8

DATA AND DATA INFRASTRUCTURES

Rob Kitchin and Tracey Lauriault

INTRODUCTION

This chapter is centrally concerned with the production of and access to digital data that are reshaping how geographical research is undertaken. The chapter starts by outlining the changing nature of data, and the relatively recent wide-scale digitization of non-digital data sources and the production of big data. It then discusses cyber-infrastructures and data infrastructures – archives and repositories – that facilitate the sharing of data for research. It concludes with a brief discussion of the emerging field of critical data studies and its relevance for geographical epistemology.

DIGITAL DATA AND GEOGRAPHICAL RESEARCH

Data are often understood as the building blocks for information and knowledge. They are the raw material produced by capturing and abstracting the world into measures and other representational forms (numbers, characters, symbols, images, sounds, electromagnetic waves, bits, etc.). Data are typically representative, wherein measures explicitly seek to denote the characteristics of a phenomena (e.g., a person's age, height, weight, opinion, habits, location, etc.). They can also be derived (e.g., produced from other data) or implied (e.g., through an absence rather than presence). In broad terms, data vary by:

- *form*, being quantitative (e.g., numeric, categories) or qualitative (e.g., words, images);
- *structure*, being structured (a defined data model with consistent measures/categories), semi-structured (no predefined data model/schema), or unstructured (no defined data model and highly variable data);
- *source*, being captured (deliberately sourced) or exhaust (inherently produced as a by-product of another process);
- *producer*, being primary (generated by a person, organization or community for a specific purpose), secondary (made available by a third party), or tertiary (third party derived or aggregated data);
- *type*, being indexical (uniquely identified, enabling linkage), attribute (representative but not indexical), or metadata (data about data that facilitate their use, such as data definitions, provenance and lineage) (Kitchin, 2014).

Data have long been generated by societies for the purposes of administration, business and science, though the term 'data' was only used for the first time in the English language in the seventeenth century (Rosenberg, 2013). From this time on, the generation, storage, and analysis of data grew enormously due to the utility and value of the insights and processes they facilitated, being used to create new knowledge, policies, innovations, and products, and to manage and regulate populations and run businesses (Poovey, 1998).

While large quantities of analog data were produced in the nineteenth and twentieth century, the amount generated was very modest compared to the present digital deluge. Traditionally, data were time-consuming and costly to produce, analyse and interpret because they involved a lot of labour – all the data had to be collected and analysed by hand (Grier, 2007). Even with the first computers, data generally had to be digitized by hand and the computational analysis was relatively slow by today's standards (taking hours or days to do what would now take milliseconds – and coding errors would mean starting again). As a consequence, data generation was narrowly focused to answer specific questions and produced using well-tested techniques in tightly controlled ways. And outside of public administration they tended to be generated periodically (e.g., annually) or on a one-off basis using sampling techniques to produce data that were hopefully representative of whole populations and had limited levels of error, bias, and uncertainty (Miller, 2010). Beyond some exceptions, such as population censuses or meteorological data collections, individual social science and physical geography datasets tended to be quite small in size, especially those that were not produced by institutions, such as those generated using surveys or interviews or fieldwork. Datasets generated in traditional ways were thus typically characterized by limited volume, sampled collection, small geographic extent, and narrow variety and framing. In the case of meteorological data, even though their global collection constituted a 'vast machine' (Edwards, 2010), they were not collected in an integrated, standardized manner or in real-time.

In contrast, new forms of big data have quite different qualities. While there has been much discussion in recent years as to what constitutes big data, Kitchin (2014) details that they have the following characteristics:

- huge in *volume*, consisting of millions or billions of records, or terabytes or petabytes in storage;
- high in *velocity*, being created continuously and in or near real-time;
- diverse in *variety*, being structured and unstructured in nature;
- *exhaustive* in scope, striving to capture entire populations or systems (n = all);
- fine-grained in *resolution* and uniquely *indexical* in identification;
- *relational* in nature, containing common fields that enable the conjoining of different datasets; and
- *flexible*, holding the traits of *extensionality* (new fields can easily be added to the method of generation) and *scalability* (datasets can expand in size rapidly).

