


Geography and the future of big data, big data and the future of geography

Dialogues in Human Geography
3(3) 255–261
© The Author(s) 2013
Reprints and permission:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/2043820613513121
dhg.sagepub.com


Mark Graham

University of Oxford, UK

Taylor Shelton

Clark University, USA

Abstract

As digital social data have become increasingly ubiquitous, many have turned their attention to harnessing these massive data sets in order to produce purportedly more accurate and complete understandings of social processes. This intervention addresses the relationships between geography and big data and their intertwined futures. We focus on the impacts of an age of big data on the discipline of geography and geographic thought and methodology, as well as how geography might provide a useful lens through which to understand big data as a social phenomenon in its own right. Ultimately, we see significant potential in big data, but remain skeptical of the prevalent discourses around it, as they tend to obscure, more than reveal, the complexity of social and spatial processes.

Keywords

big data, geographic thought, positivism

Data are now ubiquitous. Sensors and software are digitizing and storing all manner of social, economic, political, and environmental patterns and processes. As the size of these data sets has increased exponentially, many have begun to focus on how ‘big data’ can allow potentially unprecedented insights into our world. Furthermore, much of these data are geographic in nature and contain either explicit or implicit spatial information. Now that the big data produced by so many places and processes can be mapped, measured, and analyzed, many have argued that big data possesses the potential to produce fundamentally new ways of knowing, enacting, and being in the world.

While there is an effort underway to critically interrogate discourses and practices surrounding big data and related shifts in social science practices (Boyd and Crawford, 2012; Crampton et al., 2013; Floridi, 2012; cf. Miller, 2010; Savage and Burrows, 2009), there has yet to be a significant, sustained effort to understand its geographic relevance. As such, it is specifically the links between big data and

Corresponding author:

Taylor Shelton, Graduate School of Geography, Clark University,
950 Main Street, Worcester, MA 01610, USA.

Email: jshelton@clarku.edu

its *spatial* contexts, meanings, processes, and effects that we seek to interrogate in this forum.¹

Interrogating the relationship between geography and big data first requires reflecting on what we mean by geography and what we mean by big data. We are concerned with geography as an institutionalized academic discipline, as a way of thinking about the world (consistent with what has been termed the ‘spatial turn’ in the social sciences), and as a particular quality that big data can possess. Similarly, we seek to understand big data not as an entirely unified and coherent thing around which we should police boundaries but as a set of discourses, objects, and practices. It is both the assemblage of skills, information, and resources to analyze and something that can itself be subject to analysis.

The first thing we might want to ask is what are data, and at what point do they become big? As long ago as the 17th century, philosophers were employing the word ‘data’ to refer to information assumed to be factual (Taylor, 2012). In a contemporary sense, we take the word to indicate information that is collected through some form of measurement as well as the lowest level of abstraction in any sort of meaning making.

Big data, however, are more difficult to define. The modifier ‘big’ is always relative and represents a moving target. Contemporary ‘small data’ were extraordinarily large a half-century ago, and contemporary notions of big data will likely be tiny just a half-century into the future, making it difficult to define the ‘bigness’ of data by its size or volume alone. Is it therefore more useful to think of big data as being transactional? Many, including several of the commentaries that make up this forum, turn to Laney’s (2001) definition of big data as possessing the three V’s: volume, velocity, and variety. Or should big data be defined not by the data itself but by the effects that it engenders? Perhaps big data could be defined as representing a broader computational paradigm in research and practice, in which automated algorithmic analysis supplants domain expertise.

We initially posed a series of questions to our panelists to inform their presentations, which in turn guided the pieces published in this forum.² But

rather than restating these questions and attempting to answer them ourselves, we instead find it more useful to reframe the resultant discussion around two opposing, but related strands visible within our original approach to the panel: geography and the future of big data, and big data and the future of geography. That is, we are interested in understanding what the so-called ‘big data revolution’ means for the practice of geography as well as how geography offers an important perspective on the social (and spatial) implications of big data. Doing so allows us to set out what we see as the central lines of promise, possibility, and peril of the coming together of geography and big data.

Big data and the future of geography

Whatever exactly big data is, it appears as though something important has changed. We are in a moment where we are faced with both ever fewer barriers to the collection, storage, processing, and outputting of data and ever more data being created from people interacting with other people, machines, the environment, our cities, and virtual code and architectures. The flood not just of data but of the ways that we collect, collate, and use them is already having profound implications on a range of social processes and practices. But in the rush to start utilizing big data, many have neglected questions about whether or how it might be integrated into preexisting structures of scholarly knowledge production.

