How can we save social media data?

Libby Hemphill[1, 2, 3]

Roles: Conceptualization, Data Curation, Formal Analysis, Methodology, Project Administration, Resources, Software, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing

Margaret Hedstrom[1, 3, 4]

Roles: Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing

Susan Leonard[1]

Roles: Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing

[1]ICPSR, University of Michigan, Ann Arbor, Michigan, United States of America

[2]Institute for Social Research, University of Michigan, Ann Arbor, Michigan, United States of America

[3]School of Information, University of Michigan, Ann Arbor, Michigan, United States of America

[4]Museum Studies Program, University of Michigan, Ann Arbor, Michigan, United States of America

Note: This working paper is circulated for discussion and comment purposes. It has not been peer-reviewed but may be under review.

Abstract

The availability and scale of social media data offer researchers new opportunities to leverage those data for their work in broad areas such as public opinion, digital culture, labor trends, public health, and social movements. The success of efforts to save social media data for reuse by researchers will depend on aligning data management and archiving practices with evolving norms around capture, use, sharing, and security of datasets containing this new type of data. This paper presents an initial foray into understanding how established practices for managing and preserving data should adapt to new demands from social media data platforms, researchers who use and reuse social media data, and people who supply social media content and are subjects in social media data. We examine the data management practices of researchers who use social media data in research through a survey of researchers and an analysis of published articles. We present results from 73 respondents and 40 papers and discuss the data management practices described, how they differ from management of more conventional data types, and the implications for creating and maintaining stable archives for these important research resources. We discuss the similarities and differences between social media data and other types of social science research data, including other types of "found" data, and discuss the implications for data archives wishing to include social media data in their collections.

Introduction

Social media are implicated in many of contemporary society's most pressing issues, from influencing public opinion, to organizing social movements, to identifying economic trends. Increasing the capacity of researchers to understand the dynamics of such phenomena will depend on reliable, curated, discoverable and accessible social media data. To inform the development of research data infrastructure, we need to understand how researchers in this space work. This article reports on two efforts to understand those practices and to inform the

design of the Social Media Archive (SOMAR) being developed at Inter-university Consortium for Political and Social Research (ICPSR), the oldest and one of the largest archives for managing and disseminating social science data. We reviewed 40 papers in four journals that used data from Twitter to understand how authors described their research activities and then surveyed researchers about their social media data practices generally.

We ask two different, but related, questions about the use of social media data for research: how do researchers use social media data in their research; and how do researchers acquire, manage, archive and share social media data? We specifically address how social media researchers' practices may differ from what we know from previous studies of data practices and we consider how the features of social media data (e.g., scale, speed, platform dependence, ownership) influence data practices. We are particularly interested in whether social scientists are able to ask new questions and apply new methods when they use social media data and the extent to which researchers' data management practices mirror (or don't) the data practices of researchers who use and share more traditional data types such as surveys and administrative data. We discuss the properties of social media data, the types of research questions and methods reported in articles that rely on social media data, and the responses to our survey about data practices. Our goal is to uncover similarities and differences between social media data and more familiar types of data in order to discover gaps between current social science data archive models and identify where new approaches are needed. Changes are likely needed because of the combination of the unique characteristics of social media data, the new approaches to social science research that they enable, and changing attitudes toward data management and data sharing.

Data Sharing and Management Practices

Existing literature on researchers' data management practices tells us that although researchers are interested in sharing data, they rarely do so [1,2]. Receiving credit for their work and

maintaining the option or right to publish about the data first were important considerations for researchers when deciding whether to share their data [3]. Many researchers negotiated private access to their data, especially between their research groups and those headed by other primary investigators they knew and trusted, but were unwilling to share their data without restricting who could access the data and what scientific questions they were able to examine with it [1–4]. They sometimes thought of data sharing as a "gift economy" in which they traded resources among trusted parties [3,5], allowing them to barter for other resources in the process. Depositing data in an archive limits the bartering value of a particular data set, and the lack of credit, through data citation or other means that researchers receive for sharing, provides disincentive to do so.

Most researchers manage their data "privately" by storing it on local computers and hard drives [4,6]. This local management practice was common even on campuses that offered secure, scalable storage and computing resources through a centralized service [6]. These practices mean that data is at risk for loss or leakage. Many datasets were not backed up in a second or secure storage space, placing them at risk for loss through both hardware failure and unauthorized access. Privately managed data is difficult for others to discover because it is hidden behind password protected servers and file systems, not indexed or described to enable discovery, and controlled by terms and conditions that are not available or transparent. According to the literature, researchers are also reluctant to share their data openly because they fear that the date will be misused or misinterpreted [4,7,8].

Effective data preservation depends, in part, on researchers' data management practices. Good data practices throughout the research lifecycle help ensure that users other than the original researchers will be able to find, understand, and reuse the data accurately [9]. Requirements such as data management plans, guidelines like the FAIR principles, and standards for metadata and other types of documentation are intended to facilitate data management and

data sharing, reduce the potential for misuse and misinterpretation, and ease the flow of data from researchers to permanent repositories. Nevertheless, research on data management practices and researchers' attitudes toward data sharing find that following the guidelines entails considerable effort and many researchers find adherence to such guidelines burdensome and time consuming [10,11].

Earlier studies of sharing and management practices used surveys of broad populations of researchers (e.g., international [2], campus-wide [4,6]) or case studies of specific research centers and groups (e.g., [3,12]). Researchers have analyzed data management practices in many fields and disciplines with astronomy, biomedical fields, earth and environmental sciences, and social science particularly well represented [13–17]. Social media, however, produce new types of data that researchers across a number of fields are using to address new questions. Little is known about research data management practices for social media data and few guidelines exist to assist researchers' selection and acquisition of data [18,19]. Our analysis of 40 peer-reviewed publications presenting research that used data from Twitter and our survey of 73 researchers' data management practices was designed to gather insights into how social media data are used for research and what new data management challenges arise for researchers and for repositories like ICPSR that are developing guidance and services that will support this community most effectively.

Methods

We used two different approaches to better understand current practices among researchers who use social media data. First, we reviewed articles that appeared in four highly-regarded interdisciplinary journals that described acquiring, refining, and analyzing data from Twitter. Second, we surveyed researchers about their practices around collecting and sharing data from several social media applications.

We reviewed articles in order to effectively summarize current approaches to using social media data in research. In all, we reviewed 40 studies published in *First Monday*; *Information, Communication and Society*; *Journal of the Association of Information Science and Technology*; and *New Media & Society* (the full list of articles is provided in Appendix A). When analyzing papers, we focused on the research question or topic of the paper, data collection or acquisition method, data provider, data set size, sampling and transformations, analysis approaches, and technical skills required. We recognize that research based on social media data are published in many other outlets and that these four publications do not represent all of the disciplines that use social media data in research. We focused on these sources because the journals sit at the intersection of information science, computational science, and social sciences. We expected the breadth of disciplines and approaches reported in these journals to reveal a variety of methodological approaches to using Twitter data, and with them a broad range of data practices.

