

The development of social science research methods influenced by social media

Lieke A.C. Heszen

Pedagogical and Educational Sciences

Abstract

The introduction of social media during the past 15 years has introduced new ways to produce mass self-report data, while creating opportunities and challenges for social researchers. The first part of this literature review provides an overview of social media and presents contrasting views regarding what the focus of research in this area should be. Some researchers claim that the focus of such research should be on creating new methods in a world of knowing capitalism. Others believe that the focus should be on how the research process is distributed among various agents. The second part of this review focuses on how big data and social media analysis contribute to research methodology. The final part of this review explains the various challenges introduced by big data. Digital observatories, such as COSMOS, seem to deal with some of these challenges, such as volume, velocity and variety. Criticism of big data focuses mainly on ethical issues, such as privacy concerns. When using big data for social research, it is important to have a clear picture of its possibilities and limitations. The development of social media and the availability of data will be important in the future of digital social research.

Keywords: *big data, social media, analysis, methods, knowing capitalism, digitization*

Introduction

The current digitization of data and communications creates opportunities and challenges for social research. There are varying opinions regarding the actual implications of this digitization for social research (Marres, 2012). Some are afraid that the new technology and current digitization pose a threat to the already established research methodology (Crompton, 2008; Marres, 2012). These pessimists believe that the new digital technology will lead to the relocation of social research to big corporate research centers (Marres, 2012). Others believe this is an era in which social research methods should be reconsidered and that the ways researchers should relate to the new trends should be rethought (Savage & Burrows, 2007). The optimists believe that digital technology will improve social research capacities (Marres, 2012).

There are three important trends in this digitization (Marres, 2012). First of all, there is a "proliferation of new devices, genres and formats for the documentation of social life" (Marres, 2012, p. 142). Social Networking Sites (SNS) and micro-blogging services, such as Twitter, can be used at almost any given moment due to the easy access to devices such as smart phones and tablets. In addition, the World Wide Web is easily accessible. People can share their perceptions of particular events at any given moment (Edwards, Housley, Williams, Sloan, & Williams, 2013).

The second trend in digitization is "the routine generation of data about social life as a part of social life" (Marres, 2012, p. 142). The new social media technology is part of the Web 2.0 technologies (Edwards et al., 2013). Web 2.0 refers to an interactional web that is the second stage in the development of the Web (Edwards et al., 2013; Murugesan, 2007). Web 2.0 consists of a group of technologies, business strategies and social trends. Applications

such as YouTube and Facebook are also part of the Web 2.0, emphasizing the social basis of this type of web (Edwards et al., 2013; Murugesan, 2007). By using these Web 2.0 technologies, people create data as part of their daily lives, which is the second trend (Marres, 2012).

The third trend involves the new tools and online platforms that have been developed for the analysis of the new digital social data. The digital tools that have been developed for analyzing, recording and visualizing social life create the possibility for analysis of social behavior in real time. Most online platforms that have been developed also facilitate research by providing tools for analysis (Marres, 2012).

In this review, the focus will mainly be on how social media as a part of the digitization have influenced research methods in the social sciences. The possible ways of communicating through social media have expanded, introducing ways to produce mass self-report data. The analysis of this data via computational tools is called social media analysis. The interesting thing about this type of analysis is that it captures data that occur naturally in real time at a population level (Edwards et al., 2013). In contrast to these possibilities, the expansion of social media also introduces challenges for social researchers (Procter, Vis, & Voss, 2013). Some for instance believe that the digitization will possibly lead to the privatization of social research and the marginalization of (for example) sociological research (Savage & Burrows, 2009).

An example of how social media analysis can be applied in social research is an article by Procter et al. (2013) in which they analyzed Twitter tweets that were sent during the riots in England in 2011. Twitter is an open platform which enables users to share messages of a maximum of 140 characters with an unlimited number of persons who are registered as their "followers." The researchers were able to analyze how certain flows of information circulated and how significant these flows were. The available data for this experiment consisted of 2.6 million tweets. These large quantities of data are called *big data* (Burnap, Avis, & Rana, 2013; Chen et al., 2013; Procter et al., 2013).