In his seminal 2008 piece in *Wired Magazine*, Chris Anderson suggested that big data might spell the end of theory, as a wealth of readily accessible data could be all we need in order to understand any given phenomena (Anderson, 2008). We would no longer have to speculate and can simply observe and measure. Although Anderson’s piece has been widely ridiculed for its naïveté, it raises an important question: What do big data mean for how we do research and create knowledge? How does big data change how we position ourselves in relation to our object of study, our methodologies, our epistemologies, our funding sources, and the ways in which we understand truth?

As we codify ever more of what we are and what we know in digital data shadows, big data has

emerged as not just a way of describing data itself, and our increased prowess in measuring, mapping, analyzing, and visualizing, but a meme that speaks to and produces new ways of establishing truth. The increasing reliance on relatively new types of data presences in order to feed ways of measuring, models, algorithms, and information systems could mean that knowledges that are not so easily encapsulated within big data frameworks might become devalued.

Although big data presents challenges to conventional notions and practices of 'hard' science, including to the field of geographic information science (Goodchild, 2013; Gorman, 2013), our worry is that the prevalence of the big data meme might lead to a new scientific, positivist, and quantitative turn in the social sciences by reducing the space for critical, qualitative, and postpositivist research. As Ruppert (2013) points out, this concern is one that has been expressed throughout the social sciences: particularly because the push toward interdisciplinary computational approaches has the potential to obscure social processes and practices underlying data rather than elucidating them and their intermingling with research methods. Contrary to the understandings of quantification as inherently objective (Porter, 1996), data never speak for themselves, as the ways that we collect, analyze, and produce knowledge about the world are inevitably loaded with particular assumptions that limit the kinds of patterns, processes, experiences, values, and ideas that can be captured in databases.

In a world in which commercial analysis is coming to outpace academic research (Savage and Burrows, 2007), academics and others asking critical questions might therefore have less say about both the types of data being generated and the ways that those data can be accessed, ultimately narrowing the potential of knowledge production, rather than vastly expanding it. In other words, although big data approaches are necessarily handicapped by their own restrictive vision, the big data meme encourages the reproduction of a very different narrative: One in which big data allows us to objectively measure and map the world as it actually is in order to arrive at fundamental truths.

While big data might appear to restrict possibilities for the practice of critical geography, the history

of the discipline points to a different future. Geography has long been characterized by a diverse set of competing, but coexisting, methodological, and epistemological paradigms and practices, and geographers have *already* been grappling with some of the central issues raised by big data for some time (Barnes, 2013). From the Hartshorne–Schaefer debate of the 1950s (Hartshorne, 1955; Pattison, 1964; Sack, 1974; Schaefer, 1953) to the radical critique of geography's quantitative revolution (Harvey, 1973 [2009]) to the geographic information system (GIS) and society debates of the 1990s (Openshaw, 1991; Pickles, 1995; Schuurman, 2000; Sheppard, 1995), geographers have long struggled over what the appropriate ends of our scholarship should be, how we should be doing it, and how to accommodate competing claims to truth, especially in the context of new technologies opening up new methodological possibilities. These debates have been incredibly productive, however, and have given rise to work that straddles the boundaries between diverse subfields. The development of critical GIS, radical approaches to quantitative geography, and the emergence of critical physical geography all lend credence to the idea that geographers are well positioned to combine technical practice, quantitative methods, and critical scholarship (Crampton and Krygier, 2005; Sheppard, 2001, 2005; Lave et al., 2013; Pavlovskaya, 2006; Wyly, 2009).

It is precisely this history that we hope will allow the discipline to both engage with and critique big data. Utility can undoubtedly be found in the ready availability of data and the frameworks through which we analyze it. But value should also be placed on deconstructing and critiquing any self-reinforcing claims to validity and truth propagated by the big data meme.

Geography and the future of big data

How then can we utilize the tools of critical human geography to understand big data as a social and spatial phenomenon and set of practices? Using a series of examples, this section reflects on a diverse set of ways that geography might be enrolled to both advance and disrupt big data scholarship.

