# What are the Threats and Potentials of Big Data for Qualitative Research?

Author: Kathy A. Mills.

**Qualitative Research. 2017. Issn:** 1741-3109

**DOI:** 10.1177/1468794117743465

## Background

The idea of big data has captured the imagination of researchers worldwide, with a proliferation of digital media rendering extremely large datasets more rapidly searchable, analyzable, and shareable. Big data is defined here as digitally encoded information of unprecedented scope or scale about a phenomenon, which has relationality with other networked data. Such data require analysis or calculation to be put to use for scientific work or knowledge. The increased use of the internet and mobile technologies for human communication has generated big data that is connected, traceable, and more complex to analyze than conventional statistical analysis software permits (Snijders et al., 2012). Big data has been taken up in multiple ways as a "cultural", "technological", or "scholarly phenomenon", often with an aim to maximize computation capability and algorithmic precision to harvest, probe, and manipulate sizable data sets to solve problems or make claims (Boyd and Crawford, 2012: 663). At times, the pursuit of big data is driven by the epistemic assumption that expansive data sets offer superior forms of intelligence and erudition—a view that has been challenged in the existing literature.

Big data are impressive in terms of digital ubiquity because a great range of human behaviors are captured digitally. Examples include business and operational data, mobile data (e.g. location), social network data, public data, commercial data, streaming data, and sensory data from the Internet of Things (IoT). The research potentials of big data have been explored in relation to immigration control and border security (Ajana, 2015), civil strife management (Nardulli et al., 2015), human geography (Kitchin, 2013), political science (Clarke and Margettes, 2014), research on children and media cultures (Montgomery, 2015), ecometrics (O'Brien et al., 2015), global league tables in education research (Crossley, 2014), business scholarship (Frizzo-Barker et al., 2016), and sociology (McFarland et al., 2016).

Examples of big data include web and mobile analytics, visualization of large data sets, machine learning, sentiment analysis and opinion mining, computer-assisted content analysis, natural language processing, automated data aggregation and mining, and large social media networks (Parks, 2014; Lohmeier, 2014). For example, Twitter, Facebook and Apple are companies that keep big data, and some companies grant researchers access to subsets of data, such as iScience Maps™ for Twitter (See: Reips and Garaizar, 2013). Social media sites generate large bursts of data of short-lived relevance about a significant but not exhaustive number of users, with varied levels of accessibility to certain groups for differing uses.

The spread of mobile technologies has assisted the scope of big data, since these communication technologies are owned by users throughout most parts of the world. Digital metadata, like footprints, mark time and place, creating ongoing records of social, communication and location activity. Google is one of the largest keepers of big data, which provides an open-access interface called Google Insights to enable analysis. Digital data are becoming computation intensive and data intensive, and its manipulation often requires significant logistical challenges (Borgman, 2015; Meyer, 2009; Laney, 2001). Screen-scraping is used to extract information from internet sites, and data is collected and used for social purposes that range from gene sequencing to consumer behavior, and from learning analytics to predictive analytics (Bail, 2014; Siegel, 2013).

This paper asks about the role of qualitative research in a world in which there are data of massive breadth across so many fields and spheres of human activity. Within the academic community, some have argued that big data renders small-scale research, commonly used in the social sciences and humanities, potentially at risk (Alberts, 2012; Berlekamp, 2012; Meyer, 2009). If social and behavioral data, that were previously the locus of much qualitative research have been 'datified', is the role of the qualitative researcher losing a significant foothold (Strong, 2013). This paper argues that before researchers blindly follow the big data trends, questions need to be asked about the accessibility, ethics, and utility of big data. What is the scale of analysis necessary to understand phenomena in the area of research interest?