There is as yet no consensus definition of big data (Chen et al., 2013). Boyd and Crawford (2012) define big data as a phenomenon that relies on an interaction of technology, analysis and mythology. By "technology" they mean the maximization of computational power and algorithmic accuracy. The "analysis" part of the equation consists of the ability to use large datasets to find patterns. Finally, "mythology" refers to the belief in the opportunities created by big datasets (Boyd & Crawford, 2012). Moreover, three important aspects of big data have been identified: volume, velocity and variety. All three aspects create challenges for research, which will be discussed in detail in the third part of this review (Chen et al., 2013).

In this literature review, the way social media influences methodology in the social sciences will be discussed. The main focus will be on social media analysis and the second trend in digitization: the routine generation of data. The first part of this paper will discuss the context and focus of the digital research domain. Next, the discussion will turn to how exactly social media analysis and the big data generated by social media contribute to social research. In the third part of this paper the specific challenges created by big data will be discussed. Finally, a conclusion will be drawn regarding where social research methods stand at the moment and their future prospects.

The context and focus of the digital research domain

One of the most important articles written about how social research methods should develop was written by Savage and Burrows (2007), who contend that there is a *coming crisis of empirical sociology*. An important concept in this coming crisis is *knowing capitalism* (Thrift, as cited in Savage & Burrows, 2007). "Knowing capitalism" is defined as a proliferation of

information circuits that are part of various kinds of information technologies (Savage & Burrows, 2007). According to Savage and Burrows we are currently in an age of "knowing capitalism." Social and cultural digitization play an important role in this age (Savage & Burrows, 2009). Due to the changing world and circumstances, sociologists have to rethink the methods that they use in their research. If they fail to do this, sociology could become marginalized as a social research area. That is what these same authors refer to as "the coming crisis of empirical sociology" (Savage & Burrows, 2007).

According to Savage and Burrows (2007), the sample survey is no longer the most relevant way to conduct social research. The sample survey is a method that enables researchers to generalize and predict based on statistical procedures. According to Savage and Burrows the sample survey is no longer a major research method in the social sciences. The most important reason for this development in the context of this literature review is the new routinely gathered data about whole populations (i.e., big data). With this kind of information, research can be conducted without using samples. However, sample research will probably remain important due to the fact that this new data also has limits (Savage & Burrows, 2007).

Another way of looking at the development of the digital research domain is introduced by Marres (2012), who believes the development should focus on how methods will be distributed among various agents. This was also mentioned by Savage and Burrows (2007) who explained that future social research will be mainly conducted in a few select research centers. They also mention that companies such as amazon.com, can perform analyses without the use of a sample survey or the expertise of a social researcher. Savage and Burrows see this as the probable marginalization of sociology. According to Marres this is not where the focus of the research domain should be. She introduces a new topic: "the redistribution of research." It is important to determine how exactly the various actors in the process of research are related to one another, because research is a more complex process than it used to be. Because of digitization new agencies, such as SNS-users and devices, play a role in the research process. The research process is now a capacity shared by agencies and actors such as researchers, users and devices. Within this view, the question should be asked how the researcher should intervene in this transactional process of research. This remains an unanswered question, which should in the future be the focus in the digital social research domain (Marres, 2012).

The contribution of social media and big data to social research

Even though Savage and Burrows (2007) make a very interesting claim, there seems to be a flaw in their argument. The claim of the coming crisis of empirical sociology does not fully consider the expansion of accessible, free social media (Housley, Williams, Williams, & Edwards, 2013). However, in their further reflections they do briefly mention the possibility of gathering and analyzing enormous amounts of data from SNS (Savage & Burrows, 2009). Specifically social media analysis and big data create numerous opportunities for research objectives (Housley et al., 2013).

There are several positive contributions that social media analysis has made to social research methodology. First of all, by using social media, researchers can reach people who would otherwise be difficult to access. In conventional surveys, some groups tend to be underrepresented. In addition to this, researchers can study online networks in addition to offline networks. By doing this, researchers can establish additional influential factors and their place in offline networks. Finally, one could argue that the use of social media raises questions different from those asked in previous, more traditional research. These new questions mainly concern social structure, organization, change and identity. Some of the questions asked require new methods, such as social media analysis (Edwards et al., 2013). For example, the research on information flows on Twitter, as mentioned in the introduction

of this review, was only possible by the use of social media analysis. These new research methods make it possible to tackle more complex research challenges in the future (Procter et al., 2013).