Perhaps the most discussed social implication of big data has been its impact on privacy. As we produce increasingly thick data shadows through our everyday lives, it is easier than ever to identify individuals, create behavioral profiles, and make use of predictive categories. Such activities have been employed not only by government agencies but also by large private firms. In some instances, such as the US government's PRISM program, metadata alone were sufficient to build comprehensive portraits of individuals and groups. Here, we saw the power of not just large intrusive data sources but the triangulation of multiple sources to paint comprehensive pictures about human activity. However, the privacy concerns raised by these programs should not be seen as only emerging from the particular discourses and practices falling under the banner of big data, as the development of geodemographic information systems in the 1990s raised similar issues relating to the targeting of individuals and groups based on comprehensive profiles (Curry, 1997; Goss, 1995; Parker et al., 2007).

Big data will also likely have significant effects on the practice of politics. On the one hand, data-driven governance could have the potential to bypass political posturing on issues such as climate change, health care and immigration, allowing for more open and transparent understandings of complex issues. The availability of data can potentially increase openness, trust, and transparency in politics, though it can also allow candidates to tailor their messages to specific demographics without having to appeal to the populace as a whole (Kitchin, 2013). On the other hand, an increasing reliance on data in the political process opens up worries of ceding politics itself to data, tools, and machines. Allowing data and algorithms alone to determine policy can never be a value-neutral or depoliticized process, as data and the frameworks through which they are collected and presented are inherently imbued with politics, whether drawn from the creator of a particular tool or the social context in which it is embedded (Winner, 1980).

The veil of neutrality given to opaque software algorithms and databases has important implications for our trust in such technologies. Activities as simple as shopping or navigating through a city now often rely on search engines and the black-

boxed algorithms and ephemeral data that power them (Graham et al., 2013; Introna and Nissenbaum, 2000; Zook and Graham, 2007). Yet, it is those very platforms that shape how, where, and why we move, consume, act, and interact. Here there is much that critical geographers can do to deconstruct the decisions that are built into such invisible processes (Kitchin and Dodge, 2011). Given the increasing prominence of such location-based services, it is important for geographers to take an active role in asking not just 'where?' but also 'why?' and 'how?'

From promises to perils of big data

While geography has important implications for big data, and big data has important implications for geography, we should not forget that big data have their own geographies. Because data are always constructed, collected, stored, and used under uneven and variegated social, economic, and technical contexts, some people, places, and processes will always be easier to enroll into such vast sociotechnical assemblages.

But what are we to make of those that are left out or that are difficult to data-fy? What of those social processes that do not easily fit into the classification and categorization schemes and rules that frame ever more facets of everyday life? We have seen that far from being inclusive, advancements in information and communication technologies have often amplified the sociospatial unevenness of representation and participation in a range of online data sets (Crutcher and Zook, 2009; Graham, 2011; Haklay, 2013).

These presences and absences in data matter not simply because they are evidence of material inequalities manifesting themselves in digital contexts but also because digital data in turn have real, material effects in the world. For example, it is instructive to turn to the history of critical work addressing the connections between quantification, financialization, and urbanization in order to understand how different visions of data ('big' or otherwise) have long been employed to enact or reinforce particular uneven socioeconomic outcomes and depoliticize questions of knowledge production, social justice, and the distribution of resources (Light, 2003, 2011; Wyly, 2011). So while we may now have access to millions or

billions of records about individuals' movements through cities via public transportation (Batty, 2013), the potential gaps in such data, as well as their policy implications, remain unclear.

Not only are there always highly uneven data shadows (Graham, 2014), and not only do those data shadows have effects in the world, but the powerful links between big data practices and the big data meme itself will only reinforce those very issues. Material and digital ghettoization and Balkanization are produced in a recursive relationship. Such outcomes are only amplified by the popularity of the big data meme among technologists, journalists, and venture capitalists. As the data shadows of some people, places, and processes become increasingly dense and well defined, the utility of those data increases, reinforcing the alternatively virtuous or vicious cycles of data production, consumption, and usage.

As big data are increasingly embedded into social and spatial decisions, processes, and institutions, the links between signifier and signified might become ever more obfuscated. That is, as we begin to place more and more trust in big data and the software, algorithms, and machines that are used to produce and analyze such data, we may tend to lose sight of the very things that such data represent, blurring the boundaries between the ontological and epistemological (cf. González-Bailón, 2013). This obfuscation can happen for two reasons. First, because the objects of signification are often inaccessible due to size, the proprietary-ness, or the black boxing of data sets. Second, because even when data are available they can be refracted through aggregation, modeling, and filtering mechanisms. This leads to a state where it becomes the norm to make decisions based on layers of complex and ephemeral data, mediated by black-boxed algorithms, presented on nonopen hardware and software platforms (Graham, 2013). Not only, therefore, should there be concern about absences in big data but attention should also be paid to the often opaque and proprietary mechanisms through which those data are filtered, ranked, presented, and enacted.

