

# Longitudinal Data Donation Behavior and Data Omission across Four Social Media Platforms

Lion Wedel

*Weizenbaum Institute, Berlin*

Jakob Ohme

*Weizenbaum Institute, Berlin*

## Abstract

This research article presents insights from a two-wave, longitudinal data donation study across four major social media platforms: TikTok, YouTube, Facebook, and Instagram. We investigate a critical yet underexplored aspect of data donation: allowing participants to delete specific data traces before submission. Our analysis quantifies the impact of this selective omission on data completeness and, consequently, the analytical power of the resulting datasets. Furthermore, leveraging a longitudinal design, we examine the stability of donation and deletion behaviors over time in a panel setting. Findings reveal an overall increase in the platform donor rate in the second wave. However, we also observe substantial donor attrition. Notably, the omission of data traces is predominantly observed among first-time donors. Our results suggest the feasibility of longitudinal data donation research designs. For allowing participants selective data omission, a careful weighing of the trade-offs is necessary, as this practice—when utilized—significantly compromises data completeness.

**Keywords:** data donation, panel study, social media, data omission, data deletion, coverage error

## Introduction

The collection of digital trace data remains a challenge for social science researchers studying digital platforms such as TikTok, Instagram, YouTube, or Facebook (Valkenburg et al., 2025). Various methods have been established to collect digital trace data for these purposes: in addition to API access and the ability to track users' devices, data donation has received increasing attention from the research community (e.g., Hase et al., 2024). Here, researchers ask users to use their GDPR-guaranteed access to their data held by platforms to download these data (often in the form of data download packages) and to donate them to research. In addition to being user-centric

and allowing the analysis of individual trajectories across devices, a key advantage of data donation is that it provides retrospective data (Ohme et al., 2024). This means that users only donate data that has already been collected about them, relative to the method of tracking and log-data collection, where data is collected prospectively. However, data donation packages often do not provide the full user history, making repeated-donation research designs still necessary.

Ethical considerations, particularly regarding informed consent and data minimization, require that participants have the option to control the data they share (Boeschoten, Mendrik, et al., 2022). In standard data download package collection pipelines, participants can therefore delete any data point they are uncomfortable sharing (Carrière et al., 2025). Nevertheless, given the vast amounts of highly personal and sensitive digital trace data, there is an ongoing discussion about power imbalances between researchers and study participants (Gomez Ortega et al., 2023; Groot Kormelink et al., 2025). The pure volume of data makes a true informed consent hardly achievable, even if a data review step is part of the data donation flow. Yet, the retrospective collection of digital trace data allows participants to retain agency, as they donate only data that has already been collected and that they can review before donating. Therefore, data donation frameworks such as *PORT* or the *Data Donation Module* (e.g., Boeschoten et al., 2023; Pffner et al., 2024) have been developed with participants' agency in mind, enabling them to review their data points before donating data to researchers.

This study addresses the crucial yet under-researched aspect of how this data-deletion feature is used in data donation tasks, which data traces are most often omitted, and the consent errors (Boeschoten, Ausloos, et al., 2022) that result from the omission of data points. We present findings from a two-wave longitudinal study ( $n=2,318$ ) involving four major social media platforms: TikTok, YouTube, Facebook, and Instagram. The study reports on data omission behavior not only cross-sectionally but also on changes in donation success and omission patterns over time.

## Research Questions

Digital trace data can be an important part of a new social media effects paradigm (e.g., Valkenburg et al., 2024). However, for the establishment of the effects of such data on specific outcome variables, for example, attitudes or behaviors, the establishment of causality is a challenge. As with other data collection methods (such as survey research), multi-wave data donation collection can help to make stronger causal arguments, especially if

connected to self-reported outcome variables. Despite the retrospective and longitudinal nature of data download packages, collecting data across multiple waves remains necessary. First, historical coverage of watch history varies across platforms: at the time of writing, TikTok typically provides three months, Instagram only seven days, and YouTube allows user customization. This makes cross-platform comparisons inconsistent, and genuine cross-platform studies are restricted to the smallest historical coverage of the platforms of interest. Second, the length of this coverage is under constant threat of unilateral change by the platforms themselves, introducing a significant risk of data discontinuity that could compromise the integrity of any long-term or retrospective study. Third, as data download package data points commonly need to be augmented with meta- or content data (Wedel et al., 2024), highly retrospective data bear the risk that such data are not retrievable from platforms anymore, for example, due to the deletion of posts or videos (Entrena-Serrano et al., 2025). Hence, repeated data donation designs may be necessary to ensure timely augmentation, even if, within the time horizon of the data donations themselves, a single donation at the end of a multi-wave research design would suffice.

However, so far, it is an open question how multi-wave data donation collection affects the donor rates. Previous research has primarily relied on single-wave data donation (e.g., Bechmann et al., 2025; Breuer et al., 2022; Zannettou et al., 2024). Beyond data donations, research found that when requesting digital trace data for one wave of a running panel, the retention rate for the subsequent wave (without digital trace data collection) is reduced (Trappmann et al., 2023). In contrast, research on the hypothetical sharing of sensor-based data found no change in the willingness to share across multiple waves (Struminskaya et al., 2020). Although not examining the exact same subject, those results indicate that the question of how successful data donations can be in a panel-study setting is not clear-cut and remains unanswered.

Furthermore, within the media effects paradigm, multi-platform research is often encouraged, acknowledging that media effects on digital well-being or political attitudes are not contingent on a single social media platform but can accumulate through cross-platform use (Valkenburg et al., 2024). However, multi-platform data donation efforts remain rare, with few exceptions (Hase & Haim, 2024; Wedel et al., 2025). Consequently, another vital dimension when looking at donor rates is not only whether participants donate data multiple times, but also whether data donations are consistent across platforms and over time.

Hence, we first report on our multi-wave and multi-platform data donations collection in general, answering the following question by exploring a set of donation behavior rates that help to understand the feasibility of a multi-wave and multi-platform data donation collection:

**RQ1:** How do donor, attrition, recurring donor, new donor, switching donor, and added platform donor rates vary across platforms in a longitudinal data donation study?