In an examination of 26 datasets widely thought to constitute big data, Kitchin and McArdle (2016) contend that the two most important distinguishing characteristics are velocity and exhaustivity. Big data are produced continuously and are n = all samples within a domain or a platform. For example, automatic number plate recognition cameras continuously scan and capture the license plate of *every* vehicle that passes the camera, rather than being a sample at a particular time (as with a traditional traffic survey). And with a network of cameras one can both track the routes of individual vehicles and calculate city-wide patterns of traffic on a 24 hour, 365 days a year basis. Instead of producing static, often coarse, snapshots of traffic conditions, there is a quantum shift in the resolution and coverage of the data available for traffic analysis, including real-time monitoring.

High velocity and exhaustive data have been part of the physical sciences for quite some time with respect to oceanography, meteorology, and earth sciences, where sensors have been deployed to measure waves, temperature, or earthquakes. Such data are now becoming more widespread, with big data being generated across a whole range of domains of interest to geographers, being used to monitor and sense the environment, to mediate and deliver services and products. Much of the big data produced are inherently spatial data, given that the technologies that generate them automatically add georeferenced attributes (e.g., coordinates calculated using GPS or other forms of addressing). Sources of georeferenced big data include mobile phones, smartphone apps, smart meters and utilities, logistics systems, environmental monitoring, sensor networks, navigation systems, autonomous cars and heavy machinery, social media sites, travel and accommodation websites, online and offline financial transactions, surveillance and security systems, and emergency services. In many cases, the data are exhaust in nature – a by-product of systems – which can then be used to examine a variety of issues. For example, georeferenced social media data, such as Twitter, were not produced for the purposes of modelling mobility or sentiment patterns, but they are now often used in this way.

In addition to big data, there is improved access to traditional datasets as they become increasingly digitised and shared. For example, millions of documents, books, statistical surveys, public administration records, photographs, analog sound recordings and films are being transferred to digital media. In addition, local and traditional knowledge that was normally transmitted aurally by indigenous peoples, is also now being digitized and mapped (Taylor and Lauriault, 2014). These media can be combined to scale the scope of analysis, are searchable in new ways, and are open to the application of big data analytics such as data mining, pattern recognition, data visualization, statistics, and modelling. As such, data analysis that was once difficult or time consuming to perform by hand and/or required using analog technologies becomes possible in just a few microseconds, enabling more complex analysis to be undertaken. Consider, for example, the computational power now found in most smartphones. Moreover, these data are more and more being made available through new data infrastructures and cloud storage platforms and thus are open to be used in geographical research.

DATA INFRASTRUCTURES, OPEN DATA, AND APIS

Given the effort and cost of producing good quality data, traditional datasets were considered a valuable commodity, often either jealously guarded or expensively traded. Most social science data generated by academics were not archived or shared. Public administration and state data, and data generated by statistical agencies could largely only be accessed through visiting archives and often only with special permission. Access to data to undertake geographical research was therefore problematic. The internet, new archiving technologies, a massive expansion in digital data storage, data standards, and open science and open data movements have transformed data access.

The collection and storage of data have been, and continue to be, both informal and formal in nature. The informal approach consists simply of gathering data and storing them, and might best be described as data holdings or backups. Many academics have informal collections of data from their research stored on their computers and backed up on external media such as a datastick or personal cloud. These data generally lack curation and metadata, and are usually difficult to make sense of by anyone other than the person who produced them. The formal approach is much more professional in orientation, consisting of a set of curatorial practices and institutional structures that creates and manages an archive (Lauriault et al., 2007). Digital data archives or trusted digital repositories explicitly seek to be long-term endeavours, preserving the full record set – data, metadata, and associated documentation – for future reuse. Increasingly, academics are being encouraged, or compelled by funding agencies, to adopt data management plans and are lodging the digital data from their research in archives.

