.27	1.07	0.87, 1.32
Burden	0.99	0.89, 1.10	1.13	0.97, 1.33
Enjoyment	1.08	0.91, 1.28	1.07	0.87, 1.34
<b>Privacy</b>				
Privacy concerns	0.94	0.77, 1.14	0.89	0.67, 1.18
Earlier privacy intrusion	0.95	0.85, 1.05	0.83*	0.71, 0.97
<b>Technology</b>				
Affinity for technology	1.27*	1.04, 1.56	1.26	0.99, 1.60
<b>Network/App usage</b>				
Network/App usage	0.98	0.89, 1.08	0.85**	0.77, 0.95
<b>Incentive (ref 0€)</b>				
2.50€	1.92**	1.26, 2.96	1.72	0.90, 3.38
5€	2.11***	1.38, 3.26	1.82	0.97, 3.55
10€	2.47***	1.58, 3.90	3.54***	1.95, 6.71
<b>Device</b>				
Smartphone	1.06	0.75, 1.49	0.52**	0.32, 0.82
<b>Device Manufacturer</b>				
Samsung	NA		0.30***	0.18, 0.48
<b>Nagelkerke R<sup>2</sup></b>		0.09	0.22	
<b>N</b>		1,041	2,040	

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

Note. OR = odds ratios; CI = confidence interval; NA = item not available

In Model 1 (Twitter data), the data sharing method (API vs. data donation) could not be included because respondents were only randomly allocated to one of the two methods after giving consent to share their data.

**Online Appendix for the Manuscript “Linking surveys and digital trace data:  
Experiences from two studies on factors influencing data sharing behavior”**

**Section 1**

**Table A1.** Predicting Data Sharing Consent in Study 1

	<b>Facebook</b>		<b>Spotify</b>		<b>Twitter</b>	
	OR	CI	OR	CI	OR	CI
<b>Sociodemographics</b>						
Age	0.99	0.98, 1.01	0.94***	0.91, 0.97	0.97**	0.95, 0.99
Male	1.89***	1.39, 2.56	2.50**	1.36, 4.69	1.52	0.89, 2.59
Education (ref. high)						
Low	0.74	0.48, 1.15	1.29	0.49, 3.57	1.06	0.49, 2.32
Medium	0.82	0.58, 1.16	1.15	0.55, 2.47	1.15	0.64, 2.07
Income (ref. <1500€)						
1500€-3000€	1.06	0.74, 1.53	1.19	0.56, 2.55	0.67	0.35, 1.27
>3000€	1.08	0.73, 1.61	1.27	0.56, 2.90	0.62	0.32, 1.19
<b>Survey evaluation</b>						
Survey evaluation	<b>NA</b>		<b>NA</b>		1.45	0.96, 2.21
<b>Privacy</b>						
Privacy concerns	0.80**	0.68, 0.95	0.95	0.67, 1.34	<b>NA</b>	
<b>Network usage</b>						
Network usage	1.08*	1.01, 1.14	1.49***	1.25, 1.82	1.16*	1.03, 1.32
<b>Incentive (ref. postpaid)</b>						
Prepaid	<b>NA</b>		<b>NA</b>		1.01	0.61, 1.67
<b>Device</b>						
Smartphone/tablet	0.87	0.62, 1.23	0.84	0.44, 1.63	0.65	0.35, 1.21
<b>Nagelkerke R<sup>2</sup></b>		0.07		0.29		0.11
<b>n</b>		771		265		299

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

*Note.* OR = odds ratios; CI = confidence interval; NA = item not available

**Table A2.** Predicting Data Sharing Consent in Study 2

	<b>Twitter</b>		<b>Health app</b>	
	OR	CI	OR	CI
<b>Sociodemographics</b>				
Age	1.01	0.99, 1.02	0.99**	0.98, 0.99
Male	1.29	0.94, 1.77	1.26*	1.03, 1.56
Education (ref. high)				
low	1.44	0.75, 2.76	1.85**	1.19, 2.87
medium	0.88	0.64, 1.20	1.30*	1.01, 1.68
Income (ref. <1500€)				
1500€-3000€	0.88	0.65, 1.19	1.02	0.82, 1.27
>3000€	0.74	0.49, 1.12	1.10	0.80, 1.50
<b>Personality</b>				
Openness	0.80*	0.68, 0.95	0.77***	0.69, 0.87
Conscientiousness	0.99	0.86, 1.13	1.11*	1.01, 1.22
<b>Survey evaluation</b>				
Survey evaluation	1.31***	1.17, 1.47	1.18***	1.08, 1.30
<b>Survey attitudes</b>				
Value	1.10	0.96, 1.26	1.01	0.92, 1.10
Burden	0.80***	0.73, 0.88	0.91*	0.85, 0.98
Enjoyment	1.08	0.93, 1.25	1.12*	1.01, 1.24
<b>Privacy</b>				
Privacy concerns	0.75**	0.63, 0.89	0.85*	0.74, 0.97
Earlier privacy intrusion	0.95	0.86, 1.04	0.87***	0.81, 0.94
<b>Technology</b>				
Affinity for technology	1.14	0.95, 1.36	1.12*	1.01, 1.24
<b>Network/App usage</b>				
Network/App usage	0.96	0.88, 1.04	0.87***	0.83, 0.92
<b>Incentive (ref 0€)</b>				
2.50€	1.34	0.94, 1.93	1.43**	1.10, 1.86
5€	1.33	0.93, 1.90	1.54**	1.19, 2.00
10€	1.45	0.98, 2.14	1.71***	1.29, 2.25
<b>Device</b>				
Mobile	0.99	0.73, 1.34	0.99	0.81, 1.22
<b>Device Manufacturer</b>				
Samsung	NA		1.41***	1.15, 1.72
<b>Nagelkerke R<sup>2</sup></b>		0.12	0.11	
<b>n</b>		1,041	2,040	

