Journal of eScience Librarianship

Volume 10
Issue 4 *Special Issue: Research Data and Preservation (RDAP) Summit 2021*

Article 5

2021-11-10

Data Curation Implications of Qualitative Data Reuse and Big Social Research

Sara Mannheimer
Montana State University

ISSN 2161-3974

JeSLIB 2021; 10(4): e1218
https://doi.org/10.7191/jeslib.2021.1218

Journal of eScience Librarianship
putting the pieces together: theory and practice

Full-Length Paper

Data Curation Implications of Qualitative Data Reuse and Big Social Research

Sara Mannheimer

Montana State University, Bozeman, MT, USA

Abstract

Objective: Big social data (such as social media and blogs) and archived qualitative data (such as interview transcripts, field notebooks, and diaries) are similar, but their respective communities of practice are under-connected. This paper explores shared challenges in qualitative data reuse and big social research and identifies implications for data curation.

Methods: This paper uses a broad literature search and inductive coding of 300 articles relating to qualitative data reuse and big social research. The literature review produces six key challenges relating to data use and reuse that are present in both qualitative data reuse and big social research—context, data quality, data comparability, informed consent, privacy & confidentiality, and intellectual property & data ownership.

Results: This paper explores six key challenges related to data use and reuse for qualitative data and big social research and discusses their implications for data curation practices.

Correspondence: Sara Mannheimer: sara.mannheimer@montana.edu

Received: May 31, 2021 **Accepted**: September 23, 2021 **Published**: November 10, 2021

Disclosures: The author reports no conflict of interest. The substance of this article is based upon a panel presentation at RDAP Summit 2021. Additional information at end of article.

JeSLIB 2021; 10(4): e1218
https://doi.org/10.7191/jeslib.2021.1218

Abstract Continued

Conclusions: Data curators can benefit from understanding these six key challenges and examining data curation implications. Data curation implications from these challenges include strategies for: providing clear documentation; linking and combining datasets; supporting trustworthy repositories; using and advocating for metadata standards; discussing alternative consent strategies with researchers and IRBs; understanding and supporting deidentification challenges; supporting restricted access for data; creating data use agreements; supporting rights management and data licensing; developing and supporting alternative archiving strategies. Considering these data curation implications will help data curators support sounder practices for both qualitative data reuse and big social research.

Introduction

Big social data (such as social media and blogs) and archived qualitative data (such as interview transcripts, field notebooks, and diaries) are similar, but their respective communities of practice are under-connected. Research with both types of data repurpose existing data to advance discoveries in social science. However, despite these similarities, big social research has not yet been widely framed as a form of qualitative data reuse, and qualitative data reuse has only begun to be discussed through a big social data lens. This paper explores six key issues that are present in both big social research and qualitative data reuse, and outlines implications for data curation practices related to each issue. This paper suggests that by understanding shared challenges and data curation implications, these communities of practice—and the data curators who work with them—can inform each other for mutual benefit.

Background

Defining qualitative data reuse and big social research

This paper investigates the similarities between qualitative data and big social data, aiming to provide guidance for data curators to make connections between these types of data, thus enhancing our practice. This section defines qualitative data reuse and big social data research, then highlights the similarities between these definitions.

Qualitative data reuse

A key defining element of qualitative data is that they are non-numeric, although they may be analyzed to produce numeric results such as code counts and statistics (Kitchin 2014; DuBois, Strait, and Walsh 2018; Greener 2011). There are four main strategies for conducting qualitative research:

1. Unstructured, relatively open-ended interactions or information gathered from respondents, resulting in data such as solicited diaries and focus group videos.
2. Structured interviews or information solicited from respondents, resulting in data such as interview transcripts, and questionnaire responses.
3. Direct observations of behavior and environments, resulting in data such as field notes and observational records.
4. Examination of existing data such as autobiographies, found diaries, correspondence, historical documents, photographs, and home videos (Bernard et al. 1986).

The above list suggests that qualitative data can be defined by the process of creating or collecting them—that is, qualitative data are produced by qualitative research (Heaton 2004; Bernard, Wutich, and Ryan 2017).

For the purpose of this paper, taking into account definitions by Bernard (1986), Corti (1999), and Heaton (2004), I define qualitative data as follows: Qualitative data are physical objects, images, sounds, moving images, and texts that are collected and analyzed by researchers for the purpose of qualitative analysis.

The term "secondary analysis" has been used since the mid-20th century to describe a research methodology using existing data, with its earliest definitions encompassing both quantitative and qualitative data. Thorne defines qualitative secondary analysis as "the reexamination of one or more existing qualitatively derived data sets in order to pursue research questions that are distinct from those of the original inquiries" (Thorne 2004). When researchers use archived qualitative data, they repurpose previously created data, introducing new contexts, asking new research questions, and potentially gathering new data to augment the archived data. As data sharing and data publication become more common practice, the focus is not necessarily on the distinct methodology of secondary analysis, but rather on the idea of data reuse for future research of many different types. Scholars have therefore begun to increasingly use the broader term "data reuse." In 2017, Bishop and Kuula-Luumi suggest that "reuse provides an opportunity to study the raw materials of past research projects to gain methodological and substantive insights" (2017). van de Sandt et al. take a very broad view of reuse, concluding that reuse can be seen as equal to use. They define reuse as "the use of any research resource regardless of when it is used, the purpose, the characteristics of the data and its user" (2019).

Drawing on the preceding literature, this paper suggests the following working definition for qualitative data reuse:

> Qualitative data reuse is when researchers use existing qualitative data to gain new insights and produce new scholarship.

Big social research

Big social data are data derived from social media or other online environments where people share, contribute, and connect with one another. Big social data can reflect direct human interaction—usually unstructured or semi-structured data such as text, videos, and audio that are created and shared online (Olshannikova et al. 2017), or it can reflect indirect human interaction—usually structured metadata that reflects user behavior such as interactions with interfaces, or the spatial or temporal aspects of user behavior (Gandomi and Haider 2015).

Big social data can come in several formats:

- Digital self-representation data: Login data, profile pictures, biographical information
- Social interaction data: timeline posts, online forum posts, content sharing, commenting, direct messaging

- Digital relationships data: Follower/following data, "likes"
- Metadata: Timestamps, geospatial data, type of operating system, type of device, application used to post (Adapted from Olshannikova et al. 2017)

It is possible to use social media to recruit participants—conducting online ethnographies or directly contacting interview subjects via social media. However, this paper focuses on the use of big social data that is available online through web scraping, API access, or other methods that don't require direct contact with individual people.

Big social data research is most often conducted using computational social science methods. Computational social science blends theory and practice from computer science, statistics, and the social sciences, using computational methods to conduct research inquiry about society (Mason, Vaughan, and Wallach 2014, 257). Computational social science began in the 2000s, and uses methods such as topic modeling, sentiment analysis, network analysis, artificial intelligence, and deep learning techniques to support drawing conclusions from large corpora of text (Bankes, Lempert, and Popper 2002; Berkout, Cathey, and Kellum 2019).

Drawing from the preceding literature, this paper suggests the following working definition for big social research:

> Big social research is when researchers collect existing data from social media or other online social environments to gain insights and produce scholarship.

Qualitative data reuse and big social research: the connection

As illustrated above, qualitative data reuse and big social research are distinct in terms of data sources and methods of data analysis, but the two types of data also share key similarities that have implications for data curation. This paper draws upon the above definitions of qualitative data reuse and big social research:

> **Qualitative data reuse** is when researchers use existing qualitative data to gain new insights and produce new scholarship.
>
> **Big social research** is when researchers use existing data from social media or other online social environments to gain insights and produce scholarship.

These definitions help to illustrate the connection between qualitative data reuse and big social research. Both types of scholarship take data that has been created for one purpose, then repurpose the data to gain insights and produce scholarship. This paper highlights this similarity in its discussion of shared challenges.