Moreover, the paper looks at data omission in data donation behavior: Data donation frameworks include the possibility for users to review data points in data download packages before transferring them to the researcher, responding to calls for compliance with informed consent and data minimization efforts (Boeschoten et al., 2023; Carrière et al., 2025; Pfiffner et al., 2024). Recent qualitative work has shown that participants appreciate the option, even when they do not know what to search for or what to delete, or overlook the function altogether (Groot Kormelink et al., 2025). The same study identifies the platform selection stage in multi-platform data donation as relevant to privacy beliefs, consistent with findings from quantitative work (Kmetty & Stefkovics, 2025). Suggesting that if participants are uncomfortable with sharing their digital trace data from a particular platform, they seem to decide not to donate data at all, rather than selectively removing data points from that data download package. However, quantitative analysis of omission behavior is still rare, hence we test whether this option is used and how it changes the donated data structures. It is important to understand the trade-off between the increased user agency and data minimization afforded by data omission features, and the reduced data completeness that may weaken the data's explanatory power.

Importantly, the data omission feature is available to users for every wave in multi-wave studies. Theoretically, users who delete data in wave 1 may also use this feature in wave 2, which would suggest stable traits such as perceived privacy risks or mobile privacy literacy (Ohme et al., 2021). However, it is also possible that users who deleted traces in the first wave, without experiencing any negative outcomes from this behavior, waive deletion in the second wave because they may have built a higher level of trust during the field period. With a systematic analysis having been missing so far, we ask:

**RQ2a:** To what extent is the data trace omission feature used in a cross-platform, longitudinal data donation study?

**RQ2b:** Which data traces are most likely to be omitted?

Finally, research on data donation behavior has shown that significant non-response errors are introduced during data donation collection (e.g. Hase & Haim, 2024; Wedel et al., 2025). Specifically, in this work, we focus on the consent error and resulting biases: “If the respondent decides to not or only partly share this file, this results in consent error.” (Boeschoten, Ausloos, et al., 2022, p. 406) The decision to selectively omit data points especially aligns with the notion of partial donation. This specific type of consent error introduced during participants’ review and selective deletion of data from their data download package (Boeschoten, Ausloos, et al., 2022) has not yet been studied. We investigate how sociodemographic characteristics predict omission behavior and, consequently, lead to consent errors.

**RQ3:** How do sociodemographic factors (age, gender, education, political orientation) relate to data omission behavior?

## Methods

This study was part of a greater data-collection effort that employed a two-wave panel survey design with optional data donations. In this work, we report on the longitudinal patterns of user donation and deletion behavior in the context of social media data donation. The first wave of data collection was conducted from August 2024 to early September 2024. The second wave was conducted in February 2025 to early March 2025. The underlying data is available for transparency and secondary use (Wedel et al., 2026).

## Participant Recruitment

Participants were recruited in Germany through Bilendi, a non-probability online panel provider. We aimed to draw a nationally representative sample ( $N = 2000$ ) of German Internet users via a quota sampling design, with quotas for age, gender, and education in wave one. Age and gender quotas were provided by the panel provider (Bilendi), and the education quota by the German Federal Statistical Office (Statistisches Bundesamt, 2022). To account for the younger age structure of digital platforms such as TikTok or Instagram, a second sample of 18-27-year-old users ( $N=500$ ) was targeted. This resulted in an intentional oversampling of young adults by 16.67% to ensure sufficient representation of this demographic. In wave 2, all participants of wave 1 were invited to participate again.

For this study, we only consider participants who a) indicated during the survey that they use at least one of the four platforms for which we asked for data donations and b) completed the survey, thus reaching the donation

stage at the end of it. This results in 2,318 *eligible* participants in wave 1 being the basis for this study.

In general, the deviation in sample characteristics across waves was minor, with the exception of age for wave 2 (see Appendix A). Wave 2 includes an older sample: the share of 18–27-year-olds decreased by 9.87% compared to Wave 1, while the three older age groups increased by 3.11% (38–47) to 4.62% (58 and older).

## Informed Consent

Before starting data collection, ethical approval was obtained from the Research Ethics Committee of the Weizenbaum Institute. All participants provided their informed consent prior to participating in the study. The informed consent process clearly outlined the purpose of the study, the nature of the digital trace data collected, the option to selectively delete data points, data storage protocols, measures for anonymity or pseudo-anonymity, respectively, and the right to withdraw from the study at any time.

## Data and measures

The data resulting from the data collection are twofold:

First, we retrieved survey data on participants, including sociodemographic variables shown to be relevant predictors of data donation behavior in previous studies (e.g., Keusch et al., 2024; Pfiffner & Friemel, 2023; Strycharz et al., 2024). The distribution values are reported for the eligible users of wave 1 ( $n = 2,318$ ):

- **Age** was measured on a numerical scale ( $M = 39.69$ ;  $SD = 14.86$ ). Only participants aged 18 or older were eligible to the study.
- For **Gender** 1,166 (50.30%) indicated male, 1,152 (49.70%) indicated female.
- **Education** was operationalized according to the International Standard Classification of Education (ISCED) for the German education system (Statistisches Bundesamt, 2022), at three levels: low (primary education;  $n=331$  [14.28%]), medium (secondary education;  $n = 1,283$  [55.35%]) and high (tertiary education;  $n=698$  [29.72%]).
- **Political leaning** was measured on a 10-point scale (zero to nine) from left to right ( $M = 6.80$ ;  $SD = 2.65$ ).

In the survey, participants reported on which social media platforms they have used in the past month. If they indicated usage of any of the four platforms of interest (Facebook, Instagram, YouTube, TikTok) they were asked which of the platforms they would be willing to donate at the end of the survey. For the first donation, we offered an incentive of 5 Euro; for any subsequent donation, 3 Euro. The platforms were listed in random order as response options. This filter question enabled us to direct participants to a specified instance of the PORT data donation platform, which requested only the user-specific set of platforms for which they indicated willingness to donate. Each PORT instance displays the platforms in a static order; we did not find this to affect data donation behavior (Wedel et al., 2025). On PORT, participants retrieved extensive instructions on requesting, downloading, uploading, and the option to choose and delete data points upon upload (see Appendix C). After uploading their data download packages, the data were processed by a custom Python script, and the subset of traces intended for collection was extracted and reviewed by the user (see Appendix B). At this point, users could search through the data and delete single data points (or select multiple and delete them in one operation). This process was repeated in wave 2. Participants were again asked about the platforms they used and their willingness to donate data from these platforms. Hence, participants could donate data for different platforms in waves 1 and 2.

