

# After the crisis? Big Data and the methodological challenges of empirical sociology

Big Data & Society  
 April–June 2014: 1–6  
 © The Author(s) 2014  
 DOI: 10.1177/2053951714540280  
 bds.sagepub.com



Roger Burrows<sup>1</sup> and Mike Savage<sup>2</sup>

## Abstract

Google Trends reveals that at the time we were writing our article on ‘The Coming Crisis of Empirical Sociology’ in 2007 almost nobody was searching the internet for ‘Big Data’. It was only towards the very end of 2010 that the term began to register, just ahead of an explosion of interest from 2011 onwards. In this commentary we take the opportunity to reflect back on the claims we made in that original paper in light of more recent discussions about the social scientific implications of the inundation of digital data. Did our paper, with its emphasis on the emergence of, what we termed, ‘social transactional data’ and ‘digital byproduct data’ prefigure contemporary debates that now form the basis and rationale for this excellent new journal? Or was the paper more concerned with broader methodological, theoretical and political debates that have somehow been lost in all of the loud babble that has come to surround Big Data. Using recent work on the BBC Great British Class Survey as an example this brief paper offers a reflexive and critical reflection on what has become – much to the surprise of its authors – one of the most cited papers in the discipline of sociology in the last decade.

## Keywords

Sociology, class, Big Data, empirical, crisis, methods

When we wrote our article ‘The Coming Crisis of Empirical Sociology’, and an update to it a couple of years later, in the journal *Sociology* (Savage and Burrows, 2007, 2009), neither of us had come across the notion of ‘Big Data’. To be sure we now know that the term had been circulating within certain domains for quite a while (Mayer-Schönberger and Cukier, 2013: 1–18) but, at the time, it had not impinged itself on our sociological imaginations. We were clearly not alone in this, as a quick analysis using Google Trends reveals. At the time of writing the original article almost nobody was searching the web for ‘Big Data’. It was only towards the very end of 2010 that the term had begun to register, just ahead of an explosion of interest from 2011 onwards. However, although we did not articulate the issues that interested us in these terms it is clear, in retrospect, that what we have come to term ‘Big Data’ was indeed a phenomenon we were alluding to in our polemic.

Re-reading the article now it is difficult to understand why it has become one of the most widely cited in the discipline over the last few years. As we draft this

commentary Google Scholar reports almost 300 citations and we recently discovered that it is the most cited article to appear (<http://soc.sagepub.com/reports/most-cited>) in *Sociology* – the journal of original publication – in the last decade. It has been subject to various comments and critiques from a range of perspectives and has even formed the focus of a special e-issue of the journal (McKie and Ryan, 2012). What must have read as new, innovative and important in 2007, and even 2009, now reads to us as a pretty mainstream position, not just in sociology but also across the cognate social sciences more generally. However, what we had to say has been subject to a wide

<sup>1</sup>Department of Sociology, Goldsmiths, University of London, London, UK

<sup>2</sup>Department of Sociology, London School of Economics, London, UK

## Corresponding author:

Roger Burrows, Department of Sociology, Goldsmiths, University of London, New Cross, London SE14 6NW, UK.

Email: [r.burrows@gold.ac.uk](mailto:r.burrows@gold.ac.uk)



range of interpretive flexibility by those drawing upon our arguments, and so it might be worth trying to restate the core aspects of our position once again.

The paper is a polemic – an intervention aimed to alert colleagues to what we perceived at the time to be a cluster of challenges to the jurisdiction of academic sociology. The title, of course, was a play on Gouldner's classic but now little read account of the sociology of sociology, published almost 45 years ago (Gouldner, 1970). Our argument was that the recent development of the discipline – in the UK at least – had become insular and self-regarding. Despite all the talk about profound and far-reaching global social change, it seemed as if the discipline of sociology was exempted from this. Sociologists generally used and refined rather familiar methods, talked mainly to each other about esoteric theoretical pre-occupations, and had not caught up with the fact that sociology was no longer an avant-garde discipline which had attracted legions of critical students and scholars in the 1960s and 1970s (Savage, 2010) but had become fully part of the academic machine (see Burrows, 2012; Kelly and Burrows, 2012). In particular, we urged ourselves to look at the actual methods which sociologists used as they went about their research. Qualitative interviews and representative surveys which had, in the 1960s, been radical new windows onto the society had now been fully absorbed into the circuits of 'knowing capitalism' (Thrift, 2005). We took the view that rather than assume that the discipline of sociology was bound to exist, and to enjoy some kind of natural authority in offering insights into the social, it was more important to recognize the proliferation of social research that now took place, in much of which sociologists were only marginally involved.