Methods

Using the methods outlined by Creswell (2009) and in more detail in the Handbook of Research Synthesis and Meta-analysis (H. M. Cooper, Hedges, and Valentine 2019), I conducted an inductive research synthesis of the literature on qualitative data reuse and big social data research. The research synthesis consisted of the following steps: literature search, data evaluation, data analysis, and interpretation of results (H. M. Cooper, Hedges, and Valentine 2019).

Search and selection

For the literature search, I searched the library catalog and online databases using the following strings:

- "qualitative secondary analysis"
- "qualitative data reuse"
- "qualitative data archiving"
- "social media data"
- "social media data archiving"
- "big social data"

While reviewing initial articles, I identified further reading through backward and forward citation chaining (C. Cooper et al. 2017; Hu, Rousseau, and Chen 2011), a process of reviewing literature that have been cited in a particular article, as well as reviewing literature that cites that particular article. Articles were limited to those published in English.

I organized and coded approximately 300 articles. Publication dates ranged from 1934 to present, with most articles occurring in the past 30 years for qualitative data reuse, and the past 20 years for big social research.

Coding

I coded each article according to key themes, inductively creating the themes using Grounded Theory's constant comparative method (Glaser and Strauss 1967). My coding focused on (1) research objectives and methods; (2) discussions of theory, including epistemological and ethical issues; and (3) data curation practices. I focused on common themes between the literature on qualitative data reuse and the literature on big social data. Six central issues emerged in common between qualitative data reuse and big social data research—context, data comparability, data quality, informed consent, privacy and confidentiality, and intellectual property.

JeSLIB 2021; 10(4): e1218
https://doi.org/10.7191/jeslib.2021.1218

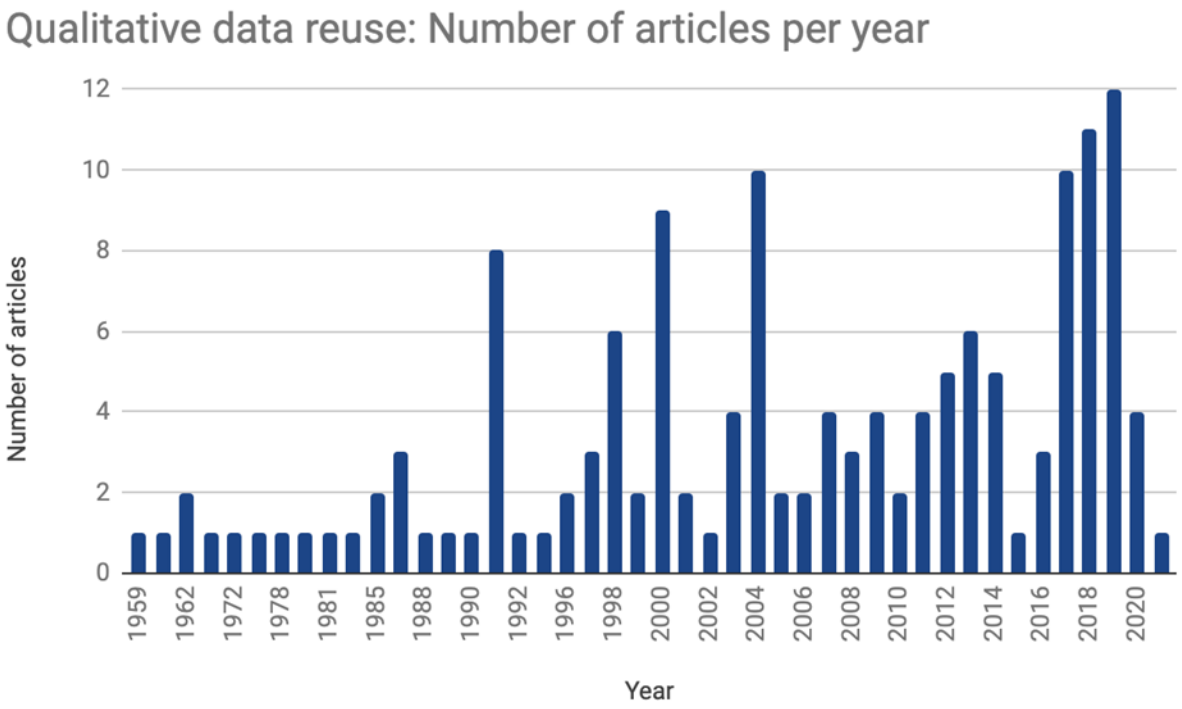

Figure 1: Qualitative data reuse: Number of articles per year

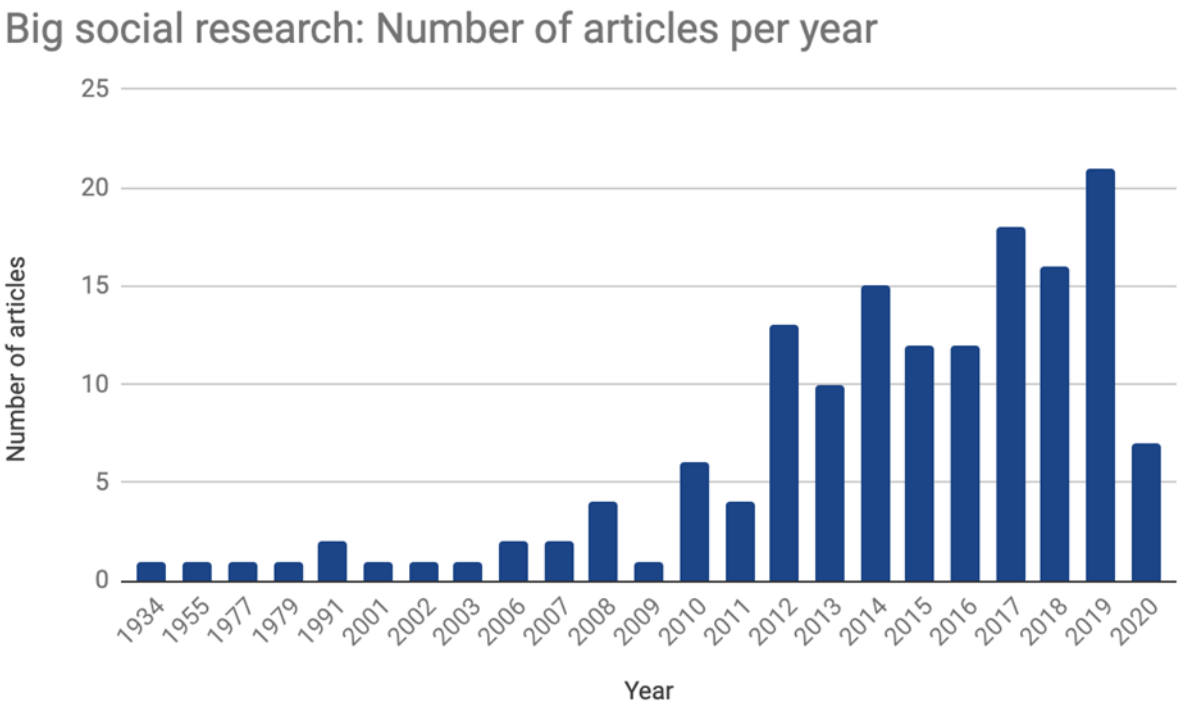

Figure 2: Big social research: Number of articles per year

Benefits of data sharing

A key idea running through the literature is the emerging consensus that data sharing is beneficial to science and society. Benefits of data sharing can be grouped into three key categories: scientific, moral, and economic benefits (Mauthner 2012). Scientific benefits include building new knowledge, new hypotheses, new methodologies, comparative research, or strengthening existing theories; promoting interdisciplinary use of data; increasing citations and scholarly impact; and providing data for the purpose of teaching students. Moral benefits include reducing burden on research subjects; facilitating more research about rare, hard-to-reach, or inaccessible respondents; and supporting transparency and accountability—in order to foster trust from the public and other researchers and to share the results of public research funding. Economic benefits include conserving time and resources, therefore supporting a higher return on investment. Each of these benefits have been further discussed in the literature (e.g., Piwowar et al. 2008; Logan, Hart, and Schatschneider 2021; Levenstein and Lyle 2018; Fienberg, Martin, and Straf 1985).

