

Big Data, Digital Technologies and Methodological Promises and Challenges for Social

Research

Suraj Beri

Abstract

The emergence of online and technological sets of data, also known as Big Data, seems both challenging and promising for sociological research. Digital technologies and online interaction via several social media sites have changed the way we interact with people and the way we experience and understand the social world. These technologies help in mapping and locating how people behave, think and interact with each other. These data sets, which are generated globally through media or online mediums and apps, provide an opportunity for social researchers to address fundamental questions about group behaviour, individual attitudes, beliefs, preferences, social identity, individual choices, tastes, and consumer behaviour. Similarly, many researchers have studied and shown in their works that digital space and technologies also shape individual habits, behaviour and thoughts on various issues. The opportunity provided by Big Data and digital technologies has simultaneously posed methodological challenges as well for sociologists to make claims about the method and knowledge of 'the social' in contemporary societies. This proliferation of a new research field, that is, digital technology and Big Data, also facilitates the emergence of new research agendas. For example, research on Big Data and digital social life involves both online and offline worlds. Engaging with these debates on Big Data science and

sociology, this article examines and reflects upon Big Data as a new form of data on human action and online sites as new field sites in a sociological sense. It would explore, through the debates and methodological studies, what kind of changes are brought by this adoption of different research tools from diverse fields. It would also examine whether this results in a new research paradigm or colonisation of sociological enterprise by other disciplines such as computer science, data analytics and software management.

Keywords: Big Data, sociology, digital social, methodology

Introduction

This article aims to understand the emerging challenges posed within social research by the multi-dimensional transformations being created by the ‘data society’. Platforms and digital infrastructures that can capture, assemble, circulate, and manage unimaginable shares, and loads of data, are becoming increasingly influential as a result of the digital revolution’s so-called *datification* process (Van Dijck, 2014). This ability allows them to shape the growing culture of measurement, ranking, forecasting, and evaluating social behaviour, product usage, popular trends, and company or brand perception, etc. This change has huge effects on policy systems, which are shifting their attention to auditing, logistics and economic circuits as the race to own and make money with data heats up. The emergence of online and technological sets of data, also known as

Big Data, seems both challenging and promising for sociological research. Digital technologies and online interaction via several social media sites have changed the way we interact with people and the way we experience and understand the social world. These technologies help us in mapping and locating how people behave, think and interact with each other. These data sets, which are generated globally through media or online mediums and apps, provide an opportunity for social researchers to address fundamental questions about group behaviour, individual attitudes, beliefs, preferences, social identity, individual choices, tastes, and consumer behaviour. Similarly, many researchers have studied and shown in their works that digital space and technologies also shape individual habits, behaviour and thoughts on various issues. The opportunity provided by Big Data and digital technologies has simultaneously posed methodological challenges as well for sociologists to make claims about the method and knowledge of 'the social' in contemporary societies. In this paper, I highlight some of the early insights available to sociologists using Big Data to study social questions, as well as some of the potential flaws and pitfalls that one may encounter if caution is not maintained and practiced in this endeavour. Thus, I conclude by presenting some changes in sociological research owing to the increasing presence, usage and production of Big Data.

Big Data and Studying Social Life

The digital revolution led to the Big Data revolution by making it possible to digitally record everything about human civilisation. The digitisation of film and records, the transition to electronic documents and the widespread use of the Internet, all contributed to the increasing centralisation of computing in everyday life (Radford and Lazer, 2020). The proliferation of computing devices, the improvement in their ability to store and handle large volumes of digital

information and the development of tools to evaluate this data have all contributed to the pervasiveness of what is known as 'Big Data'. This proliferation of new research fields, that is, digital technology and Big Data, also facilitates the emergence of new research agendas. For example, the research on Big Data and digital social life involves both online and offline worlds. Engaging with these debates on Big Data science and sociology, one can examine and reflect upon Big Data as a new form of data on human action and online sites as new field sites in a sociological sense.

The advent of Big Data, which not only includes data that is both huge and enormously complex, but also involve a bewildering variety of phenomena, marks a turning point in the history of the social sciences. From computational linguistics and ethnography to network science, machine learning and in-situ experiments, these data have spurred the development of novel techniques for reducing/simplifying data dimensionality, identifying fresh patterns and linkages and predicting outcomes. These shifts have inspired researchers to explore novel areas of social study. Big Data has attracted the attention of engineers, computer scientists, policy makers, and social scientists across industries, which might lead to the development of a thriving collaborative research agenda. Because of inherent differences in research paradigms, Big Data may not be a free-trade zone among fields. Instead, it may be, at least momentarily, a period in which computer technology and engineering outweigh social research. On the other hand, I contend that a constructive synthesis across paradigms is likely in the long term.

