

CESSDA Work Plan 2020

New Data Types

D3b: Report on the online workshop “Linking Twitter & Survey Data”

Document info

Dissemination Level	PU
Due Date of Deliverable	31/08/20
Actual Submission Date	27/08/20
Type	Report
Approval Status	Approved by Training Working Group Leader Irena Vipavc Brvar and Tools & Services Working Group Leader Mari Kleemola
Version	V1.3
Number of Pages	16
DOI	10.5281/zenodo.5041068

The information in this document reflects only the author's views and CESSDA ERIC is not liable for any use that may be made of the information contained therein. The information in this document is provided “as is” without guarantee or warranty of any kind, express or implied, including but not limited to the fitness of the information for a particular purpose. The user thereof uses the information at his/ her sole risk and liability. This deliverable is licensed under a Creative Commons Attribution 4.0 International License.



Version history

Version	Date	Comment	Revised by
0.1	25/08/2020	First draft completed	Johannes Breuer
1.0	27/08/2020	Internal peer-review by Libby Bishop (GESIS)	Johannes Breuer
1.1	25/11/2020	Addressed review comments by CESSDA MO & WGL	Johannes Breuer
1.2	26/03/2021	Addressed review comments by WGL	Johannes Breuer
1.3	22/06/2021	Minor edits based on feedback from WGL & CESSDA MO	Johannes Breuer

Author List

Organisation	Name	Contact information
GESIS – Leibniz Institute for the Social Sciences	Johannes Breuer	johannes.breuer@gesis.org

Peer-review

Organisation	Name	Contact information
GESIS – Leibniz Institute for the Social Sciences	Libby Bishop	ElizabethLea.Bishop@gesis.org

Contents

Executive Summary	4
Abbreviations and Acronyms	5
1) General information about the workshop	6
a) Organization of the workshop	6
b) Information on participants	7
c) Workshop materials	7
2) Workshop content	8
a) Study planning	9
b) Twitter data collection	10
c) Data processing	10
d) Archiving and sharing	11
3) Reflections on the workshop format, feedback from participants, and future plans	12
4) Conclusion	15
5) Suggestions for further reading	16

List of Tables

Table 1: Breakdown of participants by country	7
Table 2: Breakdown of participants by status group	7
Table 3: Timetable of the workshop	8
Table 4: Evaluation results for the online workshop	12

Executive Summary

This report summarizes the relevant information on the online workshop “Linking Twitter & Survey Data” as an additional written record to the video recording and the published slides. The online workshop “Linking Twitter & Survey Data” was conducted on June 23rd, 2020 as part of the CESSDA New Data Types Work Plan for 2020.

The first part of this report provides general information about the online workshop. The second part includes a more detailed description of the workshop content. The third part presents some reflections on the workshop format, summarizes some of the feedback from participants, and briefly describes future plans for this workshop. The fourth section of this report provides a short summary describing the outcomes of the workshop and their implications, and the fifth section contains some suggestions for further reading.

Abbreviations and Acronyms

API	Application Programming Interface
CESSDA	Consortium of European Social Science Data Archives
FAIR	Findable Accessible Interoperable Reusable
GDPR	General Data Protection Regulation
GESIS	GESIS – Leibniz Institute for the Social Sciences
GUI	Graphical user interface
ID	Identifier
REST	Representational State Transfer
SERISS	Synergies for Europe’s Research Infrastructures in the Social Sciences
ToS	Terms of Service
WP2020	Work Plan 2020

1) General information about the workshop

The online workshop “Linking Twitter & Survey Data” took place on Tuesday, June 23rd from 10 am to 5 pm CEST. The instructors for the workshop were Luke Sloan (Cardiff University), Libby Bishop (GESIS), and Johannes Breuer (GESIS). The online workshop was delivered via Zoom and contained a mixture of lectures by the instructors and interactive sessions with exercises and discussion. The topics of the lecture parts are presented in section 2 of this report. In addition to those, there were nine interactive sessions in total. The topics of these interactive sessions/exercises were the following: Introduction & research interests, New research questions (with Twitter data and linked survey and Twitter data), Questions on ethics, Informed consent, Collecting Twitter data, Evaluation disclosure potential, Summary measures, Ethical reflection on Twitter’s ToS, Final discussion & feedback. Participants worked on the exercises in groups or alone and then discussed with the whole group. Depending on the complexity of the tasks, the time for the interactive sessions varied from around 15 to about 45 minutes (including the group discussion at the end of each interactive session).

