

Rethinking Empirical Social Sciences

In (2013) *Dialogues in Human Geography* 3(3): 268-273.

Evelyn Ruppert

Goldsmiths, University of London

Abstract

I consider some arguments of social science and humanities researchers about the challenge that Big Data presents for social science methods. What they suggest is that social scientists need to engage with Big Data rather than retreat into internal debates about its meaning and implications. Instead, understanding Big Data requires and provides an opportunity for the interdisciplinary development of methods that innovatively, critically and reflexively engage with new forms of data. Unlike data and methods that social scientists have typically worked with in the past, Big Data calls for skills and approaches that cut across disciplines. Drawing on work in science and technology studies and understandings of the 'the social life of methods', I argue that this is in part due to the fragmentation and redistribution of expertise, knowledge and methods that new data sources engender including their incipient relations to government and industry and entanglements with social worlds.

Keywords

Big Data; digital devices; Interdisciplinarity; social life of methods; transactional data

The critical questions for geographers raised by the editors of this special issue are related to concerns within the social sciences more generally. While Big Data – the vast amounts of digital information generated, accumulated and stored in myriad databases and repositories, both online and offline – does present specific challenges to the geography discipline, I would suggest that it also calls for interdisciplinary approaches perhaps more than ever. There are of course many different rationales for interdisciplinarity but in the case of Big Data I will attend to two. First, the distributed relations and entanglements of ownership and expertise that make up Big Data call not only for interdisciplinary approaches but also cross-sectoral engagements between the social sciences, industry, government and business. And second, the ontological and epistemological consequences of methods that take up Big Data cut across disciplines and provide an opportunity for collaboration on the underlying theoretical propositions as well as the vexed political questions of data privacy, rights, ethics and ownership.

Ownership and Expertise

Regarding the first, a few years ago sociologists Mike Savage and Roger Burrows wrote a provocative article in the journal *Sociology*, which generated considerable debate (Savage, 2007). While not specifically addressed to Big Data, they argue that empirical sociology is facing a 'coming crisis.' The empirical methods and expertise of sociologists is being challenged by the proliferation of 'social' transactional data that is routinely collected, processed and analysed by a wide variety of private and public institutions. At one time sociologists were the recognised experts of empirical methods such as the survey and in-depth interview for generating data on the social. However, Savage and Burrows argued that these methods are becoming dated in the face of new digital data sources and methods. Other sociologists have advanced different angles on this issue. For example, while noting that social science methods have always relied on distributed expertise and relations, Celia Lury and Nina Wakeford (2012) argue that the current challenge for the social sciences is to attend to how new forms of data and conceptions of the empirical are reorganising social and cultural research and knowledge making, a point that Noortje Marres (2012) also advances.

Similar arguments have been made about the social sciences more generally. Lazer et al. (2009), for example, respond to the challenges of Big Data, albeit from a different perspective, in their call for computational social sciences. They note that while the capacity to collect and analyze Big Data has transformed disciplines such as biology and physics, publications in leading social science journals reveal that a data-driven 'computational social science' has been much slower. However, they argue that it is being done elsewhere—in internet companies such as Google and Yahoo and through the data mining and inferential analytics of national security agencies. If social scientists do not step forward then computational social science risks becoming the exclusive domain of governments and the private sector and of computing scientists. They thus ask: "What value might a computational social science—based in an open academic environment— offer society, by enhancing understanding of individuals and collectives?" (721)

For Christakis (2012), a physician and sociologist who studies social networks, one answer lies in a 'new kind of social science' that is necessary because Big Data and computational sciences are changing how we can analyse and understand individual and collective human behaviour. He relates these changes to two developments that involve convergence between the social and natural sciences: discoveries in the biological sciences such as the sequencing of the human genome, which are calling into question the origins of human behaviour and sociological concepts such as 'free will'; and a resurgence of experimentation as a social science method made possible by online platforms such as virtual laboratories. These developments, according to Christakis, call for a social science in the twenty-first century that is more collaborative and interdisciplinary.

