LSE Research Online

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

Nick Couldry and Alison Powell

Big data from the bottom up

Article (Published version)
(Refereed)

Original citation:
Couldry, Nick and Powell, Alison (2014) *Big data from the bottom up*. Big Data & Society, 1 (2). ISSN 2053-9517 DOI: 10.1177/2053951714539277

© 2014 The Authors. Published by SAGE Publications.

This article is distributed under the terms of the Creative Commons Attribution 3.0 License which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page.

This version available at: http://eprints.lse.ac.uk/57941/
Available in LSE Research Online: July 2014

LSE has developed LSE Research Online so that users may access research output of the School. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LSE Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain. You may freely distribute the URL (http://eprints.lse.ac.uk) of the LSE Research Online website.

http://eprints.lse.ac.uk

Big Data & Society
http://bds.sagepub.com/

Big Data from the bottom up

Nick Couldry and Alison Powell

Big Data & Society 2014 1:
DOI: 10.1177/2053951714539277

The online version of this article can be found at:
http://bds.sagepub.com/content/1/2/2053951714539277

Published by:
SAGE
http://www.sagepublications.com

On behalf of:

Additional services and information for *Big Data & Society* can be found at:

Email Alerts: http://bds.sagepub.com/cgi/alerts

Subscriptions: http://bds.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav

>> Version of Record - Jul 1, 2014

What is This?

Downloaded from bds.sagepub.com at London School of Economics & Political Sciences on July 17, 2014

Commentary

BIG DATA & SOCIETY

Big Data & Society
July–September 2014: 1–5
© The Author(s) 2014
DOI: 10.1177/2053951714539277
bds.sagepub.com
SAGE

Big Data from the bottom up

Nick Couldry and Alison Powell

Abstract

This short article argues that an adequate response to the implications for governance raised by 'Big Data' requires much more attention to agency and reflexivity than theories of 'algorithmic power' have so far allowed. It develops this through two contrasting examples: the sociological study of social actors used of analytics to meet their own social ends (for example, by community organisations) and the study of actors' attempts to build an economy of information more open to civic intervention than the existing one (for example, in the environmental sphere). The article concludes with a consideration of the broader norms that might contextualise these empirical studies, and proposes that they can be understood in terms of the notion of voice, although the practical implementation of voice as a norm means that voice must sometimes be considered via the notion of transparency.

Keywords

Agency, reflexivity, analytics, political economy, voice, transparency

Introduction

We are living through a transformation of governance – both its mechanisms and reference-points – which is likely to have profound implications for practical processes of government and everyday understandings of the social world. A shift is under way from discrete forms of intervention in social space based on intermittent and/or specific information-gathering to continuous processes of management based on total and unremitting surveillance (Ruppert, 2011). Both management and government increasingly are becoming predicated upon the continuous gathering and analysis of dynamically collected, individual-level data about what people are, do and say ('Big Data'). However misleading or mythical some narratives around Big Data (Boyd and Crawford, 2011; Couldry, 2013), the actual processes of data-gathering, data-processing and organisational adjustment associated with such narratives are not mythical; they constitute an important, if highly contested, 'fact' with which all social actors must deal. This article will offer a social approach to the construction and use of such data and related analytics.

The possibility of such a social approach to Big Data has, until now, been obscured by unnecessarily generalised readings of the consequences of these broad changes. Without a doubt, the information types that management and governance take as their starting-point have changed: it is digital infrastructures of collection, transmission, analysis and presentation that have made possible continuous data-mining. Compared to representative sampling, such new approaches to data collection are totalising; they are also characterised by the aggregation of multiple data sets through the use of calculation algorithms. This seemingly increased role for algorithms has led some commentators to focus on the dominance of 'algorithmic power' (Lash, 2007), an approach that leaves no room for agency or reflexivity on the part of 'smaller' actors. We posit that emerging cultures of data collection deserve to be examined in a way that foregrounds the agency and reflexivity of individual actors as well as the variable ways in which power and participation are constructed and enacted.