Is big data incommensurable or diametrically opposed to the values of qualitative research? Should big data and qualitative research be seen as complementary and suited to particular types of social questions and problems? What exemplars do we have to support the integration of qualitative methods with big data in research investigations? This paper examines both the limits and potentials of big data for qualitative researchers. After historically contextualizing big data, the article discusses assumptions about who has access to big data and who misses out. Issues of ethics and privacy are examined in a risk society, such as surveillance, data ownership, and the economics and management of big data repositories. The paper explores some of the potentials of combining the strengths of big data with those of qualitative research, such as combining automated tools for the analysis of big data with the interpretative theories or cultural frames of reference generated through qualitative research to contextualize the social context in which the data were produced.

## Big Data – A Short History and Examples

Borgman (2015) argues that big data is not necessarily a new concept, but sees the current distinction between big data and small data as somewhat analogous to the distinction made in the 1960s about big science and small science by De Solla Price

(1963). Big science, De Solla Price (1963) argues, is comprised of collaborative efforts by networks of researchers who exchange information both informally and formally worldwide. The adjective *big* essentially denotes the maturity of science, rather than the quantifiable size of scientific research projects. *Little science* refers to small-scale work by many individual researchers or small teams, rather than scientific enterprises, which aims to generate research theories or methods to address specific local research problems. Little science is typically more open to novel and diverse methodological approaches, often locally owned and analyzed.

While we are currently witnessing a "Big Data Movement" (Parks, 2014: 355) historically there have existed much larger data sets than what some currently regard as *big data*, such as census data (Boyd and Crawford, 2011). There are diary studies, such as the Mass Observation project of 1937 to the early 1950s, which aimed to record everyday life in Britain through a national panel of diarists composed of men and women. The diaries were collected and held by the research team in monthly intervals, and the diaries varied greatly in form, detail, and length. These extensive records of significant scope and duration provided a detailed view of early 20th century Britain (Bancroft et al., 2014).

A relevant example of Big Data in the 1960s is the International Time-Use Study of 1965 by Szalai (1972), which involved 2000 participants, ages 18–64 from 12 countries. Participants kept continuous logs to map time use over the course of day, which later expanded to include budget spending, wages, transportation, leisure and other dimensions of time and economy. These kinds of ‘*big data’* diary studies continued into the late 20$^{th}$ century, such as project SIGMA (1986–1994) by Coxon (1993), which used 1035 diaries to chronicle the personal experiences of gay men. While the digital turn has rendered manual, large-scale recording of daily activities virtually obsolete, situating the current big data hype within a historical context demonstrates that the largeness of big data is not the main development. Datasets generated through large networks by research participants have a long history. However, there are new political, ethical, and epistemic questions about data production, privacy, access, and ownership (Bancroft et al., 2014). A key feature of big data is that it is not intentionally engendered by researchers to test a theory, rather, new data analytics often aims to gain insights that emerge from existing data, with the analysis typically occurring after data is already collected and stored (Chandler, 2015).

**Big Data and Assumptions about Access**

There has been an assumption that big data afford easy access to large amounts of useful data. Digital data generated daily are potentially very useful, particularly the big data advantage of being almost completely networked between multiple things and people (Boyd and Crawford, 2011). However, there are hard questions about who gets access to big data, under what terms, and for what purposes (Qui, 2015). Complex research questions about human behavior and society require identification of patterns within relevant data, and these data are typically owned by individuals and organizations. Access may depend on a convergence of interests or goodwill of the owners of different data sources. The use of big data is similarly encumbered by established institutional protocols and issues of ownership, human relationships, and new implications for research ethics that are only beginning to be understood.

The issue of open access to, or sharing of data is one that has become contentious. Despite the claims about big data and calls to make data accessible to secondary users after publication, many research areas are data-poor fields where good data are hard-won and precious (Sawyer, 2008). Borgman (2015) argues that many scholars are typically rewarded for generating original data, and few would disagree that competitive research funding is often awarded to those who are gathering something fresh, or analyzing data in new ways to solve complex problems using innovative techniques. Reusable data is becoming increasingly important in research of astronomy, social media, town modelling, climate research, and dry lab research in the biosciences, to name a few fields. Yet the most competitive grants and publications are those that address topics and provide solutions for problems that require new data, or data collected in new ways.