Our survey received an "exempt" determination from University of Michigan's Institutional Review Board because we did not ask for personally identifying information and participation in the survey posed minimal risk to respondents. We recruited respondents through email lists (e.g., the Association of Internet Researchers listserv), Facebook groups (e.g., Researchers of the Socio-Technical), and investigators' individual social media accounts. The survey was open from July 31, 2018 to August 21, 2018 and received 73 responses. We offered three $100 Amazon gift cards as incentives and used random number draws to select recipients. We used Qualtrics to manage our survey and collection of responses. Our survey instrument had five main sections: general and demographic, data acquisition, data transformation, analysis and visualization, and data sharing and reuse. We restricted our demographic data collection to an investigator's affiliation (e.g., university, government lab) and position (e.g., PhD student, faculty, staff) in order to focus on the researchers' practices rather than their individual

characteristics. Prior work suggests that researchers in different age brackets and disciplines have different attitudes about data sharing [see, e.g., 3], and we expect that some of those differences are also present in the population we surveyed.

Our current goal is to understand existing data management practices so that we and others who are building capacity to archive and disseminate social media data will be cognizant of current social media research practices, be able to identify common needs, and develop services that support researchers in data acquisition, management, archiving and reuse. We reserve more explicit questions about encouraging sharing of social media for future work.

Results

Practices Reported in Publications

To understand the breadth of practices and methods among social media researchers, we collected articles published in four interdisciplinary journals where researchers reported on empirical analyses of Twitter data. Overall, we did find variety in the topics covered, methods used, and scope and scale of studies in this sample of papers. We also found that most methods sections were (understandably) brief and did not provide rich detail about the data collection or transformation processes, and none of the studies provided access to their data or analysis in supplementary materials.

Diversity of Research Areas

Social scientists use social media data to study a range of topics such as economic and consumer behavior [20,21], cultural differences [22], social capital [23,24], feminist and anti-racist movements [25–27], political activism [28–30], the relationship between social and traditional media [31–34], and the impact and reach of research [35,36]. In our analysis of research that used Twitter data we found a similar breadth of research topics, ranging from audience interactions around television shows [e.g., 37,38] to social justice movements under

hashtags such as #Ferguson [e.g., 39], and many political discussions around the world [e.g., 40,41–43]. Several studies used Twitter to characterize social networks of followers of particular hashtags, to test its effectiveness as a communication medium, or to identify characteristics of tweets associated with concepts like trustworthiness or utility. The studies in our sample often relied on data acquired from third-party distributors rather than directly from Twitter. For instance, Crimson Hexagon and Radian6 were frequently mentioned. Data sets ranged in size from just over 100 images to over 2 million tweets. In some cases, the boundaries of the data set were established by content (e.g., hashtags, keywords) and in others by the authors of the content (e.g., members of parliament, journalists). Papers also reported a variety of analytical approaches requiring wide-ranging methodological and computational expertise (e.g., qualitative grounded theory and computationally-intensive machine learning).

Survey Results

Demographics and Research Areas

The vast majority of respondents (87.7%) are affiliated with universities, with faculty (N=23) and PhD students (N=17) making up more than half (54.8%) of all respondents. Demographics are summarized in Table 1. Respondents in industry (N=5) and government or non-profit organizations (N=3) are not well represented in our survey, mostly likely because the types of email lists, online interest groups, and social networks we tapped for recruitment of subjects are more heavily populated with academic researchers.

Table 1. Survey respondents' affiliations. Percentages do not add up to 100% because of rounding.

Affiliation	% of Respondents	N
University	**86.3%**	**63**

Faculty	31.5%	23
PhD Student	23.3%	17
Master's Student	12.3%	9
University Post-Doc	9.6%	7
Undergraduate Student	5.5%	4
University Staff	5.5%	4
Industry	**6.9%**	**5**
Government or Non-profit	**4.1%**	**3**
Total		**73**

We asked respondents whether the focus of their research was on some aspect of the use or users of social media platforms themselves (e.g., Twitter) or whether they analyzed user-generated content from social media platforms to understand some other phenomenon (e.g., economic trends). Thirty-eight of our 73 respondents (52%) chose “I study social media platforms and/or social media users themselves”; 17 (23%) chose “I use social media data to study something else beyond social media”. Just six respondents chose “other” and supplied free-text answers that fell somewhere in between (e.g., “social media data as part of the agenda setting process”) or said “both”. Although the respondents as a whole used social media data from 11 different platforms (See Table 2), very few reported collecting data from two or more platforms.

Table 2: Social media platforms used to supply data for analysis. Percentages add up to more than

100% because respondents could have chosen more than one platform.		
Platform	% of Respondents	N
Twitter	39.7%	29
Facebook	28.8%	21
Instagram	11.0%	8
Reddit	11.0%	8
Wikipedia	6.8%	5
Tumblr	5.5%	4
Other	4.1%	3
Twitch	2.7%	2
YouTube	2.7%	2
Pinterest	1.4%	1

Data acquisition and analysis

We asked respondents to list tools or software they used to gather social media data. Python, the programming language, was the most frequent tool mentioned; and Python libraries such as pandas, scikit-learn, TensorFow, NLTK, NumPy, and related tools such as Jupyter notebooks were also mentioned. R or related tools (RStudio) were the next most frequent category of tools. Respondents who mentioned specific software or services listed NVivo, DiscoverText, NodeXL, TAGS, IFTTT, Social Feed Manager, Zapier, Hydrator, WebRecorder.io, and SPSS. Eleven respondents (15%) said they had paid for access to social media data.

We also asked respondents to indicate what skills they thought were important for people working with social media data to have. Their responses are summarized in Table 3.

Table 3: Skills that respondents considered important

Skill	Respondents
Web scraping	38
Python	33
R	26
Advanced statistics	24
System/server administration	10

Twenty-two respondents also provided an answer under “other” and indicated that skills such as “understanding of privacy issues/ethics of social media data,” “thoughtful engagement with the ethics and accountability of their research,” and “understanding of digital culture.” Respondents also indicated that computational skills were not always necessary. For instance, one said, “I don't think any of these are ‘necessary’ as one can perform research on social media data via qualitative means,” and another commented, “analytical skills, all the other things can come from a team.”

When asked about where those skills were acquired, 63% of respondents (N=46) said they had “learned on my own or with help online (e.g., Stack Overflow)”. The options “taught by someone on my research team” and “platform API documentation” were both chosen by 27% of respondents (N=20). Only 10% learned “in class” (N=7). Other answers included “from a book” (N=11), and “other” (N=7). Among the “other” responses, people reported learning from colleagues, staff, and students who were not members of their research team.

Data sharing and reuse

Twenty-three respondents (31.5%) said they do not make their data available to others. Thirty-four respondents (46.6%) do make their data available use repositories and websites (see Table 4). Eleven respondents chose “other” when asked “How do you make your data available to others?”, and in those responses, many mentioned restrictions on data sharing imposed by platforms or indicated that they would be willing to share data directly with researchers who asked. For instance, they indicated, “code is on GitHub, they can request data” or “they will receive an external hard drive with the data” and “We can directly share signals we calculate from that data, but not the social media data itself” or “We make data available on a case-by-case basis, given platform Terms of Service.” Respondents who used repositories or archives to share their data listed their university’s institutional repositories (N=3), Github (N=3), Figshare (N=2), and ICPSR (N=1).

Table 4: Mechanisms used by respondents to share social media data. Percentages reported are of all respondents. Because not all respondents answered the question, the percentages do not total 100%.