Challenges and implications for data curation

This paper discusses six key challenges relating to data use and reuse that are present in both qualitative data reuse and big social data research, and discusses data curation implications.

Below, I describe each of the challenges. For each challenge, I also outline data curation implications that are discussed in the literature.

Context

Issues of context are similar for both qualitative data reuse and big social research. For both types of research, there is concern that data may not be able to be properly understood outside of their original context.

When considering reuse of qualitative data, concerns center around whether data can be meaningful without the knowledge and expertise of the researchers who conducted the original research project. As Pasquetto, Borgman, and Wofford write, "removing data from their original context necessarily involves information loss" (2019, 23). Such loss includes small adjustments that may be made to the data during research, deep knowledge of the research that data creators hold but may not be able to communicate in a dataset description, and the de-contextualizing effects of deidentification efforts (Mauthner, Parry, and Backett-Milburn 1998; Fielding and Fielding 2000; Dale, Arber, and Procter 1988).

When conducting big social research, data often take the form of photos, videos, or short pieces of text, drawn from a larger context of personal and public life (Törnberg and Törnberg 2018; Boyd and Crawford 2012). This out-of-context effect is only compounded when data are amassed at a large scale. For big social

research, the researcher may never speak to the people who created the data, know their identities, or be aware of other broader contexts. Marwick and Boyd (2011) also refer to a "context collapse" in big social research, in which multiple audiences are flattened into one, making the context and viewpoint of big social research difficult to discern—to whom is a user speaking when they post on social media? This context collapse can also apply to archived qualitative data—while the original audience and context are generally more concrete, when qualitative data are published openly, the future audience is unknown.

For both big social research and qualitative data reuse, the literature suggests that full context and meaning may never be accurately understood by qualitative data reusers/big data researchers. However, using data curation strategies to communicate as much context as possible can help support meaningful data use and reuse.

Context: Data curation implications

Clear documentation

- For qualitative data: Data curators can encourage contextual documentation throughout the research process, to be published alongside qualitative data. This could include documentation about research methods and practices, consent form, IRB approval numbers information about the selection of interview subjects and interview setting, instructions given to interviewers, data collection instruments, steps taken to remove direct identifiers in the data, problems that arose during the selection and/or interview process and how they were handled, and interview roster (ICPSR 2012).

- For big social research: Data curators can encourage as much documentation as possible of the methods, communities, and platforms. Context can also be communicated through metadata such as geolocation, @-mentions, or hashtags.

- Initiatives such as Annotation for Transparent Inquiry (Karcher and Weber 2019), Open Context (Kansa and Kansa 2018), and the Data Curation Network (Johnston et al. 2018) all support researchers and data repositories in creating documentation to encourage contextual integrity for data reuse.

Archiving related data

- Repositories may also choose to archive (or link to archived versions of) web URLs, images, and other resources (Thomson and Beagrie 2016).

Data quality and trustworthiness

While challenges related to data quality exist for both qualitative data reuse and big social research, the challenges are relatively distinct for each type of data. For

qualitative data, quality issues often relate to human error. Humans throughout the process could introduce errors through simple mistakes and inaccuracies. Errors can come at various stages in the research—from research subjects, reporters or recorders of field data, researchers, and data coders (Sherif 2018).

Data quality issues for big social research have additional complexities that introduce different types of errors. Because of the automated nature of data collection and analysis, there are fewer opportunities for simple mistakes in these phases. However, quality issues can result from the element of self-performance that is more present in big social research—users are not speaking directly to the researcher, but rather to a perceived online community (Hogan 2010; Manovich 2012). Other quality issues can result from the specific environment of online social platforms. Multiple accounts from one user, fake accounts, and bots can all introduce errors, bias, and distortion (Marwick and Boyd 2011; Shah, Cappella, and Neuman 2015; Varol et al. 2017). Additionally, big social data sampling is often biased because social media APIs may not return complete data, and users of social media platforms may not be representative of society as a whole (Burgess and Bruns 2012), and some social media platforms such as Facebook and Twitter tend to be overrepresented in big social research due to ease of access (Zimmer and Proferes 2014; Stoycheff et al. 2017)

For both types of data, systematic errors can be introduced as a result of bias, and when scaling up by reusing qualitative data, combining datasets, or collecting big social data, these errors can compound (Bernard et al. 1986; Morstatter and Liu 2017; M. Hammersley and Gomm 1997; Hargittai 2015). While data curation is not a simple solution to these challenges, clear documentation, use of trustworthy repositories, and linking to related datasets are all discussed in the literature as strategies to support data quality and trustworthiness.

Data quality and trustworthiness: Data curation implications

Clear documentation

- Data curators can support documentation of the research process when sharing data, including documenting any potential bias, errors, or missing data.

Trustworthy repositories

- Data repositories and academic libraries can contribute to data quality and trustworthiness by supporting data management, curation, and metadata (Frank et al. 2017; Giarlo 2013; Yoon and Lee 2019).

- Trust in data can be enhanced by trust in the repository where it is archived. To support healthy infrastructure and long-term preservation for repositories, initiatives such as the CoreTrustSeal Trustworthy Data Repositories Requirements help repositories meet community standards for

data stewardship (CoreTrustSeal 2020).

- Data curators may also refer to the recently developed TRUST Principles, which are designed to complement the FAIR Principles to support trustworthy data stewardship for archived data (Lin et al. 2020).

Related and combined datasets

- Some researchers have attempted to create more representative datasets by blending big social data with smaller social datasets, a strategy that helps include a broader range of perspectives than are present in a single dataset (Croeser and Highfield 2020). Data curators could provide links between related datasets to support future use. However, combining datasets comes with its own set of challenges (see Data comparability, below).

Data comparability

When combining qualitative and big social datasets, researchers must determine whether each dataset can be understood to be applicable to another—also referred to as data "fit." Because qualitative research tends to produce data sets that are relatively unstructured, complex, and heterogenous (Heaton 2004), it can be difficult to combine multiple qualitative datasets. Researchers can assess the comparability (or "fit") of the data by (1) identifying the extent of missing data; (2) identifying convergence of primary and secondary research questions; (3) assessing the methods used to produce the primary data (Heaton 2004; Hinds, Vogel, and Clarke-Steffen 1997; Thorne 1994).

Comparability of big social data is additionally affected by the issue of metadata interoperability. While standardized metadata such as Data Documentation Initiative (DDI Alliance 2019) are fairly commonly used for qualitative data, big social datasets have less standardized metadata. Social media platforms may use different metadata schemas, and it can be difficult and time-consuming to combine multiple big social datasets if the metadata are not interoperable. As Acker and Kriesberg note, "there are no data models for cross-walking or mapping like-with-like across platforms, for example a tweet, a Facebook post and a YouTube video that all link to the same content or event such as a townhall livefeed" (2017, 7). While the proprietary nature of many social media platforms may continue to impede metadata interoperability, there are some models for unified metadata schemas such as Schema.org (W3 2021) that could inspire similar community efforts for social media.

Comparability is an especially important issue for both qualitative data reuse and big social research. For both types of data, combining multiple datasets can support larger-scale studies, which is a particular focus for qualitative data, but can apply to both; combining data can be used as a strategy to better understand context and to enhance data quality, which is a particular focus for big social

research, but can apply to both (see Data quality, above). The literature suggests that data curators can support data comparability by helping researchers create clear documentation, and by advocating for interoperable metadata standards.