The foregrounding of a style of sociology in which questions of method had come to be underplayed had occurred within the same time-frame as methods of sociological research – that had once defined the professional practice of the discipline – had not only become detached from the sociological mainstream but – even more problematically – had become ubiquitous outside of the academy. Our argument was that the innovative and inventive nature of sociological methods that had come to define the discipline in the post-1945 period – sampling theory, survey design, qualitative interviewing and the rest – had waned, whilst a form of, what we might call, 'synthetic sociology' practised by the canonical 'malestream' of the discipline – Bauman, Beck, Castells, Giddens and their ilk – had become ascendant. What innovations were occurring in research methods were located not in the academy but in organizations primarily engaged in generating profits. We identified various forms of 'commercial sociology' (Burrows and Gane, 2006) wherein the

methods of sociology were being routinely deployed and radically innovated in order that an increasingly reflexive capitalism could come to know itself in ever more precise ways in order to better extract value. Our sense was that these innovations in the field of 'commercial sociology' were – at the time – happening with little awareness or engagement on the part of those working in the academy. We pointed towards innovations in one area in particular – new forms of what were essentially unobtrusive methods (Webb et al., 1966) rethought for the analysis of digital data – that seemed to us to be of particular importance.

In our initial paper we focused on what we termed 'transactional' data held within large and complex commercial and government databases – data generated as a digital by-product of routine transactions between citizens, consumers, business and government (as brilliantly satirized at: <http://www.aclu.org/ordering-pizza>). Our emphasis was perhaps understandable – such transactional data were obviously a crucial part of the informational infrastructures of contemporary capitalism but, at the time, we were only just becoming aware of the sociological possibilities that it opened up. We gave less emphasis than perhaps we should have to data derived from what we were only just learning to call Web 2.0, or social media. A linked paper by Beer and Burrows (2007) published in the same year as the 'coming crisis' reviewed the, then, sparse sociological literature that did exist on what was then a very new cultural phenomenon. As was predicted in that paper, social media is now a fundamental part of popular culture and generative of not only digital transactional data – although that remains important – but also huge amounts of data actively created through myriad acts of global prosumption (Ritzer and Jurgenson, 2010) – data actively produced and consumed within the same moment by social media users. Beer and Burrows (2013) have recently updated the position somewhat to include the growth of such social media data within the original framing of the 2007 paper.

It was the emerging analysis of digital transactional and social media data that interested us. At the time this was something that few in the academy showed any awareness of or interest in and our paper was a sort of sociological call to arms. In many aspects of our work – at conferences, within fieldwork and so on – we were routinely coming across analysts working outside of the academy and, indeed, outside of the social sciences, who were producing social knowledge based upon access to, and the analysis of, such data. Our concern was that this was likely to be yet another major nail in the coffin of academic sociological claims to jurisdiction over knowledge of the social. As Osborne et al. (2008: 531) observed, at about the same time: 'professional sociologists... are not the only people who investigate,

analyse, theorise and give voice to . . . phenomena from a “social” point of view’. They go on to enumerate ‘statisticians, economists of certain persuasions, educationalists, communications analysts, cultural theorists . . . journalists, TV documentary-makers, humanitarian activists, policy makers’ (2008: 531–532) and other actors who, they argue, sometimes ‘produce better sociology than . . . sociologists themselves’ (2008: 532). Now to this list we have to add ‘data scientists’ because with the rise of ‘Big Data’ has come the emergence and rise of a new professional grouping claiming expertise over the analysis of the stuff. Google Trends again shows how in mid-2013 searches for ‘data scientist’ surpassed those for ‘statistician’ for the first time (see <http://flowingdata.com/2013/12/18/data-scientist-surpasses-statistician-on-google-trends/>)

However, ‘data scientists’ working with ‘Big Data’ offer a rather different challenge to the traditional sociological sensibility than the other professional actors enumerated above. They offer the possibility of describing the social world in a manner hitherto impossible. To state the issue baldly it is this: the majority of sociological methods – other than those based upon direct observations of actions – rely upon *accounts of actions*. Whether it is data collected as part of a survey or via interviews or within a focus group or whatever, most sociological data are based upon a sample of research participants providing discursive accounts of some prior actions. One does not need to be an ethnomethodologist to accept that the accounts we routinely provide often have as much to do with the occasioned nature of a conversation or an interaction – be it with a researcher or in everyday communication – than with any exterior ontology existing independently of that interaction (Gilbert and Abell, 1983). Many sociologists would readily accept this premise in relation to complex narrative accounts of actions but, it turns out, even quite mundane and routine reporting of numbers, events, places, times and so on are not easily reconciled with data obtained by unobtrusive methods of ‘Big Data’ digital tracing (such as: where we said – and believed – we had been in the last 48 hours compared to what the GPS on our smart phone in our pockets reveals about where we actually were and when).

