



# Perspective: Acknowledging Data Work in the Social Media Research Lifecycle

Katharina E. Kinder-Kurlanda<sup>1\*</sup> and Katrin Weller<sup>2</sup>

<sup>1</sup>CAIS – Center for Advanced Internet Studies, Bochum, Germany, <sup>2</sup>Computational Social Science Department, GESIS – Leibniz Institute for the Social Sciences, Cologne, Germany

This perspective article suggests considering the everyday research data management work required to accomplish social media research along different phases in a data lifecycle to inform the ongoing discussion of social media research data's quality and validity. Our perspective is informed by practical experience of archiving social media data, by results from a series of qualitative interviews with social media researchers, as well as by recent literature in the field. We emphasize how social media researchers are entangled in complexities between social media platform providers, social media users, other actors, as well as legal and ethical frameworks, that all affect their everyday research practices. Research design decisions are made iteratively at different stages, involving many decisions that may potentially impact the quality of research. We show that these decisions are often hidden, but that making them visible allows us to better understand what drives social media research into specific directions. Consequently, we argue that untangling and documenting choices during the research lifecycle, especially when researchers pursue specific approaches and may have actively decided against others (often due to external factors) is necessary and will help to spot and address structural challenges in the social media research ecosystem that go beyond critiques of individual opportunistic approaches to easily accessible data.

**Keywords:** social media research, epistemology, research data management, data lifecycle, digital trace data, methodology, data archiving

## OPEN ACCESS

### Edited by:

Katja Mayer,  
University of Vienna, Austria

### Reviewed by:

Fred Morstatter,  
University of Southern California, Los Angeles, United States  
Fenwick Mckelvey,  
Concordia University, Canada

### \*Correspondence:

Katharina E. Kinder-Kurlanda  
katharina.kinder-kurlanda@cais.nrw

### Specialty section:

This article was submitted to  
Data Mining and Management,  
a section of the journal  
Frontiers in Big Data

**Received:** 04 November 2019

**Accepted:** 11 November 2020

**Published:** 10 December 2020

### Citation:

Kinder-Kurlanda KE and Weller K  
(2020) Perspective: Acknowledging  
Data Work in the Social Media  
Research Lifecycle.  
Front. Big Data 3:509954.  
doi: 10.3389/fdata.2020.509954

## INTRODUCTION

Web platforms such as search engines or online shopping platforms and especially social media services have become an important source of research data across disciplines. Respective data can be content published and shared by users (e.g., texts, images and videos shared on Facebook, Instagram or Twitter) as well as users' communication and interaction networks (e.g., via followings, likes). In addition to an interest in studying user behavior within online environments (e.g., communication practices or phenomena such as edit wars on Wikipedia), another interest is in using data from online platforms to infer conclusions about broader societal issues (e.g., voting behavior or racism). User communication data from social media are often extracted by using public application programming interfaces (APIs) or web scraping, and thus are highly dependent on the platform and API structures and affordances.

Consequently, research based on web and social media data has been criticized for frequently being data-driven with research questions being tailored to data availability and accessibility (e.g.,

boyd and Crawford 2012; Kitchin 2014; Ekbia et al., 2015). The challenges of working with this dynamic, impermanent type of data are aggravated by the constant evolution of platforms accompanied by changes in data access opportunities (as outlined e.g., by Rogers (2015), or Karpf (2012) and prominently affecting research after major changes to Facebook's API as outlined by Bruns (2013). "Opportunistic" (Olteanu et al., 2019) approaches to social media data have sparked various critical reflections on consequential ethical drawbacks (e.g., Fiesler and Proferes, 2018; Franzke et al., 2020) and methodological issues of social media research and its specific epistemic quality [e.g., in sociology by Halford et al. (2018) or Schroeder (2014) or in anthropology by Boellstorff (2013)]. While these issues certainly need to be attended to, we are here contributing a perspective that may allow to better understand the epistemological challenges and drivers of social media research beyond the starting point of a specific project. We focus on everyday engagement with data as an ongoing, situated process of engaging with technology-enabled structures and affordances of tools and platforms to accomplish the everyday data work that is required to accomplish social media research. Our perspective is informed by a) our experiences when curating social media data in the data archive of the GESIS - Leibniz Institute for the Social Sciences (e.g., Kaczmirek et al. (2014), Bruns and Weller (2016), Weller and Kinder-Kurlanda (2016), Kinder-Kurlanda et al. (2017)), b) results from an interview study<sup>1</sup> on the data practices of social media researchers (e.g., Kinder-Kurlanda and Weller (2014), Weller and Kinder-Kurlanda (2015), Weller and Kinder-Kurlanda (2017)), and c) various literature on methodological, ethical and epistemological issues of social media data as well as notions from science and technology studies that view technology use (by social media users and researchers alike) as situated within the social and cultural world (Suchman, 1987) in a complex entanglement of humans, technologies, structures and organizations (Latour, 1996; Orlikowski, 2007).