Radford and Lazer (2020) talk about three forms of Big Data: digital traces, digital life and digitalised life (p. 420–21). Their classification is quite useful in analysing the different forms of Big Data. The first form refers to the digitisation of texts, books, newspapers, records, archive

materials, oral histories, the growth of electronic booking of tickets, etc., which were previously not available for a diverse set of people to access, read and interpret. These are now providing new data to sociologists. The second form of data, that is digital traces, means the collection of data, which is recorded on any technical system. For instance, messages, emails, chats, screen shots, tweets, calls records, etc. This set of data can allow sociologists to study networks among different categories of people, the flow of networks, the strength and weakness of social ties, campaigns, etc. The third form refers to the way life itself is lived through digital portals or tools. People spend a certain amount of time socialising with their family, friends, colleagues, and everyone around them. People also have stored information on sites, people find partners on marriage websites, people apply for jobs through online portals etc.

Having discussed the context of new data to study social life, it is time that sociologists focus, now, on the kind of technical skill sets and knowhow required to be able to do research with a new set of data. Broadly speaking there are three areas where this digital technology helps in conducting research: data analysis, data processing and data visualisation and prediction. For data analysis, machine learning as part of Artificial Intelligence (AI) and computer language processing is crucial. It allows to unearth hidden values and broad trends and relationships from sets of data. Text mining, Informational Retrieval (IR) and processing speech are significant elements of data analysis. In order to process data, researchers use Business Intelligence (BI) applications and cloud computing. The cloud has developed as a practical and inexpensive means of handling massive amounts of data. The findings of data analysis are often presented in easy-to-understand visual formats such as 3D presentations and Dashboard.

Big Data encompasses a wide variety of endeavours based on the numerous data-generating gadgets and resources: to increase the operational efficiency of production lines; to identify fraud

and ensure compliance with security requirements; to anticipate the maintenance of plant and machinery; and to construct predictive models based on machine learning.

Emerging Society

Throughout the globe, digital technology has become an integral element of everyday life in several ways during the last three decades. Many of us have gotten reliant on being constantly connected to the Internet. Digital devices that can connect to the Internet from almost anywhere are increasingly pervasive. The portability of smartphones and tablet computers allows them to be carried at all times. Some wearable digital devices, such as smart watches and smart bands, may be worn on the body day and night to monitor biological processes and activities. Through digital platforms and devices, people have access to news, music, daytime soaps, reality television, and motion pictures. At least in part, our emotional and professional relationships, as well as our engagement in communities, may be developed and maintained via the usage of social media such as LinkedIn, Facebook and Twitter. One can broadcast our digitised photographs and home videos to the whole globe. Using search engines such as Google and Yahoo, it is easy to locate information on the Internet. Wikipedia is now the most accessed information source in the world.

The advent of 'Big Data' provides opportunities as well as poses threats to the field of social science. Manovich (2011) came up with the term 'Big Data' to describe digital data sets that were so big that they were too big to be stored and analysed by ordinary computers and software. Since then, the term has grown to include a wider range of interesting things about these data. Beyond the sheer volume of these data, attention is drawn to their proportionality; these data are 'whole'

data sets, capturing everything within a particular field utility records or on a particular platform (Manovich, 2011); are dynamic, capturing social activity in real-time, over time; and provide information on what people do and say 'in the wild', as opposed to what they say they do and say in surveys. Because these data are digitised, there are new opportunities for data 'mining' and connecting. This facilitates the establishment of associations between previously unrelated data sets.

While Big Data has many opportunities, it also presents certain difficulties for social scientists. There are methodological hurdles to overcome before one properly addresses any of these difficulties. Indeed, unless sociologists learn to comprehend and analyse Big Data, they would not be able to fathom the breadth or depth of the ethical and methodological concerns that may arise and will occur in a wide variety of forms of Big Data. Despite the name's connotations of unity and uniformity, there is more than one type of 'Big Data', each of which may be distinguished from the other in a variety of ways. Despite widespread recognition of Big Data's research potential (e.g., Savage and Burrows, 2007), many data sets remain in the hands of governments and private corporations due to privacy and/or commercial sensitivity concerns (Tinati, Halford, Carr, and Pope, 2014: 664).

Over the last decade, social media sites such as Facebook, video-sharing sites such as YouTube, and encyclopaedias such as Wikipedia have captivated the public's imagination and have become staples in people's daily routines. Nowadays, digital technologies mediate our everyday contacts with governmental and corporate entities and organisations as well as our travels through public areas. Examples of the results of digital technology include the ways in which urban space is produced, arranged, monitored, and administered.