Also the big data routinely generated through social media contributes to social research by introducing several opportunities. By using big data, information about (for example) society can be revealed. This makes an adequate response to social change possible (Chen et al., 2013). In addition the digital data can easily be kept, categorized and analyzed (Housley et al., 2013). Moreover, this data offers more insight into responses to events than conventional data does. Big social media data for example provides more insight into societal change. The large amount of data can be linked to certain events in history, thus offering valuable information. In addition, researchers can establish how social processes develop at the level of the population (Edwards et al., 2013).

Even though social media and its big data seem to contribute to social research in a positive way, they also have several disadvantages. First of all, the big data available through social media is of lower reliability than data generated by survey research or interviewing (Dutton, 2013; Edwards et al., 2013). Secondly, social media data in itself does not usually contain adequate information. For example tweets and statuses typically consist of only a few sentences. In addition, when conducting research there is only limited data available on SNS (Edwards et al., 2013). Twitter for example only releases 1% of its data for research use (Procter et al., 2013). This makes research more dependent on external factors. Finally, the social context is still needed when social media's big data are being analyzed. When focusing on social media statuses or tweets and interpreting them, the specific social events that preceded them and the social context are still important. Otherwise researchers cannot establish a causal relationship (Dutton, 2013).

In conclusion, both social media analysis and big data seem to have made a positive contribution to social research. However, considering their disadvantages, researchers should in the future benefit from the new opportunities to analyze social media, while at the same time effectively relying on methods based on human expertise (Procter et al., 2013). Also, there is no single "appropriate" technique for all research. Researchers should be able to tackle a range of topics that require a number of different methods (Crompton, 2008).

Big data challenges

Technical challenges

As mentioned in the introduction of this review, big data presents three main challenges. The first challenge is posed by the data volume. Web 2.0 technologies have more users than ever before, creating huge amounts of data. This requires sufficient computational tools. Secondly, the velocity challenge introduces the problem of how fast the data are created and updated. Research applications should be able to store and process the data at a fast rate. The final challenge is posed by the variety of the data. Big data comes from various sources, which causes heterogeneous formats and models to exist within a database. Several types of data exist: structured, semi-structured and unstructured data. Technology can deal with each of these types separately, but has yet to find a seamless way of integrating them (Chen et al., 2013). This might make analysis and finding correlations more difficult (Burnap et al., 2013; Chen et al., 2013).

The Cardiff Online Social Media Observatory (COSMOS) deals with the volume and variety challenge. COSMOS is a new technological facility, inspired by digitization, that attempts to be a resource for all social researchers, giving them the chance to capture, analyze and visualize data gathered online, mainly on SNS (Burnap et al., 2013; Edwards et al., 2013). The exact way in which COSMOS deals with the variety of data is by creating a relational

database in which common types of data are identified. For example, both Twitter and Facebook use the concept "username." In this relational database, extraordinary types of data are also identified, such as the hashtag on Twitter and "likes" on Facebook. COSMOS deals with the volume challenge by providing access to "big social data" that it mines on a daily basis for research purposes (Burnap et al., 2013).

Ethical challenges

An additional important challenge for big data is ethical in nature. The standards that were used in the past for conventional methods might not apply to the new analysis of social media (Housley et al., 2013). For example "informed consent" poses a problem. Informed consent should actually be obtained for every new analysis with the already existent data (Edwards et al., 2013). Some researchers who use existent data consider previously obtained informed consent to remain valid (Krotoski, 2012). Due to the fact that data is already available, such researchers have the opportunity to collect information without informed consent (Dutton, 2013).

Also, many social media users are unaware of the big data research and, when publishing information online, they do not fully understand the possible ways their information will be used (Burnap et al., 2013). The data on SNS does not belong to the social media users, but to companies such as Facebook. This means that SNS can distribute this data without consent from the people who published the information (Krotoski, 2012). For example, Procter et al. (2013) received 2.6 million tweets from Twitter, including user names, without the informed consent of the users. However, some SNS do make it difficult to store this raw data without permission (Burnap et al., 2013).

