GeoJournal
DOI 10.1007/s10708-014-9600-8

Guest Editorial: What's so big about Big Data? Finding the spaces and perils of Big Data

Ryan Burns · Jim Thatcher

For scholars, entrepreneurs, and technology writers, 2012 was a watershed year for "Big Data"—here referring to massive datasets produced through the aggregation of crowdsourced, social, and other digitally available data (Mayer-Schönberger and Cukier 2013; Rasmus 2012; Nicole 2012). The "year of Big Data" (Rasmus 2012) saw a drastic rise in the use of Big Data and its related methods and technologies in fields as diverse as marketing (see: Baker 2013), healthcare (see: Cerrato 2012), international development (see: Letouzé 2012), humanitarianism (see: UN OCHA 2013), and national funding agencies (see: National Science Foundation 2012) (Thatcher 2014). More recently, geographers studying technology have also turned their gaze to 'Big Data' and its variegated spatialities (Barnes and Wilson 2014; Batty 2013; Goodchild 2013; Kitchin 2013, 2014; Crampton et al. 2013). While much data may always have been big (Dalton and Thatcher 2014b; Kitchin 2013, this issue), and the manipulation and analysis of spatial data may have a long and complex history within geography (Goodchild 2006), the ability to rapidly aggregate and analyze previously unheard of combinations of data has led to an increased focus on the relations between data and knowledge production (Boellstorff 2013; Thatcher 2014). For many, the panacea to diverse social ills has become larger data sets and quantification, in what could be construed as "naive empiricism" (Taylor 1990, 212). At worst, this view can devolve—as it has in at least one well-known case—into the "end of theory", wherein numbers have come to "speak for themselves" (Anderson 2008). As government, private industries, and academic researchers all rush to embrace Big Data, some scholars have pushed back, calling into question the "purely data-driven approach" (Kling and Pozdnoukhov 2012, 483) and its epistemological, economic, and political commitments (Batty 2012; Boyd and Crawford 2012; Burgess and Bruns 2012; Richards and King 2013; Dalton and Thatcher 2014b).

This special issue responds and contributes to those debates by calling attention to and exploring the continued importance of "small data," within the context of Big Data's continual rise to prominence. Here we conceptualize "small data" as datasets and data production and analysis methodologies that are limited in size and scope relative to Big Data, but may contain rich, contextualized data that has been produced with a particular purpose in mind. Growing out of a session at the 2012 AAG entitled "Whither Small Data?

R. Burns
Department of Geography, University of Washington, Smith Hall 408, Box 353550, Seattle, WA 98195, USA
e-mail: burnsr77@gmail.com

J. Thatcher (✉)
Urban Studies, University of Washington - Tacoma, 1900 Commerce Street, Box 358437, Tacoma, WA 98402, USA
e-mail: jethatch@uw.edu

Published online: 18 October 2014

The limits of 'big data' and the value of 'small data' studies", the issue uses the "end of theory" as a starting provocation. By analyzing the epistemic limits of Big Data and accentuating the emerging social, political, and analytic challenges posed by Big Data research and analysis, the issue seeks to (re)make and maintain a space for "small data" in current studies of society. Underlying these concerns is an inquiry into the *place of geography* in Big Data studies. With upwards of sixty-percent of currently held datasets containing spatial information (Hahmann and Burghardt 2013), geographers may hold a "home field advantage" in the study of Big Data (Farmer and Pozdnoukhov 2012). Through an emphasis on space and spatial relations, we have the opportunity to emerge as a central pillar of the Big Data research agenda. By beginning with the "end of theory," and bringing diverse perspectives together, this special issue and others like it (Graham and Shelton 2013; Crawford et al. 2014) are making room for the "hard work of theory" (Pickles 1997, 370) in Big Data research within geography and cognate disciplines.

Each paper in this special issue addresses a subset of these questions. Rather than seeing Big Data as the deterministic culmination of unerring technological progress, these papers situate Big Data and its accompanying methodologies in their contingent social and historical contexts. Miller and Goodchild examine the ways the rise of Big Data is transforming geographic research. They argue that while geographers have long had access to large, heterogeneous data, Big Data represents an evolution wherein the data and computation drive the research questions asked. This raises significant concerns for the ways geographical science is undertaken, with particularly strong implications for modelling and pattern-detection. Despite the breadth of Big Data, researchers must still be attentive to problems identified in "small data," such as selection biases and data quality. Kitchin and Lauriault are similarly concerned with how Big Data may impact geographical research, instead focusing on data infrastructures. They highlight important ways "small data" will remain crucial for both data infrastructures and for empirical research, arguing that all data is increasingly being agglomerated and treated as if it were Big Data. This new environment is characterized by a sort of epistemic hybridity wherein "small data" are subject to demands typically leveled at Big Data.