a) Organization of the workshop

The workshop was a joint effort by the CESSDA project “New Data Types” and GESIS Training¹. Thus, the workshop was part of the “New Data Types” project as well as the GESIS Training workshop program. The GESIS Training team has ample experience with planning and conducting workshops and established procedures promoting the events, registration of participants, local organization, evaluation, and the provision of certificates of participation that greatly facilitated the delivery of the workshop and presented an added value for the participants as well as the instructors. The workshop was originally planned as a two-day in-person event at GESIS in Cologne. Due to the global COVID-19 pandemic, the decision was made in April 2020 to switch to an online format.² For this switch, the instructors decided to also change the duration of the workshop from two days to one day as the expectation was that participants may be exhausted by an online workshop lasting for two full days, given that work life (especially in academia) was largely being moved into the online sphere for everybody. The collaboration with GESIS Training allowed the instructors to use a professional Zoom license, a platform with which one of the instructors, Luke Sloan, already had some experience from his university teaching. The GESIS training staff also provided support for using Zoom for the participants as well as the instructors before and during the workshop. In addition, the instructors were able to make use of the e-learning platform ILIAS for which

¹ GESIS Training is a team within the GESIS department Knowledge Transfer that is responsible for organizing the GESIS workshop program and other training events, such as the Summer School in Survey Methodology.

² However, people were already able to register for the workshop before that and those who registered before did so with the expectation of a 2-day in-person workshop in Cologne. Of course, those who had registered before the switch to an online format was made were informed about this immediately and had the opportunity to cancel their registration.

GESIS Training has a license to share literature and other materials with the participants. For its workshops, GESIS Training charges fees. This was also the case for the online workshop on “Linking Twitter & Survey Data” to cover the internal costs for the workshop (staff time & licenses) and to be able to pay the external instructor, Luke Sloan, a professional fee. At the time of the workshop the fees for a one-day workshop were 60 Euros for students, 90 Euros for other participants from academia, and 180 Euros for participants from commercial companies.³ The workshop was advertised by GESIS Training through its newsletter and Facebook page and by CESSDA through the CESSDA Training events calendar and via Twitter.

b) Information on participants

Eleven people from different academic institutions and academic disciplines and at different career stages (from Ph.D. student to full professor) from Germany and Austria participated in this one-day online workshop. Tables 1 and 2 provide a detailed breakdown of which countries the participants came from and to what status group they belonged.

Table 1: Breakdown of participants by country

Country	Number of participants
Germany	9
Austria	2

Table 2: Breakdown of participants by status group

Status group	Number of participants
Academic	5
Student	3
GESIS staff	3

As stated in the previous section, the workshop was originally planned as a two-day face-to-face event which is the main reason why the participants came from Germany and Austria, as for people from most other countries, attending the in-person workshop would have been associated with higher travel costs and time expenditure. Regarding the overall number of participants, this was well within the expectations for such a workshop based on previous events. Typically, the number of participants for in-person workshops organized by GESIS Training varies between 8 and 20 participants. The duration of the workshop as well as its interactive nature impose limits on the number of participants, even for an online event.

c) Workshop materials

The following workshop materials are publicly available:

- Slides on Zenodo: <https://zenodo.org/record/4001700#.X0ZVJtRS8-U>
- Video on Zenodo: <https://zenodo.org/record/4001700#.X0ZVJtRS8-U>

³ GESIS staff can participate for free but participation for this group is typically limited to 2-3 per course.