Drawing on our experiences in archiving and data management practice and theory, we have organized this perspective article on everyday data work along a typical research lifecycle model, distinguishing different stages throughout the research process - from collecting research data to preserving it.

## A DATA LIFECYCLE FOR SOCIAL MEDIA DATA

There are various models of digital data that allow us to understand different aspects of it. For example, Crawford and

Joler (2018), by detailing an anatomical model of the Amazon Alexa device, recently revealed the vast complexity and scale of social, environmental, economic, and political costs hidden behind seemingly simple everyday data interfaces. Within the data management literature there are various processual models of data, many of them cyclical models stressing the circle of uses and reuses of data that may retain value indefinitely (Borgman, 2019; for an overview of different data life cycle models see: Carlson, 2014). The data life-cycle model used by data archivists views such data from the perspective of those who are used to curating it as research data, with the intent to make it findable, accessible and interoperable in order to facilitate reuse (Wilkinson et al., 2016) and the reproducibility of research. The model guides archivists' and curators' assistance to researchers - what assistance researchers most likely require varies depending on the phase of the cycle model they can be allocated to. Untangling the steps that occur along a typical research lifecycle shifts the perspective from individual opportunistic strategies toward creating research environments and infrastructures that assist researchers in pursuing their quest for best practices and solutions within ephemeral structures of data and publics.

Pouchard (2015, p. 183) proposes a lifecycle model for Big Data that "combines the perspective of research with that of data curation, identifies the tasks of data management that lead to analysis, while preserving the curation aspect, and encompasses the steps necessary to handle Big Data."

Following this model we have structured this contribution along the different phases of planning, acquiring, preparing and analysing, preserving and discovering for secondary use (and then back to planning etc.). Pouchard especially stresses the importance of describing (or documenting) every step as soon as possible to reflect the source and facilitate discovery and also to prevent omission of a potentially crucial transformation of the data (p. 184) and to assess and monitor data quality throughout every step of the life cycle. Documenting and assessing the data as important elements of data management are intended to enhance research quality by eventually allowing for research to become more transparent, critiqueable and reusable - with the aim of advancing cumulative research. The cyclical model reflects the archivists' perspective. From our interviews with social media researchers we conclude that in actual research projects the different phases are better modeled as different types of activities that rely on each other, but that do not necessarily follow each other. Researchers may go back to previous steps several times or skip steps.

Documentation, ideally in addition to shared datasets, is one tangible outcome of data management. Data management principles also encourage critical reflection of all aspects of how data is collected, prepared, handled, stored and shared, which leads to continuous engagement with challenges in data quality and research ethics. Underlying this focus on critically reflecting the concrete handling of data is the assumption that such practices are epistemic and linked to specific types of knowledge being produced (Koch and Kinder-Kurlanda, 2020). Data management thus is closely linked to the best practices of applied methodologies. Nevertheless, it is often neglected

<sup>1</sup>Details on the research design of the interview study can be found in our previous publications, e.g., Weller and Kinder-Kurlanda (2015). Interviews in our first wave (starting 2013–14) were conducted as semi-structured face-to-face interviews with more than 40 researchers from different disciplines, countries and at different career levels. We are currently in the process of re-interviewing participants in a second wave (starting 2019–20).

throughout the research process because of the time and effort involved - something that also became apparent during our interviews.

Data management can be aided by data curation tools and software and, especially in the later phases of archiving and sharing, by institutions such as data repositories. However, when dealing with data from social media as an example of digital data sourced from Internet platforms, researchers are faced not only with little guidance on how to manage data, but also with the fact that existing best practices and tools from other research fields may not be transferable. What is more, best practice with regard to opening one's research to peer review often still needs to be established for the new methodologies with the different types of data. In the following we focus on the concrete steps of data handling in social media research and detail some of the challenges of revealing information about them to make research valid and to open it to scrutiny by peers.