Public areas are monitored by Closed-Circuit Television (CCTV) cameras, traffic signals and public transportation networks, building design and construction programmes and the purchasing, production and payment processes for most goods, services and public utilities are all computerised. Whether one likes it or not, and whether one chooses it or not, we are increasingly becoming digital data subjects in this age of mobile and wearable digital devices. The digital recording of images and audios with the integration of security and commercial surveillance technologies into public spaces has created enormous data that previously didn't exist. Massive data sets, sometimes referred to as 'Big Data', are produced and preserved automatically through our digital interactions with networked technology such as search engine queries, phone calls, shopping, interacting with government agencies, and banking. Users' tweets, status updates, blog entries and comments, images, videos, and so on are all examples of 'user-generated material' that may be considered part of Big Data. In addition, what is said on online platforms, profiles of the speaker and the audience, how other users reacted to the content, how many 'likes' a post or tweet attracts, comments, views, time spent on a page or retweets were generated, the time-of-day interaction occurred, the location of users, the search terms used to find the content etc. also form part of Big Data. The importance of Big Data to businesses and non-profits alike has been getting more press recently. The existence of these data raises a number of concerns about how they are used and the implications for privacy, security, and enforcement, surveillance, global development, and the economy.

In many ways, the development and success of Twitter mirrors some of the most pressing contemporary issues in sociology. Metaphorically speaking, it is indicative of a broader shift from

the ‘social as society’, at least as society confined by nation states, to the ‘social as mobility’, emerging in dynamic flows of people, objects, images, and information (Urry, 2000). To be more precise, Castells (1996) describes this as a ‘network society’, in which ‘information’ (now the key commodity) flows across time and space between loosely connected individuals and groups that form and reform fluid identities and connections that go beyond older ties of place, time, class, gender, race, etc. In this way, networks do not only mirror society, they actively participate in the formation of culture (Urry, 2000). Daily activities are the key constituents of the ‘global networks’ of multinational corporations and the ‘heterogeneous, unequal, and dynamic “global fluids”’ (Tinati, Halford, Carr and Pope, 2014: 665).

Traditional Tools of Data Collection, Analysis and Interpretation

Sociological approaches evolve with new data. Quantitative sociologists are used to dealing with tabular data, where each row represents an observation and each column is a number. Census surveys and bureaucratic reports tend to form the bulk of data. Traditional qualitative sociological works are used to deploy narrations, in-depth interviews and observations. Such techniques are impractical when dealing with enormous quantities of data. Videos, photographs, documents, networks, and bit streams cannot be processed by even the most sophisticated human coding facility because of the absence of tabular organisation and the enormous bulk of the data. Due to the absence of standardised data structure techniques, new computer-assisted methods for gathering, processing, analysing, and visualising the massive amounts of data that exist today have become vital. A generation ago, large volumes of information were characterised as thousands of surveys or one million data. Currently, it consists of the hundreds of millions of tweets sent each day and the billions of hours of video uploaded to YouTube annually. Tweets may include images,

emoticons, user locations and social relationships, among other types of information. The introduction of modern computers gave birth to the Big Data revolution, since it was via this medium that the acts of human civilisation could be recorded digitally. The digitisation of media, such as films and recordings, as well as the introduction of electronic documents contributed to the gradual shift towards performing a greater portion of social and professional activities through computers and, ultimately, the Internet. Big Data is pervasive due to the expansion of computer devices, their rising capacity to store and process huge amounts of digital information and the development of evaluation tools for this data.

New Areas of Research: Methods and Challenges

The rate of technological progress has reached a fever pitch. In terms of money and prestige, engineering solutions to real-world problems are worth more than scientific explanations and knowledge. Multiple structural factors contribute to this fact. Technical issues and their answers are increasingly being used in the service of commercial goals. The ability to take a computer science strategy and use social science only as an afterthought to colour what was found may be a deciding factor in hiring social scientists. Putting it succinctly, knowing why something works is being perceived more useful than knowing how it works. Engineers' brute-force application of machine learning to predict a range of behavioural and consumer outcomes using large data will outpace the careful sampling methodologies that social scientists are taught to undertake. The social sciences receive far less money from awarding bodies than the sciences or technical disciplines. Due to the significantly higher commitment of resources and desire for engineering applications in relation to social media and the mining of digital information, it is likely that social

science will have less of a voice in these issues and less of an opportunity to shape the conversation around them. Because of this, I expect theory to fall since it will no longer be the main focus.