In addition, privacy might be compromised by the huge amounts of data that can be gathered and linked together. The chance of revealing a participant's identity is higher than usual (Edwards et al., 2013). In addition, with this new type of data, researchers can easily violate a person's privacy by revealing their identity. This may expose the participant to a situation in which harmful information is revealed (Krotoski, 2012)

Conclusion and future prospects

Social media use has recently exploded, exposing researchers to both opportunities and challenges. The first part of this literature review focused on the context and the focus of the digital research domain. Savage and Burrows (2007) claimed that, in a world of "knowing capitalism," the focus should be on new methods. Marres (2012) in contrast believes that the focus in the new digital research domain should be on how the research process is distributed among various agents (Marres, 2012). One could conclude that there are various views on where the focus should be with respect to this research domain.

The second part of this review focused on social media analysis and big data's contribution to digital social research. Considering the advantages and disadvantages of social media analysis and big data, researchers should in the future benefit from the new opportunities to analyze social media, while at the same time effectively relying on methods based on human expertise (Procter et al., 2013). The final part of this review explained the various challenges introduced by big data: volume, velocity and variety. Digital observatories, such as COSMOS, seem to deal with some of these challenges. Criticism of big data mainly revolves around ethical concerns. The question of whether researchers can use all accessible data for research is unresolved (Burnap et al., 2013).

In the future, the development of methods will mostly depend on how social media will develop (Edwards et al., 2013). The generation of data will increase (Chen et al., 2013). However, future access to this data is not guaranteed. An open society and stable data are critical for social media analysis (Edwards et al., 2013). Even when these data streams remain

accessible, the querying and indexing of the data should be optimized for research. A lot of work still remains, even though much has already been achieved (Chen et al., 2013).

This literature review provides an overview of how the social media have already influenced the development of social science research methods. However, due to the fact that social media and the web continue to evolve at a rapid pace, we can only speculate about further development. One thing seems clear for now: The new digital social research methods can be a useful complement to traditional research methods (Edwards et al., 2013).

References

- Boyd, D., & Crawford, K. (2012). Critical questions for big data. *Information, Communication & Society, 15*, 662-679. doi:10.1080/1369118X.2012.678878
- Burnap, P., Avis, N. J., & Rana, O. F. (2013). Making sense of self-reported socially significant data using computational methods. *International Journal of Social Research Methodology, 16*, 215-230. doi:10.1080/13645579.2013.774174
- Chen, J., Chen, Y., Du, X., Li, C., Lu, J., Zhao, S., & Zhou, X. (2013). Big data challenge: A data management perspective. *Frontiers of Computer Science, 7*, 157-164. doi:10.1007/s11704-013-3903-7
- Crompton, R. (2008). Forty years of sociology: Some comments. *Sociology, 42*, 1218-1227. doi:10.1177/0038038508096942
- Dutton, W. H. (2013). The social shaping of digital research. *International Journal of Social Research Methodology, 16*, 177-195. doi:10.1080/13645579.2013.774171
- Edwards, A., Housley, W., Williams, M., Sloan, L., & Williams, M. (2013). Digital social research, social media and the sociological imagination: Surrogacy, augmentation and re-orientation. *International Journal of Social Research Methodology, 16*, 245-260. doi:10.1080/13645579.2013.774185
- Housley, W., Williams, M., Williams, M., & Edwards, A. (2013). Special issue: Introduction. *International Journal of Social Research Methodology, 16*, 173-175. doi:10.1080/13645579.2013.774164
- Krotoski, A. K. (2012). Data-driven research: Open data opportunities for growing knowledge, and ethical issues that arise. *Insights, 25*, 28-32. doi:10.1629/2048-7754.25.1.28
- Marres, N. (2012). The redistribution of methods: On intervention in digital social research, broadly conceived. *The Sociological Review, 60*, 139-165. doi:10.1111/j.1467-954X.2012.02121.x
- Murugesan, S. (2007). Understanding Web 2.0. *IT professional, 9*(4), 34-41. doi:10.1109/MITP.2007.78
- Procter, R., Vis, F., & Voss, A. (2013). Reading the riots on Twitter: Methodological innovation for the analysis of big data. *International Journal of Social Research Methodology, 16*, 197-214. doi:10.1080/13645579.2013.774172

Savage, M., & Burrows, R. (2007). The coming crisis of empirical sociology. *Sociology*, *41*, 885-899. doi:10.1177/0038038507080443

Savage, M., & Burrows, R. (2009). Some further reflections on the coming crisis of empirical sociology. *Sociology*, *43*, 762-772. doi:10.1177/0038038509105420