The trend toward large repositories of open access research data, such as national data commons, is becoming a requirement of some funding bodies, with data citation indexes potentially benefiting from the development of metrics services associated with the use of these data repositories. Conversely, a sizable proportion of big data will remain proprietary data. Certain data cannot be released by law, embargo periods may apply that delay the use of data beyond its period of relevance, and individual human research data may be too sensitive (Borgman, 2015). Additionally, entities that have the power to release data may see that the risks and the hidden costs necessary to make the data useable and interpretable outweigh the benefits of releasing data for use by others. Big data require vast amounts of attention to maintaining and organizing metadata to be able to reuse the data. Furthermore, big data rendered in digital form are potentially more short-lived than cultural artifacts and even paper records, due to the rapid change of technologies and the software used to store and to analyze them. In addition, the further from its origin that data are extracted and applied, the further data are open to issues of ethics, access, and decontextualization.

Researchers who have attempted to share big data have also demonstrated the difficulties that arise when making data available to other researchers in ethically responsible ways, such as by de-identifying the original data for use by other researchers. Daries and colleagues (2014) shared data generated from MOOCs—massive open online courses—with the twofold goal to permit other researchers to: (1) reproduce the outcomes of the analysis and, (2) perform new analyses beyond the initial research. They were required to de-identify the data to protect student privacy under the district regulatory regime, but when they compared statistics on the original data set and the de-identified data, there were major discrepancies. For example, the original study found that 5% of the students enrolled had received certificates, while the curated data set cut that percentage by half.

Such modifications distort the ‘truth’ of the original dataset considerably, with analysis of the modified data sets resulting in incorrect statistics due to the incompleteness of the information, and opening up critical uncertainties about the utility of the open, reusable data for replication or innovative analysis by other researchers (Daries et al., 2014). This case illustrates the fundamental tension between generating and curating data sets that meet the ethics requirement of anonymity, while providing useful, openly accessible data to advance new knowledge.

Similarly, there are big data divides entangled with issues of access: between those who have access to big data and those who don’t, between those who have the computational expertise and means to analyze it and those who don’t, and between those who need essential community collections of big data to answer their questions and those who simply don’t. Even when researchers have easier access to shared big data via national data repositories, there is the problem of misinterpretation and misuse. Mayer-Schönberger and Cukier (2013) warn that underlying data used to generate knowledge may be big, but could be used inappropriately. Big data, like small data, may be biased, misused, or be misleading, and fail to capture what authorities purport that it quantifies. However, when compared to the use of small data, the consequences of misuse will be much larger. And while big data has potential for optimizing and advancing the efficiency of research and scholarship, more than ever before there is the need for reason, theorization, problem-solving, originality, and social justice in determining what questions can be served by the data, and whose interests they serve.

The ready supply of big data does not mean that rare datasets no longer exist or are no longer needed in many fields of research. There will always be the need for difficult-to-obtain data. A ready supply of statistics and the vast scale of data in the digital world is not particularly useful for answering the kinds of research questions that people in the social sciences are asking. For example, how can big data help us understand remote Indigenous communities and their cultural beliefs and epistemologies? How can we study rare chromosomal disorders through big data, since there are very few people in the world who have these conditions? (Gilmore, 2014). While many rural contexts cannot escape from digital transformation, there are likely to be digital data research 'black spots'—research about people who have limited access to the Internet, such as those in remote rural areas, the very elderly, the very ill, the disabled, children, refugees, people living in infrastructure that has been destroyed by natural disasters, and babies too young to leave a digital footprint, to name a few examples.

Most importantly, the value of data is not tied to the data itself, but to what questions can be answered by those data. Even when researchers think they are asking valuable questions about data and publish the findings, millions of research articles that are openly accessible to the public on Google scholar remain uncited by people beyond their own research teams for years, raising questions about the cost of making raw data reusable to produce more uncited papers that are competing for attention in a data-overloaded world. Creative use of big data requires being able to ask the significant questions within a research field at relevant times in history, while finding a fit between readily available data and the most pressing human problems to forge new frontiers of theory.