So here we have a conundrum. We have disciplines in the social sciences largely based on accounts of our actions being confronted by new data able to track, trace, record and sense our complex interactions with the social world. The metricization of social life derivable from the analysis of Big Data begins to reveal patterns of social order, movement and engagement with the world – and on such a scale – that it might demand nothing less than a fundamental *re-description* of what it is that needs to be explained and understood by the social sciences. Hence our concern was to begin

to rethink the descriptive power of the social sciences – to reinvigorate a sociological imagination able to grasp the complexities of the data and to visualize, map and otherwise represent it in ways that could claim back a distinctive jurisdiction over the study of the social (Burrows, 2011).

We have often been heartened by the critical response of the sociological community to our arguments, and there have been a range of positive, inventive and creative methodological developments of late that seem to be re-edifying our discipline (Adkins and Lury, 2012; Back and Puwar, 2013; Lury and Wakeford, 2013), not least the establishment of this innovative journal. However, in the space we have left in this commentary, rather than reviewing these developments we thought we would offer some critical reflections on a major project that Savage has recently been involved in that raises a number of issues of relevance to the matters already discussed. The Great British Class Survey (GBCS) is a telling instance of the new significance of a certain type of ‘Big Data’, and it is possible to learn certain lessons from the project that are revealing in a number of ways for the arguments we made in 2007.

The GBCS is Big Data only in relative terms. It is a hybrid project, which spliced together a fairly conventional social survey (providing accounts of actions), with a high-profile web platform hosted by the BBC asking a battery of questions about respondents’ economic, social and cultural capital (see <http://www.bbc.co.uk/science/0/21970879>). Its innovative features were that it was designed to be entirely answered through an interactive format; respondents who completed the 20-minute web survey were rewarded with a ‘coat of arms’ indicating the amount of economic, social and cultural capital they had. As it was a web survey, there was no upper limit to the number of respondents, or any controls about what kinds of people answered the survey.

In the end, an unusually large sample of 325,000 respondents completed the GBCS between January 2011 and June 2013, which easily dwarves any other survey on social class ever conducted in the UK. Of course, this is ‘small’ data by the standards of many of the new data sources that are discussed in this journal, but nonetheless it certainly challenges standard sociological repertoires. The research team, led by Mike Savage and Fiona Devine, analysed the data alongside a small nationally representative survey and argued that seven latent classes could be detected to link the measures of cultural, social and economic capital. This ‘seven class model’, with an ‘elite’ at the top, a ‘precariat’ at the bottom, and a variety of more diffuse groups within the middle ranks, was elaborated in an academic publication in April 2013 (Savage et al., 2013). When the initial results were published, they

were extensively publicized by the BBC on their news and digital platforms, so permitting a much greater media exposure than usual, at the heart of this strategy was the ‘class calculator’, a fast quiz which could be answered in a minute and which gave approximations to which of the seven classes respondents fell into. This is therefore certainly an interesting case study to explore what it reveals about new the repertoires of social research. Here we can make four major points.

Firstly, we can see that the ‘impact’ of social research is now fundamentally bound up with media circuits which operate largely autonomously of academic circuits. The impact of the GBCS was not due to the academic quality of the research – which is contested – but was fundamentally associated with a whole series of ‘data intermediaries’. These included BBC’s Lab UK who hosted the GBCS, the BBC’s own journalists who promoted the news story, and the web design team who invented the ‘class calculator’ which allowed people to find out their class by answering five questions, and which was answered by seven million people. This team won a Guardian prize for data journalism. The impact of the story was dependent also on the use of the social media in spreading and disseminating the findings. We can also see how digitization is used not only in generating but also in disseminating the data and findings, leading to a more complex research process itself.

Secondly, the story of the GBCS shows powerfully the stakes and tensions involved in social scientific research. Almost instantaneously, after the release of the news story, there was a critical reaction from social scientists who criticized the data analysis and reasserted the primacy of orthodox sociological repertoires, notably those based on the statistical analysis of large-scale nationally representative sample surveys (e.g. Mills, 2014). There was also a strong defence of existing models of class, enshrined in the National Statistics Socio-Economic Classification (Rose, 2013). The study thus became part of a battle over the ‘politics of method’ (on which see Savage, 2010). These defenders of orthodox repertoires speak from powerful positions, and with an institutional apparatus – organized around major national sample surveys – which is well funded and continues to command considerable authority. And they are also able to correctly identify problems – albeit ones that are also acknowledged by Savage et al. (2013) – regarding inference from non-representative samples. To this extent, there is no simple eclipse of old by new methods. Rather, we are seeing a methodological battle, the outcome of which is currently uncertain, but is likely to end up with greater demarcation and differentiation between the domains of ‘old’ and ‘new’ methods. And this politics of method intensifies when new modes of analysis are more directly in competition with the substantive foci

of traditional orthodox areas of research – such as the study of social class.