A data infrastructure is a digital means for storing, sharing, linking together, and consuming data holdings and archives across the internet. Over the past two decades considerable effort has been expended on creating a variety of related science and research data infrastructures: catalogues, directories, portals, clearinghouses, and repositories (Lauriault et al., 2007). These terms are often used interchangeably, though they are slightly different types of entities. Catalogues, directories, and portals are centralized resources that may detail and link to individual data archives (e.g., the Earth Observation Data Management Service of the Canada Centre for Remote Sensing), or data collections held by individual institutions (e.g., the Australian National Data Service), or are federated infrastructures which provide the means to access the collections held by many institutions (e.g., the US National Sea Ice Data Center). They might provide fairly detailed inventories of the datasets held, and may act as metadata aggregators, but do not necessarily host the data (e.g., the GeoConnections Discovery Portal). Single-site repositories host all the datasets in a single site, accessible through a web interface, though they may maintain back-up or mirror sites in multiple locations (e.g., the UK Data Archive). A federated data repository or clearinghouse can be a shared place for storing and accessing data (e.g., NASA's Global Change Master Directory). It might provide some data services in terms of search and retrieval, and data management and processing, but each holding or archive has been produced independently and may not share data formats, standards, metadata, and policies. Nevertheless, the repository seeks to ensure that each archive meets a set of requirement specifications and uses audit and certification to ensure data integrity and trust among users (Dasish, 2012).

A cyber-infrastructure is more than a collection of digital archives and repositories. It consists of a suite of dedicated networked technologies, shared services (relating to data management and processing), analysis tools such as data visualizations (e.g., graphing and mapping apps), metadata and sharing standards (e.g., ISO19115, web map services (WMS), web feature services (WFS), semantic interoperability and structured vocabularies, and shared policies (concerning access, use, intellectual property rights, etc.) which enable data to be distributed, linked together, and analysed (Cyberinfrastructure Council, 2007). Such cyber-infrastructures include those implemented by national statistical agencies and national or international spatial data infrastructures (SDIs) that require all data stored and shared to comply with defined parameters in order to maximize data interoperability and ensure data quality, fidelity, and integrity that promote trust. The objectives of SDIs are to ensure that users from multiple sectors and jurisdictions can seamlessly reuse these data and link them into their systems. The Arctic Spatial Data Infrastructure, for example, aligns with global, regional, and national geospatial data contexts such as the Infrastructure for Spatial Information in Europe, the United Nations Committee of Experts on Global Geospatial Information Management, Global Earth Observation System of Systems and the Canadian Geospatial Data Infrasctucture , is open data, is International Standards Organization and Open Geospatial Consortium standards based, and is an official spatial data collaboration between Canada, Denmark, Finland, Iceland, Norway, Russia, Sweden and the USA.

The benefits of creating data infrastructures include the following: ready access to more data; the data have improved quality and integrity through the adoption of standards, protocols, and policies; combining datasets can produce new insights; there are scales of economy through sharing resources and avoiding replication; and there is improved return on investment for research funders by enabling data reuse. Indeed, data infrastructures have become a significant source of secondary and tertiary data for geography scholars, significantly increasing ease and speed of access to useful data.

Nonetheless, there are still access issues with respect to both traditional and big data. Traditional data produced by academia, public institutions, non-governmental organizations, and private entities can be restricted in use to defined personnel or be available only for a fee or under license. In recent years there has been an attempt to change this situation by making data produced using funding from the public purse open in nature. Pollock (2006) contends that 'data is open if anyone is free to use, reuse, and redistribute it – subject only, at most, to the requirement to attribute and/or share-alike'. There are, however, different understandings as to what openness means, with some open data archives having some restrictions with respect to use and reuse, reworking, and redistribution. Many jurisdictions now have open data repositories that make some of their administrative and operational data freely available for analysis and reuse, thus fostering transparency and accountability in public services and enabling the creating of an open data economy. However, there is a wide variation in how extensive these repositories are, and many are little more than data dumps, more like data holdings than archives in their organization and operation (Lauriault and Francoli, 2017). Similarly, there has been a strong drive to promote open science in which the data generated from publicly funded research is made available for reuse, sharing the data via institutional repositories or dedicated disciplinary data infrastructures (Borgman, 2015).