Data comparability: Data curation implications

Clear documentation

- For both qualitative data reuse and big social research, data curators can support comparability by encouraging researchers who publish data to include clear documentation to address missing data, research questions, and methods.

Metadata standards

- For both types of data, data curators can adapt existing standards such as DDI to support better data comparability (DDI Alliance 2019).

- The research and data curation communities can advocate for interoperable metadata standards that can be adopted by social media platforms themselves, potentially including existing models such as Schema.org metadata (W3 2021).

Informed consent

The issue of informed consent is similar with qualitative data reuse and big social research. In the case of shared qualitative data, some researchers are now including consent for data sharing and archiving in consent agreements. In fact, the 2019 revision of the Common Rule includes the idea of broad consent, in which participants agree to "future storage, maintenance, or research uses" of their data (U.S. Department of Health and Human Services 2017), and some IRBs now suggest language to support data reuse (Cornell Research Services 2019; Elman, Kapiszewski, and Lupia 2018). However, broad consent is not a perfect solution, especially when viewed through the lens of feminist and post-colonial theories, which consider power structures between researchers and research subjects. There is concern that broad consent could expose respondents to risk and reduce their agency, since the data may be used to ask any number of future research questions (Mauthner and Parry 2013). Tiered consent models could provide a middle ground, supporting more granular consent options than broad consent. In the tiered consent model, participants are given choices about the specifics of data sharing. For instance, a consent form could allow participants to opt out of sharing any of their data—while still participating in the study; the consent form could give participants the option to share only a subset of their data; or the consent form could allow participants to share their data only with reusers who meet certain criteria (Meyer 2018).

In the case of big social research, social media terms of service may include user

agreements that address data availability for research purposes. However, users generally don't read terms of service (Obar and Oeldorf-Hirsch 2020), and even if they do, they are not informed of the nature and extent of research that may be conducted with their data. The U.S. Health and Human Services' Secretary's Advisory Committee on Human Research Protections suggested in 2015 that guidance should be developed regarding consent standards for big data research, including methods such as focus groups or community advisory boards that could help big data researchers identify representative concerns of participant populations (Secretary's Advisory Committee on Human Research Protections 2015). However, such guidance is not codified in the Common Rule. Some have suggested that IRBs should review big social research even it is not yet mandated by law (Schneble, Elger, and Shaw 2018). In practice, most big social research is classified as exempt by IRBs (Metcalf 2016).

Some projects have developed technology-mediated strategies to address the issue of consent for big social research. Two examples are pop-up messages gauging participants' willingness to share certain types of data on Facebook (Hutton and Henderson 2013), and software that provides structures to "ask participants (as normal procedure within qualitative and quantitative studies) if the researcher may retrieve and use the data in a specific research project" (Bechmann and Vahlstrup 2015). However, these strategies are rarely used. Such strategies are also made more difficult by the large scale and networked nature of big social data. For example, even if one user consents to their social media posts being used for research purposes, they may @-mention other members of their network or link to other profile or group pages; these other users would therefore be part of the research dataset, without having consented to the research (Mneimneh et al. 2021).

The literature suggests that if data curators can reach investigators early in the research process, they can help provide guidance for alternative consent strategies for qualitative data reuse and big social research.

Informed consent: Data curation implications

Alternative consent strategies for qualitative data reuse

- If data curators can connect with researchers early in the research process, they can help researchers draft broad consent language to support data reuse (Kirilova and Karcher 2017).

- Researchers, curators, and IRBs can also work together to support tiered consent models, allowing research participants to select the level of data sharing with which they are comfortable.

Alternative consent strategies for big social research

- If data curators can connect with researchers early in the research process,

they can encourage strategies such as focus groups, community advisory boards, or software-supported strategies for obtaining individual informed consent within social media platforms.

Privacy and confidentiality

While privacy is a major issue for both qualitative data reuse and big social research, some specific elements of these concerns are distinct between the two types of data.

For qualitative data reuse, deidentification strategies are used to support data sharing. However, some argue that deidentification may compromise the integrity and quality of the data or remove important contextual information (Fielding 2004; Martyn Hammersley 1997; Stenbacka 2001). Moreover, deidentification may not be guaranteed to prevent deductive disclosure based on other contextual information—exactly the kind of contextual information that is necessary to understand and reuse the data in the first place (Mauthner, Parry, and Backett-Milburn 1998; Tsai et al. 2016). Qualitative researchers often study sensitive issues such as domestic abuse, substance use, and sexual practices (DuBois, Strait, and Walsh 2018); reidentification of such data could lead to additional social or physical harm for participants.

For big social data, some researchers argue that such data are public by nature, and deidentification is therefore unnecessary (Zimmer 2010; Wilkinson and Thelwall 2011). For example, in 2016, researchers scraped profiles from the online dating service OkCupid and released the data without any attempt at deidentification (Kirkegaard and Bjerrekær 2016), asserting that the data were "already public" and required no special privacy considerations or user consent (Zimmer 2016). However, researchers are increasingly considering privacy when using big social data. Nissenbaum's theory of contextual integrity (2009), which suggests that expectations of privacy are context-dependent, has been widely used to understand privacy online. The literature suggests that people's strategies for protecting their privacy online are constantly changing and adapting, depending on a variety of factors, including physical environment, perceived audience, social status, motivation, and technologies or social media platforms in use (Palen and Dourish 2003). The idea of contextual integrity can explain why users might be fine with publicly sharing information in one context, but feel more protective of that same information in a different context (Reuter et al. 2019).

Even if researchers intend to deidentify shared big social data, the practice of deidentification may be difficult (Zimmer 2010; Schneble, Elger, and Shaw 2018). Comparing the identifiability of traditional qualitative research with that of big social research, Chu et al. point out that while it is common in qualitative studies to directly quote respondents in order to support key findings and highlight ideas of interest, the full-text indexing of social media platforms may cause any direct quotes to be easily identifiable (Chu et al. 2019).

For both qualitative data reuse and big social research, privacy should be more carefully considered when the research involves vulnerable populations (Clark et al. 2018), for whom reidentification could be especially damaging. In 1991, Sieber wrote that surveillance "is not a legitimate use of shared data and may be damaging to science" (Sieber 1991, 148). However, the intervening decades have seen a rise in technology-mediated surveillance. In the case of big social data, advertisers track social media user activities (Oboler, Welsh, and Cruz 2012), employers review the online presence of potential hires (Duffy and Chan 2019), and social media may be used by law enforcement for surveillance purposes (Jules, Summers, and Mitchell 2018). In the European Union, the General Data Protection Regulation (GDPR) went into effect in 2018 and includes the "right to be forgotten"—that is the opportunity for internet users to request their data be removed from online spaces (Voigt and von dem Bussche 2017). While the GDPR is a step forward for ethical online data practices, the ramifications for big social research are still not fully clear (Greene et al. 2019; Vestoso 2018).

To address some of the privacy challenges reviewed above, data curation and data repository services have been developed to provide deidentification support, restricted data access, and data use agreements.

Privacy & confidentiality: Data curation implications

<u>De-identification procedures</u>

- Data curators can support deidentification procedures such as deleting names or replacing with pseudonyms; removing potentially identifying details about participants' lives and experiences; amalgamating or aggregating data.

<u>Restricted access</u>

- Data repositories may support data embargo for a period of time or restrict access to the data.

<u>Data use agreements</u>

- Data curators and repositories can provide customizable data use agreements that dictate the conditions required for other researchers to access and reuse the data. The data use agreement includes terms that the user must agree to follow if they download the data. For example, the agreement may stipulate that the data be used for academic research purposes, that the research be approved by an institutional review board, or that the researcher not attempt to reidentify the data (ICPSR 2018; QDR 2019).