- Video on CESSDA Training YouTube channel:
<https://www.youtube.com/watch?v=IG1jJ3WP2h8>
- Interactive Jupyter Notebooks to demonstrate several tools for collecting Twitter data: <https://github.com/jobreu/twitter-linking-workshop-2020>

While GESIS Training does not record its online events, an exception was made for this workshop as it was part of the CESSDA WP2020 project “New Data Types” for which one of the deliverables required to make the slides as well as a recording of this workshop publicly available so that also people who did not participate in the live event can access and use the materials. The workshop was recorded using the respective function in Zoom. To protect participants’ privacy, the interactive parts of the workshop were not recorded. As the exercises (including their introduction and discussion) and discussions with/among participants were not recorded, only the lecture parts are included in the recording. Accordingly, while the duration of the workshop itself (incl. breaks) was 7 hours, the duration of the recording is 1 hour and 36 minutes. Due to the recording options in Zoom and the necessity to switch the host roles during the workshop the video thumbnails of some of the participants were visible during the lecture parts. This required extensive post-editing before the video recording could be published.

2) Workshop content

After a brief introduction of the workshop topic and the instructors, the structure of the workshop followed the typical research process for studies working with linked survey and Twitter data and covered the following phases:

- a) Study planning
- b) Twitter data collection
- c) Data processing
- d) Archiving and sharing

The data analysis phase was not covered as the choice of appropriate analysis methods always depends on the type of data that are used, the research questions of the study, and the methodological background of researchers as well as the standards in their respective fields. Table 3 presents the detailed timetable of the workshop.

Table 3: Timetable of the workshop

Time	Section/Topic
09:45 – 10:00	Arrival
10:00 – 10:30	Introduction & Outline
10:30 – 11:15	Study Planning
11:15 – 11:30	Coffee Break
11:30 – 12:00	Ethical Issues
12:00 – 12:45	Informed Consent
12:45 – 13:45	Lunch Break
13:45 – 14:45	Collecting Twitter Data

14:45 – 15:15	Processing Data
15:15 – 15:45	Data Security & Reduction
15:45 – 16:00	Coffee Break
16:00 – 16:45	Data Archiving & Sharing
16:45 – 17:00	Discussion, Evaluation, & Wrap-Up

The workshop drew on recent experiences from different studies that the instructors were involved in and provided guidance on how to address the practical and ethical issues associated with linking survey and Twitter data in the different phases of the research process listed above. Key points from the presentations are summarized in the following sections.

a) Study planning

When planning a study that links survey and Twitter data it is important to keep in mind that there are different types of Twitter data. Researchers can work with textual data from tweets, metadata (time, location, likes, etc.) or network data (follower or tweet networks). Importantly, what type of data is needed depends on the specific research questions at hand. While using data from Twitter can help to avoid or alleviate some of the issues related to self-report data, such as social desirability or problems with recollection, Twitter data have their own sets of limitations, including the lack of information about individuals or missing outcome variables of interest. Linking data from surveys and Twitter is a way to combine the unique strengths of the two data types and overcome some of their respective limitations. There are different forms of linking Twitter and survey data, each of which have their own advantages and disadvantages. For example, data can be linked on the individual or aggregate level. When planning a study, researchers also need to take into account the ethical challenges of working with Twitter data. These include aspects like the (potentially) disclosive nature of these data or the fact that using the data for research is different from the original purpose of the content. Another important topic in this context is informed consent. If researchers combine survey and Twitter data from the same individuals, the survey can be used to obtain informed consent. Of course, the informed consent needs to adhere to relevant legal regulations (such as GDPR in Europe) and needs to satisfy ethical standards (e.g., defined by Institutional Review Boards). Al Baghal et al. (2020)⁴ have developed a good template for an informed consent for studies linking survey and Twitter data.