**Big Data and Privacy in a "Risk Society"**

The use of big data is inextricably linked to a range of ethical, social, and political complexities in a "risk society"—a society that systematically deals with the "hazards and insecurities induced and introduced by modernization itself" (Beck, 1992: 21). A feature of modernization over the past few decades has been rapid digital and technological development and rationalization, which has become associated with the rising threat of manufactured risks. These manufactured risks are associated with a significant level of human agency, as opposed to external risks, such as natural disasters, that were the central concern of the previous industrial era (Giddens, 1999). While social theorists such as Beck and Giddens have provided social commentary about the risk society since the 1990s, big data calls into play consideration of new privacy risks that have increased in complexity with technological advances—a manufactured risk of societies' own making.

An example of the divisive potentials for continuous monitoring and *uberveillance* through the transmission of personal and product data is the use of RDIF (radio frequency identification) tags–implanted microchips that are as small as a grain of rice, and embedded in clothing, products, credit cards, and passports (Hayles, 2009), and as microchip implants. These subdermal technologies in pets–and approved for use in humans by the Food and Drug Administration–are associated with a kind of physical body monitoring that is called *uberveillance* (Michael and Michael, 2014). Consumers, products and things in built environments become "nodes in a web of algorithms" (Hayes, 2014: 50) that are often designed, not to make a contribution to knowledge, but to produce sales-relevant data from everyday human behaviour (Rust, 2017).

Even in countries where technologies are less accessible, mobile telecommunication or wireless services have expanded, as marketers use a plethora of applications that draw on Global Information Systems (GIS) services to inform and enhance the capabilities of the device and user experiences (Hayes, 2014: 50). Other data technologies that have implications for privacy include Smartcards, national identification schemes, genetic testing and potential discrimination, and biometric imaging data, including retina scans, biometric passports, fingerprinting techniques, voice recognition, hand geometry, DNA sampling, facial scanning, and digital imagery (Crompton, 2002).

Data are always generated for a range of social purposes, such as to leverage profits (e.g. online purchasing data), track web usage and create user profiles (cookies), manage organizations, gain broader social networks or popularity, or to govern societies (Kitchin, 2014). Data are not always used in emancipatory or empowering ways for research participants or for society, particularly if the they are used for secondary purposes. Data that were previously private are no longer protected, due to the digital capabilities and interests of corporations, through *dataveillance*, digital footprints, online profiles, corporate governance other data-driven decision-making (Kitchin, 2014). No data generation and analysis can be free from complex ethical concerns, and the technological changes to the production, sharing, and management of data raises new issues and pitfalls for social scientists who wish to harness big data for their own scholarly purposes.

In an era in which big data can be used by corporations, states and nations, the issue of privacy for individuals, and similarly, the ethical implications for researchers wishing to reuse big data should not be forgotten. Controversies over the secretive use of big data have suggested that there is a dark side; namely, protecting individual privacy that requires more than minor tweaks to current protocols. Likewise, government policy on privacy laws often lag behind (Crompton, 2002).

Users of big data, both in the scholarly community and the public sphere, need to be held accountable for the potential harm to individuals. In society more broadly, individuals producing risk will also be exposed to risk (Beck, 1992). The distribution of risk more broadly in society, like wealth, is uneven as is the knowledge to avert or mitigate risk. But all risks need not be perceived negatively. As Giddens (1999: 29) argues, “active risk-taking is a core element of a dynamic economy and an innovative society”.

Logistically then, what concrete recommendations can qualitative researchers consider to avoid ethical pitfalls when negotiating the big data landscape? A set of core questions can be used to guide big data researchers at the beginning and throughout the research process (Dekas and McCune, 2015):

1. How comfortable are the participants in this particular community with the use of data analytics?
2. Would participants consider that the research is a normal part of operations, and presents no unmanaged risks to them?
3. Do the participants have sufficient trust that the researcher and any collaborating organisation have their best interests at heart?
4. Would the participants feel any hint of violation if they learned about the study findings and conclusions?
5. Can participants trust the analytic processes to be unbiased?