Thirdly, the GBCS itself elicited respondents who were relatively ‘elite’ and hence helped facilitate the mobilization of the kind of people who were themselves delineated in the GBCS itself. The fact that it was relatively well-educated and well-off respondents who were inclined to do the GBCS makes it possible to provide quite fine-grained analyses of this very group, so allowing a recursive loop between the producers and consumers of the research to be set in place (see Mike Savage’s inaugural lecture laying out this argument: <http://www.lse.ac.uk/newsAndMedia/videoAndAudio/channels/publicLecturesAndEvents/player.aspx?id=2118>). Furthermore, this recursive loop tightens over time, with those who responded to the GBCS after the initial media interest in April 2013 being more elite than the initial respondents. We thus see how a politics of ‘Big Data’ can help bring about a self-referential performativity in which the educated upper and middle classes are given a new mirror in which to look at themselves. We also need to note, however, that the working class tends to appear in such research repertoires as ghostly, invisible figures, and hence we emphasize the very real limitations of relying on sources such as the GBCS for comprehensive accounts of class relations.

Fourthly, and related to this, the GBCS shows a fundamentally different temporal structure of the research process to that embedded in conventional social scientific methods. Standard methods demarcate fieldwork and the acquisition of data from its analysis. Even though longitudinal designs allow return to fieldwork at later time periods, nonetheless it is still essential to demarcate these two operations. With the GBCS, this proved far more problematic. The publicizing of the findings in April 2013 caused a whole new wave of respondents to complete the web survey, the result being that the sample size more than doubled afterwards.

Even more than this, the public reaction to the news story is itself revealing for understanding the dynamics of social class. The widespread criticisms of classification systems which the public articulated on blogs and on the social media is revealing of a politics of class, as mobile, fluid and uncertain. The critical and ironic response to the news story, for instance by the ‘Emergent Service Workers Party’ (<http://spacehijackers.org/eSWP>), who ironically deployed one of the class labels as a badge to hold anarchist street parties outside Google HQ, is particularly interesting. In a further way, the massive deployment of GBCS results on the social media allows the tweets generated following the release to be analysed (<http://www.dhirajmurthy.com/a-quick-analysis-of-tweets-linking-to-the-great-british-class-survey/>).

Finally, the GBCS allowed us access to repertoires that are hardly possible using conventional measures.



An example is the fact that we can record the ‘real time’ when respondents submitted their replies to the GBCS. Precisely because the interview was not done by an interviewer, the timing becomes a matter of interest as it was the choice of the respondent and therefore revealing of their temporal organization. If we look at which classes are over-represented at different times of the day, we can see the ‘precariat’ were more likely to submit the GBCS in the early hours of the morning, whereas between 5.00 and 6.00 it was the ‘elite’ who took over from them, perhaps indicating their relatively early start to their working day. These differences recur at the end of the working day, between 17.00 and 19.00. By contrast, some classes, such as the established middle classes or the emergent service workers, rarely vary from the norm.

What follows from these observations is that if the GBCS is judged purely by the criterion of orthodox social science, then it can be found wanting since it does not approximate to extensively validated and legitimated methods. However, the GBCS data allows possible repertoires of research that are not possible using orthodox methods – such as using crowdsourcing techniques to generate additional data, allowing more specific analyses of time periods, allowing much more refined and granular analysis as a result of the larger sample size, and so forth. Having said this, the GBCS is controversial in part because it is hybrid, both in its conceptual framing and partly because it uses a (small) national representative survey alongside a large web survey. In straddling both old and new modes of social scientific research, it is unsurprising that it should attract especial attention and be subject to turf warfare.

What is evident here, therefore, is that the use of new data sources involves a contestation over the social itself (see further Ruppert et al., 2013). It is not that ‘Big Data’ (or, to be exact in the case of the GBCS, large scale digital data) readily provides better understandings of the social as it is understood within orthodox, variable-centred social science. Instead, it permits a different kind of more temporally and spatially specific set of analyses which allow a more granular conception of the social to be delineated. This is more attuned to ‘outliers’ and to particular cases, rather than to aggregate banded groups.