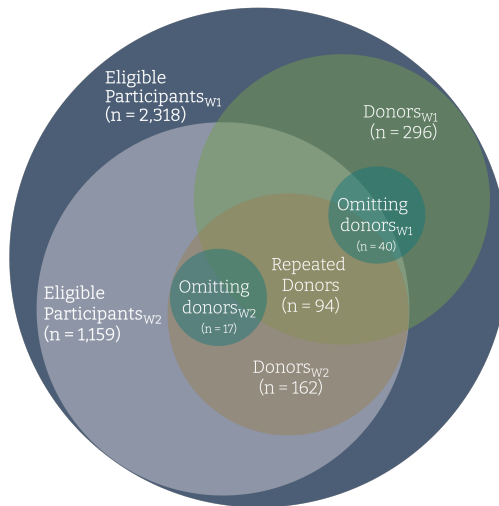


Figure 1: Visualizations of user sets emerging in a two-wave data donation study.

Second, based on the data donations, we derive several measures related

to user-level data donation behavior and data trace omission behavior. The action of omitting specific data points is closely related to what participants see when reviewing the data they are about to donate. Only if participants have digital traces recorded, they can omit them. To enforce a minimal baseline of activity in the DDPs, only data donations that contain platform-specific watch history data were considered successful data donations, since no matter how passive one uses a platform, if one uses it at all, one must be exposed to content. This decision ensures a focus on a consistent, foundational aspect of passive platform use, which is highly relevant to empirical data donation research. We employ the following terminology and metrics<sup>1</sup> to evaluate the behavior of cross-wave donation based on the sets of participants that occur between waves (see Figure 1):

**Donor Rate:** This metric quantifies the fraction of donors for the respective wave. It represents the proportion of participants who donated in a wave ( $\text{Donors}_{W_x}$ ), relative to all participants in that wave ( $\text{Participants}_{W_x}$ ). A higher donation rate indicates a more successful compliance of participants. The donor rate can also be calculated at the platform level, as the number of donors for a specific platform divided by the total number of participants using that platform.

$$\text{Donor Rate} = \frac{|\text{Donors}_{W_x}|}{|\text{Participants}_{W_x}|}$$

**Donor Attrition Rate:** This metric quantifies the decline in continued participation of existing donors. It refers to the proportion of participants who donated data for a specific platform in wave 1 ( $\text{Donors}_{W_1}$ ) but subsequently chose not to donate data for that same platform in wave 2 ( $\text{Non-donors}_{W_2}$ ) or did not participate in wave 2 altogether ( $\text{Non-participants}_{W_2}$ ). A high attrition rate indicates a decline in sustained engagement for a particular platform.

$$\text{Donor Attrition Rate} = \frac{|\text{Donors}_{W_1} \cap (\text{Non-donors}_{W_2} \cup \text{Non-participants}_{W_2})|}{|\text{Donors}_{W_1}|}$$

**Recurring Donor Rate:** This metric measures the proportion of wave 2 donors who previously donated in wave 1 for the same platform. It represents the proportion of participants who donated to the same platform in both waves relative to all donors of those platforms in wave one. A high recurring donor rate suggests strong representation from existing donors in wave 2.

---

<sup>1</sup>The notation  $|X|$  is used throughout this section to denote the total number of participants in group  $X$ .

$$\text{Recurring Donor Rate} = \frac{|\text{Donors}_{W1} \cap \text{Donors}_{W2}|}{|\text{Donors}_{W2}|}$$

**New Donor Rate:** This metric quantifies the influx of fresh donors. It represents the proportion of participants who did not donate in wave 1 (Non-donors<sub>W1</sub>) but subsequently chose to donate in wave 2 (Donors<sub>W2</sub>), relative to all donors in wave 2 (Donors<sub>W2</sub>). A higher new donor rate indicates successful recruitment of previously unengaged users.

$$\text{New Donor Rate} = \frac{|\text{Non-donors}_{W1} \cap \text{Donors}_{W2}|}{|\text{Donors}_{W2}|}$$

**Switching Donor Rate:** This metric quantifies the fraction of donors donating for different platforms across the two waves on the platform level. It represents for each platform the proportion of participants who switched to a platform by donating in wave 2 for it after donating in wave 1 for a different platform. A higher switching donor rate indicates that a platform was a prioritized target to switch towards from other platforms.

$$\text{Switching Donor Rate} = \frac{|\text{Donors}_{W1, \text{Non-A}} \cap \text{Donors}_{W2, A}|}{|\text{Donors}_{W2, A}|}$$

**Added Platform Donor Rate (To A):** This metric quantifies the fraction of donors who were active in wave 1, but did not donate to platform A, and then added Platform A to their donation portfolio in wave 2, expressed as a proportion of all donors to Platform A in wave 2. A higher rate indicates that a platform was a prioritized target to be added.

$$\text{Added Platform Donor Rate (To A)} = \frac{|\text{Donors}_{W1} \setminus \text{Donors}_{W1, A} \cap \text{Donors}_{W2, A}|}{|\text{Donors}_{W2, A}|}$$

**Data trace and point:** Data trace refers to a subset of a data donation package. For example, the watch history or the likes. A data point is then a single entry in such a data trace (e.g., one watch or a like instance).

**Omitting Donor Rate:** This metric quantifies the extent of data omission among active donors for a specific platform. It represents the proportion of participants who omitted at least one data point for a given platform in wave x (Omitting donors<sub>Wx</sub>), relative to all participants who donated to that respective platform in the same wave (Donors<sub>Wx</sub>). A higher omitting donor rate suggests a greater tendency for selective data sharing among active donors for the respective wave and platform.

$$\text{Omitting Donor Rate}_{W_x} = \frac{|\text{Donors}_{W_x} \cap \text{Omitters}_{W_x}|}{|\text{Donors}_{W_x}|}$$

**Average omitted points (per trace):** This refers to the mean of the fraction of data points omitted by users for a specific data trace or aggregated across data traces.

## Analysis

We will first examine overall donation behavior across waves, assessing changes in the various donation behavior rates listed above by platform. Next, our analysis will extend to the omission of data trace levels, scrutinizing specific data deletions within packages to identify frequently omitted data traces and to examine how omission behavior changes over time by reporting the omitting donor rates and the average number of omitted points. Finally, we examine whether data omission is biased toward certain user groups with respect to age, gender, education, and political leanings, using t-tests and Chi-Square Tests.