Scholars such as Thomas Piketty (2014), Mike Savage *et al.* (2015) in their works provide a novel revitalisation of social imagination that can be useful to reconfigure sociological methods and approaches to analyse digitalised social lives. Their work has not only led to important theoretical debates and opened important new lines of inquiry in the academic world, but it has also been widely used by the general public and influenced political and social policy around the world. In any case, they point to a novel social scientific strategy for data collection, which may have far-reaching implications for the development of Big Data analytics in the future. Both Piketty and Savage *et al.* draw on different methods of data collection and data analysis that reconstruct the connection between data, method and theory in a manner that is comparable to, but distinct from, the assemblages of Big Data analytics.

There are three aspects of social life, where Big Data can contribute to sociological relocating and remodelling of framework:

- a. Social Behaviour: passive, direct, real-time desires, wishes

New domains to explore, analyse data other than government's census surveys and traditional social surveys are opened by Big Data. More often than not social scientists study what people do and what people say about any idea, event or norm. But during the collection of data, scholars often end up relying more on what people say than what people do due to a variety of factors, such as time constraint, easier to ask people than to observe them doing something, not enough

resources or accessibility to spend longer time with respondents etc. Big Data can prove really helpful in addressing these perennial problems plaguing social research. By obtaining the real-time data and people's behaviour recorded via social media platforms, electronic equipment and digital tools, sociologists can better understand and measure the pattern of social conduct and its manifold implications. Scholars can utilise the strength of Big Data and digital tools in early stages of their research, that is, when they are searching for the baseline data in the field of people's usage of any policy or participation in any social, cultural behaviour. During this stage, if they can figure out the broad range of behaviour, opinion or real-time practices, then it would aid further in identifying key variables on which one can focus for detailed analysis.

b. Online Groups and Communities

Because there are so many online groups and communities that allow data to flow and be shared, it is important to study how data is managed and processed through data infrastructures. Data about people with specific characteristics becoming constitutive members of a group, their interaction pattern, interaction themes, collective likes, anxieties, aspirations, shared goals, tasks, and so on can be extremely rewarding. Data analytics featuring social groups or a category of people with certain key social, cultural attributes, characteristics, and specific behaviour patterns provide another opportunity to study collective behaviour (Mertia, 2020). People's actions are reflected in the information that makes up Big Data. After this data is analysed, decisions can be made (such as the appropriateness of any policy, mechanisms of welfare distribution) to facilitate more precise targeting of individuals with specific characteristic patterns.

c. Activism, Campaigns and Movements

Big Data techniques and projects require, at a minimum, the presence of seemingly seamless digital infrastructures, techno-managerial expertise, a sizable user base, and measurable market or public policy outcomes. Through online real time data and people's behaviour sociologists can follow their political values, perceptions and campaigns.

In addition to having the largest number of software engineers, the fastest growing mobile Internet user base and market and national government programmes for creating a 'Digital India', 'Startup India' and one hundred 'Smart Cities', India is also home to the largest population of people who speak languages other than English. Despite this, the country's technological infrastructure is highly decentralised, with nearly half of the population still lacking access to broadband Internet. The sheer number of social media users and the size of the market these social media giants i.e. Facebook, Instagram, WeChat, TikTok etc. hold is enormous. According to a recent estimate by the Government of India (2021), there are over 530 million WhatsApp users, around 448 million YouTube users, around 410 million who use Facebook, around 210 million who use Instagram, and Twitter comes last with 15 million users.

These individuals' actions further solidify the aforementioned patterns within the information banks' databases. As a result, Big Data analytics and human beings become intertwined in a self-referential cycle. Similar criticisms to those levelled at simplification have been levelled at the widespread dissemination of data records and analytical insights, with many studies highlighting the exaggeration of some 'realities' while downplaying others.

When viewed through a different lens, however, the ever-increasing number of people consuming digital media, primarily on mobile devices, demonstrates inventiveness in developing new ways