London School of Economics, London, UK

Corresponding author:
Nick Couldry, Department of Media and Communications, London School of Economics, Houghton St., London WC2A 2AE, UK.
Email: N.Couldry@lse.ac.uk

Creative Commons (CC-BY): This article is distributed under the terms of the Creative Commons Attribution 3.0 License (http://www.creativecommons.org/licenses/by/3.0/) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (http://www.uk.sagepub.com/aboutus/openaccess.htm).

Downloaded from bds.sagepub.com at London School of Economics & Political Sciences on July 17, 2014

This more agent-focused inquiry into the consequences of algorithmic calculation's deep embedding in everyday life has been foreshadowed in some earlier debates, notably Beer's (2009) response to Lash's (2007) argument that 'algorithmic power' has changed the nature of hegemony. As Beer (2009: 999) noted, sociology must also 'focus ... on those who engage with the software in their everyday lives'. Such a focus does not come naturally within Lash's broadly philosophical formulations of issues in social theory which foreground 'a collapse of ontology and epistemology' (Lash, 2006: 581), and a new power-laden regime of 'facticity' (Lash, 2007: 56) in which 'there is no time, nor space ... for reflection' (Lash, 2002: 18). If that were right, why pay close attention to what actors say when they 'reflect' on their position in the social world? But this analytic closure is unhelpful.

Needed instead is a more open enquiry into what actual social actors, and groups of actors, are doing under these conditions in a variety of places and settings. Without denying of course the 'generative' importance of algorithms (Lash, 2007: 71) when embedded in modes of calculation, processing and rule, we need to remember that social actors are often themselves aware of being classified. Even if they are not privy to the details of when, by whom, and how they have been classified, that this has happened is something of which they are aware, and indeed one of the main 'facts' they have to deal with as social actors. We need to become sensitive to what Beer (2009: 998) has called people's 'classificatory imagination' and, over the longer term, the wider 'social imaginaries' (Mansell, 2012; Taylor, 2005) that may be emerging around these new cultures of data collection.

Beer goes on helpfully to distinguish three levels of resulting empirical research: first, regarding the 'organizations that establish and activate Web 2.0 applications'; second, regarding the 'software infrastructures and their applications on the web'; and third, regarding how the first two levels 'play out in the lives of those that use (or do not use) particular web applications' (2009: 998). We would like in this short article to build particularly on Beer's third level, and on the lessons of our own empirical researches, to map out some more detailed and concrete ways of researching the everyday uses of data and analytics from a social perspective. The result is to open up a much wider and more varied space of agency and reflexivity than allowed for in philosophical accounts. The likely outcome may be no less critical of Big Data's implications, but will develop critique through a more nuanced characterisation of 'Big Data' as a variegated space of action, albeit one very different from the spaces in which pre-digital social actors operated.

Doing social analytics

Our first example of a more agent-focused account of Big Data is what has been called 'social analytics' (see Couldry et al., forthcoming, for a much more detailed account). A social analytics approach is an explicitly sociological treatment of how analytics get used by a range of social actors. Such an approach aims to capture how particular actors reflect upon, and adjust, their online presence and the actions that feed into it, through the use of 'analytics'. 'Analytics' here is used broadly to cover both basic analytics (the automated measurement and counting installed within the operation of digital platforms and associated websites, apps and tools) and the adjustments made by actors themselves in response to such measurement and counting operations. Platforms that count and sort online data, such as Google and Facebook, work automatically via algorithms, often allowing users only limited degrees of manual adjustment (van Dijck, 2013). Other adjustments around those operations may take direct digital form (a website redesign) or organisational form (an adjustment in an organisation's management of its resources). In all these cases, the variable use of analytics is a social process involving reflection, monitoring and adjustment.

By 'social actors' we mean actors with social ends over and above the basic aim of generating and analysing data (usually for profit): that basic aim in itself is of little sociological interest. The broader sociological interest starts when there is some tension, actual or potential, between the aims that social actors are trying to achieve and the interpretations of their activities that analytics generate. This use of the term 'social analytics' encompasses, but goes beyond, the everyday 'technical' use of the term 'analytics' to mean the measurement and reporting of internet data. The mutual intertwining of human and material agency is hardly a new insight (Pickering, 1995: 15–20), but it acquires a special interest when analytics' operations are opaque to non-expert social actors who must work hard to acquire control over them.