The situation with respect to private companies is somewhat different. Data produced by companies are valuable assets that produce products for profit and provide competitive advantage. Consequently, companies are often reluctant to openly share their data with others. Since big data are predominately produced by the private sector, access to them is usually restricted, available for a fee, and/or subject to proprietary licensing. In some cases, a limited amount of the data might be made available to researchers or the public through application programming interfaces (APIs). For example, Twitter allows a few companies to access its firehose (stream of data) for a fee for commercial purposes, but researchers are restricted to a 'gardenhose' (about 10 percent of public tweets), a 'spritzer' (about 1 percent of public tweets), or different subsets of content ('white-listed' accounts) (boyd and Crawford, 2012). One consequence of limited access to big data in geography and academia more broadly is that some datasets, such as Twitter, Flickr, and Instagram, receive disproportionate attention, sometimes being repurposed to examine issues for which they are ill-suited (Kitchin et al., 2017).

It should also be noted, that while data infrastructures can improve the quality and usability of a dataset, data quality – how clean (error- and gap-free), objective (bias-free), and consistent (few discrepancies) the data are – and veracity – the extent to which they accurately (precision) and faithfully (fidelity, reliability) represent what they are meant to – remain significant issues. And while some have argued that big data do not need to meet the same standards of data quality and veracity as traditional data because their exhaustive nature removes sampling biases and compensates for any errors, gaps, or inconsistencies in the data or weakness in fidelity (Mayer-Schönberger and Cukier, 2013) the maxim 'garbage in, garbage out' still holds. Indeed, big data can be full of dirty (through instrument bias), fake (through false accounts), or biased (due to the non-representative nature of the demographic being measured) data, or be data with poor fidelity and constitute weak proxies (the data are often exhaust and are being repurposed) (Crampton et al., 2013). A number of experiments conducted by national statistical agencies to assess the suitability a number of big data sources for official reporting purposes (e.g., debit cards and credit cards for household spending) have demonstrated that while some big data sources can augment knowledge, these data cannot replace existing approaches because of data quality issues, standards, completeness, privacy issues, costs, proprietary ownership, and sustainability (Vale, 2015). When using digital data for geographic analysis, therefore, it is important to consider the integrity and fitness for purpose of the data (Miller and Goodchild, 2015). The data deluge might lead to enormous quantities of data with which to make sense of the world, but that does not mean that they will automatically lead to greater insights Sometimes it is more profitable to work a narrow seam of high quality data than to open-pit mine a dirty deposit.

CRITICAL DATA STUDIES AND GEOGRAPHIC SCHOLARSHIP

Corresponding to what has been termed the 'data revolution' or 'data deluge' has been a more sustained analytical and critical focus on data, databases, data infrastructures, and how they are produced and employed. While there has long been some critical attention applied to data, in the main this has been concerned with technical issues such as quality, interoperability, and representativeness, rather than the politics and praxes of data generation, processing, analysis, and application. Instead, the theoretical and analytical focus has been targeted at the production and use of information and knowledge. Data were largely cast as pre-analytic and pre-factual in nature, that which exists prior to interpretation and argument – that is, they are benign, neutral, objective, and non-ideological in essence, capturing the world as it is subject to technical constraints (Gitelman and Jackson, 2013). In other words, it is only the uses of data that are political, not the data themselves.

Critical data studies takes the politics and praxes of data as its central concern. Drawing on critical social theory, it posits that data are never simply neutral, objective, independent, raw representations of the world, but rather are situated, contingent, relational, contextual, and do active work in the world. Data are constitutive of the ideas, techniques, technologies, people, systems, and contexts that conceive, produce, process, manage, and analyse them (Bowker, 2005). In other words, data do not pre-exist their generation; they do not arise from nowhere and their generation is not inevitable. Their generation, processing, and analysis are shaped by protocols, standards, measurement processes, design decisions, disciplinary norms, and institutional politics, and their use is framed contextually to try and achieve certain aims and goals. As Gitelman and Jackson (2013: 2) put it, 'raw data is an oxymoron' as 'data are always already cooked'.