Intellectual property and data ownership

Qualitative data are the shared intellectual property of the research participants and the researchers. For researchers to publish the text of participant responses, participants must either waive their rights or license their responses for use in the research study (Parry and Mauthner 2004). Participants may agree to data publication when signing the consent agreement; however, if the consent agreement did not specifically include data publication and reuse, publishing the data may not be allowable. In some cases, contacting participants for re-consent may be possible (Resnik 2009). Some also suggest that if data are sufficiently deidentified, it may be ethical to publish data without explicit consent from participants (DuBois, Strait, and Walsh 2018).

While universities generally claim ownership over research data created by affiliated researchers (Steneck 2007), strategies for addressing intellectual property and data ownership may vary according to how and where the data were collected. For example, when collecting data from Indigenous communities, additional considerations and guidelines come into play. Communities who participate in research are increasingly contributing to the development of protocols that inform the ethical use of data, "allowing contributors, as collectives, to have a say in how their data actually gets used" (Carroll et al. 2021). The CARE Principles for Indigenous Data Governance (Research Data Alliance International Indigenous Data Sovereignty Interest Group 2019) and The First Nations Principles of OCAP® (FNIGC 2010) provide guidance for supporting responsible data stewardship when conducting research with Indigenous communities.

Big social data sharing is made more complex by the fact that these data are often controlled by private, for-profit companies. In 2018, Facebook CEO Mark Zuckerberg testified before Congress, saying, "every piece of content that you share on Facebook, you own, and you have complete control over who sees it and—and how you share it, and you can remove it at any time" (Washington Post 2018). However, under United States law, intellectual property on social media is still a gray area (Blank 2018; Bosher and Yeşiloğlu 2019). Even if the contents of social media posts are the intellectual property of the users who posted them, social media companies may still implement terms of service that govern the behavior of users, developers, researchers, and archivists (Puschmann and Burgess 2014). Some social media companies have tried to prevent web scraping on their sites by invoking the Computer Fraud and Abuse Act (Neuburger 2020), thus far unsuccessfully. Social media terms of service may also prevent sharing big social data in the same manner as other research data. One example of data sharing restrictions is the case of Twitter, whose Terms of Service dictate that only Tweet ID numbers may be openly shared. In response, tools have been developed such as Documenting the Now's Hydrator tool, which uses the Twitter API to pull complete metadata for shared Tweet IDs (Summers 2017).

Data curators can support intellectual property challenges through rights management guidance, data licensing, and alternative archiving strategies.

Intellectual property: Data curation implications

Rights management for both big social research and qualitative data reuse

- Data curators and data repositories can help researchers with rights management—understanding how they can and cannot reuse shared data.

- For big social research, data curators can help researchers navigate terms of service to collect, archive, and share data in accordance with these terms.

Data licensing for qualitative data

- For qualitative data, data curators can encourage researchers to consider data licensing as part of initial consent agreements, and again at the point of data archiving and sharing.

Alternative archiving strategies for big social data

- If raw data cannot be archived, data repositories can archive associated information such as data workflows and code that can allow future users to replicate the data collection and analysis process (Hemphill, Leonard, and Hedstrom 2018).

- Data repositories maybe able to archive representative metadata such as lists of TweetIDs.

- Data curators can encourage inclusion of tools such as the Twitter Hydrator as part of the data deposit, to support usability for the archived data (Kinder-Kurlanda et al. 2017).

Conclusions and Future Research

Big social research and qualitative data reuse both have the potential to reveal large-scale insights about human behavior. However, epistemological, ethical, and legal challenges arise when reusing qualitative data, conducting research with big social data, and sharing or archiving big social data. This paper outlines six key challenges gleaned from the literature: context, data quality and trustworthiness, data comparability, informed consent, privacy and confidentiality, and intellectual property. Data curators can benefit from understanding these six key challenges and examining data curation implications. Data curation implications from these challenges include developing strategies for: providing clear documentation; linking and combining datasets; supporting trustworthy repositories; using and advocating for metadata standards; discussing alternative consent strategies with researchers and IRBs; understanding and supporting deidentification challenges; supporting restricted access for data; creating data use agreements; supporting rights management and data licensing; developing and supporting alternative

archiving strategies. These data curation practices can help mitigate some of the challenges that are present with both data types. Future research could be done interviewing qualitative researchers, big social researchers, and data curators to verify and further investigate the challenges that have been discussed here, and to support data curation strategies that can support shared challenges. By investigating issues in qualitative data reuse and big social research side by side, data curation practices can be extended to support sounder practices for both qualitative data and big social research.

Acknowledgments

Many thanks to Vivien Petras at Humboldt University of Berlin for her guidance and support on this paper.

Disclosures

The content of this article is based upon a panel presentation at RDAP Summit 2021 titled "Supporting Responsible Research with Big Social Data by Connecting Communities of Practice" available at https://osf.io/e4u7v.

References

Acker, Amelia, and Adam Kriesberg. 2017. "Tweets May Be Archived: Civic Engagement, Digital Preservation and Obama White House Social Media Data." Proceedings of the *Association for Information Science and Technology* 54(1): 1–9. https://doi.org/10.1002/pra2.2017.14505401001

Bankes, Steven, Robert Lempert, and Steven Popper. 2002. "Making Computational Social Science Effective: Epistemology, Methodology, and Technology." *Social Science Computer Review* 20(4): 377–388. https://doi.org/10.1177/089443902237317

Bechmann, Anja, and Peter Bjerregaard Vahlstrup. 2015. "Studying Facebook and Instagram Data: The Digital Footprints Software." *First Monday* 20(12). https://doi.org/10.5210/fm.v20i12.5968

Berkout, Olga V., Angela J. Cathey, and Karen Kate Kellum. 2019. "Scaling-up Assessment from a Contextual Behavioral Science Perspective: Potential Uses of Technology for Analysis of Unstructured Text Data." *Journal of Contextual Behavioral Science* 12(April): 216–224. https://doi.org/10.1016/j.jcbs.2018.06.007

Bernard, H. Russell, Pertti J. Pelto, Oswald Werner, James Boster, A. Kimball Romney, Allen Johnson, Carol R. Ember, and Alice Kasakoff. 1986. "The Construction of Primary Data in Cultural Anthropology." *Current Anthropology* 27(4): 382–396. https://doi.org/10.1086/203456

Bernard, H. Russell, Amber Wutich, and Gery W. Ryan. 2017. *Analyzing Qualitative Data: Systematic Approaches*. Second edition. Los Angeles: SAGE.

Bishop, Libby, and Arja Kuula-Luumi. 2017. "Revisiting Qualitative Data Reuse: A Decade On." *SAGE Open* 7(1): 2158244016685136. https://doi.org/10.1177/2158244016685136

Blank, Jeff. 2018. "IP Law in the Age of Social Media." Northeastern University Graduate Programs (blog). May 8, 2018. https://www.northeastern.edu/graduate/blog/intellectual-property-and-social-media

Bosher, H., and S. Yeşiloğlu. 2019. "An Analysis of the Fundamental Tensions between Copyright and Social Media: The Legal Implications of Sharing Images on Instagram." *International Review of Law, Computers & Technology* 33(2): 164–186. https://doi.org/10.1080/13600869.2018.1475897

Boyd, Danah, and Kate Crawford. 2012. "Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon." *Information, Communication & Society* 15(5): 662–679. https://doi.org/10.1080/1369118X.2012.678878

Burgess, Jean, and Axel Bruns. 2012. "Twitter Archives and the Challenges of 'Big Social Data' for Media and Communication Research." *M/C Journal* 15(5). http://journal.media-culture.org.au/index.php/mcjournal/article/view/561