## PLANNING A PROJECT

The main objective of planning activities is to determine the specific research questions, which will then lead to selecting suitable methodological approach(es) and to identifying the data required for answering the research questions. Researchers often return to this planning stage from later stages to refine research questions and requirements for data, and particularly the data acquisition phase may be closely entangled with the planning phase. When working with social media data, refining initial research questions and refining data collection strategies may be necessary if the "ideal" dataset turns out not to be accessible (Mayr and Weller, 2017), and iterations are needed to define a question and look for suitable data. For example, in our interviews with social media researchers we witnessed that researchers may find that a chosen API has unexpected restrictions so that only a limited amount of data is available, may then decide to employ a different method (mixed, explorative) while necessarily adjusting the research question (Weller and Kinder-Kurlanda, 2015).

Such an iterative process is not per se different from working in other areas or with other data. And usually documentation does not report the full iteration of planning and selecting data to collect; discarded ideas that could not be realized in practice are usually not described. However, in social media research with its greater innovation in methods and lack of standard epistemologies this neglect is contributing to the notion of opportunism. The extended tinkering with APIs, the search for interesting data, or the struggle with cumbersome website designs that make scraping difficult are typically hidden in deference to the presentation of only the final, carefully chosen dataset in a published paper. Not revealing the iterative processes of planning, however, makes assessing the quality of the research design and the appropriateness of the selected data source difficult. Decisions in the planning phase made to ensure ethical standards are also not typically part of published material and not regularly asked for in review processes for journals and conferences. This is not to say that decisions

taken in the iterative processes of planning social media research are always well-thought out, valid, ethically reflective and methodologically sound - but putting the focus on these processes highlights that it is here where important decisions are being taken that fundamentally determine a study's validity and deserve to become more transparent. First steps toward transparency and for reducing the current challenges could be better documentation practices (including developing standards for documenting research design decisions) and a stronger focus on the specific research design choices during peer review. This will be of particular relevance for the processes of data acquisition.

## Acquiring Data

Social media data can be collected in different ways (e.g., requested from an API, bought from a reseller or collected via screenshots and copy-paste) and can occur in various shapes (e.g., text, image, video, network data). Decisions about how to collect what data from which sources are iteratively developed during the planning phase, as mentioned above. But as we have seen throughout our interviews, they also depend on the researcher's capacities and skills (such as programming skills or financial resources), on the collections system's technical affordances (such as the available server capacity or (third party) tools), on platform providers' (legal or technical) restrictions imposed on data collection possibilities and on ethical considerations about users' privacy and (lack of) consent to research. The question of whether a particular research design based on social media data might be problematic for, desired or expected by, or go against the aims of specific individual users or groups of users of a platform are often difficult to answer.

In a more fine-grained view we saw that it is a challenge during this phase to untangle how single choices in the data acquisition process may have influenced the validity of the data. Deciding upon the exact selection criteria (e.g., search terms used to retrieve single posts, the chosen time period for data collection, the focus on certain languages) may indicate a specific focus, limit the scope of the research, or induce errors if the aim is to infer knowledge about whole populations (Olteanu et al., 2019; Sen et al., 2019). However, the level of detail of documentation required to understand and reconstruct the data acquisition process from the perspective of a (reviewing) peer or secondary user goes beyond what is feasible to include in the "methods" section of a paper (see Zimmer and Proferes, 2014 or Hemphill et al., 2019). To mitigate these challenges, supplementary material and additional publication formats could help to fill this gap. Some journals and conferences now feature dataset papers as a specific genre and to encourage code sharing. However, there are currently no tools or standards for documenting the data acquisition process, that allow, for example, to describe the rationale for choosing specific search terms or collection periods or to record other critical information, such as server downtimes during the data collection phase. And even if authors have described their data acquisition in great detail, there are aspects that go beyond what they can deliver, e.g., third party tools that act as black boxes and add an additional

layer of uncertainty. Approaches such as the proposal of “datasheets for datasets” (Geburu et al., 2018) can help to fill this gap in the future. Such approaches should be used in test cases by different communities involved in social media research to then discuss and ideally agree on shared solutions. This will be an important step forward, as we have seen that there is no consensus (or often even little debate) amongst researchers about what level of understanding or reconstruction by peers should be enabled beyond the effort currently required to document enough detail of acquisition activities. Making decisions about which information to provide about acquisition activities are made under conditions of great uncertainty. The availability of choices is often dependent on tools provided by hard-to-influence third parties which determine what can be done and what features data has. Features may also change over time and some information may be hard or impossible to document as it is unknown.