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

Note. OR = odds ratios; CI = confidence interval; NA = item not available

In Model 1 (Twitter data), the data sharing method (API vs. data donation) could not be included because respondents were only randomly allocated to one of the two methods after giving consent to share their data.

**Table A3.** Description of Measures in Study 1

<b>Item/Indicator</b>	<b>Response options/scale</b>	<b>Construct</b>	<b>Notes</b>
Device used to answer online survey	Smartphone, PC, Tablet	Device	Based on operating system captured by online survey tool
Would you tell us in which year you were born?	Year (4 digits)	Age	Age calculated as year of survey – year of birth
Please indicate your gender.	Male, female (+ do not want to indicate)	Gender	Only male & female included in analyses
What is your monthly net household income? (in Euros)	< 500, 500 – 750, 750 – 1000, 1000 – 1250, 1250 + 1500, 1500 – 2000, 2000 – 2500, 2500 – 3000, 3000 – 4000, 4000 – 5000, 5000 – 7500, 7500 – 10000, > 10000	Income	Collapsed into three categories: < 1500 €, 1500€ - 3000 €, > 3000€
What is your highest educational degree?	13 options reflecting the German educational system	Education	Collapsed into three categories: low, medium, high
How was the questionnaire... interesting?	Not at all, rather not, partly, rather, very	Survey evaluation	Mean score used
How was the questionnaire... diversified?	Not at all, rather not, partly, rather, very	Survey evaluation	Mean score used
How was the questionnaire... important for science?	Not at all, rather not, partly, rather, very	Survey evaluation	Mean score used

How was the questionnaire... long?	Not at all, rather not, partly, rather, very	Survey evaluation	Mean score used (reverse-coded)
How was the questionnaire... difficult?	Not at all, rather not, partly, rather, very	Survey evaluation	Mean score used (reverse-coded)
How was the questionnaire... too personal?	Not at all, rather not, partly, rather, very	Survey evaluation	Mean score used (reverse-coded)
I do not care about my privacy on the Internet.	Fully disagree (1) agree to fully agree (5)	Privacy concerns	Mean score used (reverse-coded)
I am concerned about how much data there is on the Internet about me.	Fully disagree (1) agree to fully agree (5)	Privacy concerns	Mean score used
Incentive condition	Paid regardless of consent (prepaid) or only if consented (postpaid)	Incentive condition	Coded as 0 if no visits/uses were tracked; log-transformed with $\log(x + 1)$
Tracked Twitter use	Number of total tracked website visits and app uses before the date the online survey was answered	Twitter use	Coded as 0 if no visits/uses were tracked; log-transformed with $\log(x + 1)$
Tracked Facebook use	Number of total tracked website visits and app uses before the date the online survey was answered	Facebook use	Coded as 0 if no visits/uses were tracked; log-transformed with $\log(x + 1)$
Tracked Spotify use	Number of total tracked website visits and app uses before the date the online survey was answered	Spotify use	Coded as 0 if no visits/uses were tracked; log-transformed with $\log(x + 1)$

**Table A4.** Description of Measures in Study 2

<b>Item/Indicator</b>	<b>Response options/scale</b>	<b>Construct</b>	<b>Notes</b>
Device used to answer online survey	Smartphone, PC, Tablet	Device	Based on information's from the user agent
In which year you were born?	Year (4 digits)	Age	Age calculated as year of survey – year of birth
Please indicate your gender.	Male, female	Gender	-
What is your average personal monthly net income?	< 1000, 1001 – 1500, 1501, 1500 – 2000, 2001 – 2500, 2501 – 3000, 3001 – 3500, 3501 - 4000, 4001 – 4500, > 4501	Income	Collapsed into three categories: < 1500 €, 1500€ - 3000 €, > 3000€
What is your highest general education degree?	9 categories based on German school system	Education	Collapsed into three categories: low, medium, high
I complete tasks thoroughly.  I am comfortable, inclined to laziness.  I am efficient and work fast.  I am conscientious.	5-point scale from fully agree to fully disagree	Openness	Mean score used
I am interested in many things.  I am profound, like to think about things.  I have an active imagination, I am imaginative.	5-point scale from fully agree to fully disagree	Conscientiousness	Mean score used