Ultimately, this piece, and the larger forum which it introduces, has pointed to the promises and perils of new data practices in academia and the

wider world, as well as the discursive power of big data as a meme. As ever more people point to the myriad ways big data will influence the doing, creating, making, and enacting of geography, we hope that geographers of all stripes are able to productively and critically employ big data to address long-standing questions of social justice, inequality, and our relationship to the environment, among other concerns. Unfortunately, our hopes are outweighed by fears about the persistent unevenness of representation, limited possibilities for participation, barriers to research and their implications for governance, privacy, and our ways of knowing the world.

And yet, the futures of geography and big data are still to be made. It is at this moment of change and potential that structured and sustained research, conversation, and critique is necessary, not just into big data presences but also absences that are both deliberately and unknowingly omitted. We believe that a broader conversation into the big data meme itself and the ways that it is able to redirect and displace attention, conversation, resources, and practices away from other pressing issues will not only allow us to avoid the most problematic implications of big data but also work toward a more productive integration of big data with existing research paradigms. In other words, we need sustained inquiry into the question of whose interests big data practices and the big data meme ultimately serves. Our hope is that this forum acts as a beginning that will not only shape the debate but also inform the practices of geography within big data and big data within geography.

Notes

1. This forum has its origins in a panel we organized at the 2013 meeting of the Association of American Geographers in Los Angeles ('More data, more problems? Geography and the future of "big data"').
2. The prompts offered were: What unique contributions, if any, can geography make to wider debates about big data?; Does big data present any unique challenges to geographic thought and methodology?; What are the potentials for critical scholarship using, and about, big data?; What are the social, political, and economic implications of big data?; and Whose interests does the move toward widespread analysis and use of big data serve?

References

- Anderson C (2008) The end of theory: the data deluge makes the scientific method obsolete. *Wired Magazine* 16(7). Available at http://www.wired.com/science/discoveries/magazine/16-07/pb_theory
- Barnes T (2013) Big data, little history. *Dialogues in Human Geography* 3(3): 297–302.
- Batty M (2013) Big data, smart cities, and city planning. *Dialogues in Human Geography* 3(3): 274–279.
- Boyd D and Crawford K (2012) Critical questions for big data: provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society* 15(5): 662–679.
- Crampton J and Krygier J (2005) An introduction to critical cartography. *ACME: An International e-Journal for Critical Geographies* 4(1): 11–33.
- Crampton JW, Graham M, Poorthuis A, Shelton T, Stephens M, Wilson MW, et al. (2013) Beyond the geotag: situating ‘big data’ and leveraging the potential of the geoweb. *Cartography and Geographic Information Science* 40(2): 130–139.
- Crutcher M and Zook M (2009) Placemarks and waterlines: racialized cyberscapes in post-Katrina google earth. *Geoforum* 40(4): 523–534.
- Curry MR (1997) The digital individual and the private realm. *Annals of the Association of American Geographers* 87(4): 681–689.
- Floridi L (2012) Big data and their epistemological challenge. *Philosophy & Technology* 25(4): 435–437.
- Goodchild M (2013) The quality of big (Geo)data. *Dialogues in Human Geography* 3(3): 280–284.
- González-Bailón S (2013) ‘Big data’ and the capillaries of human geography. *Dialogues in Human Geography* 3(3): 292–296.
- Gorman S (2013) The danger of a big data episteme and the need to evolve GIS. *Dialogues in Human Geography* 3(3): 285–291.
- Goss J (1995) ‘We know who you are and we know where you live’: the instrumental rationality of geodemographic systems. *Economic Geography* 71(2): 171–198.
- Graham M (2011) Time machines and virtual portals: the spatialities of the digital divide. *Progress in Development Studies* 11(3): 211–227.
- Graham M (2013) Virtual geographies and urban environments: big data and the ephemeral, augmented city. In: M Acuto and W Steele (eds) *Global City Challenges: Debating a Concept, Improving the Practice*. London, UK: Palgrave, pp. 117–139.
- Graham M (2014) The knowledge based economy and digital divisions of labour. In: V Desai and R Potter (eds) *Companion to Development Studies*, 3rd ed. London, UK: Hodder (in press).
- Graham M, Zook M, and Boulton A (2013) Augmented reality in the urban environment: contested content and the duplicity of code. *Transactions of the Institute of British Geographers* 38(3): 464–479.
- Haklay M (2013) Neogeography and the delusion of democratisation. *Environment and Planning A* 45(1): 55–69.
- Hartshorne R (1955) ‘Exceptionalism in geography’ re-examined. *Annals of the Association of American Geographers* 45(3): 205–244.
- Harvey D (1973 [2009]) *Social Justice and the City*. Athens, GA and London, UK: University of Georgia Press.
- Introna LD and Nissenbaum H (2000) Shaping the web: why the politics of search engines matters. *The Information Society* 16(3): 169–185.
- Kitchin R (2013) Big data and human geography: opportunities, challenges and risks. *Dialogues in Human Geography* 3(3): 262–267.
- Kitchin R and Dodge M (2011) *Code/Space: Software and Everyday Life*. Cambridge, MA: The MIT Press.
- Laney D (2001) 3D data management: controlling data volume, velocity, and variety. [Gartner Group], <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf> (last accessed 18 November 2012).
- Lave R, Wilson MW, Barron E, Biermann C, Carey M, Doyle M, et al. (2013) Critical Physical Geography. *The Canadian Geographer*.
- Light JS (2003) *From Warfare to Welfare: Defense Intellectuals and Urban Problems in Cold War America*. Baltimore, MD and London, UK: Johns Hopkins University Press.
- Light JS (2011) Discriminating appraisals: cartography, computation, and access to federal mortgage insurance in the 1930s. *Technology and Culture* 52(3): 485–522.
- Miller HJ (2010) The data avalanche is here. Shouldn’t we be digging? *Journal of Regional Science* 50(1): 181–201.