## PREPARING AND ANALYSING DATA

Preparation and analysis of data comprise activities such as preprocessing, cleaning, labeling, sorting, and filtering. Some of these activities are closely tied to the data acquisition process and often there is again deliberation, as several attempts of trial and error, going back and forth to tease out specific aspects in the data (and hide others discarded as “noise”) are being made. However, with this phase the focus is on the processes of working with social media data after they have been collected: the final analyses as well as preliminary steps that prepare the data for being interpretable. Especially data preprocessing steps often rely on supporting tools such as dictionaries of terms, or labeling algorithms. These may need to be improved or adapted, as it is not always possible or feasible to design own tools. Existing methods and tools for data preparation and analysis may have certain limitations, e.g., tools for detecting sentiments or opinions in texts or users’ gender and age based on profile photos are limited in accuracy - while the exact performance in a specific use case may be difficult to assess [see e.g., Sen et al. (2020) for a comparison of different opinion mining approaches].

The limitations of the available tools may be well known to the researcher by the end of a project, but they are difficult to publish and there is little incentive to do so. In our interviews we found that some researchers were even concerned that studies that they themselves saw to be limited in scope and analytic value due to the limitations of both data and tools were perceived as much more general and powerful by the media or even other researchers. Critical voices have long warned that with Big Data there is a risk of over-interpretation of observed phenomena (e.g., boyd and Crawford, 2012). This observation is also closely related to reflections on the role the platform affordances play in shaping users’ actual behavior (Langlois, Redden and Elmer, 2015; Wu and Taneja, 2020) which is rarely factored into analyses and interpretation of results, as well as to discussions of ethical challenges of automatic content and user classification that may be biased toward specific user groups.

Finding ways to support researchers to reveal the complexities and limitations of concrete data preparation and analysis activities would enable them to work toward a shared understanding of the epistemic power of data and tools. First steps toward this could be releases of exemplary datasets and replication studies that put them into broader contexts or studies revealing the influence of platform affordances and their impact on user behavior. Synthetic data or sandboxes as safe spaces for data work with datasets specifically prepared for experimental purposes could allow for secure exploration of the consequences of different analyses. Such approaches can also help to better assess the consequences of specific analyses and to think through their ethical dimensions.

## PRESERVING DATA AND SECONDARY USAGE

As other studies (e.g., Hemphill et al., 2019) as well as our own (e.g., Weller and Kinder-Kurlanda, 2015) have shown, many social media researchers are willing to preserve and share collected data. However, sharing “officially” and transparently not only requires effort but also certainty about the legal and ethical limitations that apply for a specific dataset. There is hence little (but growing) evidence of official sharing (for an overview see Thomson, 2016; Acker and Kreisberg, 2019) but also an unmeasured “gray market” in which data is shared informally amongst researchers within the same group or field (Weller and Kinder-Kurlanda, 2015).

Preserving and sharing research data enables others to reuse these data for their own research. While other fields, especially social science survey research, have a long tradition of secondary data usage, the relative ease of obtaining access to (certain) types of social media data and the lack of social media data repositories accessible for researchers have led to little albeit growing secondary data usage. From an archivist’s perspective social media data sharing is difficult to accomplish due to the lack of documentation tools, the enhanced importance of privacy protection and the restrictions on sharing imposed by social media platform providers. Reproducibility and making data accessible to non-programmers requires new ways to share tools, scripts, code and documentation. Finally, the lack of consensus about what information is required to achieve reproducibility (if that is the goal) or reusability adds uncertainty to developing documentation tools and defining archiving standards. Sustainable solutions in this research phase will depend highly on consolidated efforts of different players in the research landscape, including infrastructure institutions and publishers. These efforts should also explore novel approaches for creating data access. For example, datasets could be made reusable without distributing the actual dataset but rather by submitting analysis scripts to a secure space where the data is stored, and then receiving the aggregated results.