## Results

### Cross-wave donation behavior (RQ1)

Except for Facebook, all platforms showed an increase in their donor rate from wave 1 to wave 2. The donor rate (i.e., the fraction of platform users donating) increased for Instagram from 8.2% to 9.9%, for TikTok from 5.8% to 7.2%, and, notably, for YouTube from 7.8% to 10.6% (see Table 1). Facebook emerged as an anomaly due to a technical error during data collection: The company adjusted the structure of the data download package during field time, which prevented our processing script from retrieving meaningful wave 2 donation data for an unknown number of donations. Although this invalidates the comparison of cross-wave rates for Facebook (see Tables 1 and 2), we still report the platform's meaningful wave 1 results for the remainder of the article.

The high donor attrition rate for each platform (see Table 1) shows that a vast majority of wave 1 donors either did not participate in wave 2 or did not donate in wave 2. While the donor attrition rates for Instagram (73.85%) and YouTube (71.22%) are close to 70%, the rate for TikTok (83.67%) is nearly as high as the - due to a technical error artificially inflated - rate for Facebook (85.58%). This indicates that TikTok donors from wave 1 were

Table 1: Absolute number of donors and donor rates in wave 1 and 2 by platform. Fractions are reported as a proportion of the total number of eligible participants for each platform.

<b>Platform</b> <b>(Participants<sub>W1</sub>, Participants<sub>W2</sub>)</b>	<b>Donor Rate</b> <b>Wave 1 (n)</b>	<b>Donor Rate</b> <b>Wave 2 (n)</b>	<b>Donor</b> <b>Attrition</b> <b>Rate (n)</b>
Facebook (1520, 775)	6.8% (104)	4.8%* (37)	85.58% (89)
Instagram (1585, 708)	8.2% (130)	9.9% (70)	73.85% (96)
TikTok (845, 346)	5.8% (49)	7.2% (25)	83.67% (41)
YouTube (1815, 900)	7.7% (139)	10.6% (95)	71.22% (99)
Any (2318, 1159)	12.81% (297)	13.98% (162)	68.01% (202)

\* The decline for Facebook is attributed to a technical issue that invalidated an unknown portion of Facebook donations during data collection for wave 2.

more likely to switch platforms in wave 2 or to abandon donating altogether than Instagram and YouTube donors. This matches our observation that the second wave was prone to dropout among younger users (minus 9.87% of 18 to 27 participants and an increased fraction of 38+ year old participants). Given TikTok’s younger age structure, a higher attrition rate is expected. Overall, the majority of wave 1 donors does not participate or donate again; however, wave 2 participants are generally more likely to donate.

Next, we examine the composition of the wave 2 donor sample across different social media platforms, focusing on the recurring donor, new donor, switching donor, and expanding donor rates (see Table 2).

Excluding the technically-affected Facebook data, which is reported in Table 2 for transparency, we focus on the remaining platforms. YouTube and Instagram earlier showed the lowest attrition rates. Consequently, both platforms exhibited the highest recurring donor rates of 48.57% (Instagram) and 42.1% (YouTube). Instagram’s notably higher recurring donor rate stems from a combination of a slightly lower new donor rate (34.29% vs. 37.89%) and a higher added platform donor rate (15.71% vs. 10.53%). YouTube, conversely, has a clear advantage in the switching donor rate (9.47% vs. 1.4%). This indicates that, although both platforms are similar in attracting entirely new donors in wave 2, donors who previously donated to a different plat-

Table 2: Platform comparison of the different rates that explain the composition of the second-wave donor sample. For the *Any* condition, the switching donor and added platform donor rates are not computed since they depend on platform distinction.

<b>Platform</b> <b>(Donors<sub>W1</sub>, Donors<sub>W2</sub>)</b>	<b>Recurring Donor</b> <b>Rate (n)</b>	<b>New Donor</b> <b>Rate (n)</b>	<b>Switching Donor</b> <b>Rate (n)</b>	<b>Added Platform Donor</b> <b>Rate (n)</b>
Facebook (104, 37)	40.54% (15)	45.95% (17)	5.41% (2)	8.11% (3)
Instagram (130, 70)	48.57% (34)	34.29% (24)	1.4% (1)	15.71% (11)
TikTok (49, 25)	32% (8)	36% (9)	4% (1)	28% (7)
YouTube (139, 95)	42.1% (40)	37.89% (36)	9.47% (9)	10.53% (10)
Any (297, 162)	58.02% (94)	41.98% (68)	-	-

form are more likely to switch entirely to YouTube than to Instagram, where they are more likely to add the platform to their donation portfolio. For YouTube, the composition of wave 2 is more sustainable, attracting new, old, same platform, and old different platform donors alike - resulting ultimately in the highest wave 2 donor rate of all platforms. TikTok exhibits a pattern similar to Instagram, but with an even more extreme emphasis on expansion over switching. The TikTok wave 2 donor sample consists largely of new (36%) and recurring donors (32%). The remaining 32% of donors come mainly from expanding donors (28%) rather than from switching donors (4%). If participants migrated to TikTok, they did so to expand their donation portfolio rather than switch away from another platform.

This closer look at the wave 2 donor samples compositions shows that while the wave 2 donor samples tend to be majorly composed from wave 1 donors (recurring and added platform donor rate), for all platforms, a substantial number (over a third) of donors stemmed from entirely new donors.

### **Data trace and point omission behavior (RQ2)**

Next, we examined the data trace level and the corresponding data-point omission behavior (RQ2a and RQ2b). In general, only a moderate number of donors omitted parts of any trace in the first place: 40 participants in wave

1 (13,47%) and 17 in wave 2 (10,49%). If participants omitted data points, their omission behavior was spread across the various traces collected per platform (see Figure 2). Some traces were unaffected; depending on the research objective, data omission may not be problematic.

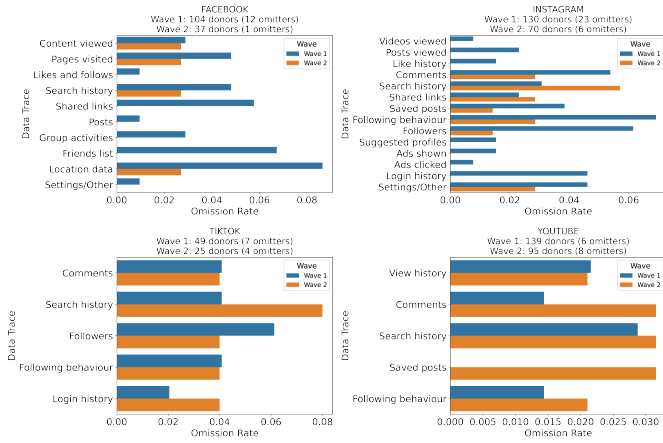


Figure 2: Fraction of donors omitting per trace. Each graph shows the omission rates for the corresponding digital traces on a single platform, as indicated by the subfigure title.