Similarly, databases and repositories are not simply neutral, technical means of assembling and sharing data. Rather they are complex socio-technical systems that are embedded within a larger institutional and political landscape, shaped by institutional and organizational cultures and practices, systems of thought, financial regimes, and legal and regulatory requirements (Ruppert, 2012; Kitchin, 2014). As we can attest from personal experience, creating and running a repository involves a lot of institutional, political, and personnel work – negotiating, debating, cajoling, lobbying, and wheeling-and-dealing – with a number of stakeholders. Moreover, databases and repositories are expressions of knowledge/power, shaping what questions can be asked, how they are asked, how they are answered, how the answers are deployed, and who can ask them (Ruppert, 2012).

Critical data studies seeks to unpack how data are always already cooked, and how they are constitutive of power/knowledge. As the field has emerged it has sought to examine a range of related processes and issues. For example, there are now a number of studies that have sought to: document the data practices utilized to handle and store data; chart the politics of open data and of sharing and accessing of data repositories; detail the genealogies, temporalities, and spatialities of datasets and archives; uncover the workings and economies of data markets and the commercial trading of data; plot the governmentalities that shape the generation and use of data; and examine and debate the ethics of producing, sharing, and extracting value and utility from data (see the journal *Big Data & Society*). In addition to critical data scholars, some First Nations groups are advancing the concept of 'data sovereignty' (Phillips, 2017), whereby First Nations and Inuit are demanding that data about them be returned, are questioning data indicators models that claim to measure their health and well-being (which, in fact, measure ills rather than resiliency, indigenous knowledge assets, and community health), and are revisiting the power/knowledge of researcher and research subject when it comes to data collected about them and their local and traditional knowledge. In Canada, the First Nations Information Governance Council's principles of ownership, control, access and possession and the collective licensing of local and traditional knowledge (Canadian Internet Policy and Public Interest Clinic, 2016) are two examples of this fledgling movement.

Geographers have been at the forefront of developing the field of critical data studies. For example, in an influential position paper, Dalton and Thatcher (2014) set out seven provocations needed to provide a comprehensive critique of the new regimes of data:

1. Situate data regimes in time and space.
2. Expose data as inherently political and whose interests they serve.
3. Unpack the complex, non-deterministic relationship between data and society.
4. Illustrate the ways in which data are never raw.
5. Expose the fallacies that data can speak for themselves and that big data will replace small data.
6. Explore how new data regimes can be used in socially progressive ways.
7. Examine how academia engages with new data regimes and the opportunities of such engagement.

In terms of operationalizing these provocations, Kitchin (2014) has suggested unpacking what he terms 'data assemblages', while Kitchin and Lauriault (2014) discuss Michel Foucault's idea of the 'dispositif' and Ian Hacking's 'dynamic nominalism' to make sense of the politics and praxes of data production and use. Crampton et al. (2013) outline the shortcomings of spatial big data and associated data analytics and forward an alternative epistemology. Leszczynski and Crampton (2016) outline what they term 'data anxieties'; that is, a paradoxical concern that spatial big data are simultaneously insufficient for tasks at hand, while also being excessive and over-sufficient. Shelton (2017) details the politics and ethics of big data in underpinning urban science and new urban imaginaries. GIScience scholars, such as Harvey Miller and Michael Goodchild (2015), have been critically and empirically examining data-driven geography. And there are numerous other contributions.

These accounts draw to a large degree on critical GIS scholarship that has sought to expose the politics of GIS and their deployment, and to formulate and practise a more situated and reflexive form of GIScience (which itself is rooted in feminist critiques of science) (Schuurman, 2000). The latter behoves GIS practitioners to explicitly recognize and account for their positionality (with respect to their knowledge, experience, beliefs, and aspirations), how their research and practices are framed within disciplinary debates and institutional politics and ambitions, how their data are cooked and hold certain characteristics (relating to cleanliness, completeness, consistency, veracity, and fidelity), and how their analytical methods have opportunities and pitfalls that affect methods, models, findings, and interpretations. We believe all geographical scholarship should be mindful of such epistemological considerations. In other words, in practising digital geography – generating or sourcing digital data from data infrastructures and using digital methods – one should critically reflect on and account for their politics and praxes. This is not to say all digital geography should be undertaking critical data studies, but rather that it takes heed of its ontological, epistemological, and ethical observations.