## Conclusion

Returning to the arguments of Savage and Burrows (2007) with the GBCS in mind, we can see that the crisis of empirical sociology is far reaching indeed. New data sources bring with them different modes of addressing the public, mobilizing expertise, conceptualizing the social, and research methodology. Several elements of the GBCS method are similar to those

usually understood in relation to Big Data in encouraging a kind of interactive, dynamic, recursive (i.e. responses mediated by other responses, media reports), and digitally circulated set of social inscriptions. The GBCS performs and produces sociality as much as it describes it.

This crisis, we suggest, is one which does not unite experts in a quest to explore the potential of new modes of Big Data, but instead is likely to polarize and divide. We can anticipate that this will lead to different research orientations which do not engage with each other, as well as moments of intense contestation, as with the GBCS. Of one thing, however, we can be certain. Big Data does challenge the predominant authority of sociologists and social scientists more generally to define the nature of social knowledge. It permits a dramatically increased range of other agents to claim the social for their own. It is for this reason that we maintain that we were right in our 2007 paper that sociologists need to be prepared to intervene in the world of Big Data in order to ensure we command a voice in this new terrain.

## Declaration of conflicting interest

The authors declare that there is no conflict of interest.

## Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

## References

- Adkins L and Lury C (2012) *Measure and Value*. Oxford: Wiley Blackwell.
- Back L and Puwar N (2013) *Live Methods*. Oxford: Wiley Blackwell.
- Beer D and Burrows R (2007) Sociology and, of and in Web 2.0: Some initial considerations. *Sociological Research Online*. [www.socresonline.org.uk/12/5/17.html](http://www.socresonline.org.uk/12/5/17.html).
- Beer D and Burrows R (2013) Popular culture, digital archives and the new social life of data. *Theory, Culture & Society* 30: 47–71.
- Burrows (2011) Visualisation, digitalisation and the ‘descriptive turn’ in contemporary sociology. In: Heywood I and Sandywell B (eds) *The Handbook of Visual Culture*. London: Berg, pp. 572–588.
- Burrows R (2012) Living with the H-index? Metric assemblages in the contemporary academy. *Sociological Review* 60(2): 355–372.
- Burrows R and Gane N (2006) Geodemographics, software and class. *Sociology* 40(5): 793–812.
- Gilbert GN and Abell P (1983) *Accounts and Action*. Aldershot: Gower.
- Gouldner A (1970) *The Coming Crisis of Western Sociology*. London: HEB.

- Kelly A and Burrows R (2012) Measuring the value of sociology? Some notes on the performative metricisation of the contemporary academy. In: Adkins L and Lury C (eds) *Measure and Value: A Sociological Review Monograph*. Oxford: Wiley-Blackwells, pp. 130–150.
- Lury C and Wakeford N (2013) *Inventive Methods: The Happening of the Social*. London: Routledge.
- Mayer-Schönberger V and Cukier K (2013) *Big Data: A Revolution that will Transform How We Live, Work and Think*. London: John Murray.
- McKie L and Ryan L (eds) (2012) Introduction to e-special issue: Exploring trends and challenges in sociological research. *Sociology* 46(6): 1–7.
- Mills C (2014) The Great British Class Fiasco: A comment on Savage et al. *Sociology*. Epub ahead of print 14 March 2014. DOI: 10.1177/0038038513519880.
- Osborne T, Rose N and Savage M (2008) Reinscribing British sociology: Some critical reflections. *Sociological Review* 56(4): 519–534.
- Ritzer G and Jurgenson N (2010) Production, consumption, prosumption: The nature of capitalism in the age of the digital ‘prosumer’. *Journal of Consumer Culture* 10: 13–36.
- Ruppert E, Law J and Savage M (2013) Reassembling social science methods: The challenge of digital devices. *Theory, Culture & Society* 30(4): 22–46.
- Rose D (2013) Little solidarity over the question of social class. *The Guardian*, 5 April.
- Savage M (2010) *Identities and Social Change in Britain Since 1940: The Politics of Method*. Oxford: Oxford University Press.
- Savage M and Burrows R (2007) The coming crisis of empirical sociology. *Sociology* 41(5): 885–899.
- Savage M and Burrows R (2009) Some further reflections on the coming crisis of empirical sociology. *Sociology* 43(4): 765–775.
- Savage M, Devine F, Cunningham N, et al. (2013) A new model of social class? Findings from the BBC’s Great British Class Survey experiment. *Sociology* 47(2): 219–250.
- Thrift N (2005) *Knowing Capitalism*. London: Sage.
- Webb EJ, Campbell DT, Schwartz RD, et al. (1966) *Unobtrusive Methods: Nonreactive Research in the Social Sciences*. Chicago: Rand McNally.