I appreciate artistic and aesthetic impressions.  I have little artistic interest.			
How concerned are you about your privacy in general?	Fully disagree (1) agree to fully agree (4)	Privacy concerns	Mean score used
Please indicate if you have ever felt that your privacy has been violated by ...	Doctors, Apps, Browser, Social Networks, Market research, Government (yes/no)	Privacy Intrusion	Mean score used
Surveys are very important.  Opinion polls make society more democratic.	Fully disagree (1) agree to fully agree (7)	Survey Value	Mean score used
Surveys will only tell you about more important things held.  Nowadays, far too many surveys are conducted.	Fully disagree (1) agree to fully agree (7)	Survey Burden	Mean score used
I enjoy participating in surveys.  Surveys are distracting and interesting.	Fully disagree (1) agree to fully agree (7)	Survey Enjoyment	Mean score used
Incentive	0€, 2.5€, 5€, 10€	Incentive	
How often do you open or use Twitter?	Several times a day, Daily, Several times a week, Several times a month,	Twitter Usage	

	Once a month, Less than once a month, Never		
How often do you open or use Samsung Health?  How often do you open or use Apple Health?	Several times a day, Daily, Several times a week, Several times a month, Once a month, Less than once a month, Never	Health App Usage	
Manufacturer of Smartphone owned	Apple, Samsung	Manufacturer	

**Table A5.** Descriptive Statistics of Study 1 (Survey 1)

Characteristic	Twitter, N = 307 <sup>1</sup>
Age	42.0 (13.3)
Gender	
Female	101 (33.1%)
Male	204 (66.9%)
Unknown	2
Education	
Low	48 (15.8%)
Medium	96 (31.7%)
High	159 (52.5%)
Unknown	4
Income	
<1500€	95 (31.2%)
1500€-3000€	108 (35.5%)
>3000€	101 (33.2%)
Unknown	3
Survey evaluation	4.0 (0.6)
Twitter visits	187.1 (907.9)
Incentive condition	
Prepaid	154 (50.2%)
Postpaid	153 (49.8%)
Device	
Desktop	242 (78.8%)
Mobile (tablet/smartphone)	65 (21.2%)

<sup>1</sup>Mean (SD); n (%)

**Table A6.** Descriptive Statistics of Study 1 (Survey 2)

Characteristic	Facebook, N = 785 <sup>1</sup>	Spotify, N = 281 <sup>1</sup>
Age	44.8 (12.7)	36.4 (12.7)
Unknown	0	1
Gender		
Female	396 (50.5%)	134 (47.7%)
Male	388 (49.5%)	147 (52.3%)
Unknown	1	0
Education		
Low	142 (18.4%)	33 (12.4%)
Medium	295 (38.2%)	70 (26.3%)
High	336 (43.5%)	163 (61.3%)
Unknown	12	15
Income		
<1500€	224 (28.6%)	76 (27.1%)
1500€-3000€	321 (40.9%)	117 (41.8%)
>3000€	239 (30.5%)	87 (31.1%)
Unknown	1	1
Privacy concerns	3.7 (0.9)	3.6 (0.9)
Platform visits	3,799.8 (8,267.9)	87.7 (278.6)
Device		
Desktop	586 (74.6%)	203 (72.2%)
Mobile (tablet/smartphone)	199 (25.4%)	78 (27.8%)

<sup>1</sup>Mean (SD); n (%)