Watts (2007) observes that computational social science presents an opportunity to rethink academic disciplines and their relations. Due to the relative complexity and heterogeneity of social phenomena, he suggests that the social sciences have been

less successful than the physical and biological sciences in providing explanatory and coherent theoretical accounts of, for example, the complexities of collective social behaviour. However, new sources of data and computational analytics are now making it possible to study complex phenomena such as social networks in ways never before imagined or possible.

Writing from the perspective of digital humanities, Manovich (2011) contends that computational developments and an emerging 'big data society' are leading to new divisions of labour and expertise between disciplines, academics, users, business, governments and industry. He suggests that there are three new 'data classes': "those who create data (both consciously and by leaving digital footprints), those who have the means to collect it, and those who have expertise to analyze it" (10). Regarding the second, limited access to the massive amounts of transactional social data that is continuously generated means corporations are dominating data-driven social science and humanities. But even when access is possible, there is a large gap between the kinds of analytics that are possible and those who have the capacities and knowledge to undertake such analytics. Consequently, social analytics are being done more by computer scientists rather than social and humanities scientists. For Manovich collaboration is certainly one way forward but ultimately social and humanities scientists need to acquire and develop the skills to do data analysis and experiment with the visualisation tools necessary to manage and interpret Big Data. Otherwise, the humanities and social sciences will become ever-more alienated from the creative power of software analytics in formatting their working practices (Dodge, Kitchin, and Zook, 2009).

This challenge for the social sciences has also been recognised in the UK by the British Academy (2012) in a recently published policy paper. The Academy suggests that university graduates have a "skills deficit" and are ill equipped to deal with the challenges presented by the increase in volumes of data generated by a digital society. This deficit undermines the future of UK social sciences in providing robust analyses of large and complex databases. Lazer et al. (2009) also underscore this in their call for training new scholars and collaboration between "computationally literate social scientists and socially literate computer scientists" (10-11).

What the British Academy is calling for are skills that are not only valuable for the future of empirical social science research but also for the training of social scientists to meet the analytic questions, demands and requirements of industry, business, and governments. Regarding the first, the call for social scientists with the skills to engage with Big Data has been sounded by the technology industry. To date, engagement with academics has primarily involved computer scientists and mathematicians while collaboration has only just started with social scientists. As Prabhakar Raghavan, VP of strategic technologies at Google and former head of Yahoo Labs has noted, the biggest obstacle in growing the social sciences is that there are not enough people trained and interested in the issues that Big Data raise (Mann, 2012). He argues that while the computer scientist knows *what* people are doing (shallow), what is needed is the capacity to understand *why* people do things such as purchase a product or read a story (deep). To do so, industry needs "people

who can straddle the disciplines” (n.p.). The question that this raises is whether the social sciences will be compelled to supply the answer to Google and Yahoo or develop critical perspectives from which to question the motives and interests that drive the quest for Big Data.

Governments are also experimenting with Big Data from social media, mobile phones and browsers. International governmental organisations such as the United Nations also identify the potential of Big Data sources such as call logs, mobile-banking transactions, and online user-generated content (blogs, tweets, searches) as a source of actionable information for decision makers to track development progress, identify crises, improve social protection, and understand where existing policies and programmes require adjustment (UNECE, 2006). But for other government authorities Big Data constitutes not just an opportunity but also a challenge. Like some social scientists, national statisticians also see Big Data as a challenge to their traditional role as the generators and authoritative sources of official statistics. According to one of many reports, statisticians are concerned that Big Data could “beat official statistics on timeliness and relevance” and national statistical organisations “could slowly lose their reputation and relevance unless they get on board” (Conference of European Statisticians, 2013: 2). For them, taking up Big Data requires a new generation of tools and methods and rethinking fundamental statistical principles, which need to be undertaken in collaboration with the academy.