In the case of TikTok, the highly interesting watch history is never omitted, nor are engagement activities such as liking and sharing. This is likely due to those data points containing only a timestamp and a link, and therefore, at first glance, not disclosing any information about the participants. What is omitted, on the other hand, are data points that convey meaning without an additional click on a link: followers and following (usernames), comments, and searches (both include text written by the donor).

A similar pattern is observed across the other platforms, indicating that data traces on followers/followings, comments, and searches, as well as login/location histories, are likely perceived as especially sensitive. Donors focus on data traces that carry interpretable metadata, not on those hidden behind a link and post hoc augmentation.

Although the distribution of omission behavior across traces is heterogeneous, once users decide to omit a trace, they omit a large fraction of that trace (see Figure 3). Participants who donated in both waves for the same platform and omitted data (orange and green bars) are extremely rare (1 per platform), but they deleted between 50 and 100% of the data points in the traces they omitted. For donors who donated only for one wave of the two and also omitted (blue and red bars), the average fraction of omitted

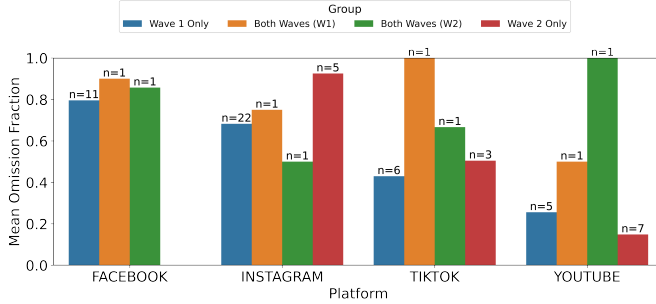


Figure 3: Average fraction of omitted data points per user across waves. The figure illustrates the severity of omission behavior by displaying the mean proportion of a data trace omitted by users, categorized by platform and cross-wave omission patterns. Bar heights represent the mean omission fraction, while the sample size ( $n$ ) labeled on each bar indicates the number of cases meeting those specific criteria (e.g., Facebook, Wave 1 only).

points is very platform dependent, but always substantial: Participants who omit data trace of their Facebook donation omitted 79.60% of those traces (e.g., friends), for Instagram donations it is 68.24% to 92.55%, wave-to-wave. TikTok shows a lower average of omitted data points, at 42.96% (wave 1) and 50.47% (wave 2). If donors omit YouTube traces, they are the least invasive compared to the other platforms (wave 1: 25.60%, wave 2: 14.84%).

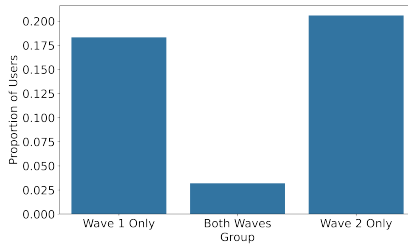


Figure 4: Distribution of omitting donors across waves relative to the number of donors donating only in wave 1, 2 or in both.

Looking at cross-wave omission behavior, we make a relevant observation: only three users (3.16%) who donated in both waves omitted data in both (see Figure 4). Therefore, the omission of specific traces seems to be a characteristic of first-time donors. 18.32% of wave 1-only donors omitted data points ( $n=37$  of 202). Among participants who donated in wave 2 for the first time, 20.60% ( $n=14$  out of 65) omitted at least one data point. This finding indicates either an increase in trust or a systematic dropout of

less-trustful users between waves 1 and 2.

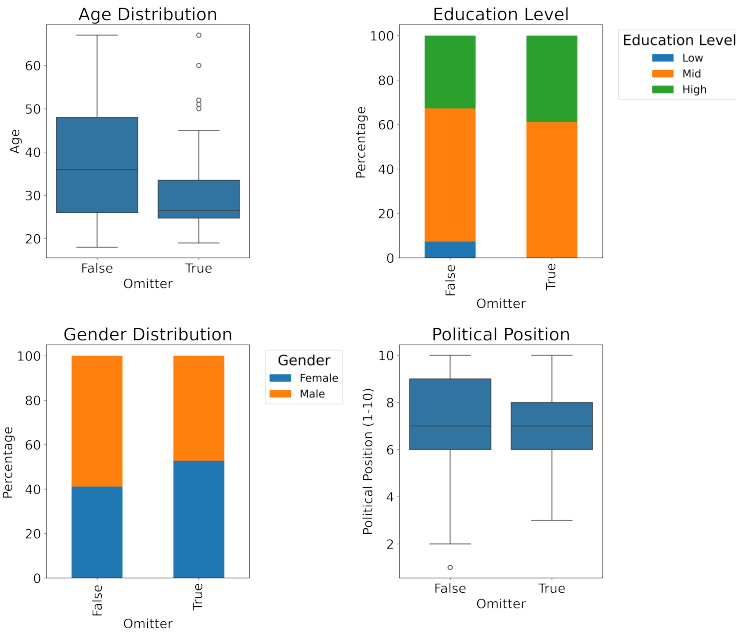


Figure 5: Comparison of sociodemographic characteristics between participants who omit and those who do not when donating.

### Consent error due to omission behavior (RQ3)

Finally, we examine differences in four sociodemographic variables between donors who omitted and those who did not omit ( $n = 364$ ; see Figure 5).

A independent t-test shows that omitting donors ( $M = 31.36$ ,  $SD = 12.19$ ) compared to non-omitting donors ( $M = 37.80$ ,  $SD = 13.33$ ) are significantly younger,  $t(362) = -2.76$ ,  $p = 0.006$ . In contrast, the level of education ( $\chi^2(2) = 3.02$ ,  $p = 0.221$ ) and gender ( $\chi^2(1) = 1.35$ ,  $p = 0.245$ ) do not differ significantly between omitters and non-omitters at the 0.05 significance level. An independent t-test comparing donors who omitted data ( $M = 6.94$ ,  $SD = 1.96$ ) with non-omitting donors ( $M = 6.90$ ,  $SD = 2.49$ ) did not find a significant difference in political position,  $t(2) = 0.10$ ,  $p = 0.917$ .