While these ethical questions could be asked of qualitative research, there are new complexities for maintaining privacy, gaining informed consent, developing trust, and managing new risks in an era of open access and big data. Whether using big data or small data, researchers need to engage in critical practice to reposition and redirect the operation of big data. We may apply Foucault's (1997: 44) principles of critique to the ethical use of data and learn: "…how not to be governed like that, by that, in the name of those principles, with such and such an objective in mind and by means of such procedures, not like that, not for that, not by them."

Big data is used for particular objectives, and uses particular algorithms and analytic procedures that may cause participants discomfort, serve particular interests that may or may not be aligned with those of the research subjects, and be interpreted in ways that participants may or may not trust—these are the ultimate ethical touchstones (Bassett, 2015).

**What are the potentials of big data for qualitative researchers?**

Qualitative researchers are well-positioned to generate research questions, and to select, curate, interpret and theorize big data away from reductionist claims (Strong, 2013). Bail (2014: 467) sees that the slow uptake of big data in fields such as cultural sociology is astounding:

> …inattention to big data among cultural sociologists is doubly surprising since it is naturally occurring—unlike survey research or cross-sectional qualitative

> interviews—and therefore critical to understanding the evolution of meaning structures *in situ*. That is, many archived texts are the product of conversations between individuals, groups, or organizations instead of responses to questions created by researchers who usually have only post-hoc intuition about the relevant factors in meaning making—much less how cultural evolves in 'real time'.

Texts are a central form of data in qualitative research, whether as interview transcripts, observations, field notes, or primary documents. Of course, qualitative research is not homogenous or monolithic, but includes a diverse array of methodologies, each with their distinctive strengths. For example, participatory action research (PAR) remains distinct from other qualitative methodologies because of the position of the action researcher in relation to personal practice, the role of the participants (Gibson, 2002), and its striving for research to be democratic, empowering, and potentially life-enhancing for those involved (Koch et al., 2002; MacDonald, 2012). However, most strands of qualitative researcher can potentially make greater use of relevant text-based data across a range of modes that have proliferated since the rise of the internet, with open access archives of web pages, books, news reports, legislation, community archives, town hall meetings, social media content, audio archives, podcasts, and images.

Researchers can combine the analysis of big data patterns with interviews, focus groups, and ethnographic observations of online users to make the connections between large data trends, and rich complementary data from individual users or cases. For example, social network analysis, a form of big data analytics, builds on the theory that the relationships among the interacting units are important, the relational ties between the actors facilitate the flow of information, and that these networks are both enabling and constraining of the user actions (Todd, 2008).

Researchers of youth and media studies who are not big data analysts have demonstrated complementary understandings about these relational ties through multisite ethnographies, illustrating how such online networks are friendship-driven (Ito et al., 2008). The participants in these spaces are also united by affinities or shared identities and interests that are not dependent upon shared race, age, or other social geographies (Gee, 2005). Online social media activity typically supports face-to-face friendships, but not exclusively (Mills, 2016), and networking behavior is influenced by reciprocity (e.g. reciprocal retweeting), and to facilitate deliberate sharing of ideas to a wider network.

There are potentials for taking a small data approach to questions that seek to understand the motivations and beliefs of social actors, combining big data with conventional qualitative methods, such as ethnographic interviews and first-hand data or on-the-ground observation of how the data is produced *in situ*, such as in homes, schools, workplaces, and recreational sites.  Context-specific qualitative analysis of the social practices of particular online communities can be combined with the big data analytics showing trends and mapping the dynamics of large networks (Lohmeier, 2014).