All of these arguments for interdisciplinarity and collaboration are not a call for turning social scientists into statisticians or computer scientists, but for ‘socialising’ what could otherwise become a positivist science of individuals and societies or lead to re-inscribing a division between quantitative and qualitative methods. Retreating and engaging in internal debates within social science disciplines cannot achieve this, as Savage and Burrows (2007) also warn. Instead, it means to explore methods of doing immersive interdisciplinary data work by innovatively, critically and reflexively engaging with new forms of data. This calls for experimenting with various data sources and techniques, innovating methods, and working with researchers in computing and other sciences. In this way, social scientists can contribute to our understanding of the ontological effects of Big Data (who we are as individuals and societies) and also the epistemological effects of the very methods that we deploy to analyse and interpret it (how we are known).

The social life of methods

As this last point suggests, not only do we need to rethink methodological repertoires but also the theories and concepts of selves and societies that make them up. While concern has been expressed about the absence of social-scientific theoretical insights and reflections on the consequences of Big Data (Snijders, Matzat, and Reips, 2012), there are many examples that attend to what has been more commonly referred to as digital data. For example, Latour et al. argue that new digital data sources challenge the long-standing theoretical proposition that individuals and collectives constitute two levels. Instead, the availability of digital

traces allows us to think of one level, the crowd of elements that make up an individual, which they argue are more complex than aggregates. In this way they demonstrate how particular 'datascapescapes' and theories of social order are closely tied and thus the availability of rich digital traces requires rethinking classical questions of social order and social theory (Latour, Jensen, and Venturini, forthcoming).

This relation between data, methods and theories is one that has also been taken up by other researchers in sociology, anthropology and science and technology studies. One version is referred to as the 'social life of methods,' which is an approach that understands methods as having a 'double life:' they are simultaneously embedded in and shaped by social worlds, and can in turn become agents that act in and shape those worlds (Law, Ruppert, and Savage, 2011). In regards to the proliferation of digital data, the social life of methods draws attention to how digital devices are not only materially implicated in the production and performance of contemporary sociality but also our methods, theories and knowledge of it (Ruppert, Law, and Savage, 2013). For one, devices are enacting novel forms of sociality by redoing the very mediums of social life and are inventive of new forms of sociality and being. In this view, devices are not just tools but enrolled in and constitutive of social relations. From social networking sites, search engines, blogs, wikis to online purchasing, e-Government, and open data, all of these can be understood as modes of instantiating social relations and 'making up' individuals and collectivities in ways that also blur the boundaries and distinctions between online and offline worlds.

For instance, devices such as Twitter organise new forms of sociality and ways for people to interact, communicate, connect and know about themselves and others. But at the same time while making up the very stuff of selves and social relations, this device and the lively data it generates is also giving rise to new methods that materialise versions of those lives. Various actors such as users, researchers, governments, industry or businesses deploy methods to analyse this data. For instance, Twitter gives rise to various knowledge practices or methods: academic researchers, data journalists and police surveillance units develop combinations of analytical procedures (algorithms, software), infrastructures (computers, networks) and personnel (analysts, IT experts) to analyse the data that it generates (Ruppert, Law and Savage, 2013).

These practices also make the data recursive (Beer and Burrows, 2013). For example, designers, as a result of analyses of data on usage, reconfigure social media platforms. Users also adjust and alter their conduct as a result of learning from and interacting with others, or from changes in privacy protocols or scandals such as PRISM. Thus how users perform and the digital traces they leave are dynamic. Data thus modulates in relation to these changes as well as reconfigurations of analytic devices such as API's and other tools for accessing, scraping and analysing data. Devices, users, and data are thus unstable, dynamic and lively. This understanding of dynamic social relations could of course also apply to spatial relations. Geolocated social media devices such as Google Maps placemarks, Flickr photos, Wikipedia entries and geocoded tweets are generative of new spatial relations as well as the methods of human geography (Crampton et al., 2012).