Three papers in the issue focus on Big Data's roles in humanitarianism, crisis, and international development. Burns argues that within discourses of humanitarianism, Big Data has created a new epistemology centered around crises and their victims. From this epistemology has sprung a new set of practices and social relations that run counter to the ways in which "digital humanitarian" organizations promote and market themselves. Similarly, Crawford and Finn link the conceptualization of crises with how emergency managers engage Big Data. Their article demonstrates the ways in which crisis data are performative and selective, posing important concerns for privacy. The Big Data generated around a crisis is necessarily partial in its perspective, presenting challenges for the basic ontologies, epistemologies, and ethics of crisis response. Similar limitations are demonstrated by Taylor and Shroeder whose article examines how development agencies and policymakers have made use of mobile phone records to alter and adapt their approaches to development in the global South. They argue that while mobile phone records provide a wealth of new information, this does not remedy a lack of *important* and *relevant* data, since Big Data may not address contextually-specific needs. In short, development "problems" should still guide the process of identifying data needs in development projects.

Crampton examines the ethics of Big Data through the lens of state surveillance. In particular, he looks at the 2013 release of US National Security Administration documents by Edward Snowden in order to show the ways Big Data participates in the production of biopolitical subjects. Borrowing from Richards and King (2013), Crampton seeks to explore the paradoxes of transparency, identity, and power in Big Data surveillance practices, with attention to the social and political processes enacted in this new environment. Perhaps most imperatively, he identifies three specific areas in which researchers and citizens can intervene into these processes to enact a more positive world.

Despite their diverse intellectual backgrounds and objects of analysis, the authors of this special issue have raised several critical points for the study of Big Data moving forward. First, Big Data is never simply a replacement for "small data," but instead constitutes a distinct perspective for observing and understanding the world. "Small data" retains an empirical value distinct from Big Data sets, able to answer unique questions impossible to ask of many aggregated, quantified data sets. Neither can be reduced to the other. Big Data can never mean the "end of theory" as

its very existence necessitates the reflexive study of the kinds of knowledge it can (and cannot) produce.

Second, Big Data has changed and will continue to change what privacy means at both an individual and societal level. Conceptualizations of privacy must be reconsidered in light of both state and NGO actors' abilities to access, analyze, and act upon Big Data. Most importantly, the authors here demonstrate the need for researchers, on the one hand, to redefine privacy in light of Big Data and, on the other, to simultaneously uncover and enact practices that improve the experience of privacy in day-to-day life.

Finally, the current conceptualizations of Big Data are both far-reaching and diffuse; they have yet to coalesce into strategic, widely-accepted theorizations. In organizing this issue, it became clear that even amongst a small set of authors working from a single set of prompts, Big Data, its influence upon society, and its meaning in day-to-day life will differ radically depending on the research focii contested as important, distinctive, or superfluous. What one author clearly demonstrates as a fundamental concern to epistemology stemming directly from Big Data analysis, another accepts as a prerequisiste for consideration of another fundamental focus. On the one hand, the empirical and theoretical strength of these diffuse arguments attests to the intrinsic importance Big Data has come to play in what we know and how we know it; however, on the other, it suggests weaknesses in the extant conceptual value of Big Data. Individual researchers tack between fields of interest, but there remains a gap between those precise, empirical formulations and a broad-based theoretical critique of epistemology and ontology in Big Data. We therefore suggest the further development of a Critical Data Studies research program (Dalton and Thatcher 2014a, b) in which multiple research questions can be asked of Big Data, focusing on its influences upon society 'big' and 'small.'

This special issue provides one vision of how such a research program could coalesce disparate perspectives, topical interests, and political commitments into a focused Critical Data Studies research program. First, this issue challenges notions that 'bigger' is necessarily more desirable when it comes to knowledge production. Instead, researchers must navigate the complex epistemological connections between data and social phenomena. Second, this issue foregrounds important questions around the social, political, economic, and ethical implications of new data infrastructures and analytical technologies. This acknowledges that data are never 'raw', constituting a distinct plethora of practices, epistemologies, and political interests. Data is always a social product, and it is a Critical Data Studies responsibility to interrogate the particular forms this takes across different contexts. Third, this issue suggests researchers should be attentive to the spatialities of data, both as context for data production and as the material through which data and technologies are produced. In other words, *geography* should be central to our inquiries into the operations and purposes of data.

References

Anderson, C. (2008). The end of theory: The data deluge makes the scientific method obsolete. *Wired*. http://www.wired.com/science/discoveries/magazine/16-07/pb_theory. Last accessed Dec 17, 2012.

Baker, S., (2013, January 5). Can social media sell soap? *The New York times*. http://www.nytimes.com/2013/01/06/opinion/sunday/can-social-media-sell-soap.html. Last accessed Oct 13, 2014.

Barnes, T., & Wilson, M. (2014). Big data, social physics and spatial analysis: The early years. *Big Data & Society*, *1*(1), 1–14.

Batty, M. (2012). Editorial: Smart cities, big data. *Environment and Planning B: Planning and Design*, *39*(2), 191–193.

Batty, M. (2013). Big data, smart cities and city planning. *Dialogues in Human Geography, 3*(3), 274–279.

Boellstorff, T. (2013). Making big data, in theory. *First Monday*, 18(10). http://firstmonday.org/ojs/index.php/fm/article/view/4869/3750. Last accessed Oct 13, 2014.