The results of these kinds of analysis of complementary data sets are needed to inform the selection of mathematical models, such as to understand the micro-process of social media networks (Snijders et al., 2012). Bringing together teams of qualitative researchers with big data analysts could further elucidate the social basis or micro-mechanisms of online tie-formation, such as how they differ for particular individuals, groups, and networks across different social media platforms. Qualitative research is useful for generating and refining theories to help explain the data. This is vital because, what scholars regard as basic data always has an element of arbitrariness, and data are not truth in themselves. They are simply sources of evidence that can be used to assert a certain view of reality. Similarly, data always belong to somebody, whether they are big or small, and they are constructed *in situ or onsite* and must be recovered accordingly (Christians and Carey, 1989).

A potential contribution of the qualitative researcher is to investigate the underlying micro-processes that contribute to network characteristics or patterns in big data that are often selected on the basis of “mathematical tractability” and explanation, rather than on understanding human social processes (Snijders et al., 2012: 2). For example, big data analytics can tell what users do online, where activities are located, or what users purchased, but it cannot explain why users access certain sites, nor understand participants' subjective perceptions and feelings about their purchasing practices (Borgman, 2015). Qualitative methodologies can contribute the challenge of developing techniques to measure the implicit meanings that occur in-between strings of words, and the "preconscious cultural scripts or frames that shape how people understand the world" (Bail, 2014: 467).

Certainly, big data is radically changing the way research is done in some disciplines, and there is a clear shift toward the development of research infrastructure, products and services to increase the sharing of data for secondary uses, as Boyd and Crawford (2012:13) contend:

> There is a deep government and industrial drive toward gathering and extracting maximal value from data, be it information that will lead to more targeted advertising, product design, traffic planning or criminal policing. But we do think there are serious and wide-ranging implications for the operationalization of big data, and what it will mean for future research agendas.

**Conclusion**

The current hype that guilds discourses of big data on the world wide web is illustrated by this provocative quote from Anderson (2008) published in Wired:

> …massive amounts of data and applied mathematics replace every other tool that might be brought to bear. Out with every theory of human behavior, from

> linguistics to sociology. Forget taxonomy, ontology, and psychology. Who knows why people do what they do? The point is they do it, and we can track and measure it with unprecedented fidelity. With enough data, the numbers speak for themselves.

This is an expression of blind faith in mathematics or automated computation and big data, and the unfounded dismissal of whole disciplines of research speaks volumes. It also highlights a dismissal of research that aims to understand human perceptions, values, reasoning and *emic* or insider perspectives of participants—hallmarks of much qualitative research.

Qualitative researchers often recognize that we are increasingly living in a digital environment in which significant traces of our social action can be captured and analyzed with little or no involvement on our part. Social networking data abounds, amidst the Internet of Things (IoT), material-discursive relations in which objects communicate with one another with little human intervention, providing significant amounts of data about social behavior. Corporations and governments increasingly mine data to make decisions that are also new opportunities for generating knowledge of national security, and for insights into consumer behavior for markets. However, qualitative researchers are also cautious about uncritically touting the potentials of big data when it oversimplifies issues of human behavior, motivations, and emotions. Something is lost when the complexities of human reasoning are datified or determined by the data scientists. Strong (2013: 340) cautions:

> There is a danger that the avalanche of digital data about our lives is used in a way that underestimates the nature of the human condition…often applying arid predictive analytics in a reductionist manner to predict behavior.

It has been long recognized that despite their superior speed of mathematical computations, computers are unable to capture the subtlety, creativity, and personality that real human beings demonstrate across social contexts (Silver, 2012). The fact is that we cannot understand human behavior and social action, without contextualizing the data and understanding the environment in which the data about social action occurs. This is why qualitative research has an important place to trace and probe the complex interactions between the social context and social action, as well as the subjectivity of human actors, whether dealing with big data or little data.

## References

Ajana B (2015) Augmented borders: Big data and the ethics of immigration control. *Journal of information, Communication and Ethics in Society* 13(1): 58–78.

Alberts B (2012) The end of ‘small science’? *Science* 337(6102): 1583.

Anderson C (2008) The end of theory: The data deluge makes the scientific method obsolete. *Wired*. (accessed 25 May 2017).