**Table A7.** Descriptive Statistics of Study 2

Characteristic	Twitter, N = 1,060 <sup>1</sup>	Health App, N = 2,073 <sup>1</sup>
Age	36.2 (13.1)	39.3 (14.4)
Gender		
female	349 (32.9%)	1,250 (60.3%)
male	711 (67.1%)	823 (39.7%)
Education		
high education	510 (48.1%)	568 (27.4%)
low education	55 (5.2%)	139 (6.7%)
middle education	495 (46.7%)	1,366 (65.9%)
Income		
<1500€	588 (55.8%)	1,219 (59.4%)
1500€-3000€	310 (29.4%)	602 (29.3%)
>3000€	155 (14.7%)	232 (11.3%)
unknown	7	20
Openness	5.4 (1.0)	5.6 (1.0)
unknown	0	1
Conscientiousness	5.3 (1.1)	5.0 (1.1)
Privacy concerns	2.7 (0.8)	2.7 (0.8)
unknown	1	1
Survey evaluation	5.5 (1.4)	5.6 (1.2)
unknown	1	3
Survey Value	5.6 (1.2)	5.3 (1.3)
unknown	0	1
Survey Burden	5.0 (1.6)	5.1 (1.4)
unknown	0	3
Survey Enjoyment	5.8 (1.2)	5.7 (1.2)
unknown	1	2
Earlier privacy intrusion	2.5 (1.5)	2.6 (1.5)
Affinity for Technology	3.9 (0.9)	3.3 (1.0)
Twitter/health app usage	3.1 (1.6)	5.5 (1.9)
unknown	9	5
Incentive		
0€	300 (28.3%)	521 (25.1%)
2.50€	270 (25.5%)	552 (26.6%)
5€	287 (27.1%)	567 (27.4%)
10€	203 (19.2%)	433 (20.9%)
Device		
Desktop/Tablet	751 (71.0%)	1,328 (64.1%)
Smartphone	307 (29.0%)	744 (35.9%)
unknown	2	1

<sup>1</sup>Mean (SD); n (%)

**Online Appendix Section 2 – Informed consent language [Translated from German to English by the authors]**

**Study 1 – Twitter**

Since social media play an increasingly important role in society, we would like to know who uses Twitter and how people use Twitter. We are also interested in combining the answers from people, and also your responses from the survey with publicly available information from your Twitter account.

Would you be willing to provide us with your Twitter username for this research project so that we can link your Twitter data with your responses from this survey for scientific purposes?

Of course, your data will be treated confidentially and not used for commercial purposes. Your Twitter name will not be mentioned in any publication and all Twitter data will be protected by us with the same care as the data from the survey. You can find more information on how we process the data here [link to website with information].

**Study 1 – Facebook**

For many people, Facebook is an important source of information. As you probably know, the display of news items on Facebook is highly personalized. Since Facebook provides virtually no information about this, it is unclear how this selection is made.

As independent scientific researchers, we are interested in how the personalized display of messages on Facebook works. To this end, we cooperate with researchers from the LMU Munich who have developed a browser plugin (for Firefox and Chrome) that collects public posts in the news feed of individual users. We would like to link the data we already have from the survey and web tracking with data on the public posts in your Facebook news feed.

Would you be willing to install this browser plugin?

- Yes, I am willing to install the browser plugin to capture public posts in my Facebook News Feed.
- No, I am not willing to install the browser plugin to capture the public posts in my Facebook News Feed.

The plugin only records posts from your news feed that have actually been publicly shared on Facebook and can, therefore, be seen by any Facebook user. Private posts, such as status updates from friends or private messages, are not recorded. Login codes and passwords are also not recorded. In addition, you can view the data collected from your news feed at any time and delete it if necessary. You can find more detailed information on data protection for the browser plugin here [link to a website with information].

**Study 1 – Spotify**

In this project, we are not only interested in online news, but also in the use of entertainment services on the Internet. To gain an impression of how people consume music on the Internet, we are cooperating with researchers at the University of Leuven in Belgium, who have developed an online app that collects data on the use of Spotify. We would like to combine the data we already have from this survey as well as from the web tracking with data on your Spotify usage.

Would you be willing to use this online app once to share some of your Spotify usage data with us?

The app collects the following data:

- The last 50 songs you've played
- The contents of your public playlists
- Spotify's assessment of your short-, medium- and long-term preferences (artists & genres)

We will not have access to your login information (username or password) at any time and no settings in your account will be changed. The data will be stored and used for scientific purposes only. More detailed information about data privacy for the Spotify app can be found here [link to a website with information].

- Yes, I am willing to use the online app once to share some of my Spotify usage data.
- No, I am not willing to use the online app once to share some of my Spotify usage data.

**Study 2 – Twitter**

Social media play an increasingly important role in our society. We are therefore interested in how people communicate publicly via Twitter. To this end, we want to collect and link data from your public Twitter profile for scientific purposes.

In order to combine this data with your survey data, we would like to ask for your consent. If you agree, you will receive detailed instructions on the necessary steps on the next screen.

When analyzing the data, we absolutely ensure that all data protection regulations are complied with and that no data is passed on to third parties.

Your consent is of course voluntary. You can revoke it at any time by sending an email to [email address].

**Study 2 – Health App**

Health plays a major role in our society. We are therefore interested in how people move in their everyday lives and what sport activities they engage in. For this purpose, we want to collect data from your Health App (Apple / Samsung Health).

In order to combine this data with your survey data, we would like to ask for your consent. If you agree, you will receive detailed instructions on the necessary steps on the next screen.

When analyzing the data, we absolutely ensure that all data protection regulations are complied with and that no data is passed on to third parties.

Your consent is of course voluntary. You can revoke it at any time by sending an email to [email address].