Mechanism	% of respondents	Respondents
I don't make my data available.	31.5%	23
I make my data available.	46.6%	34
In a repository or archive	15.1%	11
Through a personal website	11.0%	8
Through journal or conference site	8.2%	6
Through a University affiliated website	6.8%	5
Through a third-party data provider	5.4%	4

Other	15.1%	11

We also asked whether people had prepared data for reuse within their research groups (N=17), by others outside their groups (N=14), or not at all (N=28). The majority of respondents had not received requests for their data or prepared their data for replication. Table 5 summarizes the results of these questions about preparation and requests for reuse or replication. When preparing for replication, respondents most often indicated that they provided code (e.g., Jupyter notebooks, R scripts) for analysis and filtered or cleaned datasets that contained only the data reported in a publication. When preparing for sharing, respondents anonymized datasets, published tweet IDs, cleaned the data, and wrote documentation about their analysis process (e.g., README files, documentation).

Table 5: Preparation and requests for reuse and replication

	Yes	Respondents	No	Respondents
Have you ever prepared your data especially for reuse?	28.8%	21	38.4%	28
Have you ever prepared your data especially for replication?	23.3%	17	52.1%	38
Has anyone ever contacted you, or your team, to request access to your social media data set?	15.1%	11	54.8%	40

Summary of findings

Through our analysis of 40 papers that used Twitter data and our survey of researchers who use social media data we reached three tentative conclusions. First, researchers used Twitter data to address a wide variety of issues ranging from characterizing the social networks of Twitter users to analyzing the content of tweets associated with particular hashtags, political issues, events, and other phenomena. Some of these studies used Twitter data as a new

source for insights into long-standing questions about social, behavioral, political, and economic issues, while other studies attempted to understand the impact of Twitter as a new form of communication.

Second, using social media data for research requires more technical skills and familiarity with a wider variety of tools for assembling data sets from social media feeds and platform API's, cleaning data, and analysis than research using more established sources, like surveys, and methods such as regression analysis. Most researchers gained these skills through informal means including practice and on-line help use, assistance from co-workers, and consulting documentation. It appears that a single individual rarely possesses the full complement of conceptual, analytical, computational, and technical skills needed to work with social media data; rather, these skills are distributed across different members of research teams.

Third, we found similarities and differences between data management and data sharing practices of researchers using social media data and other social scientists. Researchers using social media data seem to focus their data management efforts on acquisition of data from social media platforms or third-party providers and on making the data usable for their own analyses, with less emphasis on making the data reusable by others. We found that they raise concerns similar to those of other social scientists about sharing their data and ethical issues such as privacy and misinterpretation of data. Whether these differences are a consequence of unique characteristics of social media data, the new affordances of social media for novel paths of inquiry, the relative immaturity of social media research, or other factors is the topic of our discussion below.

Discussion

What makes social media data different?

Social media data consists of user-generated content that they create, share, or react to and system-generated data, such as timestamps, account information, and click streams. In the most general sense, user-generated content is produced by individuals for the purpose of communicating, sharing, or disseminating information to a selected audience of friends or followers or to the broad community of users of a particular platform. System-generated data is collected automatically by the platform providers and used for monitoring system performance, targeting advertising, monitoring and moderating content, and internal research.

Typically, researchers acquire data directly from one or more social media platforms or submit requests to these private entities for data sets that meet specific criteria. The data are proprietary with differing terms of service depending on the platform of origin, which may place limits on researchers' requests to obtain access, customize data, link content to account information, share data with others, and archive the data. Social media data are updated constantly and usually delivered as raw feeds that generally require programming before analysis; historical data (sometimes as recent as two weeks old) are often more difficult or costly to access than live streams. They consist of system-generated metadata (e.g., user account age, date the content was created), user-generated content (e.g., the text of a tweet or Facebook post), and pointers to resources that live elsewhere (e.g., photos, videos, URLs). The platforms are unwilling to provide access to the proprietary algorithms that structure the streaming data into meaningful feeds. Social media data often include personally identifiable information (PII) that users may have shared with the understanding that only people who they authorized would see their user-generated content [44].

Researchers may also gain access to "digital trace data" [45], such as clickstream data that tracks users' behavior on a platform or the behavior of others referencing their content. The

Social Science One partnership with Facebook [46] in which Facebook compiles data about users' reactions (e.g., "like", "angry") to URLs shared on the platform but not the posts in which the URLs were shared is one example of this type of trace data.

Data structures, scale, and speed

One challenge social media data present is the difficulty in describing what constitutes a "collection of social media data" or a "social media data set" [47]. Researchers and archives must know what it is they are proposing to collect, share, and archive, and the answer for social media data is not straightforward.

Should a social media data set include only the content from the social media platform (e.g., a tweet record from Twitter's API) or the social media content and the content it references? Platform terms of service also attempt to restrict what platform data users can do with data they have collected, and researchers modify the data collected in order to comply with these terms. For instance, Twitter's Developer Policy, the agreement governing programmatic access to the site's content, states that people sharing Twitter content "will only distribute or allow download of Tweet IDs, Direct Message IDs, and/or User IDs" [48]. Does this then mean that Twitter datasets include only these items, and archives will be accepting and caring only for lists of identifiers rather than the content of the tweets? Tweets can be deleted from the platform at any time, by the author or by Twitter, and therefore, these shared lists of IDs are insufficient for reconstructing the original data sets. Research suggests that tweets in these ID collections persist at rates varying from 30%–80% over four years [49]; collections that contain only IDs are most likely incomplete.

Data from both the articles we reviewed and the responses to our survey suggest researchers use different approaches to data collection (e.g., purchasing from third-party data resellers, writing bespoke applications to collect data through APIs). Researchers then rarely describe the particulars of those collection methods or the transformations they perform on the data to

prepare it for analysis. The inability to judge the quality or understand the provenance of a single research group's effort presents additional challenges for other research groups to reuse the data [19,50].

Data practices: finding, curating, sharing, and storing data

Data management practices for structured survey, polling, and administrative data have matured over the last 50 years, and reuse of data beyond the original investigators is common in social science research. Researchers learn that the design of a good survey includes documenting the sampling frames and response rates, developing codebooks, and ensuring that explicit obligations to protect privacy and confidentiality are met [51]. Although researchers have less control over the structure, quality, accuracy, and completeness of statistical and administrative data, they can use a combination of documentation, statistical techniques, and prior experience with canonical data sets (e.g., census data, economic indicators) to detect errors or estimate reliability of data sets [52–54]. Repositories for social science data provide training, advice, and curation services for these more common types of social science data.

Sound data management practices, scalable curation, and archiving processes rely on documentation about the collection or creation of a data set or collection, its internal structure, transformations performed on the data, and many field-specific ontologies, metadata schema, quality control measures, and the like. When researchers create or collect their own data through surveys, interviews, experiments, and observation, they make choices about the quantity, structure, granularity, scope, and other aspects of the data as part of the research design. By documenting these decisions, data collections are more amenable to validation, replication and reuse by others. Administrative records, such as police reports, financial transactions, and unemployment claims, and statistical data such as censuses are common types of data that social scientists also use to address research questions. Unlike surveys, experiments, interviews, and observations, where researchers design and then create or collect

data to address a particular research question, statistical, administrative and other transactional data are not created explicitly for research. These types of data have been characterized as "found" [55,56] or "non-designed" [57] data because they were not collected originally to address a particular research question. Rather, researchers discover data, assess its suitability for their research questions, and then manipulate the data for the specific purposes of their own research. We are not the first to use the term "found" for these and similar data. See, for instance, Harford [56] on "found" data in our digital traces or McOverton, et al. [55] on "found" data in non-probability samples.