of visualising, gaining access to and disseminating media. India is experiencing a surge in innovation and investment thanks to the country's ability to host both domestic and international technology start-ups. In the changing industrial and manufacturing sector, that is, at the level of supply chain, application of Big Data enabling cost benefit and effective quality maintenance is just one example of the many social and economic sectors where opportunities exist for Big Data to be successfully utilised. The retail industry is a good place to use AI since it generates massive amounts of data on its customers. Conventional wisdom holds that banking and financial services data is put to good use in the context of meeting clients' financial service requirements in a secure and efficient manner. Insurance-related matters such as accounting, claims handling, fraud detection, wellness evaluations, etc. are now growingly being addressed by AI. Computer-aided diagnostics and predictive analysis can meet a critical societal need because of the massive amount of data available in the healthcare sector. E-Government encompasses many different types of public services, including but not limited to public distribution, record-keeping, tax collecting, infrastructure development and use, and so on. In the field of media, these techniques can provide a cross-section of the public with useful information tailored to their own interests and requirements. The information generated by the manufacturing sector's many processes is used to enhance efficiency and cut down on wasteful downtime. Security services, including disaster management and mitigation, and the protection of public assets and information, deal with vast quantities of data. In agriculture and farming, statistics on weather, sowing, soil, plant diseases, and land ownership are all valuable sources of information. Analysis of education-related data is helpful for several reasons, including improving content delivery and meeting the need for specialists, pedagogies, educational institutions, and instructors to compare, analyse and examine data to generate course contents.

Even as the Indian economy expands and the political landscape shifts, unleashing new aspirational energies, it is important to keep in mind that the social context is defined by extreme hierarchies of religion, caste, gender, and socioeconomic status.

Some scholars fear that this excessive reliance on digital tools and data confines sociology into a defensive posture, disadvantaging social scientists from a new and more powerful data aggregation that is beginning to play an essential role in the creation of knowledge and understanding in the 21st century. Moreover, sociologists need to participate in Big Data analytics since the field is evolving whether we like it or not. As part of changes brought through educational policies, the government can begin new digital capacity-building programmes at the school, college and university level to introduce Big Data technology, its methods and application so that students can utilise them and generate, process and analyse huge data sets available in the public digital domain.

Conclusion

The advent of new digital data sources has made it possible to record and analyse human activity in unprecedented detail and speed, both in the present and in the past. Sociology is essential to Big Data analytics because of the theoretical, methodological and empirical knowledge they provide to the study of society. In this reflexive account, I have briefly discussed the greater significance of sociological research methods in the field of Big Data analytics if it is approached through the lens of the three core frames that I have discussed in the paper. While maintaining our dedication to critical thinking about ‘data’, methodological rigour and theoretical interpretation, this approach would allow sociologists to investigate the potential of Big Data for sociological study. It is becoming increasingly apparent that sociological knowledge and understanding must form part of

the assemblage, that sociologists may bring their specialised domain expertise to bear on shaping the agendas that drive Big Data analytics and the modes of investigation and interpretation that arise in play. New data assemblages powered by Big Data are, increasingly, emerging to impact government policy, commercial practices and popular representations of the social world. It is high time sociologists pay serious attention to this new ‘social’ taking shape around us.

Works Cited

- Burns, Ryan. 2015. ‘Rethinking Big Data in Digital Humanitarianism’, *GeoJournal*, 80(4): 477–90.
- Castells, M. 1996. *The Rise of Network Society*. Oxford: Blackwell.
- Dijk, Van. 2014. ‘Datafication, Dataism and Dataveillance. Big Data between the Scientific Paradigm and Ideology’, *Surveillance & Society*, 12(2):197–208.
- Halford, Susan and Savage, Mike. 2017. ‘Speaking Sociologically with Big Data’, *Sociology*, 51(6): 1132–48.
- Lupton, Deborah. 2015. *Digital Sociology*. Oxon: Routledge.
- Manovich, L. 2011. ‘Trending: The promises and Challenges of Big Social Data’, *Debates in Digital Humanities*, 1–17.

- Mertia, Sandeep. 2020. *Lives of Data: Essays on Computational Cultures from India*. Amsterdam: Institute of Network Cultures.
- Piketty, Thomas. 2014. *Capital in the Twenty-First Century*. Cambridge: Harvard University Press.
- Radford, Jason and Lazer, David. 2020. 'Big Data for Sociological Research', In George Ritzer and Wendy Murphy (eds.) *The Wiley Blackwell Companion to Sociology*, 419–43. New Jersey: Wiley Blackwell.
- Savage, Mike and Burrows, R. 2007. 'The Coming Crisis of Empirical Sociology', *Sociology*, 41(5): 885–99.
- Savage, Mike and Devine F. et al. 2015. *Social Class in the 21st Century*. London: Penguin.
- Tinati, Ramine, Halford, Susan, Leslie Carr and C. Pope. 2014. 'Big Data: Methodological Challenges and Approaches for Sociological Analysis', *Sociology*, 48(4): 663–81.
- Urry, J. 2000. 'Mobile Sociology', *British Journal of Sociology*, 51(1): 185–203.