Carroll, Stephanie Russo, Edit Herczog, Maui Hudson, Keith Russell, and Shelley Stall. 2021. "Operationalizing the CARE and FAIR Principles for Indigenous Data Futures." *Scientific Data* 8(1): 108. https://doi.org/10.1038/s41597-021-00892-0

Chu, Kar-Hai, Jason Colditz, Jaime Sidani, Michael Zimmer, and Brian Primack. 2019. "Re-Evaluating Standards of Human Subjects Protection for Sensitive Health Data in Social Media Networks." *Social Networks* November. https://doi.org/10.1016/j.socnet.2019.10.010

Clark, Karin, Matt Duckham, Marilys Guillemin, Assunta Hunter, Jodie McVernon, Christine O'Keefe, Cathy Pitkin, et al. 2018. "Advancing the Ethical Use of Digital Data in Human Research: Challenges and Strategies to Promote Ethical Practice." *Ethics and Information Technology* November. https://doi.org/10.1007/s10676-018-9490-4

Cooper, Chris, Andrew Booth, Nicky Britten, and Ruth Garside. 2017. "A Comparison of Results of Empirical Studies of Supplementary Search Techniques and Recommendations in Review Methodology Handbooks: A Methodological Review." *Systematic Reviews* 6(1): 234. https://doi.org/10.1186/s13643-017-0625-1

Cooper, Harris M., Larry V. Hedges, and Jeff C. Valentine, eds. 2019. *Handbook of Research Synthesis and Meta-Analysis*. 3rd edition. New York: Russell Sage Foundation.

CoreTrustSeal. 2020. "Core Trustworthy Data Repositories Requirements." 2020. https://web.archive.org/web/20200408004456/https://www.coretrustseal.org/why-certification/requirements

Cornell Research Services. 2019. "IRB Consent Form Templates." 2019. https://researchservices.cornell.edu/forms/irb-consent-form-templates

Corti, Louise. 1999. "Text, Sound and Videotape: The Future of Qualitative Data in the Global Network." *IASSIST Quarterly* 23(2): 15. https://doi.org/10.29173/iq726

Creswell, John W. 2009. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. 3rd ed. Thousand Oaks, Calif: Sage Publications.

Croeser, Sky, and Tim Highfield. 2020. "Blended Data: Critiquing and Complementing Social Media Datasets, Big and Small." In *Second International Handbook of Internet Research*, edited by Jeremy Hunsinger, Matthew M. Allen, and Lisbeth Klastrup, 669–690. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-024-1555-1_15

Dale, Angela, Sara Arber, and Michael Procter. 1988. *Doing Secondary Analysis*. Crows Nest, Australia: Allen & Unwin.

DDI Alliance. 2019. "Data Documentation Initiative." 2019. https://ddialliance.org

DuBois, James M., Michelle Strait, and Heidi Walsh. 2018. "Is It Time to Share Qualitative Research Data?" *Qualitative Psychology* 5(3): 380–393. https://doi.org/10.1037/qup0000076

Duffy, Brooke Erin, and Ngai Keung Chan. 2019. "'You Never Really Know Who's Looking': Imagined Surveillance across Social Media Platforms." *New Media & Society* 21(1): 119–138. https://doi.org/10.1177/1461444818791318

Elman, Colin, Diana Kapiszewski, and Arthur Lupia. 2018. "Transparent Social Inquiry: Implications for Political Science." *Annual Review of Political Science* 21(1): 29–47. https://doi.org/10.1146/annurev-polisci-091515-025429

Fielding, Nigel. 2004. "Getting the Most from Archived Qualitative Data: Epistemological, Practical and Professional Obstacles." *International Journal of Social Research Methodology* 7(1): 97–104. https://doi.org/10.1080/1364557032000164069 9

Fielding, Nigel, and Jane L. Fielding. 2000. "Resistance and Adaptation to Criminal Identity: Using Secondary Analysis to Evaluate Classic Studies of Crime and Deviance." *Sociology* 34(4): 671–689. https://doi.org/10.1177/S0038038500000419

Fienberg, Stephen E., Margaret E. Martin, and Miron L. Straf, eds. 1985. Sharing Research Data. Washington, DC: The National Academies Press. https://doi.org/10.17226/2033

FNIGC. 2010. "The First Nations Principles of OCAP®, a Registered Trademark of the First Nations Information Governance Centre (FNIGC)." Akwesasne, ON: First Nations Information Governance Centre. https://fnigc.ca/ocap-training

Frank, Rebecca D., Zui Chen, Erica Crawford, Kara Suzuka, and Elizabeth Yakel. 2017. "Trust in Qualitative Data Repositories." Proceedings of the *Association for Information Science and Technology* 54(1): 102–111. https://doi.org/10.1002/pra2.2017.14505401012

Gandomi, Amir, and Murtaza Haider. 2015. "Beyond the Hype: Big Data Concepts, Methods, and Analytics." *International Journal of Information Management* 35(2): 137–144. https://doi.org/10.1016/j.ijinfomgt.2014.10.007

Giarlo, Michael. 2013. "Academic Libraries as Data Quality Hubs." *Journal of Librarianship and Scholarly Communication* 1(3): eP1059. https://doi.org/10.7710/2162-3309.1059

Glaser, Barney G., and Anselm L. Strauss. 1967. *Discovery of Grounded Theory: Strategies for Qualitative Research*. Hawthorne, NY: Aldine.

Greene, Travis, Galit Shmueli, Soumya Ray, and Jan Fell. 2019. "Adjusting to the GDPR: The Impact on Data Scientists and Behavioral Researchers." *Big Data* 7(3): 140–162. https://doi.org/10.1089/big.2018.0176

Greener, Ian. 2011. *Designing Social Research: A Guide for the Bewildered*. 1 Oliver's Yard, 55 City Road, London EC1Y 1SP United Kingdom: SAGE Publications Ltd. https://doi.org/10.4135/9781446287934

Hammersley, M., and R. Gomm. 1997. "Bias in Social Research." *Sociological Research Online* 2(1): 1–13. https://doi.org/10.5153/sro.55

Hammersley, Martyn. 1997. "Qualitative Data Archiving: Some Reflections on Its Prospects and Problems." *Sociology* 31(1): 131–142. https://doi.org/10.1177/0038038597031001010

Hargittai, Eszter. 2015. "Is Bigger Always Better? Potential Biases of Big Data Derived from Social Network Sites." *The ANNALS of the American Academy of Political and Social Science* 659(1): 63–76. https://doi.org/10.1177/0002716215570866

Heaton, Janet. 2004. *Reworking Qualitative Data*. London: SAGE Publications.

Hemphill, Libby, Susan H Leonard, and Margaret Hedstrom. 2018. "Developing a Social Media Archive at ICPSR." In Proceedings of *Web Archiving and Digital Libraries* (WADL'18). New York: ACM. https://hdl.handle.net/2027.42/143185

Hinds, Pamela S., Ralph J. Vogel, and Laura Clarke-Steffen. 1997. "The Possibilities and Pitfalls of Doing a Secondary Analysis of a Qualitative Data Set." *Qualitative Health Research* 7(3): 408–424. https://doi.org/10.1177/104973239700700306

Hogan, Bernie. 2010. "The Presentation of Self in the Age of Social Media: Distinguishing Performances and Exhibitions Online." *Bulletin of Science, Technology & Society* 30(6): 377–386. https://doi.org/10.1177/0270467610385893

Hu, Xiaojun, Ronald Rousseau, and Jin Chen. 2011. "On the Definition of Forward and Backward Citation Generations." *Journal of Informetrics* 5(1): 27–36. https://doi.org/10.1016/j.joi.2010.07.004

Hutton, Luke, and Tristan Henderson. 2013. "An Architecture for Ethical and Privacy-Sensitive Social Network Experiments." *SIGMETRICS Performance Evaluation Review* 40(4): 90–95. https://doi.org/10.1145/2479942.2479954