Social media data are a new type of "found" data, and practices around its use in research and its curation, dissemination, and reuse are immature. Social media data have broad disciplinary applications and uses, and with that breadth comes wide variety in data practices. Many of the challenges these practices pose for archiving and sharing are common to research data generally and are not unique to social media data (e.g., reluctance to share data, resource limitations, and risky data storage). Others, though, are more pronounced for social media data (e.g., determining what constitutes a "collection" or "data set", scaling methods of curation, documenting data transformations). However, even these practices that seem new to social science have useful analogs among other types of data that are used for research but weren't first collected to support research.

The processes of finding social media data and preparing it for use in research are frequently conducted computationally. Our respondents indicated that experience with computational skills such as programming, web scraping, and server administration are necessary for research that uses social media data. These skills are used at each stage of the data lifecycle (e.g., Python scripts for collecting from the platform APIs, Jupyter and R notebooks for cleaning and analyzing data). The computational processes involved in research with social media data present both challenges and opportunities for documenting workflow and preserving data

provenance. Because the processes are captured in the code and/or notebooks, they are technically available for collection and preservation. However, code and notebooks are not document types most archives are structured for or experienced at handling.

Researchers who use social media data showed a reluctance to share data for reasons that are similar to those expressed in other studies of researchers' attitudes toward data sharing [2,6]. The resources, both computational and human, required to collect, transform, and manage social media data are non-trivial and norms for recognizing this effort through citation, some share in authorship, or other means are nascent at best. Even when social media researchers are willing to share data upon request or distribute it through a website or repository, they are seeking guidance on how to document their data. No shared metadata standard for social media exists. Recent efforts by ad hoc groups of researchers have not gained traction (e.g., Open Collaboration Data Factories [58]) nor produced proposals for metadata and documentation standards (e.g., Documenting Social Media Datasets [59]).

These efforts and respondents' comments highlight that documenting social media data poses challenges in part because of the difficulty in describing the provenance of the data. For instance, the specific hashtags used to search for data through the Twitter API may change over the course of a project (e.g., a study of health care policy discussions begins by collecting #aca tweets, expands to include #obamacare and #trumpcare tweets as those hashtags emerge). Documentation of the provenance of a social media data set should include the specific search terms, dates those terms were used, data returned that matched the query, and tracking of any subsequent transformations of the data, including the software and scripts used.

Finally, even among this computationally-savvy group, researchers engage in risky data storage practices (e.g., using personal laptops instead of secured servers). Storing data on individual laptops increases risks of data loss and unauthorized access. Choosing to store locally rather than using university data services is a common practice among academic researchers [4,6],

and is not unique to social media data users. Though they eschewed university data services, many respondents reported using university license agreements for software (e.g., MAXQDA, NVivo).

Ethical considerations in social media data management

Social media data also raise a host of new legal and ethical challenges. Private companies own and control the algorithms that underpin every aspect of how social media platforms operate, and they establish the terms and conditions for individuals who use these platforms in terms of personal privacy, proper use, intellectual property, and content limitations. Although platform users have some options for setting privacy and other use preferences, research has shown that privacy policies are ineffective at actually informing users about terms [60] and users make choices about sharing that depend on context [61,62]. Social media users share sensitive and highly personal information, but it is unclear whether they are aware that this information could be harvested, archived, and reused without their explicit authorization. The responses to our survey indicate that researchers who use social media data are seeking guidance on how to prevent disclosure of individual identities and sensitive information, protect privacy, and conform to unclear and sometimes contradictory ethical guidelines and contractual obligations.

Implications for archives

The breadth and diversity of practices present challenges for archiving, in part because the secondary uses may differ dramatically from the primary use of each data set. In addition, the context of reuse is fundamentally different from that of the social media platform where a user posted, responded to, or shared content originally. We discuss three ways in which social media differ data differ enough from the more familiar types of data that established archiving policies and practices will need adjustment.

Acquisition and manipulation of social media data

Most data archives acquire research data either directly from a researcher or research team at the end of their project or obtain data from administrative or statistical agencies on a regular cycle. Typically, these deposits include some documentation that explains how the data were acquired and organized into a data set or collection of data sets. Social media data, however, are first acquired by researchers from the social media platforms through their APIs or sites or by way of special access negotiated with the platform providers or through third party distributors. All of these mechanisms for acquiring social media data place terms and conditions on what content and system-generated metadata can be downloaded, how the data can be used, and whether it can be shared with others.

We learned from our survey that researchers use a variety of tools to acquire data and further manipulate the data to make it useful for their particular research questions. Placing restrictions on the conditions of use and reuse is not new to social media data, nor is the practice of cleaning and manipulating data prior to analysis. Nevertheless, it appears from our survey that researchers have greater challenges ascertaining the scope, depth, granularity, and temporality of the data they acquire from social media platforms and third parties, raising questions about the ability to benchmark social media data against some reality or ground truth. We also noted that the data are acquired and manipulated computationally. These new acquisition and research practices suggest that traditional notions of documentation may be inadequate, and that reuse of social media data by others will require much richer documentation of provenance, explicit documentation of the terms and conditions for acquiring the data, and documentation or deposit of the software and scripts used to acquire and manipulate the data.

Technical and conceptual challenges

Social media data are complex objects that live in networks of relationships and linkages between user-generated content, metadata, external references, external content, and system-

generated metadata. Compared to most types of archived data collections, social media data are especially voluminous and dynamic. For example, researchers may decide not to download linked content to comply with terms and conditions or for practical reasons such as limiting storage requirements or improving the performance of the scripts used to scrape data from APIs. This means that linked content, which was available on the original platform, may have been deleted or changed by the time a researcher wishes to reuse the data. Current methods for curation are unlikely to scale for social media data, and they will remain ineffective and unaffordable without new tools and workflows for the currently laborious processes of metadata extraction and creation, quality control, and detection of disclosure risk [47].

Privacy, confidentiality and ethical use of social media data

Established practices for informed consent, confidentiality and privacy protection, anonymization, and preventing deductive disclosure of individual identities are starting points for considering the ethical responsibilities that repositories incur when they acquire social media data. Nevertheless, new questions are arising about the appropriate use of social media data because of changing assumptions about consent, disclosure, persistence, and control over user-generated content. The terms and conditions for posting, sharing, and deleting content on social media platforms are governed by user agreements, platform terms of service, and individual configurations of privacy and other settings, as well as ever changing norms about what is appropriate to post in the first place, who “owns” personal data, and how decisions are made about distribution, deletion and disposition of social media data, and regulations such as the General Data Protection Regulation (GDPR) in the European Union [63,64] .

The results of our survey suggest that researchers are seeking guidance on many of the issues we have discussed. Collaboration between repositories, such as SOMAR, that are developing new archiving capacity for social media data and researchers who are encountering myriad conceptual, technical, and ethical questions as they bring innovative methods and new types of

data sources into their research seems necessary for tackling this complex challenge while building on the knowledge and experience of both researchers and curators. It is worth noting that in our survey students constitute the largest single group engaged in research using social media data. Aiming services and training at students in the beginning of their careers may be more effective than trying to reeducate more senior scholars with entrenched habits.