Mills, K (2016) *Literacy Theories for the Digital Age: Social, Critical, Multimodal, Spatial, Material and Sensory Lenses.* Bristol: Multilingual Matters.

Bail C (2014) The cultural environment: measuring culture with big data. *Theory and Society* 43(3–4): 465–482.

Bancroft A, Karels M, Murray OM, Zimpfer, J. (2014) Not being there: Research at a distance wtih video, text and speech. *Studies in Qualitative Methodology* 13: 137–153.

Bassett C (2015) Plenty as a response to austerity? Big Data expertise, cultures and communities. *Eurpoean Journal of Cultural Studies* 18(4-5): 548–516.

Beck U (1992) *Risk society: Towards a new modernity.* London, UK: SAGE.

Berlekamp E (2012) Small science: Radical innovation. *Science* 338(6109): 882.

Borgman C (2015) *Big data, little data, no data: Scholarship in the networked world* Cambridge, MA: MIT Press.

Boyd D and Crawford K (2011) Six provocations for big data. *Conference of the Oxford Internet Institute – a decade in internet time: Symposium on the dynamics of the internet and society.* Oxford, UK, 1–17.

Boyd D and Crawford K (2012) Critical questions for big data. *Information, Communication and Society* 15(15): 5.

Chandler D (2015) A world without causation: Big Data and the coming of age of posthumanism. *Millennium: Journal of International Studies* 43(3): 833–851.

Christians G and Carey JW (1989) The logics and aims of qualitative research. In: Stempel GHI and Westley BH (eds) *Research methods in mass communication.* Englewood Cliffs, NJ: Prentice Hall, 354–374.

Clarke A and Margettes H (2014) Governments and citizens getting to know each other? Open, closed, and big data in public management reform. *Policy and Internet* 6(4): 393–497.

Coxon T, Davies PM, Hunt AJ, et al. (1993) Research note: Strategies in eliciting sensitive sexual information: The case of gay men. *The sociological review* 41(3): 537–555.

Crompton M (2002) Biometrics and Privacy. *Privacy Law and Policy Report* 9(3): 53–58.

Crossley M (2014) Global league tables, big data and the international transfer of educational research modalities. *Comparative Education* 50(1): 15–26.

Daries JP, Reich J, Waldo J, et al. (2014) Quality social science research and the privacy of human subjects require trust. *Communications of the ACM* 57(9): 56–63.

De Solla Price D (1963) Little science, big science and beyond. Columbia Univerity Press: New York, NY.

Dekas K and McCune EA (2015) Conducting ethical research with big and small data: Key questions for practitioners. *Industrial and Organizational Psychology* 8(4): 463–467.

Foucault M (1997) What is Critique? In: Lotringer S and Hochroth L (eds) *The politics of truth.* New York, NY: Semiotext(e), 41–82.

Frizzo-Barker J, Chow-White PA, Mozafari M, et al. (2016) An empirical study of the rise of big data in business scholarship. *International Journal of Information Management* 36: 403–413.

Gee J (2005) Semiotic social spaces and affinity spaces In: Barton D and Tusting K (eds) *Beyond communities of practice: Language power and social context.* New York: Cambridge, 214–232.

Gibson M (2002) Doing a doctorate using a participatory action research framework in the context of community health. *Qualitative Health Research* 12(4): 546–588.

Giddens A (1999) *Runaway world: How globalization is reshaping our lives.* London, UK: Profile Books.

Gilmore L (2014) Understanding chromosome disorders and their implications for special educators. *Cypriot Journal of Educational Sciences* 9(1): 14–24.

Hayes A (2014) Uberveillance: Where wear and educative arrangement. In: Michael MG and Michael K (eds) *Urbervillance the the social implications of microchip implants emerging technologies* Hershey, P.A.: Information Science Reference.

Hayles NK (2009) RFID: Human agency and meaning in information-intensive environments. *Theory, Culture and Society* 26(2-3): 47–72.