One key variable in such research is what is measured and analysed, the 'object' of analytics. The underlying data's relationship to an organisation's online presence may be more or less direct: direct if the data is literally about that organisation's online presence (numbers of unique users, their characteristics, types of interaction with online content); or indirect if the data is not about an organisation's online presence, but is generated or presented online, becoming part of how that organisation is judged by online visitors (online reviews, debates). The closeness, or distance, of the relation between the object of data analysis and the general aims and practice of social actors

Downloaded from bds.sagepub.com at London School of Economics & Political Sciences on July 17, 2014

clearly will shape the degree of tension and reflexivity that exists over the implementation of analytics. At one end of the spectrum will be cases where analytics are used directly to support other mechanisms of power (e.g. performance management); at the other end will be cases where what is at stake in the use of analytics is the broad redefinition of an organisation's aims and performance, with no direct impact on the evaluation or management of individuals. In the former case, social analytics may merge into the study of management and power; in the latter case, social analytics may be something closer to a phenomenology of how social actors and organisations with social aims appear to themselves, and to the world, under digital conditions.

Other variables when doing social analytics will include the degree of technical expertise of the actors involved, including the degree to which they can go beyond merely using off-the-shelf analytics to customising them, or perhaps even developing their own analytic tools and data-collection designs. Financial and other resources will also affect how far the processes which social analytics studies can develop, or get blocked, for example, if the staff to do the analytic work that would enable a richer re-evaluation of an organisation's digital presence cease to be available. Expertise and resources are, of course, variables in any fieldwork setting.

Within these basic parameters, however, social analytics promise a rich vein of inquiry into the conditions of data use and analytics use, from the perspective of social actors who are not principally experts in relation to data or algorithms, but who look to them to do certain work towards other ends. It has so far been explored in the context of community and civic activism, but it has the potential to be expanded to many more areas.

Data as media

For media scholars more generally, the shift to a data-rich environment poses challenges for a robust understanding of how agency and expression might still work within that environment. The critical tradition in media and communications has largely been concerned with the operation of power in the construction of systems of symbolic mediation – for example, the function of ideological systems (in the Marxist tradition) or the Gramscian concept of hegemony. These strategies have allowed media and communication scholars to 'work backwards' through systems of symbolic mediation in order to understand the process and initial starting points of mediated 'messages'. This focus on the symbolic quality of media messages allows us to examine power relationships from several different vantage points. Within traditional broadcast media forms we can observe how the symbolic control of mediated messages solidifies control and results in things like propaganda, but we can also see how alternative media producers can wrest control of ideas and their representation to challenge that kind of hegemony.

Broadcast models have however been overtaken, for important purposes, by models of mass self-communication. Whereas institutionalised mass media is structured to disseminate messages from one to many, mass self-communication is structured to invite continual input of data by individuals. This reorganisation of media production initially seemed to promise a reconfiguration of the top-down production of ideology and the bottom-up resistance to it, but as political–economic analyses have developed, we are beginning to see how such shifts have also led to the production of data replacing the production of audiences.

If the exemplary product of institutionalised mass media is propaganda, the exemplary product of mass self-communication is data. A mass media apparatus requests information to be disseminated from the one to the many; its economic model uses this information to generate an audience whose attention can be sold to an advertiser. In the mass self-communication model individuals are still part of an aggregate product to be sold, but instead of their attention on a single message produced for broadcast, it is their individual acts of communication that comprise the 'Big Data' and drive much media value-extraction.

Early critics of mass self-communication noted that the model encouraged individuals to create 'content' that was then sold to others in order to capture their attention (Terranova, 2000; van Dijck, 2013). However, 'content' is still expressive, even when it is sold to capture attention. A more complicated issue concerns the data that is produced, often unwittingly, which now generates much of the value in the newest iteration of the contribution economy. Many everyday activities now produce data without requiring human meaning-construction (or even basic consent). The rise of sensor networks has meant that increasingly individuals are producing not 'content' composed of messages containing intrinsic or constructed meaning, but mere data – temperature readings, status updates, location coordinates, tracks, traces and check-ins. Not one of these individual data-types is necessarily meaningful in itself – but taken together, either through aggregation, correlation or calculation, such data provide large amounts of information. The difference between this and the 'content' that mass self-communication promises to distribute is that the meaning of data is made not semantically (through expression and interpretation) but through processing – especially the matching of metadata (Boellstorf, 2013). Big Data sets are composed of numerous pieces of information that can