Conclusion

Research that relies on data from social media covers a wide range of topics, allows new research questions to be formulated and addressed, and creates opportunities to address old questions in novel ways. The data management practices employed for working with social media data resemble the processes for other types of social science data, especially other types of “found” data such as censuses, police records, and other administrative records. However, for other found data, documentation and storage standards are generally agreed upon, and data archives around the globe offer guidance for researchers working with such data. Standards for social media data are nascent, and archives are just beginning to offer support.

Researchers who use social media data also mirror other researchers in their reluctance to share data without ensuring credit for their work, awareness of who will reuse the data, and confidence that the data will not be used inappropriately. Social media data are an uneasy fit in existing data archives due to differences in scale, speed, platform dependence, structure, and ownership. An archive that facilitates the preservation and reuse of social media data will need to contend with additional challenges in documenting data and its provenance, in describing what constitutes a “dataset” in this space, and in ensuring appropriate protections for personal and sensitive information. We can save social media data if researchers, social media platforms, and repositories all attend to these new conceptual, technical, and ethical challenges.

Acknowledgments

- Joshua Guberman
- Saul Hankin
- Rebekah Small

References

1. Kennan MA, Markauskaite L. Research data management practices: a snapshot in time. Int J Digit Curation [Internet]. 2015;10(2):69–95. Available from: http://www.ijdc.net/index.php/ijdc/article/view/329

2. Tenopir C, Allard S, Douglass K, Aydinoglu AU, Wu L, Read E, et al. Data sharing by scientists: practices and perceptions. PLoS One [Internet]. 2011 Jun 29;6(6):e21101. Available from: http://dx.doi.org/10.1371/journal.pone.0021101

3. Wallis JC, Rolando E, Borgman CL. If we share data, will anyone use them? Data sharing and reuse in the long tail of science and technology. PLoS One [Internet]. 2013 Jul 23;8(7):e67332. Available from: http://dx.doi.org/10.1371/journal.pone.0067332

4. Akers KG, Doty J. Disciplinary differences in faculty research data management practices and perspectives. Int J Digit Curation [Internet]. 2013 Nov 19 [cited 2018 Jun 19];8(2). Available from: http://www.ijdc.net/article/view/8.2.5

5. Hilgartner S, Brandt-Rauf SI. Data access, ownership, and control: toward empirical studies of access practices. knowledge [Internet]. 1994 Jun 1;15(4):355–72. Available from: https://doi.org/10.1177/107554709401500401

6. Whitmire AL, Boock M, Sutton SC. Variability in academic research data management practices: implications for data services development from a faculty survey. Programmirovanie [Internet]. 2015;49(4):382–407. Available from: https://doi.org/10.1108/PROG-02-2015-0017

7. Kim Y, Stanton JM. Institutional and individual factors affecting scientists' data-sharing behaviors: a multilevel analysis. J Assoc Inf Sci Technol [Internet]. 2016 Apr 4;67(4):776–99. Available from: http://doi.wiley.com/10.1002/asi.23424

8. Cragin MH, Palmer, Carole L, Carlson JR, Witt M. Data sharing, small science and institutional repositories. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences [Internet]. 2010 Sep 13;368(1926):4023–38. Available from: https://doi.org/10.1098/rsta.2010.0165

9. Wilkinson MD, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data [Internet]. 2016 Mar 15;3:160018. Available from: http://dx.doi.org/10.1038/sdata.2016.18

10. Tenopir C, Dalton ED, Allard S, Frame M, Pjesivac I, Birch B, et al. Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide. PLoS One [Internet]. 2015 Aug 26;10(8):e0134826. Available from: http://dx.doi.org/10.1371/journal.pone.0134826

11. Sayogo DS, Pardo TA. Exploring the determinants of scientific data sharing: understanding the motivation to publish research data. Gov Inf Q [Internet]. 2013 Jan 1;30:S19–31. Available from: http://www.sciencedirect.com/science/article/pii/S0740624X12001529

12. Mayernik MS. Research data and metadata curation as institutional issues. J Assn Inf Sci Tec [Internet]. 2016 Apr 30;67(4):973–93. Available from: http://doi.wiley.com/10.1002/asi.23425

13. Field D, Sansone S-A, Collis A, Booth T, Dukes P, Gregurick SK, et al. 'Omics data sharing. Science [Internet]. 2009 Oct 9;326(5950):234–6. Available from: http://dx.doi.org/10.1126/science.1180598

14. Faniel IM, Kriesberg A, Yakel E. Social scientists' satisfaction with data reuse. J Assoc Inf Sci Technol [Internet]. 2016 Jun 4;67(6):1404–16. Available from: http://doi.wiley.com/10.1002/asi.23480

15. Federer LM, Lu Y-L, Joubert DJ, Welsh J, Brandys B. Biomedical data sharing and reuse: Attitudes and practices of clinical and scientific research staff. PLoS One [Internet]. 2015 Jun 24;10(6):e0129506. Available from: http://dx.doi.org/10.1371/journal.pone.0129506

16. Pepe A, Goodman A, Muench A, Crosas M, Erdmann C. How do astronomers share data? Reliability and persistence of datasets linked in AAS publications and a qualitative study of data practices among US astronomers. PLoS One [Internet]. 2014 Aug 28;9(8):e104798. Available from: http://dx.doi.org/10.1371/journal.pone.0104798

17. Kim Y, Adler M. Social scientists' data sharing behaviors: investigating the roles of individual motivations, institutional pressures, and data repositories. Int J Inf Manage [Internet]. 2015 Aug 1;35(4):408–18. Available from: http://www.sciencedirect.com/science/article/pii/S0268401215000432

18. Kinder-Kurlanda K, Weller K, Zenk-Möltgen W, Pfeffer J, Morstatter F. Archiving information from geotagged tweets to promote reproducibility and comparability in social media research. Big Data & Society [Internet]. 2017 Nov 1;4(2):2053951717736336. Available from: https://doi.org/10.1177/2053951717736336

19. Driscoll K, Walker S. Big data, big questions| working within a black box: transparency in the collection and production of big Twitter data. Int J Commun Syst [Internet]. 2014 Jun 16 [cited 2018 Mar 16];8(0):20. Available from: http://ijoc.org/index.php/ijoc/article/view/2171

20. Antenucci D, Cafarella M, Levenstein M, Ré C, Shapiro MD. Using social media to measure labor market flows [Internet]. National Bureau of Economic Research; 2014. (Working Paper Series). Available from: http://www.nber.org/papers/w20010. Cited 6 May 2019.