⁴ 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. <https://doi.org/10.1177/0894439319828011>

b) Twitter data collection

There are different ways of collecting/accessing social media data in general, each with their own sets of strengths and limitations (see Breuer et al., 2020⁵). For Twitter data, the most commonly used options are buying the data from resellers or collecting the data via the Twitter APIs. If researchers wish to link survey and Twitter data on the individual level, using the Twitter APIs to collect the data themselves is typically the better option as this ensures that the data from the people who answered the survey and agreed to have their Twitter data collected and linked are included. Twitter offers different APIs which can be used by researchers: the REST API that provides information about users and a limited number of historical tweets from users, and the Streaming API that allows the collection of tweets in real time. Notably, while these APIs can provide useful data for researchers, the type and quantity of the data that can be accessed through them is limited and these limitations can be changed by Twitter. In addition, the ToS of these APIs regulate what researchers can do with the data and these regulations can change as well. APIs can even be shut down completely. Despite these risks and drawbacks, the Twitter APIs are a rich and convenient source of data as they provide different types of data in a structured format (typically as JSON files). There are many free (and often also open-source) tools available for collecting data - via the Twitter APIs or otherwise. They differ in several regards, including whether they have a GUI or whether they require programming skills. The Social Media Lab at Ryerson University curates a list of tools called the "Social Media Research Toolkit" that provides an extensive overview.⁶ For this workshop, the instructors have also created a set of interactive Jupyter Notebooks through which it is possible to try out some of the tools/packages for collecting Twitter data without the need to install any software locally.⁷ These were created and are hosted on via GESIS Notebooks service⁸ which can be used and accessed by anyone.

c) Data processing

When working with Twitter data, researchers and archivists need to keep in mind that they are highly disclosive most of the time. Surveys typically promise anonymity and assuring this with (linked) Twitter data can be a challenge. Importantly, the disclosure risks are not only associated with obvious data types, such as user profile information or tweet texts. Depending on the API request, a single tweet can come with over 150 associated attributes that can, in combination, be used to identify users, even when the actual username or tweet text are not available. To address this issue, researchers need to employ or develop appropriate measures

⁵ 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>

⁶ Social Media Research Toolkit, <https://socialmediadata.org/social-media-research-toolkit/> (date of access: 23/11/2020)

⁷ Interactive notebooks for the workshop Linking Twitter & Survey Data, <https://github.com/jobreu/twitter-linking-workshop-2020> (date of access: 27/08/2020)

⁸ GESIS Notebooks, <https://notebooks.gesis.org/> (date of access: 27/08/2020)

and processes. Sloan et al. (2020)⁹ recommend four principles for maintaining security and maximizing privacy when working with linked Twitter and survey data: 1) Systematic processing, 2) Data reduction, 3) Controlled access, and 4) Data deletion. For example, it is advisable to keep the full survey and Twitter data separate and only combine what is necessary for a given analysis. In addition, using only derived variables and summary measures instead of the raw data can be a viable solution for minimizing disclosure risks while maintaining the analytical value of the data.

d) Archiving and sharing

The reasons for archiving linked survey and Twitter data are generally the same as those for archiving research data in general: increasing the transparency and, thus, the reproducibility and replicability of research. In addition to that, sharing these data can help alleviate disparities in data access. Also, the principles for sharing these data are also the same: The data should be FAIR – Findable, Accessible, Interoperable, and Reusable. Two key challenges with regard to archiving and sharing linked survey and Twitter data are their disclosive nature (see previous section) and the restrictions imposed by the Twitter (API) ToS. As these do not allow the sharing of large amounts of tweets, a common practice is to share only tweet IDs which can be used to collect the tweets (again) – a process called rehydration. While this increases user privacy as tweets and accounts that are deleted cannot be used in future analyses, this reduces reproducibility and potentially also the reuse value of the data. Another way of increasing data privacy is to employ access restrictions as offered by most archives. Regarding data linking specifically, the Twitter Developer Policy¹⁰ explicitly states that matching Twitter data with other data on the individual level requires “express opt-in consent” from the individuals. Beyond adhering to Twitter ToS there are also other ethical considerations. While there are no universal ethical guidelines for archiving such data, some resources that present recommendations for archiving Twitter data (e.g., Williams et al., 2017¹¹) or social media data in general (e.g., Appendix A of the SERISS WP6-D3 Report¹²) can provide some helpful guidance.