## DISCUSSION

Social media research has in the past been criticized for being data-driven. While this criticism is legitimate, it obscures two

important things: First, whether its starting point is with theory, with a question or with data does not alone determine the quality of a piece of research. Second, our focus on individual data work activities and the way in which researchers accomplish these, making use of tools and platforms not always under their control shows that “data-driven” is not just a research design choice but the result of very complex, not necessarily well-understood constraints in the research process that “drive” research questions into specific directions. In this perspective paper we have taken a closer look at the different phases in the social media research lifecycle. We looked at how researchers “meticulously explore, ‘quibble’, test, touch, adapt, adjust, pay attention to details and change them until a suitable arrangement (material, emotional, relational) has been reached” (Mol 2015, p.111) with data, tools and questions. From the perspective of research best practice and transparency it does not matter where a specific research project starts but how the challenges posed by the complex ecosystems of platforms and data economies are navigated along the way, how open researchers are toward improving and innovating in methodology.

The social media ecosystem and its influences are already being discussed intensely in the community (e.g., Bruns, 2019). Throughout its different phases, research with social media data is embedded into a complex ecosystem of interconnections between, first, (commercial) social media service providers, platforms’ affordances that may change over time, and third-party tools and methods that support data collection, preparation and analysis, second, social media users that come from different backgrounds, have different (and evolving) usage practices and expectations on ethical usage of “their” data, and the various formats this social media data can appear in, and third, research institutions and infrastructures, from publishers, conferences, data repositories to community-sourced tools and platforms that often challenge structures built around more traditional methods. Within this setting, research is a more faceted process than simply accessing an API and working with the returned data. Research work can be divided into several phases that require iteratively going back and forth between different tasks. From a research data management perspective, research may be divided into phases of planning a research design, acquiring data, preparing and analysing data, preserving data and discovering secondary datasets. In their everyday data work, all of these phases require social media researchers to actively make decisions about how they engage with data - and these decisions have an impact on research quality and validity. However, many of these decisions are hidden—they happen unnoticed by the broader research community and often remain undocumented by the researcher.

Documentation of actual everyday research data work in its different phases and decision-making can liberate researchers from hiding the everyday messiness of working with social media data. Critically reflecting on each step and documenting also those ideas that were not pursued for good reasons can help to

uncover where researchers have the ability to actively make choices in data work, and where decisions depend on external factors that require attention from the research community as a whole.

As a guiding principle for actions aimed at improving social media research best practice by furthering transparency and data re-use, we suggest to focus on finding ways to allow for flexibility in documentation, for facilitating communication between researchers, and for safe spaces to explore and access existing datasets. Documentation requirements as presented to researchers need to be flexible both in the sense of required quantity as well as in the sense of allowing for a variety of formats such as code or screenshots to be shared. Concerning a further development of research infrastructure our perspective suggests that a greater flexibility in systems, tools and platforms that facilitate documentation, and to allow making available additional material and information about the research process may enable recording information critical to quality and validity. More easily accessible and user-friendly documentation features may also allow defining general requirements or even standards not only for documentation of social media data but also for how to accomplish the different tasks of everyday data work. Facilitating personal exchange between primary and secondary researchers allows addressing the different levels of detail required to understand research. Setting up formal and informal communication channels between primary and secondary users of a dataset could enable conversations about research design decisions and the details of data work. From a data archivist’s perspective such communications would pave the path to being able to distinguish between and eventually better describe the best practice requirements of different research methodologies. If these are combined with new ways of exploring and accessing datasets (like synthetic data for exploratory purposes or submitting analyses scripts to remotely secured datasets) this may open up new ways of understanding data work across stages in the data life cycle.

## DATA AVAILABILITY STATEMENT

The datasets analyzed in this article are not publicly available. Requests to access the datasets should be directed to [katharina.kinder-kurlanda@gesis.org](mailto:katharina.kinder-kurlanda@gesis.org).

## AUTHOR CONTRIBUTIONS

The article was written by both co-authors with equal contribution. KK-K contributed expertise in Science and Technology Studies, epistemology, data management and archiving. KW contributed expertise in social media research, information science, data management and archiving.