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, *15*(5),662–679.

Burgess, J., & Bruns, A. (2012). Twitter archives and the challenges of "big social data" for media and communication research. *M/C Journal, 15*(5). http://journal.media-culture.org.au/index.php/mcjournal/article/viewArticle/561. Last accessed Aug 20, 2014.

Cerrato, P., (2012, November 1). Big data analytics: Where's the ROI? *Information week: Healthcare*. http://www.informationweek.com/healthcare/clinical-information-systems/big-data-analytics-wheres-the-roi/d/d-id/1107170? Last accessed Oct 13, 2014.

Crampton, J., Graham, M., Poorthuis, A., Shelton, T., Stephens, M., Wilson, M., & Zook, M. (2013). Beyond the geotag: Situating "Big Data" and leveraging the potential of the geoweb. *Cartography and Geographic Information Science, 40*(2), 130–139.

Crawford, K., Miltner, K., & Gray, M. L. (2014). Critiquing big data: Politics, ethics, and epistemology. *International Journal of Communication, 8*, 1663–1672.

Dalton, C., & Thatcher, J. (2014a). Inflated granularity: The promise of big data and the need for a critical data studies. http://meridian.aag.org/callforpapers/program/AbstractDetail.cfm?AbstractID=56048. Last accessed Oct 13, 2014.

Dalton, C., & Thatcher, J. (2014b). What does a critical data studies look like, and why do we care? Seven points for a critical approach to "Big Data." *Society and Space Open Site*. http://societyandspace.com/material/commentaries/craig-dalton-and-jim-thatcher-what-does-a-critical-data-studies-look-like-and-why-do-we-care-seven-points-for-a-critical-approach-to-big-data/. Last accessed Aug 5, 2014.

Farmer, C. J., & Pozdnoukhov, A. (2012). Building streaming GIScience from context, theory, and intelligence. In *Proceedings of the Workshop on GIScience in the Big Data Age*. Columbus, OH.

Goodchild, M. F. (2006). GIScience ten years after ground truth. *Transactions in GIS, 10*(5), 687–692.

Goodchild, M. F. (2013). The quality of big (geo)data. *Dialogues in Human Geography*, *3*(3), 280–284.

Graham, M., & Shelton, T. (2013). Geography and the future of big data, big data and the future of geography. *Dialogues in Human Geography, 3*(3), 255–261.

Hahmann, S., & Burghardt, D. (2013). How much information is geospatially referenced? Networks and cognition. *International Journal of Geographical Information Science*, *27*(6), 1171–1189.

Kitchin, R. (2013). Big data and human geography: Opportunities, challenges and risks. *Dialogues in Human Geography, 79*(1), 1–14.

Kitchin, R. (2014). Big data, new epistemologies and paradigm shifts. *Big Data & Society, 1*, 1–12.

Kling, F., & Pozdnoukhov, A. (2012). When a city tells a story: Urban topic analysis. In *Proceedings of ACM SIGSPATIAL 2012* (pp. 482–485). New York: ACM Press.

Letouzé, E. (2012). Big data for development: Challenges & opportunities. *UN Global Pulse*. http://www.unglobalpulse.org/projects/BigDataforDevelopment. Last accessed Oct 6, 2014.

Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. New York: Houghton Mifflin Harcourt Publishing Company.

National Science Foundation. (2012). NSF announces interagency progress on administrations big data initiative. National Science Foundation. http://www.nsf.gov/news/news_summ.jsp?cntn_id=125610. Last accessed Oct 13, 2014.

Nicole, K., (2012, December 5). Big data needs eye candy to go mainstream. Here's the iPad App to Do It. *Forbes*. http://www.forbes.com/sites/siliconangle/2012/12/05/big-data-needs-eyecandy-to-go-mainstream-heres-the-ipad-app-to-do-it/. Last accessed Oct 13, 2014.

Pickles, J. (1997). Tool or science? GIS, technoscience, and the theoretical turn. *Annals of the Association of American Geographers*, *87*(2), 363–372.

Rasmus, D. (2012). Why big data won't make you smart, rich, or pretty. *Fast Company*. http://www.fastcompany.com/1811441/why-big-data-won%E2%80%99t-make-you-smart-rich-or-pretty. Last accessed Aug 19, 2014.

Richards, N., & King, J. (2013). Three paradoxes of big data. *Stanford Law Review Online, 66*(41). http://www.stanfordlawreview.org/online/privacy-and-big-data/three-paradoxes-big-data. Last accessed July 6, 2014.

Taylor, P. (1990). Editorial comment: GKS. *Political Science Quarterly*, *9*(3), 211–212.

Thatcher, J. (2014). Living on fumes: Digital footprints, data fumes, and the limitations of spatial big data. *International Journal of Communication, 8*, 1765–1783.

United Nations Office for the Coordination of Humanitarian Affairs. (2013). *Humanitarianism in the network age*. UN Office for the Coordination of Humanitarian Affairs. https://ochanet.unocha.org/p/Documents/WEBHumanitarianism%20in%20the%20Network%20Age%20vF%20single.pdf. Last accessed April 15, 2013.