⁹ 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. <https://doi.org/10.1177/1556264619853447>

¹⁰ Twitter Developer Policy, <https://developer.twitter.com/en/developer-terms/policy> (date of access: 22/06/2021)

¹¹ Williams, M. L., Burnap, P., & Sloan, L. (2017). Towards an ethical framework for publishing twitter data in social research: Taking into account users’ views, online context and algorithmic estimation. *Sociology*, 51(6), 1149–1168. <https://doi.org/10.1177/0038038517708140>

¹² SERISS Report on legal and ethical framework and strategies related to access, use, re-use, dissemination and preservation of social media data, https://seriss.eu/wp-content/uploads/2019/11/D6.3-Report-on-legal-and-ethical-framework-and-strategies..._FINAL.pdf (date of access: 27/08/2020).

3) Reflections on the workshop format, feedback from participants, and future plans

The workshop was originally planned as a face-to-face event for two days. Due to the COVID-19 pandemic, the format was changed to an online workshop. As the instructors expected an online workshop for two full days to be too exhausting since during the pandemic people had/have to participate in many online meetings and may experience fatigue, the duration was reduced to one full day. The platform that was used for this workshop was Zoom which had an easy-to-use breakout room function that was convenient to use for some of the interactive sessions. While this was certainly different from a face-to-face setting, the discussion and interaction in this online workshop worked quite well. It was good to have multiple instructors. That way it was possible that one person monitored the chat while another one was presenting or talking.

The workshop was evaluated using the standard GESIS Training evaluation forms. Eight of the eleven participants filled out the evaluation form. A tabular summary of the evaluation questions is presented in Table 4.

Table 4: Evaluation results for the online workshop

Question	Scale	Number of responses	Mean
The content was well structured.	1 – Strongly disagree 2 – Disagree 3 – Neither/Nor 4 – Agree 5 – Strongly Agree	8	4.4
The course's content was in line with the course description	1 – Strongly disagree 2 – Disagree 3 – Neither/Nor 4 – Agree 5 – Strongly Agree	8	4.4
A link between theory and practice was made.	1 – Strongly disagree 2 – Disagree 3 – Neither/Nor 4 – Agree 5 – Strongly Agree	8	4.3
The practical exercises have enabled me to apply the treated procedures and methods myself.	1 – Strongly disagree 2 – Disagree 3 – Neither/Nor	8	3.4

	4 – Agree 5 – Strongly Agree		
The course materials were useful.	1 – Strongly disagree 2 – Disagree 3 – Neither/Nor 4 – Agree 5 – Strongly Agree	8	4.3
The instructors are scientifically competent.	1 – Strongly disagree 2 – Disagree 3 – Neither/Nor 4 – Agree 5 – Strongly Agree	8	5.0
The instructors are didactically competent.	1 – Strongly disagree 2 – Disagree 3 – Neither/Nor 4 – Agree 5 – Strongly Agree	8	4.6
My previous substantial knowledge was sufficient.	1 – Strongly disagree 2 – Disagree 3 – Neither/Nor 4 – Agree 5 – Strongly Agree	8	4.0
My previous technical knowledge was sufficient.	1 – Strongly disagree 2 – Disagree 3 – Neither/Nor 4 – Agree 5 – Strongly Agree	8	3.9
I learned a lot in the course.	1 – Strongly disagree 2 – Disagree 3 – Neither/Nor 4 – Agree 5 – Strongly Agree	8	4.1
How do you rate the amount of material covered? Was it...	1 – Too little 2 – Rather little 3 – Just right	8	3

	4 – Rather a lot 5 – Too much		
How do you rate the opportunities for discussion? Was it...	1 – Too little 2 – Rather little 3 – Just right 4 – Rather a lot 5 – Too much	8	3.4
How do you rate the amount of time for practical exercises? Was it...	1 – Too little 2 – Rather little 3 – Just right 4 – Rather a lot 5 – Too much	8	2.6
How do you rate the level of difficulty of the course? Was it...	1 – Too easy 2 – Rather easy 3 – Just right 4 – Rather difficult 5 – Too difficult	7	2.9
How do you rate the duration of the course? Was it...	1 – Too short 2 – Rather short 3 – Just right 4 – Rather long 5 – Too long	8	2.9
How do you rate the pace of the course? Was it...	1 – Too slow 2 – Rather slow 3 – Just right 4 – Rather fast 5 – Too fast	8	3.1
How satisfied are you with this course overall?	1 – Very dissatisfied 2 – Dissatisfied 3 – Neither/Nor 4 – Satisfied 5 – Very satisfied	8	4.3