## REFERENCES

- Acker, A., and Kreisberg, A. (2019). Social media data archives in an API-driven world. *Arch. Sci.* 20, 105. <https://doi.org/10.1007/s10502-019-09325-9>
- Boellstorff, T. (2013). Making Big Data, in theory. *Fire Mater.* 18, 10. <https://doi.org/10.5210/fm.v18i10.4869>
- Borgman, C. L. (2019). The lives and after lives of data. *Harvard Data Sci. Rev.* 1 (1), 1–10. doi:10.1162/99608f92.9a36bdb6
- boyd, d., and Crawford, K. (2012). Critical questions for Big Data: provocations for a cultural, technological, and scholarly phenomenon. *Inf. Commun. Soc.* 15 (5), 662–679. doi:10.1080/1369118X.2012.678878
- Bruns, A. (2013). Faster than the speed of print: reconciling 'Big Data' social media analysis and academic scholarship. *Clin. Hemorheol. Microcirc.* 18, 10. doi:10.5210/fm.v18i10.4879
- Bruns, A. (2019). After the 'APocalypse': social media platforms and their fight against critical scholarly research. *Inf. Commun. Soc.* 22 (11), 1544–1566. doi:10.1080/1369118X.2019.1637447
- Bruns, A., and Weller, K. (2016). "Twitter as a first draft of the present: and the challenges of preserving it for the future," in Proceedings of the 8th ACM conference on web science, (New York, NY: ACM), 183–189. doi:10.1145/2908131.2908174
- Carlson, J. (2014). "The use of life cycle models in developing and supporting data services," in *Research data management: practical strategies for information professionals* (West Lafayette, IN: Purdue University Press), 63–86.
- Crawford, K., and Joler, V. (2018). Anatomy of an AI System-The Amazon Echo as an anatomical map of human labor, data and planetary resources. AI Now Institute and Share Lab, 7. Available at: <https://anatomyof.ai/> (Accessed July 17, 2020).
- Ekbia, H., Mattioli, M., Kouper, I., Arave, G., Ghazinejad, A., Bowman, T., et al. (2015). Big Data, bigger dilemmas: a critical review. *J. Assn. Inf. Sci. Tech.* 66 (8), 1523–1545. doi:10.1002/asi.23294
- Fiesler, C. and Proferes, N. (2018). "Participant" perceptions of twitter research ethics. *Social. Media Soc.* 4 (1). doi:10.1145/3170427.3186321
- franzke, a. s., Bechmann, A., Zimmer, M., and Ess, C. M., The Association of Internet Researchers (2020). Internet research: ethical guidelines 3.0. Available at: <https://aoir.org/reports/ethics3.pdf>.
- Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daume'e, III, H., et al. (2018). "Datasheets for datasets," in Proceedings of the 5th workshop on fairness, accountability, and transparency in machine learning, Stockholm, Sweden. Available at: <https://arxiv.org/abs/1803.09010> (Accessed October 01, 2020).
- Halford, S., Weal, M., Tinati, R., Carr, L., and Pope, C. (2018). Understanding the production and circulation of social media data: towards methodological principles and praxis. *New Media Soc.* 20 (9), 3341–3358. doi:10.1177/1461444817748953
- Hemphill, L., Hedstrom, M., and Leonard, S. (2019). How can we save social media data? Available at: <https://deepblue.lib.umich.edu/handle/2027.42/149013> (Accessed October 02, 2020).
- Kaczmarek, L., Mayr, P., Vatrupu, R., Bleier, A., Blumenberg, M. S., Gummer, T., et al. (2014). Social media monitoring of the campaigns for the 2013 German bundestag elections on Facebook and twitter. GESIS working papers 31. Available at: <https://www.ssoar.info/ssoar/handle/document/38195> (Accessed October 02, 2020).
- Karpf, D. (2012). Social science research methods in internet time. *Inf. Commun. Soc.* 15 (5), 639–661. doi:10.1080/1369118x.2012.665468
- Kinder-Kurlanda, K. E., and Weller, K. (2014). "I always feel it must be great to be a hacker!": the role of interdisciplinary work in social media research," in Proceedings of the 2014 ACM conference on Web Science (New York, NY: ACM), 91–98. doi:10.1145/2615569.2615685
- Kinder-Kurlanda, K. E., Weller, K., Zenk-Möltgen, W., Pfeffer, J., and Morstatter, F. (2017). Archiving information from geotagged tweets to promote reproducibility and comparability in social media research. *Big Data Soc.* 4 (2), 1–14. doi:10.1177/2053951717736336
- Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data Soc.* 1 (1), 1–12. doi:10.1177/2053951714528481