ICPSR. 2012. "Guide to Social Science Data Preparation and Archiving: Introduction." 2012. http://www.icpsr.umich.edu/files/deposit/dataprep.pdf

———. 2018. "Restricted Data Use Agreement for Restricted Data from the Inter-University Consortium for Political and Social Research (ICPSR)." University of Michigan: Inter-university Consortium for Political and Social Research (ICPSR). https://www.icpsr.umich.edu/files/ICPSR/pdf/ICPSRRestrictedDataUseAgreement_2018.pdf

Johnston, Lisa R., Jake Carlson, Cynthia Hudson-Vitale, Heidi Imker, Wendy Kozlowski, Robert Olendorf, Claire Stewart, et al. 2018. "Data Curation Network: A Cross-Institutional Staffing Model for Curating Research Data." *International Journal of Digital Curation* 13(December): 125–140. https://doi.org/10.2218/ijdc.v13i1.616

Jules, Bergis, Ed Summers, and Vernon Jr. Mitchell. 2018. "Ethical Considerations for Archiving Social Media Content Generated by Contemporary Social Movements: Challenges, Opportunities, and Recommendations." *Documenting the Now White Paper*. https://www.docnow.io/docs/docnow-whitepaper-2018.pdf

Kansa, Sarah Whitcher, and Eric C. Kansa. 2018. "Data Beyond the Archive in Digital Archaeology: An Introduction to the Special Section." *Advances in Archaeological Practice* 6(2): 89–92. https://doi.org/10.1017/aap.2018.7

Karcher, Sebastian, and Nicholas Weber. 2019. "Annotation for Transparent Inquiry: Transparent Data and Analysis for Qualitative Research." *IASSIST Quarterly* 43(2): 1–9. https://doi.org/10.29173/iq959

Kinder-Kurlanda, Katharina, Katrin Weller, Wolfgang Zenk-Möltgen, Jürgen Pfeffer, and Fred Morstatter. 2017. "Archiving Information from Geotagged Tweets to Promote Reproducibility and Comparability in Social Media Research." *Big Data & Society* 4(2). https://doi.org/10.1177/2053951717736336

Kirilova, Dessi, and Sebastian Karcher. 2017. "Rethinking Data Sharing and Human Participant Protection in Social Science Research: Applications from the Qualitative Realm." *Data Science Journal* 16(0): 43. https://doi.org/10.5334/dsj-2017-043

Kirkegaard, Emil O. W., and Julius D. Bjerrekær. 2016. "The OKCupid Dataset: A Very Large Public Dataset of Dating Site Users." *Open Differential Psychology*. https://openpsych.net/forum/showthread.php?tid=279

Kitchin, Rob. 2014. *The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences*. Los Angeles, California: SAGE Publications.

Levenstein, Margaret C., and Jared A. Lyle. 2018. "Data: Sharing Is Caring." *Advances in Methods and Practices in Psychological Science* 1(1): 95–103. https://doi.org/10.1177/2515245918758319

Lin, Dawei, Jonathan Crabtree, Ingrid Dillo, Robert R. Downs, Rorie Edmunds, David Giaretta, Marisa De Giusti, et al. 2020. "The TRUST Principles for Digital Repositories." *Scientific Data* 7(1): 144. https://doi.org/10.1038/s41597-020-0486-7

Logan, Jessica A. R., Sara A. Hart, and Christopher Schatschneider. 2021. "Data Sharing in Education Science." *AERA Open* 7(January). https://doi.org/10.1177/23328584211006475

Manovich, Lev. 2012. "Trending: The Promises and the Challenges of Big Social Data." In *Debates in the Digital Humanities*, edited by Matthew K. Gold, 460–75. Minneapolis, MN: University of Minnesota Press. https://doi.org/10.5749/minnesota/9780816677948.003.0047

Marwick, Alice E., and Danah Boyd. 2011. "I Tweet Honestly, I Tweet Passionately: Twitter Users, Context Collapse, and the Imagined Audience." *New Media & Society* 13(1): 114–133. https://doi.org/10.1177/1461444810365313

Mason, Winter, Jennifer Wortman Vaughan, and Hanna Wallach. 2014. "Computational Social Science and Social Computing." *Machine Learning* 95(3): 257–260. https://doi.org/10.1007/s10994-013-5426-8

Mauthner, Natasha S. 2012. "'Accounting for Our Part of the Entangled Webs We Weave': Ethical and Moral Issues in Digital Data Sharing." In *Ethics in Qualitative Research*, by Tina Miller, Maxine Birch, Melanie Mauthner, and Julie Jessop, 157–175. London: SAGE Publications Ltd. https://doi.org/10.4135/9781473913912.n11

Mauthner, Natasha S., and Odette Parry. 2013. "Open Access Digital Data Sharing: Principles, Policies and Practices." *Social Epistemology* 27(1): 47–67. https://doi.org/10.1080/02691728.2012.760663

Mauthner, Natasha S., Odette Parry, and Kathryn Backett-Milburn. 1998. "The Data Are Out There, or Are They? Implications for Archiving and Revisiting Qualitative Data." *Sociology* 32(4): 733–745. https://doi.org/10.1177/0038038598032004006

Metcalf, Jacob. 2016. "Big Data Analytics and Revision of the Common Rule." *Communications of the ACM* 59(7): 31–33. https://doi.org/10.1145/2935882

Meyer, Michelle N. 2018. "Practical Tips for Ethical Data Sharing." *Advances in Methods and Practices in Psychological Science* 1(1): 131–144. https://doi.org/10.1177/2515245917747656

Mneimneh, Zeina, Josh Pasek, Lisa Singh, Rachel Best, Leticia Bode, Elizabeth Bruch, Ceren Budak, et al. 2021. "Data Acquisition, Sampling, and Data Preparation Considerations for Quantitative Social Science Research Using Social Media Data." Preprint. PsyArXiv. https://doi.org/10.31234/osf.io/k6vyj

Morstatter, Fred, and Huan Liu. 2017. "Discovering, Assessing, and Mitigating Data Bias in Social Media." *Online Social Networks and Media* 1(June): 1–13. https://doi.org/10.1016/j.osnem.2017.01.001

Neuburger, Jeffrey D. 2020. "HiQ Files Opposition Brief with Supreme Court in LinkedIn CFAA Data Scraping Dispute." *The National Law Review X* (182). https://www.natlawreview.com/article/hiq-files-opposition-brief-supreme-court-linkedin-cfaa-data-scraping-dispute

Nissenbaum, Helen. 2009. *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford University Press. http://www.sup.org/books/title/?id=8862

Obar, Jonathan A., and Anne Oeldorf-Hirsch. 2020. "The Biggest Lie on the Internet: Ignoring the Privacy Policies and Terms of Service Policies of Social Networking Services." *Information, Communication & Society* 23(1): 128–147. https://doi.org/10.1080/1369118X.2018.1486870

Oboler, Andre, Kristopher Welsh, and Lito Cruz. 2012. "The Danger of Big Data: Social Media as Computational Social Science." *First Monday* 17 7). https://doi.org/10.5210/fm.v17i7.3993

Olshannikova, Ekaterina, Thomas Olsson, Jukka Huhtamäki, and Hannu Kärkkäinen. 2017. "Conceptualizing Big Social Data." *Journal of Big Data* 4(1): 3. https://doi.org/10.1186/s40537-017-0063-x

Palen, Leysia, and Paul Dourish. 2003. "Unpacking 'Privacy' for a Networked World." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 129–136. CHI '03. Ft. Lauderdale, Florida, USA: Association for Computing Machinery. https://doi.org/10.1145/642611.642635