What then can we say about social science methods that engage with such digital devices and data to represent and enact social worlds? For one, we can think about them as made up of people and things that get bundled or assembled (Ruppert, Law and Savage, 2013). Basically we could think of all methods in this way. Surveys for instance assemble questionnaires, sampling decisions, privacy protocols, methodological rules and repertoires, clipboards, interviewers, statistical procedures, analytic software, and subjects who answer questions. And critically the researcher assembles combinations and configures these in relation to his/her research purposes. What is different with digital devices is that many of these elements are configured and decided by others and for specific purposes usually not research. For these reasons data is often referred to as a by-product of devices and devices configured for purposes, such as takes place in transaction or communication. Social science researchers are thus reassembling their methods and the elements that make them up and becoming dependent on devices and data embedded in the performance of social and cultural lives – and in large part this is not of their making. This challenges longstanding assumptions about methods, from questions of ethics, expertise, knowledge and representation to the oft-separated mix of methods, theories, culture and the social.

The consequences are manifold but a key one is that methods that draw upon these various digital devices and the Big Data they generate are not simply technical procedures established a priori. Methods are not outside of but inside and come out of the very dynamic relations of cultural and social life that devices mobilise, materialise and instantiate. At the same time, both the devices and data are configured in large part by numerous corporate and industry actors. Much concern has been waged about this mediating and configuring work of online platforms and the consequences for the data they generate and which is then sometimes accessible, albeit usually in restricted ways, for social science research (e.g., boyd and Crawford, 2012). As Marres (2012) notes, some argue this is leading to the privatization of social research where corporations and information technology firms become more significant players, a trend that Thrift (2005) has called “knowing capitalism.” However, another way of thinking about this draws from science and technology studies to understand all methods as being made up of extended relations between various actors and things (think of the survey example noted above). So rather than a displacement to corporate culture, digital devices and data are perhaps redistributing the division of labour (Marres, 2012) and leading to the reassembling of social science methods (Ruppert, Law, and Savage, 2013).

One challenge then is to theoretically account for the distribution and fragmentation of the different agencies at work in both the making and performance of social worlds and the digital traces and methods through which they are known. Because the data generated by various devices are the material of methods and research being conducted across disciplines and approaches (geography, anthropology, sociology, political science, and so on), I think this is a particularly unique opportunity for interdisciplinary engagements. For example, geographers Crampton et al. (2012) argue that studies of the geoweb need to go “beyond the geotag” and take into account social media data that is not explicitly geographic and to

interrogate the contexts of data production. They argue for approaches that go “beyond the proximate” by building on the work of sociologists and in particular social network analysis to take into account forms of relational, cultural or linguistic distances. They also suggest drawing on understandings of science and technology studies to attend to actants such as code platforms.

There are many theoretical assumptions that demand rethinking such as those suggested by Latour: who are the subjects of digital data and devices? Methods that involve devices such as surveys conceive of the speaking and self-accounting subject and depend upon eliciting individual and usually verbal accounts. Digital devices register what people do and depend on recording and tracing their behaviour, how they exchange, move, connect and interact. Digital traces are not derived from conscious interventions of knowing researchers, but are by-products of online purchasing, browsing, communicating, and so on. And in many instances such as social network analysis, data does not necessarily relate to specific individuals but instead transactions, not on the factors that affect individual behaviour, but on the spatial flows of behaviours and contacts: contagion, pollution, influence, etc. (Ruppert, Law, and Savage, 2013). The play of dynamic transactions opens up inquiries that are non-individualist and non-humanist accounts of the social.

Rather than empirical social sciences facing a crisis, I think Big Data present a challenge and opportunity. Social science methods are now more in and of social worlds, not standing outside and detached from them as objects or subjects of inquiry. This presents an opportunity to be reflexive of the epistemological and ontological effects of methods as part of and not detached from social worlds. It is an opportunity for interdisciplinary approaches and engagements to understand the incipient relations to government, industry and social actors that Big Data presents. Taken together, these entanglements call for a critique from within, especially as Big Data raises political questions about privacy, rights, ethics and ownership of data, and because social science methods are inevitably bound up with the decisions, priorities, objectives and values of other actors.