- Koch, G., and Kinder-Kurlanda, K. (2020). Source criticism of data platform logics on the internet. *Histor. Soc. Res.* 45 (3), 270–287.
- Langlois, G., Redden, J., and Elmer, G. (2015). "Chapter 1. introduction: compromised data — from social media to Big Data," in *Compromised. Data: from social media to big data*. (New York, London: Bloomsbury Academic), 1–13. doi:10.5040/9781501306549
- Latour, B. (1996). On actor-network theory: a few clarifications. *Soziale Welt.* 47, 369–381.
- Mayr, P., and Weller, K. (2017). "Think before you collect: setting up a data collection approach for social media studies," in *The SAGE handbook of social media research methods*. Editors S. Luke and A. Quan-Haase (SAGE Publications), 107–124.
- A. Mol, I. Moser, and J. Pols (Editors) (2015). *Care in practice: on tinkering in clinics, homes and farms* (Bielefeld, Germany: Transcript Verlag), Vol. 8.
- Olteanu, A., Castillo, C., Diaz, F., and Kiciman, E. (2019). Social data: biases, methodological pitfalls, and Ethical boundaries. *Front. Big Data* 2, 13. doi:10.3389/fdata.2019.00013
- Orlikowski, W. (2007). Sociomaterial practices: exploring technology at work. *Organ. Stud.* 28 (9), 1435–1448.
- Pouchard, L. (2015). Revisiting the data lifecycle with Big Data curation. *Int. J. Dig. Curat.* 10, 2. doi:10.2218/ijdc.v10i2.342
- Rogers, R. (2015). "Digital methods for web research," in *Emerging trends in the social and behavioral sciences* Editors R.A. Scott and S.M. Kosslyn (John Wiley & Sons Inc.). doi:10.1002/9781118900772.etrds0076
- Schroeder, R. (2014). Big Data and the brave new world of social media research. *Big Data Soc.* 1, 2. doi:10.1177/2053951714563194
- Sen, I., Flöck, F., and Wagner, C. (2020). "On the reliability and validity of detecting approval of political actors in tweets," in The conference on empirical methods in natural language processing (EMNLP 2020) (Association for Computational Linguistics (ACL)), 1413–1426.
- Sen, I., Flöck, F., Weller, K., Weiss, B., and Wagner, C. (2019). A total Error framework for digital traces of humans. arXiv [preprint]. Available at: <https://arxiv.org/abs/1907.08228> (Accessed October 02, 2020).
- Suchman, L. (1987). *Plans and situated actions. The problem of human-machine communication*. Cambridge, MA, New York, NY: Cambridge University Press.
- Thomson, S. D. (2016). Preserving social media. DPC Technology Watch Report. Available at: <http://dpconline.org/publications/technology-watch-reports> (Accessed October 02, 2020).
- Weller, K., and Kinder-Kurlanda, K. E. (2015). "Uncovering the challenges in collection, sharing and documentation: the hidden data of social media research? in standards and practices in large-scale social media research: papers from the 2015 ICWSM workshop," in Proceedings ninth international AAAI conference on web and social media Oxford University, Ann Arbor, MI, May 26–29, 2015, 28–37.
- Weller, K., and Kinder-Kurlanda, K. E. (2016). A manifesto for data sharing in social media research, in Proceedings of the 8th ACM Conference on Web Science (New York, NY: ACM), 166–172. doi:10.1145/2908131.2908172
- Weller, K., and Kinder-Kurlanda, K. E. (2017). *To share or not to share? Ethical challenges in sharing social media-based research data*, in *Internet research ethics for the social age* Editors M. Zimmer and K. E. Kinder-Kurlanda (New York, NY: Peter Lang), 115–129.
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., and Baak, A. et al. (2016). The FAIR guiding principles for scientific data management and stewardship. *Sci. Data* 3 (1), 1–9. doi:10.1038/sdata.2016.18
- Wu, A. X., and Taneja, H. (2020). Platform enclosure of human behavior and its measurement: using behavioral trace data against platform episteme. *New Media Soc.* 2020, 1–18. doi:10.1177/1461444820933547
- Zimmer, M., and Proferes, N. J. (2014). A topology of Twitter research: disciplines, methods, and ethics. *Aslib J. Inf. Manag.* 66 (3), 250–261. doi:10.1108/ajim-09-2013-0083

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Kinder-Kurlanda and Weller. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.