21. Asur S, Huberman BA. Predicting the future with social media. In: 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology [Internet]. 2010. p. 492–9. Available from: http://dx.doi.org/10.1109/WI-IAT.2010.63

22. Hochman N, Schwartz R. Visualizing instagram: tracing cultural visual rhythms. In: proceedings of the workshop on Social Media Visualization (SocMedVis) in conjunction with the sixth international AAAI conference on Weblogs and Social Media (ICWSM--12) [Internet]. 2012. p. 6–9. Available from: http://www.aaai.org/ocs/index.php/ICWSM/ICWSM12/paper/viewFile/4782/5091

23. Ellison NB, Vitak J, Gray R, Lampe C. Cultivating social resources on social network sites: Facebook relationship maintenance behaviors and their role in social capital processes. J Comput Mediat Commun [Internet]. 2014 Jul 1 [cited 2016 Apr 22];19(4):855–70. Available from: http://onlinelibrary.wiley.com/doi/10.1111/jcc4.12078/abstract

24. Gil de Zúñiga H, Jung N, Valenzuela S. Social media use for news and individuals' social capital, civic engagement and political participation. J Comput Mediat Commun [Internet]. 2012 Apr 1;17(3):319–36. Available from: http://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.2012.01574.x/abstract

25. Dixon K. Feminist online identity: analyzing the presence of hashtag feminism. Journal of Arts and Humanities [Internet]. 2014 Aug 3;3(7):34–40. Available from: http://theartsjournal.org/index.php/site/article/view/509

26. Freelon D, McIlwain C, Clark M. Beyond the hashtags: #Ferguson, #Blacklivesmatter, and the online struggle for offline justice. 2016; Available from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2747066. Cited 6 May 2019.

27. Brock A. From the blackhand side: Twitter as a cultural conversation. J Broadcast Electron Media [Internet]. 2012 Oct 1;56(4):529–49. Available from: https://doi.org/10.1080/08838151.2012.732147

28. Freelon D. Discourse architecture, ideology, and democratic norms in online political discussion. New Media Society [Internet]. 2015 May 1;17(5):772–91. Available from: http://nms.sagepub.com/content/17/5/772

29. Roback A, Hemphill L. I'd have to vote against you: issue campaigning via Twitter. In: Proceedings of the 2013 conference on Computer supported cooperative work companion [Internet]. New York, NY, USA: ACM; 2013 [cited 2018 Mar 16]. p. 259–62. (CSCW '13). Available from: https://dl.acm.org/citation.cfm?doid=2441955.2442016

30. Boulianne S. Social media use and participation: a meta-analysis of current research. Inf Commun Soc [Internet]. 2015 May 4;18(5):524–38. Available from: http://dx.doi.org/10.1080/1369118X.2015.1008542

31. Shapiro MA, Hemphill L. Politicians and the policy agenda: Does use of Twitter by the U.S. Congress direct New York Times content? Policy & Internet [Internet]. 2017 Mar 1;9(1):109–32. Available from: http://dx.doi.org/10.1002/poi3.120

32. Soroka S, Daku M, Hiaeshutter-Rice D, Guggenheim L, Pasek J. Negativity and positivity biases in economic news coverage: traditional versus social media. Communic Res [Internet]. 2018 Oct 1;45(7):1078–98. Available from: https://doi.org/10.1177/0093650217725870

33. Papacharissi Z, de Fatima Oliveira M. Affective news and networked publics: the rhythms of news storytelling on #Egypt. J Commun [Internet]. 2012;62(2):266–82. Available from: http://onlinelibrary.wiley.com/doi/10.1111/j.1460-2466.2012.01630.x/full

34. Jungherr A. The logic of political coverage on Twitter: temporal dynamics and content. J Commun [Internet]. 2014 Apr 1;64(2):239–59. Available from: http://onlinelibrary.wiley.com/doi/10.1111/jcom.12087/abstract

35. Thelwall M, Haustein S, Larivière V, Sugimoto CR. Do altmetrics work? Twitter and ten other social web services. PLoS One [Internet]. 2013 May 28;8(5):e64841. Available from: http://dx.doi.org/10.1371/journal.pone.0064841

36. Haustein S, Bowman TD, Holmberg K, Tsou A, Sugimoto CR, Larivière V. Tweets as impact indicators: examining the implications of automated “bot” accounts on Twitter. J Assn Inf Sci Tec [Internet]. 2016 Jan 1;67(1):232–8. Available from: http://dx.doi.org/10.1002/asi.23456

37. Williams A, Gonlin V. I got all my sisters with me (on Black Twitter): second screening of How to Get Away with Murder as a discourse on Black Womanhood. Inf Commun Soc [Internet]. 2017 Jul 3;20(7):984–1004. Available from: https://doi.org/10.1080/1369118X.2017.1303077

38. Boukes M, Trilling D. Political relevance in the eye of the beholder: determining the substantiveness of TV shows and political debates with Twitter data. First Monday [Internet]. 2017 Apr 3;22(4). Available from: https://uncommonculture.org/ojs/index.php/fm/article/view/7031

39. Barnard SR. Tweeting #Ferguson: mediatized fields and the new activist journalist. New Media & Society [Internet]. 2017 Jun 19;1461444817712723. Available from: https://doi.org/10.1177/1461444817712723

40. Aelst PV, Erkel P van, D’heer E, Harder RA. Who is leading the campaign charts? Comparing individual popularity on old and new media. Inf Commun Soc [Internet]. 2017 May 4;20(5):715–32. Available from: http://dx.doi.org/10.1080/1369118X.2016.1203973

41. Engesser S, Ernst N, Esser F, Büchel F. Populism and social media: how politicians spread a fragmented ideology. Inf Commun Soc [Internet]. 2017 Aug 3;20(8):1109–26. Available from: https://doi.org/10.1080/1369118X.2016.1207697

42. Zhang Y, Wells C, Wang S, Rohe K. Attention and amplification in the hybrid media system: the composition and activity of Donald Trump’s Twitter following during the 2016 presidential election. New Media & Society [Internet]. 2017 Dec 4;1461444817744390. Available from: https://doi.org/10.1177/1461444817744390

43. Zelenkauskaite A, Niezgoda B. “Stop Kremlin trolls:” Ideological trolling as calling out, rebuttal, and reactions on online news portal commenting. First Monday [Internet]. 2017;22(5). Available from: http://uncommonculture.org/ojs/index.php/fm/article/view/7795

44. Marwick AE, Boyd D. I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. New Media & Society [Internet]. 2010 Jul 7;13(1):114–33. Available from: https://doi.org/10.1177/1461444810365313

45. Howison J, Wiggins A, Crowston K. Validity issues in the use of social network analysis with digital trace data. Journal of the Association for Information Systems [Internet]. 2011 [cited 2019 Apr 26];12(12):2. Available from: https://aisel.aisnet.org/jais/vol12/iss12/2/

46. Our Facebook Partnership [Internet]. Social Science One. Available from: https://socialscience.one/our-facebook-partnership. Cited 2019 Apr 1.

47. Voss A, Lvov I, Thomson SD. Data storage, curation and preservation. In: Sloan L, Quan-

Haase A, editors. The SAGE Handbook of Social Media Research Methods [Internet]. 55 City Road, London: SAGE Publications Ltd; 2017. p. 161–76. Available from: https://methods.sagepub.com/book/the-sage-handbook-of-social-media-research-methods

48. Developer Policy [Internet]. Twitter. 2017. Available from: https://developer.twitter.com/en/developer-terms/policy.html. Cited 2019 Jan 14.