References

- Beer, D., and R. Burrows (2013) Popular Culture, Digital Archives and the New Social Life of Data. *Theory, Culture & Society* 30 (4): 47-71.
- boyd, d., and K. Crawford (2012) Critical Questions for Big Data. *Information, Communication & Society* 15 (5): 662-679.
- British Academy (2012) *Society Counts: Quantitative Skills in the Social Sciences*. London.
- Christakis, N. A. (2012) A New Kind of Social Science for the 21 St Century. *Edge*. Avail. at <http://edge.org/conversation/a-21st-century-change-to-social-science>.

- Conference of European Statisticians (2013) 'What Does "Big Data" Mean for Official Statistics?', Sixty-first Plenary Meeting. Geneva: United Nations Economic Commission for Europe (UNECE).
- Crampton, J. W., M. Graham, A. Poorthuis, T. Shelton, M. Stephens, M. W. Wilson, and M. Zook (2013) Beyond the Geotag: Situating "Big Data" and Leveraging the Potential of the Geoweb. *Cartography and Geographic Information Science (CaGIS)* 40 (2): 130-139.
- Dodge, M., R. Kitchin, and M. Zook (2009) How Does Software Make Space ? Exploring Some Geographical Dimensions of Pervasive Computing and Software Studies. *Environment and Planning A* 41: 1283-1293.
- Latour, B., P. Jensen, and T. Venturini (forthcoming) The Whole Is Always Smaller Than Its Parts: A Digital Test of Gabriel Tarde's Monads. *British Journal of Sociology*.
- Law, J., E. Ruppert, and M. Savage (2011) 'The Double Social Life of Methods.' *CRESC Working Paper Series, Paper No. 95*.
- Lazer, D., A. Pentland, L. Adamic, S. Aral, A.-I. Barabási, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, T. Jebara, G. King, M. Macy, D. Roy, and M. V. Alstytne (2009) Computational Social Science. *Science* 323: 721-723.
- Lury, C., and N. Wakeford eds (2012) *Inventive Methods: The Happening of the Social*. London: Routledge.
- Mann, R. (2012) Five Minutes with Prabhakar Raghavan: Big Data and Social Science at Google. *Impact of Social Sciences Blog*: LSE Public Policy Group. Avail. at <http://blogs.lse.ac.uk/impactofsocialsciences/2012/09/19/five-minutes-with-prabhakar-raghavan/>
- Manovich, L. (2011) Trending: The Promises and the Challenges of Big Social Data. Avail. at: http://www.manovich.net/DOCS/Manovich_trending_paper.pdf.
- Marres, N. (2012) The Redistribution of Methods: On Intervention in Digital Social Research, Broadly Conceived. *The Sociological Review* 60: 139-165.
- Ruppert, E., J. Law, and M. Savage (2013) Reassembling Social Science Methods: The Challenge of Digital Devices. *Theory, Culture & Society, Special Issue on The Social Life of Methods* 30 (4): 22-46.
- Savage, M., and R. Burrows (2007) The Coming Crisis of Empirical Sociology. *Sociology* 41 (5): 885-899.

Savage, M. (2007) Changing Social Class Identities in Post-War Britain: Perspectives from Mass-Observation. *Sociological Research Online* 12 (3).

Snijders, C., U. Matzat, and U. Reips (2012) "Big Data": Big Gaps of Knowledge in the Field of Internet Science. *International Journal of Internet Science* 7: 1-5.

Thrift, N. (2005) *Knowing Capitalism*. London: Sage Publications Ltd.

UNECE (2006) Conference of European Statisticians Recommendations for the 2010 Censuses of Population and Housing. Prepared by UNECE in Cooperation with the Statistical Office of the European Communities (Eurostat). New York and Geneva.

Watts, D. J. (2007) A Twenty-First Century Science. *Nature* 445: 489.