In addition to the formal evaluation, the instructors asked participants to indicate one thing the instructors should start to do or do more of in the workshop, one thing they should continue to do, and one thing they should stop doing or do less of. The most frequent

comment across the two evaluation forms was that participants would like more time for the interactive parts (discussions and exercises), especially for the part on data collection (for which the interactive Jupyter notebooks were created: see above). Importantly, the feedback provided by the participants can inform the design and delivery of future similar events. In general, the topic of linking survey and Twitter data seems to be quite popular and of interest to many researchers. An indicator of this is that the instructors were invited to offer the workshop again in March 2021 at the University of Graz in Austria and as part of the GESIS Training workshop program in June 2021.

4) Conclusion

The online workshop “Linking Twitter & Survey Data” provided researchers with knowledge and resources on different aspects of linking data from Twitter and surveys, including study planning, data collection and processing, and archiving and sharing. The evaluations for the workshop were very positive and the invitations the instructors received to repeat the workshop show that there is also interest in this topic and need for training among researchers. While the number of participants was low, albeit within the range of expectations and experiences for such events, this allowed the instructors to include extended hands-on sessions. Within those, participants were able to engage in in-depth discussions among themselves as well as with the instructors and get practical experience with some of the steps involved in the linking of Twitter and survey data. In addition to the training that the participants received, the online workshop was also an opportunity to network with (other) researchers who are working or planning to work with linked Twitter and survey data, both for the participants as well as the instructors. From the perspective of the CESSDA “New Data Types” project, the close collaboration with Luke Sloan is a substantial benefit, given his expertise and number of relevant publications in this area. Collaborating with him and other prominent researchers working with social media and other new data types can increase the visibility of CESSDA activities in this domain and help in promoting the output of the “New Data Types” project as well as future work in this area. One immediate benefit of the joint workshop was that Luke Sloan also contributed to Deliverable D5 of the CESSDA WP2020 project “New Data Types” which is a paper on informed consent for linked survey and social media data. In addition to contributing himself, he also recruited Tarek Al Baghal from the University of Essex as an external coauthor for the paper who has extensive expertise and is well-known in the fields of survey methodology and research involving social media data. Beyond the networking and collaboration aspect, the more tangible outcomes of the workshop which are the materials produced for it (slides, video recording, and code for the interactive Jupyter notebooks) are reusable and likely to be of interest to many researchers who are active in the emerging field of linking surveys with social media data.

5) Suggestions for further reading

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. <https://doi.org/10.1177/0894439319828011>

Bishop, L., & Gray, D. (2017). Chapter 7: Ethical challenges of publishing and sharing social media research data. In K. Woodfield (Eds.), *Advances in Research Ethics and Integrity* (pp. 159–187). Emerald Publishing Limited. <https://doi.org/10.1108/S2398-601820180000002007>

Breuer, J., Al Baghal, T., Sloan, L., Bishop, L., Kondyli, D., & Linardis, A. (in press). Informed consent for linking survey and social media data: Differences between platforms and data types. *IASSIST Quarterly*.

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>

Kinder-Kurlanda, K., Weller, K., Zenk-Möltgen, W., Pfeffer, J., & Morstatter, F. (2017). Archiving information from geotagged Tweets to promote reproducibility and comparability in social media research. *Big Data & Society*, 4(2), 205395171773633. <https://doi.org/10.1177/2053951717736336>

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. <https://doi.org/10.1177/1556264619853447>

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>

Williams, M. L., Burnap, P., & Sloan, L. (2017). Towards an ethical framework for publishing Twitter data in social research: Taking into account users' views, online context and algorithmic estimation. *Sociology*, 51(6), 1149–1168. <https://doi.org/10.1177/0038038517708140>