49. Zubiaga A. A longitudinal assessment of the persistence of twitter datasets. J Assoc Inf Sci Technol [Internet]. 2018 Aug 14;69(8):974–84. Available from: http://doi.wiley.com/10.1002/asi.24026

50. Weller K, Kinder-Kurlanda KE. A manifesto for data sharing in social media research. In: Proceedings of the 8th ACM Conference on Web Science [Internet]. ACM; 2016. p. 166–72. Available from: https://dl.acm.org/citation.cfm?doid=2908131.2908172

51. Wolf C, Joye D, Smith TW, Fu Y-C. The SAGE Handbook of Survey Methodology [Internet]. SAGE Publications; 2016. 740 p. Available from: https://market.android.com/details?id=book-R9nangEACAAJ

52. Alvarez MR. Computational social science: discovery and prediction [Internet]. Cambridge University Press; 2016. 404 p. Available from: https://market.android.com/details?id=book-MqqzCwAAQBAJ

53. Massey CG, Genadek KR, Alexander JT, Gardner TK, O'Hara A. Linking the 1940 U.S. Census with modern data. Historical Methods: A Journal of Quantitative and Interdisciplinary History [Internet]. 2018 Oct 2;51(4):246–57. Available from: https://doi.org/10.1080/01615440.2018.1507772

54. Randall S, Coast E. The quality of demographic data on older Africans. DemRes [Internet]. 2016 Jan 21;34:143–74. Available from: http://www.demographic-research.org/volumes/vol34/5/

55. Mc Overton JC, Young TC, Overton WS. Using "found" data to augment a probability sample: procedure and case study. Environ Monit Assess [Internet]. 1993 May;26(1):65–83. Available from: http://dx.doi.org/10.1007/BF00555062

56. Harford T. Big data: a big mistake? Significance [Internet]. 2014 Dec 1;11(5):14–9. Available from: http://doi.wiley.com/10.1111/j.1740-9713.2014.00778.x

57. Weinberg D, Abowd JM, Belli RF, Cressie N, Folch DC, Holan SH, et al. Effects of a government-academic partnership: Has the NSF-Census Bureau Research Network helped improve the U.S. statistical system? 2018 Apr 26; Available from: https://doi.org/10.1093/jssam/smy023

58. OCDX-Specification [Internet]. Open Community Data Exchange. Available from: https://github.com/OCDX/OCDX-Specification. Cited 2019 Apr 1.

59. DocNow [Internet]. Documenting the Now. Available from: https://www.docnow.io/. Cited 6 May 2019.

60. Schaub F, Balebako R, Cranor LF. Designing effective privacy notices and controls. IEEE Internet Comput [Internet]. 2017;1–1. Available from:

http://dx.doi.org/10.1109/MIC.2017.265102930

61. Acquisti A, Brandimarte L, Loewenstein G. Privacy and human behavior in the age of information. Science [Internet]. 2015 Jan 30;347(6221):509–14. Available from: http://dx.doi.org/10.1126/science.aaa1465

62. Fiesler C, Proferes N. “Participant” perceptions of Twitter research ethics. Social Media + Society [Internet]. 2018 Jan 1;4(1):2056305118763366. Available from: https://doi.org/10.1177/2056305118763366

63. Politou E, Alepis E, Patsakis C. Forgetting personal data and revoking consent under the GDPR: challenges and proposed solutions. J Cyber Secur [Internet]. 2018 Jan 1;4(1). Available from: https://academic.oup.com/cybersecurity/article/4/1/tyy001/4954056

64. Mostert M, Bredenoord AL, Biesaart MCIH, van Delden JJM. Big data in medical research and EU data protection law: challenges to the consent or anonymise approach. Eur J Hum Genet [Internet]. 2016 Jul;24(7):956–60. Available from: http://dx.doi.org/10.1038/ejhg.2015.239

Appendix A: Full list of articles included in methods review

Aelst PV, Erkel P van, D'heer E, Harder RA. Who is leading the campaign charts? Comparing individual popularity on old and new media. Inf Commun Soc [Internet]. 2017 May 4;20(5):715–32. Available from: http://dx.doi.org/10.1080/1369118X.2016.1203973

Albrechtslund A-MB. Negotiating ownership and agency in social media: community reactions to Amazon's acquisition of Goodreads. First Monday [Internet]. 2017;22(5). Available from: http://journals.uic.edu/ojs/index.php/fm/article/view/7095/6161

Barnard SR. Tweeting #Ferguson: Mediatized fields and the new activist journalist. New Media & Society [Internet]. 2017 Jun 19;1461444817712723. Available from: https://doi.org/10.1177/1461444817712723

Bender S. "Happy to provide the knives": Governmentality and threats of violence via social media in the case of Roosh V and Return of Kings. First Monday [Internet]. 2017 Feb 15;22(3). Available from: https://journals.uic.edu/ojs/index.php/fm/article/view/6945

Berg J. The dark side of e-petitions? Exploring anonymous signatures. First Monday [Internet]. 2017 Jan 20;22(2). Available from: http://uncommonculture.org/ojs/index.php/fm/article/view/6001

Bock MA, Figueroa EJ. Faith and reason: An analysis of the homologies of Black and Blue Lives Facebook pages. New Media & Society [Internet]. 2017 Nov 16;1461444817740822. Available from: https://doi.org/10.1177/1461444817740822

Boukes M, Trilling D. Political relevance in the eye of the beholder: Determining the substantiveness of TV shows and political debates with Twitter data. First Monday [Internet]. 2017 Apr 3;22(4). Available from: https://uncommonculture.org/ojs/index.php/fm/article/view/7031

Chen W, Tu F, Zheng P. A transnational networked public sphere of air pollution: analysis of a Twitter network of PM2.5 from the risk society perspective. Inf Commun Soc [Internet]. 2017 Jul 3;20(7):1005–23. Available from: http://dx.doi.org/10.1080/1369118X.2017.1303076

Chow-White P, Struve S, Lusoli A, Lesage F, Saraf N, Oldring A. "Warren Buffet is my cousin": shaping public understanding of big data biotechnology, direct-to-consumer genomics, and 23andMe on Twitter. Inf Commun Soc [Internet]. 2018 Mar 4;21(3):448–64. Available from: https://doi.org/10.1080/1369118X.2017.1285951

D'heer E, Vandersmissen B, De Neve W, Verdegem P, Van de Walle R. What are we missing? An empirical exploration in the structural biases of hashtag-based sampling on Twitter. First Monday [Internet]. 2017 Jan 16;22(2). Available from: http://www.firstmonday.dk/ojs/index.php/fm/article/view/6353/5758

Duong JA, Zeller F. Tracking the imagined audience: a case study on Nike's use of Twitter for B2C interaction. First Monday [Internet]. 2017;22(5). Available from: https://uncommonculture.org/ojs/index.php/fm/article/view/6607/6190

Engesser S, Ernst N, Esser F, Büchel F. Populism and social media: how politicians spread a fragmented ideology. Inf Commun Soc [Internet]. 2017 Aug 3;20(8):1109–26. Available from: https://doi.org/10.1080/1369118X.2016.1207697

Enli G, Simonsen C-A. "Social media logic" meets professional norms: Twitter hashtags usage by journalists and politicians. Inf Commun Soc [Internet]. 2018 Aug 3;21(8):1081–96. Available from: https://doi.org/10.1080/1369118X.2017.1301515

Gutiérrez-Martín A, Torrego-González A. The Twitter games: media education, popular culture and multiscreen viewing in virtual concourses. Inf Commun Soc [Internet]. 2018 Mar 4;21(3):434–47. Available from: https://doi.org/10.1080/1369118X.2017.1284881