Those results point towards omission behavior appearing less dependent on sociodemographic characteristics (with the exception of age) and political leaning, but rather on a habituation effect, since the descriptive analysis shows that omission is primarily a first-time donor behavior.

## Discussion

This study provides evidence on multi-wave and cross-platform data donation behavior with a focus on data point omission.

Our findings show that donor rates increase from wave 1 to wave 2 across platforms, despite the fact that the participants to whom data donation samples are typically skewed [e.g., younger participants, (e.g., Wedel et al., 2025; Welbers et al., 2024)] are underrepresented in wave 2. These results suggest that repeated participation in a data donation study increases participants' likelihood of donating, potentially through greater trust or a better understanding of the process, thereby mitigating other typical biases (e.g., age). However, our observations are driven not only by the same donors donating to the same platforms in both waves, but also, to a large extent, by participants donating for the first time in wave 2 or switching platforms. In total, a slight majority of the donors in wave 2 have previously donated for a platform in wave 1 (58.02%). The switching of platforms is a novel finding in data donation research. While it is beyond the scope of this study to fully explore the reasons, future research should be aware that free platform choice across consecutive waves can lead to cross-platform inconsistencies in retrieved trace data. High attrition rates exceeding 70% across platforms are concerning for multi-wave data donation collection plans.

One cause of the above is likely our research design, in which the donation was voluntary (with an added incentive) and participants could independently choose the platforms for which they donated each wave. Our incentive of 5 to 3 euros per data donation is in a similar range to the estimated incentive participants received simply for completing the survey, and thus may have provided little additional value. Our results, therefore, show how participants behave when they have a high degree of choice and a comparatively low incentive. The results may improve if donations are mandatory to receive an incentive and platforms are fixed to first-wave donations in subsequent waves. Our results could therefore be considered a conservative estimate of the success of a multi-wave cross-platform data donation collection.

Furthermore, we find clear indicators that donors tend to omit those data points where content is easily identified, that is, they contain written text by the participant (comments, searches) or user names of the following and whom they follow - providing empirical evidence for previous findings of vignette studies (Pffnner & Friemel, 2023). Mere exposure data that consist just of URLs is less often or not at all (TikTok) omitted. If donors omit data, participants often omit large parts of the selected data traces - specifically for

Facebook and Instagram - this might be connected to those platforms being traditional social media platforms, where traces are likely more personal altogether, compared to more media consumption-oriented platforms like YouTube and TikTok. Hence, data trace omission can result in severely incomplete data donations.

Omitting data is a first-donor characteristic in both waves, suggesting, as in our findings on cross-wave donation behavior, either an increase in trust and understanding of the process or self-selection across waves. Additionally, donors who omit data points are significantly younger. Other factors that commonly impact data donation behavior (gender, education, political leaning Wedel et al., 2025) do not translate similarly to omission behavior. The role of age might point towards omitting donors having higher digital literacy. Data point omission features, therefore, exacerbate existing biases in data donation data collection, with donor samples already biased toward younger participants (e.g., Strycharz et al., 2024; Welbers et al., 2024).

The combination of low other biases and the finding that omission is a first donor characteristic leads us to a similar conclusion as for the general donation behavior: repeated participation in data donation collection seems to mitigate typical biases observed in data donation collections. This “getting used to it” effect, for donation and omission behavior alike, is a promising indicator for future multi-wave data donation research, despite its challenges.

## **Limitations**

The study has several limitations beyond the freedom of choice in the platform(s) for which the participants donate, as previously noted. First, it is conducted in a single country, and we cannot conclude that findings are valid in other country contexts. Second, the change in the structure of the Facebook data download package has introduced an anomaly in the data, which renders our estimates of retention rates conservative; however, our first-wave results across the analysis remain valid and interpretable. The change in data download packages by platforms is a problem that needs to be addressed by platform regulators (Hase et al., 2024). Finally, it is possible that our unrestricted research design, which allowed participants to choose freely whether to donate, led to self-selection bias that affected our results. A more targeted quantitative research on omission behavior in general is needed that circumvents the above limitations.

### **Implications for future data donation studies**

The study has two major implications for future data donation studies. First, multi-wave data donation studies are feasible, as a substantial yet minor share of users still donate their data in subsequent waves. The second wave donation rates are higher, and omission rates are lower, pointing to an increase in data collection efficiency with additional waves. Importantly, the voluntary nature of our research design of donating data from multiple platforms leads to users switching platforms over time and possible self-selection effects. We suggest stricter incentive structures and consistent platform choices for each user.

Second, the study finds that the option to delete data points is used by a minority of donors (13.47%). This stresses the relevance of the retrospective nature of data donations and the consequential option of selective omission to improve informed consent as an important strength of the method. However, we also find that users who delete data points do so in substantial numbers. This presents the challenge of receiving a few highly incomplete data donations, which complicate cross-user analysis and conflict with data minimization principles in data collection. Researchers must carefully balance the trade-off between offering data omission features—resulting in higher donor rates but less complete data—and restricting such options, which may lower participation but yield more comprehensive datasets. This decision should be guided by the research question and the sensitivity of the requested data. Future research should study this more systematically. Overall, a transparent approach that explains to users that these data are strictly necessary for the research design and that more extensive deletion can reduce the research's explanatory power might be a viable option for successful data donation studies.

### **Funding Statement**

This publication was supported by the Weizenbaum Institute (grant number 16DIII141), funded by the Federal Ministry of Research, Technology and Space (BMFTR) and the State of Berlin.