Ito M, Horst HA, Bittanti M, et al. (2008) *Living and learning with new media: Summary of findings from the Digital Youth Project.* Chicago, Illinois: John P. and Catherine T. MacArthur Foundation.

Kitchin R (2013) Big data and human geography: Opportunities, challenges and risks. *Dialogues in human geography* 3(3): 262–267.

Kitchin R (2014) *The data revolution: Big data, open data, data infrastructures and their consequences.* London, UK: SAGE.

Koch T, Selim P and Kralik D (2002) Enhancing lives through the development of a community-based particpatory action research program. *Journal of Clinical Nursing* 11: 109–117.

Laney D (2001) 3D data management: Controlling data volume, velocity, and variety. *Application Delivery Strategies.* Stamford, CT: Meta Group.

Lohmeier C (2014) The researcher and the never-ending field: Reconsidering big data and digital ethnography. In: Hand M and Hillyard S (eds) *Big data? Qualitative approaches to digital research: Studies in Qualitative Methdology.* Bingly, UK: Emerald Group Publishing Limited, 75–89.

MacDonald C (2012) Understanding Participatory Action Research: A qualitative research methodology option. *Canadian Journal of Action Research* 13(2): 34–50.

Mayer-Schonberger V and Cukier K (2013) *Big data: A revolution that will transform how we live, work and think.* Boston, MA: Houten, Mifflin and Harcourt.

McFarland D, Lewis C and Goldberg A (2016) Sociology in the era of big data: The ascent of forensic social science. *The American Sociologist* 47: 12–35.

Meyer ET (2009) Moving from small science to big science: Social and organizational impediments to large scale data sharing. In: Jankowski N (ed) *E-research: Transformations in Scholarly Practice.* New York, NY: Routledge, 147–159.

Michael MG and Michael K (2014) *Uberveillance and the social implications of microchip implants emerging technologies.* Hershey, PA: Information Science Reference.

Montgomery K (2015) Children's media culture in a big data world. *Journal of Children and Media* 9(2): 266–271.

Nardulli PF, Althaus SL and Hayes M (2015) A progressive, supervised-learning approach to generating rich civil strife data. *Sociological Methodology* 45(1): 148–183.

O'Brien DT, Sampson RJ and Winship C (2015) Ecometrics in the age of big data: Measuring and assessing "broken windows" using large-scale administrative records. *Sociological Methodology* 45(1): 101–147.

Parks MR (2014) Big data in communication research: Its contents and discontents. *Journal of Communication* 64: 355–360.

Qui JL (2015) Reflection on big data: Just because it is accessible doesn't make it ethical. *Media, Culture and Society* 37(7): 1089–1094.

Reips U and Garaizar P (2013) *iScience Maps.* Bilbao, Spain: University of Deusto.

Rust H (2017) *Virtuelle Bilderwolken: Eine qualitative Big Data-Analyse der Geschmackskulturen im Internet.* Springer: VS Verlag für Sozialwissenschaften.

Sawyer S (2008) Data wealth, data poverty, data science and cyber infrastructure. *Prometheus* 26(4): 355–371.

Siegel E (2013) *Predictive analytics: The power to predict who will click, buy, lie, or die.* Hoboken, NJ: John Wiley and Sons.

Silver N (2012) *The signal and the noise: Why so may predication fail – but some don't.* New York, NY: Penguin Press.

Snijders C, Matzat U and Reips U (2012) Big data: Big gaps of knowledge in the field of internet science. *International Journal of Internet Science* 7(1): 1–5.

Strong C (2013) The challenge of big data: What does it mean for the qualitative research industry? *Qualitative Market Research: An International Journal* 14(7): 336–343.

Szalai A (1972) *The use of time: Daily activities of urban and suburban populations in twelve countries.* Mouton, Netherlands: Den Haag.

Todd RJ (2008) Youth and their virtual networked words: Research findings and implications for school libraries. *School Libraries Worldwide* 14(2): 19–34.