Halse SE, Tapia A, Squicciarini A, Caragea C. An emotional step toward automated trust detection in crisis social media. Inf Commun Soc [Internet]. 2018 Feb 1;21(2):288–305. Available from: https://doi.org/10.1080/1369118X.2016.1272618

Haustein S, Peters I, Sugimoto CR, Thelwall M, Larivière V. Tweeting biomedicine: an analysis of tweets and citations in the biomedical literature. J Assn Inf Sci Tec [Internet]. 2014 Apr 26;65(4):656–69. Available from: http://doi.wiley.com/10.1002/asi.23101

Hermida A, Hernández-Santaolalla V. Twitter and video activism as tools for counter-surveillance: the case of social protests in Spain. Inf Commun Soc [Internet]. 2018 Mar 4;21(3):416–33. Available from: https://doi.org/10.1080/1369118X.2017.1284880

Honig CDF, MacDowall L. Spatio-temporal mapping of street art using Instagram. First Monday [Internet]. 2017 Feb 11;22(3). Available from: http://journals.uic.edu/ojs/index.php/fm/article/view/7072

Jordan K. Examining the UK higher education sector through the network of institutional accounts on Twitter. First Monday [Internet]. 2017;22(5). Available from: http://ojphi.org/ojs/index.php/fm/article/view/7133

Liang H, Shen F, Fu K-W. Privacy protection and self-disclosure across societies: a study of global Twitter users. New Media & Society [Internet]. 2016 May 12;19(9):1476–97. Available from: https://doi.org/10.1177/1461444816642210

Lycarião D, dos Santos MA. Bridging semantic and social network analyses: the case of the hashtag #precisamosfalarsobreaborto (we need to talk about abortion) on Twitter. Inf Commun Soc [Internet]. 2017 Mar 4;20(3):368–85. Available from: https://doi.org/10.1080/1369118X.2016.1168469

Matamoros-Fernández A. Platformed racism: the mediation and circulation of an Australian race-based controversy on Twitter, Facebook and YouTube. Inf Commun Soc [Internet]. 2017 Jun 3;20(6):930–46. Available from: https://doi.org/10.1080/1369118X.2017.1293130

McGregor SC, Lawrence RG, Cardona A. Personalization, gender, and social media: gubernatorial candidates' social media strategies. Inf Commun Soc [Internet]. 2017;20(2):264–83. Available from: https://www.tandfonline.com/doi/abs/10.1080/1369118X.2016.1167228

Molyneux L, Holton A, Lewis SC. How journalists engage in branding on Twitter: individual, organizational, and institutional levels. Inf Commun Soc [Internet]. 2018 Oct 3;21(10):1386–401. Available from: https://doi.org/10.1080/1369118X.2017.1314532

Ogan C, Varol O. What is gained and what is left to be done when content analysis is added to network analysis in the study of a social movement: Twitter use during Gezi Park. Inf Commun Soc [Internet]. 2017 Aug 3;20(8):1220–38. Available from: https://doi.org/10.1080/1369118X.2016.1229006

Oz M, Zheng P, Chen GM. Twitter versus Facebook: comparing incivility, impoliteness, and deliberative attributes. New Media & Society [Internet]. 2017 Dec 31;1461444817749516. Available from: https://doi.org/10.1177/1461444817749516

Parsloe SM, Holton AE. #Boycottautismspeaks: communicating a counternarrative through cyberactivism and connective action. Inf Commun Soc [Internet]. 2018 Aug 3;21(8):1116–33. Available from: https://doi.org/10.1080/1369118X.2017.1301514

Patton DU, Lane J, Leonard P, Macbeth J, Smith Lee JR. Gang violence on the digital street: Case study of a South Side Chicago gang member's Twitter communication. New Media & Society [Internet]. 2016 Jan 25;19(7):1000–18. Available from: https://doi.org/10.1177/1461444815625949

Raynauld V, Richez E, Boudreau Morris K. Canada is #IdleNoMore: exploring dynamics of Indigenous political and civic protest in the Twitterverse. Inf Commun Soc [Internet]. 2018 Apr 3;21(4):626–42. Available from: https://doi.org/10.1080/1369118X.2017.1301522

Rehman S, Lyons K, McEwen R, Sellen K. Motives for sharing illness experiences on Twitter: conversations of parents with children diagnosed with cancer. Inf Commun Soc [Internet]. 2018 Apr 3;21(4):578–93. Available from: https://doi.org/10.1080/1369118X.2017.1299778

Romney M, Johnson RG, Roschke K. Narratives of life experience in the digital space: a case study of the images in Richard Deitsch's single best moment project. Inf Commun Soc [Internet]. 2017 Jul 3;20(7):1040–56. Available from: https://doi.org/10.1080/1369118X.2016.1203976

Sachdeva S, McCaffrey S, Locke D. Social media approaches to modeling wildfire smoke dispersion: spatiotemporal and social scientific investigations. Inf Commun Soc [Internet]. 2017 Aug 3;20(8):1146–61. Available from: https://doi.org/10.1080/1369118X.2016.1218528

Shin J, Jian L, Driscoll K, Bar F. Political rumoring on Twitter during the 2012 US presidential election: rumor diffusion and correction. New Media & Society [Internet]. 2016 Mar 8;19(8):1214–35. Available from: https://doi.org/10.1177/1461444816634054

Skrubbeltrang MM, Grunnet J, Tarp NT. # RIPINSTAGRAM: Examining user's counter-narratives opposing the introduction of algorithmic personalization on Instagram. First Monday [Internet]. 2017;22(4). Available from: http://www.firstmonday.dk/ojs/index.php/fm/article/view/7574

Su LY-F, Cacciatore MA, Liang X, Brossard D, Scheufele DA, Xenos MA. Analyzing public sentiments online: combining human- and computer-based content analysis. Inf Commun Soc [Internet]. 2017 Mar 4;20(3):406–27. Available from: https://doi.org/10.1080/1369118X.2016.1182197

Williams A, Gonlin V. I got all my sisters with me (on Black Twitter): second screening of How to Get Away with Murder as a discourse on Black Womanhood. Inf Commun Soc [Internet]. 2017 Jul 3;20(7):984–1004. Available from: https://doi.org/10.1080/1369118X.2017.1303077

Yang S, Quan-Haase A, Rannenberg K. The changing public sphere on Twitter: Network structure, elites and topics of the #righttobeforgotten. New Media & Society [Internet]. 2016 Jun 14;19(12):1983–2002. Available from: https://doi.org/10.1177/1461444816651409

Young R, Tully M, Dalrymple KE. #Engagement: use of Twitter chats to construct nominal participatory spaces during health crises. Inf Commun Soc [Internet]. 2018 Apr 3;21(4):499–515. Available from: https://doi.org/10.1080/1369118X.2017.1301518

Zelenkauskaite A, Niezgoda B. “Stop Kremlin trolls:” Ideological trolling as calling out, rebuttal, and reactions on online news portal commenting. First Monday [Internet]. 2017;22(5). Available from: http://uncommonculture.org/ojs/index.php/fm/article/view/7795

Zhang Y, Wells C, Wang S, Rohe K. Attention and amplification in the hybrid media system: the composition and activity of Donald Trump’s Twitter following during the 2016 presidential election. New Media & Society [Internet]. 2017 Dec 4;1461444817744390. Available from: https://doi.org/10.1177/1461444817744390