Downloaded from bds.sagepub.com at London School of Economics & Political Sciences on July 17, 2014

be cross-compared, aggregated and disaggregated and made very finely grained, not things whose creators necessarily endowed with meaning. In mining the data, more insights are made available about more aspects of everyday life but no opportunity is provided for these insights to be folded back into the experience of everyday life. In this context, is there any scope, as Boellstorf urges, for integrating the epistemic perspectives of ethnography back into the calculative logic of meta-data?

All along, the political economy of personal data, as anticipated by Gandy (1993), has been concerned with value created through the aggregation and calculation of individual traces. Even if we leave aside the expressive quality of individual acts of communication online, the production of data as a by-product of everyday life practices enacts a particular political economics of media, undertaken within a situation of pervasive surveillance and generalised authoritarianism (Cohen 2012). But the potential disconnect between system and experience, phenomenology and political economy, can be overcome by examining on the ground agents' strategies for building alternative economies of information. Such alternative economies are being developed in several areas related to environment and sustainability, including projects that use data sources to make provenance and supply chains visible, and those that encourage individuals and communities to collect data as a means to make environmental issues visible by challenging conventional data collection.

Academic projects like Wikichains (Graham, 2010) and start-up companies like Provenance.it (2013) aggregate various forms of data about the production, distribution and supply chains of manufactured objects as a means of drawing attention to their long-term ecological and economic costs. While Provenance.it remains anchored in a consumer-based economic model, it does illustrate how alternative modes of data collection and analysis could shift agency and representation, especially if it permitted for greater reflexivity. Similarly, NGOs like Mapping for Change (2013) have supported individuals and community groups in gathering environmental data (like air quality and noise) as a means of engaging with gaps and flaws in official data. These actions intervene in efforts to use such environmental data within top-down governance processes. As Gabrys (2014) identifies, such citizen science efforts must be enfolded and imagined in processes of environmental governance or 'biopolitics 2.0'. These examples illustrate two ways that an alternative economics of information might employ calculation of multiple data sources or generation of alternative sources to illustrate or critique power relations, although they also illustrate the ambiguity of accountability within these processes.

Voice, transparency and power

The rise of analytics presents a significant normative challenge for scholars, activists and others who seek to understand how humanity, sociability and experience are represented. The daily practices of grappling with data and with the consequences of data analyses generate new questions about what and whose power gets exercised through such practices, and to what degree such exercises of power are satisfactorily made accountable. One approach to these challenges is through attention to problems of voice (Couldry, 2010). Voice, understood as a value for social organisation (Couldry, 2010: ch. 1), involves taking into account agents' practices of giving an account of themselves and their conditions of life. The value of voice is essential to the workings of any models so far developed of democratic institutions, but it is not immediately compatible with a world saturated with the automated aggregation of analytic mechanisms that are not, even in principle, open to any continuous human interpretation or review.

While the notion of voice insists upon organisational processes being accountable to the subjectivities and expressiveness of all, the movement towards more casual, automatic sensing and its calculative rather than epistemic logic seems to eliminate this accountability. Yet clearly something similar to 'voice' is required in this new world, and this is not just a matter of democracy: 'we have no idea', wrote Paul Ricoeur, 'what a culture would be where no one any longer knew what it meant to narrate things' (Couldry, 2010: 1, quoting Ricoeur, 1984: 29). At present, the proxy for voice in the algorithmic domain is the notion that data gathering processes ought to be transparent, and the logic of calculation revealed. A focus on transparency could begin to foreground notions of accountability in data calculation, ownership and use.