## References

- Bechmann, A., Kroman Brems, M., Olesen, M. K., Walter, J. G., & Wegmann, D. (2025). *Data donation as a method for investigating trends and challenges in digital media landscapes at national scale: The Danish population's use of YouTube as an illustrative case*. Nordicom, University of Gothenburg. Retrieved May 26, 2025, from <https://urn.kb.se/resolve?urn=urn:nbn:se:norden:org:diva-13593>
- 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., Mendrik, A., van der Veen, E., Vloothuis, J., Hu, H., Voorvaart, R., & Oberski, D. L. (2022). Privacy-preserving local analysis of digital trace data: A proof-of-concept. *Patterns*, 3(3), 100444. <https://doi.org/10.1016/j.patter.2022.100444>
- Boeschoten, L., Schipper, N. C. de, Mendrik, A. M., Veen, E. van der, 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., Johannes Breuer, Kmetty, Z., Zoltán Kmetty, Haim, M., Mario Haim, Stier, S., & Sebastian Stier. (2022). User-centric approaches for collecting Facebook data in the ‘post-API age’: Experiences from two studies and recommendations for future research. *Information, Communication & Society*, 1–20. <https://doi.org/10.1080/1369118x.2022.2097015>  
MAG ID: 4284961432
- Carrière, T. C., Boeschoten, L., Struminskaya, B., Janssen, H. L., De Schipper, N. C., & Araujo, T. (2025). Best practices for studies using digital data donation. *Quality & Quantity*, 59(S1), 389–412. <https://doi.org/10.1007/s11135-024-01983-x>
- Entrena-Serrano, C., Degeling, M., Romano, S., & Çetin, R. B. (2025, June 12). *TikTok's Research API: Problems Without Explanations*. arXiv: 2506.09746 [cs]. <https://doi.org/10.48550/arXiv.2506.09746>
- Gomez Ortega, A., Bourgeois, J., Hutiri, W. T., & Kortuem, G. (2023). Beyond data transactions: A framework for meaningfully informed data donation. *AI & SOCIETY*, 40(2), 1–18. <https://doi.org/10.1007/s00146-023-01755-5>
- Groot Kormelink, T., Houwing, E., Struminskaya, B., Boeschoten, L., de Schipper, N., & Welbers, K. (2025). Meaningful informed consent? How participants experience and understand data donation. *Information, Communication & Society*, 0(0), 1–18. <https://doi.org/10.1080/1369118X.2025.2540915>
- Hase, V., Ausloos, J., Boeschoten, L., Pfiffner, N., Janssen, H., Araujo, T., Carrière, T., Vreese, C. de, Haßler, J., Loecherbach, F., Kmetty, Z., Möller, J., Ohme, J., Schmidbauer, E., Struminskaya, B., Trilling, D., Welbers, K., & Haim, M. (2024). Fulfilling data access obligations: How could (and should) platforms facilitate data donation studies? *Internet Policy Review*, 13(3). Retrieved September 23,

- 2024, from <https://policyreview.info/articles/analysis/fulfilling-data-access-obligations>
- Hase, V., & Haim, M. (2024). Can We Get Rid of Bias? Mitigating Systematic Error in Data Donation Studies through Survey Design Strategies. *Computational Communication Research*, 6(2). <https://doi.org/10.5117/CCR2024.2.2.HASE>
- Keusch, F., Pankowska, P. K., Cernat, A., & Bach, R. L. (2024). Do You Have Two Minutes to Talk about Your Data? Willingness to Participate and Nonparticipation Bias in Facebook Data Donation. *Field Methods*, 36(4), 279–293. <https://doi.org/10.1177/1525822X231225907>
- Kmetty, Z., & Stefkovics, Á. (2025). Validating a willingness to share measure of a vignette experiment using real-world behavioral data. *Scientific Reports*, 15(1), 9319. <https://doi.org/10.1038/s41598-025-92349-2>
- Ohme, J., Araujo, T., Boeschoten, L., Freelon, D., Ram, N., Reeves, B., & Robinson, T. N. (2024). Digital Trace Data Collection for Social Media Effects Research: APIs, Data Donation, and (Screen) Tracking. *18*, 124–141. <https://doi.org/10.1080/19312458.2023.2181319>
- Ohme, J., Araujo, T., Vreese, C. H., & Piotrowski, J. T. (2021). Mobile data donations: Assessing self-report accuracy and sample biases with the iOS Screen Time function. *Mobile Media & Communication*, 9(2), 293–313. <https://doi.org/10.1177/2050157920959106>
- Pfiffner, N., & Friemel, Thomas. 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. (2024). Data Donation Module: A Web Application for Collecting and Enriching Data Donations. *Computational Communication Research*, 6(2), 1. <https://doi.org/10.5117/CCR2024.2.4.PFIF>
- Statistisches Bundesamt. (2022). *Internationale Bildungsindikatoren im Ländervergleich* (Tabellenband Ausgabe 2022). Statistische Ämter des Bundes und der Länder. Wiesbaden. Retrieved October 18, 2024, from <https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bildung-Forschung-Kultur/Bildungsstand/Publikationen/Downloads-Bildungsstand/bildungsindikatoren-1023017227004.html>
- Struminskaya, B., Toepoel, V., Lugtig, P., Haan, M., Luiten, A., & Schouten, B. (2020). Understanding Willingness to Share Smartphone-Sensor Data. *Public Opinion Quarterly*, 84(3), 725–759. <https://doi.org/10.1093/poq/nfaa044>
- Strycharz, J., Meppelink, C., Zarouali, B., Araujo, T., & Voorveld, H. (2024). The blind spot in data donations: Who is (not) willing to donate digital data in social scientific research. *Computational Communication Research*, 6(2). <https://doi.org/10.5117/CCR2024.2.3.STRY>
- Trappmann, M., Haas, G.-C., Malich, S., Keusch, F., Bähr, S., Kreuter, F., & Schwarz, S. (2023). 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>
- Valkenburg, P. M., Beyens, I., Vaate, N. B. de, Janssen, L., & Wal, A. van der. (2024, February 12). 14. Person-Specific Media Effects. In T. Araujo & P. Neijens (Eds.), *Communication Research into the Digital Society: Fundamental Insights from the Amsterdam School of Communication Research* (pp. 233–246). Amsterdam University Press. Retrieved May 26, 2025, from <https://www.degruyterbrill.com/document/doi/10.1515/9789048560608-015/html>
- Valkenburg, P. M., van der Wal, A., Siebers, T., Beyens, I., Boeschoten, L., & Araujo, T. (2025). It is time to ensure research access to platform data. *Nature Human Behaviour*, 9(1), 1–2. <https://doi.org/10.1038/s41562-024-02066-5>
- Wedel, L., Mayer, A.-T., Fan, Y., Gaisbauer, F., & Ohme, J. (2026). *Multi-Platform Social Media Data Donation Behavior Dataset*. Retrieved January 19, 2026, from <https://doi.org/10.34669/WL.RD/6>
- Wedel, L., Ohme, J., & Araujo, T. (2024). Augmenting Data Download Packages – Integrating Data Donations, Video Metadata, and the Multimodal Nature of Audio-visual Content. *methods, data, analyses*, 19(2), 11–45. <https://doi.org/10.12758/mda.2024.08>
- Wedel, L., Ohme, J., Mayer, A.-T., Gaisbauer, F., & Fan, Y. (2025). The platform matters: Cross-platform differences in data donation willingness, behavior, and bias. *Communication Methods and Measures*, 0(0), 1–25. <https://doi.org/10.1080/19312458.2025.2605946>
- Welbers, K., Loecherbach, F., Lin, Z., & Trilling, D. (2024). Anything you would like to share: Evaluating a data donation application in a survey and field study. *Computational Communication Research*, 6(2). <https://doi.org/10.5117/CCR2024.2.5.WELB>
- Zannettou, S., Nemes-Nemeth, O., Ayalon, O., Goetzen, A., Gumjadi, K. P., Redmiles, E. M., & Roesner, F. (2024). Analyzing User Engagement with TikTok’s Short Format Video Recommendations using Data Donations. *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, 1–16. <https://doi.org/10.1145/3613904.3642433>