Parry, Odette, and Natasha S. Mauthner. 2004. "Whose Data Are They Anyway?: Practical, Legal and Ethical Issues in Archiving Qualitative Research Data." *Sociology* 38(1): 139–152. https://doi.org/10.1177/0038038504039366

Pasquetto, Irene V., Christine L. Borgman, and Morgan F. Wofford. 2019. "Uses and Reuses of Scientific Data: The Data Creators' Advantage." *Harvard Data Science Review* 1(2). https://doi.org/10.1162/99608f92.fc14bf2d

Piwowar, Heather A., Michael J. Becich, Howard Bilofsky, and Rebecca S. Crowley. 2008. "Towards a Data Sharing Culture: Recommendations for Leadership from Academic Health Centers." *PLoS Medicine* 5(9). https://doi.org/10.1371/journal.pmed.0050183

Puschmann, Cornelius, and Jean Burgess. 2014. "Metaphors of Big Data." *International Journal of Communication* 8(June): 20. https://ijoc.org/index.php/ijoc/article/view/2169

QDR. 2019. "General Terms and Conditions of Use." Syracuse University: The Qualitative Data Repository. https://qdr.syr.edu/content/general-terms-and-conditions-use

Research Data Alliance International Indigenous Data Sovereignty Interest Group. 2019. "CARE Principles of Indigenous Data Governance." Global Indigenous Data Alliance, September. https://www.gida-global.org

Resnik, D. B. 2009. "Re-Consenting Human Subjects: Ethical, Legal and Practical Issues." *Journal of Medical Ethics* 35(11): 656–657. https://doi.org/10.1136/jme.2009.030338

Reuter, Katja, Yifan Zhu, Praveen Angyan, NamQuyen Le, Akil A Merchant, and Michael Zimmer. 2019. "Public Concern About Monitoring Twitter Users and Their Conversations to Recruit for Clinical Trials: Survey Study." *Journal of Medical Internet Research* 21(10): e15455. https://doi.org/10.2196/15455

Sandt, Stephanie van de, Sünje Dallmeier-Tiessen, Artemis Lavasa, and Vivien Petras. 2019. "The Definition of Reuse." *Data Science Journal* 18(June): 22. https://doi.org/10.5334/dsj-2019-022

Schneble, Christophe Olivier, Bernice Simone Elger, and David Shaw. 2018. "The Cambridge Analytica Affair and Internet-Mediated Research." *EMBO Reports* 19(8): e46579. https://doi.org/10.15252/embr.201846579

Secretary's Advisory Committee on Human Research Protections. 2015. Attachment A: Human Subjects Research Implications of "Big Data." https://www.hhs.gov/ohrp/sachrp-committee/recommendations/2015-april-24-attachment-a/index.html

Shah, Dhavan V., Joseph N. Cappella, and W. Russell Neuman. 2015. "Big Data, Digital Media, and Computational Social Science: Possibilities and Perils." The ANNALS of the American Academy of Political and Social Science 659(1): 6–13. https://doi.org/10.1177/0002716215572084

Sherif, Victoria. 2018. "Evaluating Preexisting Qualitative Research Data for Secondary Analysis." *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* 19(2). https://doi.org/10.17169/fqs-19.2.2821

Sieber, Joan E. 1991. "Introduction: Sharing Social Science Data." In *Sharing Social Science Data: Advantages and Challenges*, edited by Joan E. Sieber, 1–18. SAGE Publications.

Stenbacka, Caroline. 2001. "Qualitative Research Requires Quality Concepts of Its Own." *Management Decision* September. https://doi.org/10.1108/EUM0000000005801

Steneck, Nicholas H. 2007. "Chapter 6. Data Management Practices." Introduction to the Responsible Conduct of Research. Office of Research Integrity: Department of Health and Human Services. https://doi.org/10.1037/e638422011-001

Stoycheff, Elizabeth, Juan Liu, Kunto A. Wibowo, and Dominic P. Nanni. 2017. "What Have We Learned about Social Media by Studying Facebook? A Decade in Review." *New Media & Society* 19(6): 968–980. https://doi.org/10.1177/1461444817695745

Summers, Ed. 2017. "The Catalog and the Hydrator." Documenting the Now. August 21, 2017. https://news.docnow.io/the-catalog-and-the-hydrator-3299eddfe21e

Thomson, Sara Day, and Neil Beagrie. 2016. "Preserving Social Media." *Digital Preservation Coalition*. https://doi.org/10.7207/twr16-01

Thorne, Sally. 1994. "Secondary Analysis in Qualitative Research: Issues and Implications." In *Critical Issues in Qualitative Research Methods*, edited by Janice M. Morse, 263–729. London: Sage.

———. 2004. "Secondary Analysis of Qualitative Data." In *The SAGE Encyclopedia of Social Science Research Methods*, edited by Michael Lewis-Beck. Thousand Oaks, CA: SAGE Publications.

Törnberg, Petter, and Anton Törnberg. 2018. "The Limits of Computation: A Philosophical Critique of Contemporary Big Data Research." *Big Data & Society* 5(2): 2053951718811843. https://doi.org/10.1177/2053951718811843

Tsai, Alexander C., Brandon A. Kohrt, Lynn T. Matthews, Theresa S. Betancourt, Jooyoung K. Lee, Andrew V. Papachristos, Sheri D. Weiser, and Shari L. Dworkin. 2016. "Promises and Pitfalls of Data Sharing in Qualitative Research." *Social Science & Medicine* 169(November): 191–198. https://doi.org/10.1016/j.socscimed.2016.08.004

U.S. Department of Health and Human Services. 2017. "Attachment C - Recommendations for Broad Consent Guidance." https://www.hhs.gov/ohrp/sachrp-committee/recommendations/attachment-c-august-2-2017/index.html

Varol, Onur, Emilio Ferrara, Clayton A Davis, Filippo Menczer, and Alessandro Flammini. 2017. "Online Human-Bot Interactions: Detection, Estimation, and Characterization." In *Proceedings of the Eleventh International AAAI Conference on Web and Social Media* (ICWSM 2017), 10. Montreal, Quebec, Canada: AAAI Publications. https://aaai.org/ocs/index.php/ICWSM/ICWSM17/paper/view/15587

Vestoso, Margherita. 2018. "The GDPR beyond Privacy: Data-Driven Challenges for Social Scientists, Legislators and Policy-Makers." *Future Internet* 10(7): 62. https://doi.org/10.3390/fi10070062

Voigt, Paul, and Axel von dem Bussche. 2017. The EU General Data Protection Regulation (GDPR). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-57959-7

W3. 2021. "Schema.Org." https://schema.org

Washington Post. 2018. "Transcript of Mark Zuckerberg's Senate Hearing." April 10, 2018. https://www.washingtonpost.com/news/the-switch/wp/2018/04/10/transcript-of-mark-zuckerbergs-senate-hearing

Wilkinson, David, and Mike Thelwall. 2011. "Researching Personal Information on the Public Web: Methods and Ethics." *Social Science Computer Review* 29(4): 387–401. https://doi.org/10.1177/0894439310378979

Yoon, Ayoung, and Yoo Young Lee. 2019. "Factors of Trust in Data Reuse." *Online Information Review*. https://doi.org/10.1108/OIR-01-2019-0014

Zimmer, Michael. 2010. "'But the Data Is Already Public': On the Ethics of Research in Facebook." *Ethics and Information Technology* 12(4): 313–325. https://doi.org/10.1007/s10676-010-9227-5

————. 2016. "OkCupid Study Reveals the Perils of Big-Data Science." Wired, May 14, 2016. https://www.wired.com/2016/05/okcupid-study-reveals-perils-big-data-science

Zimmer, Michael, and Nicholas John Proferes. 2014. "A Topology of Twitter Research: Disciplines, Methods, and Ethics." *Aslib Journal of Information Management* 66(3): 250–261. https://doi.org/10.1108/AJIM-09-2013-0083