Notions of transparency have been discussed with respect to government production and use of data (Tkacz, 2012). Yet despite pledging to make public data collection transparent, governments like the US and the UK in fact collect much more information via surveillance projects and partnerships with information technology companies. With the reform of the USA's National Security Administration, perhaps more attention will begin to be paid to the data collection practices of the technology sector, making more of them visible. This kind of transparency goes part of the way to establishing accountability, but it still fails to address accountability and reflexivity. A refined concept of transparency that is sensitive to the meaning that data trails might form (even if it cannot be sensitive to the meaning inherent in their production) might go some way to addressing this. This is a

Downloaded from bds.sagepub.com at London School of Economics & Political Sciences on July 17, 2014

tricky proposal: unless and until the unconscious production of data can be conceived of as a form of expression, the philosophical basis for such an expansive transparency will be difficult to establish. One possible way to proceed might be to highlight not just the risks of creating and sharing data but the opportunities as well. The practices of social analytics and citizen science have the potential to establish these opportunities, ambiguous as they may be.

We hope that, as the debates about Big Data and society continue and their democratic stakes become clearer, the values implicit in the terms 'voice' and 'transparency' will themselves begin to converge in more satisfying ways than are at present possible.

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

Beer D (2009) Power through the algorithm? Participatory web cultures and the technological unconscious. *New Media & Society* 11: 985–1002.

Boellstorff T (2013) *Making Big Data, in Theory*. First Monday, [S.l.], September 2013. ISSN 13960466. Available at: http://firstmonday.org/ojs/index.php/fm/article/view/4869/3750 (accessed 27 January 2014).

boyd d and Crawford K (2011) Critical questions for Big Data: Provocations for a cultural, technological and scholarly phenomenon. *Information, Communication and Society* 15(5): 662–679.

Castells M (2009) *Communication Power*. Oxford: Oxford University Press.

Cohen J (2012) *Configuring the Networked Self*. New Haven: Yale University Press.

Couldry N (2010) *Why Voice Matters*. London: Sage.

Couldry N (2013) *A Necessary Disenchantment: Myth, Agency and Injustice in a Digital Age*. Inaugural lecture at LSE. Available at: http://www.lse.ac.uk/newsAndMedia/videoAndAudio/channels/publicLecturesAndEvents/player.aspx?id=2120 (accessed 21 November 2013).

Couldry N, Fotopoulou A and Dickens L (forthcoming). *Real Social Analytics: A Contribution Towards the Phenomenology of a Digital World*.

Gabrys J (2014) Programming environments: Environmentality and citizen sensing in the smart city. *Environment and Planning D: Society and Space* 32(1): 30–48.

Gandy O (1993) Toward a political economy of personal information. *Critical Studies in Mass Communication* 10(1): 70–97.

Graham M (2010) *'WikiChains: Encouraging Transparency in Commodity Chains' Research Project*. Available at: http://www.oii.ox.ac.uk/research/projects/?id=75 (accessed 30 May 2014).

Lash S (2002) *Critique of Information*. London: Sage.

Lash S (2006) Dialectic of information? A response to Taylor. *Information Community & Society* 9(5): 572–581.

Lash S (2007) Power after hegemony: Cultural studies in mutation. *Theory, Culture & Society* 24(3): 55–78.

Mansell R (2012) *Imagining the Internet*. Oxford: Oxford University Press.

Mapping for Change (2013) *Services: Citizen Science*. Available at: http://www.mappingforchange.org.uk/services/citizen-science/ (accessed 30 May 2014).

Pickering A (1995) *The Mangle of Practice*. Chicago: Chicago University Press.

Provenance.it (2013) *About Provenance*. Available at: https://www.provenance.it/about (accessed 30 May 2014).

Ricoeur P (1984) *Time and Narrative*, Vol. 2. Chicago: Chicago University Press.

Ruppert E (2011) Population objects: Interpassive subjects. *Sociology* 45(2): 218–233.

Taylor C (2005) *Modern Social Imaginaries*. Durham, NC: Duke University Press.

Terranova T (2000) Free labor: Producing culture for the digital economy. *Social Text* 18(2): 33–58.

Tkacz N (2012) From open source to open government: A critique of open politics. *Ephemera* 12(4). Available at: http://www.ephemerajournal.org/contribution/open-source-open-government-critique-open-politics-0 (accessed 30 May 2014).

Van Dijck J (2013) *The Culture of Connectivity*. Oxford: Oxford University Press.

Downloaded from bds.sagepub.com at London School of Economics & Political Sciences on July 17, 2014