## Appendix A Overview on survey quotas

Comparison of population, target, and achieved quotas. Age and Gender quotas were provided by the panel provider (Bilendi), and the education quota is based on data from the German Federal Statistical Office (Statistisches Bundesamt, 2022). Columns may not sum up to 100 due to rounding.  $N_1 = 2,318$ ;  $N_2 = 1,159$

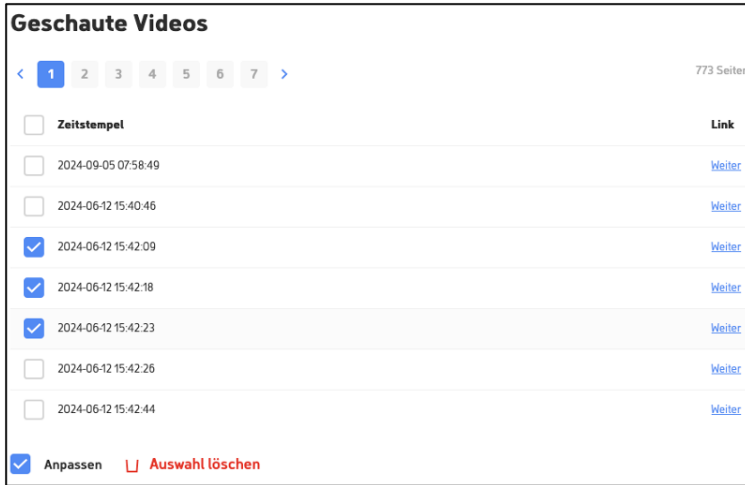
Variable	Values	Population (%)	Target (%)	Wave 1 (%)	Wave 2 (%)
Age	18-27	16.63	33.00	34.12 (n=791)	24.25 (n=281)
	28-37	19.93	15.94	16.31 (n=378)	14.15 (n=164)
	38-47	18.75	15.00	15.44 (n=358)	18.55 (n=215)
	48-57	22.89	18.31	18.03 (n=418)	22.35 (n=259)
	58 and older	21.80	17.44	16.09 (n=373)	20.71 (n=240)
Education (ISCED)	high	30.50		30.11 (n=698)	29.68 (n=344)
	mid	54.80		55.35 (n=1283)	55.74 (n=646)
	low	14.70		14.28 (n=331)	14.58 (n=169)
	I do not know	-		0.26 (n=6)	-
Gender	male	50.57		50.30 (n=1166)	53.67 (n=622)
	female	49.42		49.70 (n=1152)	46.33 (n=537)

# Appendix B Overview on collected traces for each Platform DDP

Overview of collected traces for each platform DDP.

Trace	Variables in DDP			
	YouTube	TikTok	Instagram	Facebook
View history	Timestamp, Link, Title, Account link, Account name	Timestamp, Link	Timestamp, Account name, Type [post, ad, video]	Timestamp, Account name, Type [Profile, Content]
Reactions (Likes, Hearts, Thumbs-up, etc.)	-	Timestamp, Link	Timestamp, Reaction Type, Account name, Link	Timestamp, Event, Reaction Type, Type [Post, Page]
Comments	Timestamp, Comment-ID, Account-ID, Price, Overarching Comment-ID, Video-ID, Text	Timestamp, Comment	Timestamp, Text, Account name, Type [Post, Reel]	-
Search history	Timestamp, Link, Search term or Link	Timestamp, Search term	Timestamp, Type [Account, Phrase-Word], Search term for direct messages	Timestamp, Search term for direct messages
Shares	-	Timestamp, Shared Content, Link, Method	Timestamp, Link, Type [sent, received], Conversation partner	Timestamp, Link, Type [sent, received], Conversation partner
Favorites/ Saved/ Playlists	Playlist name, Video-IDs, Timestamps	Timestamp, Link [Video, Sticker, Effect, Hashtag, Sound]	Timestamp, Account name, Link	-
Hashtags used	-	Timestamp, Hashtag	-	-
Posts	-	Timestamp, Likes, View restrictions, Sound, Location	-	Timestamp, Event, Text
Subscriptions/ Following	Account-ID, Account link, Account name	Timestamp, Account name	Timestamp, Account name, Link	Timestamp, Account name
Subscribers/ Followed by	-	Timestamp, Account name	Timestamp, Account name, Link	-
Friends	-	-	-	Timestamp, Account name
Blocked	-	Timestamp, Account name	-	-
Suggested Accounts	-	-	Timestamp, Account name	-
Watched Live-stream	-	Timestamp, Comments	-	-
Gifts made	-	Timestamp, Price, Account name	-	-
Group activity	-	-	-	Timestamp, Group
Ads clicked on	-	-	Timestamp, Type, Account name	-
Login history	-	Timestamp, Device Model, Device System, Network type	Timestamp, Device, Type	-
Location data	-	Timestamp, Region	Location	Location
Settings/ Other	-	Personalized Ads, App Language, Interests, Ad interests categories	Your topics, Privacy settings, Ad preferences	Ads preferences, Registration date

## Appendix C Data trace omission feature



Example screenshot of the omission feature for the TikTok watch history. Translation from top to bottom and left to right: Seen Videos, Timestamp, Link, Further, Adjust, Delete Selection