- Openshaw S (1991) A view on the GIS crisis in geography, or, using GIS to put Humpty-Dumpty back together again. *Environment and Planning A* 23(5): 621–628.
- Parker S, Uprichard E, and Burrows R (2007) Class places and place classes: geodemographics and the spatialization of class. *Information, Communication & Society* 10(6): 902–921.
- Pattison WD (1964) The four traditions of geography. *Journal of Geography* 63(5): 211–216.
- Pavlovskaya M (2006) Theorizing with GIS: a tool for critical geographies?" *Environment and Planning A* 38(11): 2003–2020.
- Pickles J (ed) (1995) *Ground Truth: The Social Implications of Geographic Information Systems*. New York, NY: Guilford Press.
- Porter TM (1996) *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton, NJ: Princeton University Press.
- Ruppert E (2013) Rethinking empirical social sciences. *Dialogues in Human Geography* 3(3): 268–273.
- Sack R (1974) Chorology and spatial analysis. *Annals of the Association of American Geographers* 64(3): 439–452.
- 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): 762–772.
- Schaefer FK (1953) Exceptionalism in geography: a methodological examination. *Annals of the Association of American Geographers* 43(3): 226–249.
- Schuurman N (2000) Trouble in the heartland: GIS and its critics in the 1990s. *Progress in Human Geography* 24(4): 569–590.
- Sheppard E (1995) GIS and society: towards a research agenda. *Cartography and Geographic Information Science* 22(1): 5–16.
- Sheppard E (2001) Quantitative geography: representations, practices, and possibilities. *Environment and Planning D: Society and Space* 19(5): 535–554.
- Sheppard E (2005) Knowledge production through critical GIS: genealogy and prospects. *Cartographica* 40(4): 5–21.
- Taylor L (2012) What's the big idea? Available at: <http://linnettaylor.wordpress.com/2012/10/23/whats-the-big-idea/> (23 October 2012).
- Winner L (1980) Do artifacts have politics? *Daedalus* 109(1): 121–136.
- Wyly E (2009) Strategic positivism. *The Professional Geographer* 61(3): 310–322.
- Wyly E (2011) Positively radical. *International Journal of Urban and Regional Research* 35(5): 889–912.
- Zook M and Graham M (2007) The creative reconstruction of the Internet: Google and the privatization of cyberspace and DigiPlace. *Geoforum* 38(6): 1322–1343.