Acknowledgements

The research for this chapter was enabled by a European Research Council Advanced Investigator Award, 'The Programmable City' (ERC-2012-AdG-323636). The chapter draws on two previously published papers: Kitchin, R. and Lauriault, T. 2015 and 2018.

REFERENCES

Borgman, C. (2015) *Big Data, Little Data, No Data: Scholarship in the Networked World*. Cambridge, MA: MIT Press.

Bowker, G. (2005) *Memory Practices in the Sciences*. Cambridge, MA: MIT Press.

boyd, d. and Crawford, K. (2012) 'Critical questions for big data', *Information, Communication and Society*, 15(5): 662–679.

Canadian Internet Policy and Public Interest Clinic (2016) 'A proposal: An open licensing scheme for traditional knowledge'. Available online: https://cippic.ca/en/TK_Open_Licensing_Proposal (accessed 1 February 2018).

Crampton, J., Graham, M., Poorthuis, A., Shelton, T., Stephens, M., Wilson, M.W. and 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.

Cyberinfrastructure Council (2007) Cyberinfrastructure vision for 21st century discovery. Available online: http://www.nsf.gov/pubs/2007/nsf0728/index.jsp?org=EEC%20National%20Science%20Foundation (accessed 1 February 2018).

Dalton, C. and Thatcher, J. (2014) 'What does a critical data studies look like, and why do we care?'. *Society and Space*, 12 May. Available online: http://societyandspace.org/2014/05/12/what-does-a-critical-data-studies-look-like-and-why-do-we-care-craig-dalton-and-jim-thatcher (accessed 1 February 2018).

Dasish (2012) 'Roadmap for preservation and curation in the social sciences and humanities'. Available online: http://dasish.eu/publications/projectreports/D4.1_-_Roadmap_for_Preservation_and_Curation_in_the_SSH.pdf/ (accessed 1 February 2018).

Edwards, P. (2010) *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming*. Cambridge, MA: MIT Press.

Gitelman, L. and Jackson, V. (2013) 'Introduction', in L. Gitelman (ed.), *'Raw Data' is an Oxymoron*. Cambridge, MA: MIT Press. pp. 1–14.

Grier, D.A. (2007) *When Computers Were Human*. Princeton, NJ: Princeton University Press.

Kitchin, R. (2014) *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. London: Sage.

Kitchin, R. and Lauriault, T. (2015) Small data in the era of big data. *GeoJournal* 80(4): 463–475.

Kitchin, R. and Lauriault, T.P. (2018) 'Towards critical data studies: Charting and unpacking data assemblages and their work', in J. Eckert, A. Shears and J. Thatcher (eds), *Geoweb and Big Data*. Lincoln: University of Nebraska Press. pp. 3–20.

Kitchin, R. and McArdle, G. (2016) 'What makes big data, big data? Exploring the ontological characteristics of 26 datasets', *Big Data & Society*, 3(1). DOI: 10.1177/2053951716631130.

Kitchin, R., Lauriault, T.P. and Wilson, M. (eds) (2017) *Understanding Spatial Media*. London: Sage.

Lauriault, T.P., Craig, B.L., Taylor, D.R.F. and Pulsifier, P.L. (2007) 'Today's data are part of tomorrow's research: Archival issues in the sciences', *Archivaria*, 64: 123–179.

Lauriault, T.P. and Francoli, M. (2017) 'Openness, transparency, participation', in R. Kitchin, T.P. Lauriault and M.W. Wilson (eds), *Understanding Spatial Media*. London: Sage. pp. 188–203.

Leszczynski, A. and Crampton, J. (2016) 'Introduction: Spatial Big Data and everyday life', *Big Data & Society*, 3(2). DOI: 10.1177/2053951716661366.

Mayer-Schönberger, V. and Cukier, K. (2013) *Big Data: A Revolution that will Change How We Live, Work and Think*. London: John Murray.

Miller, H.J. (2010) 'The data avalanche is here. Shouldn't we be digging?', *Journal of Regional Science*, 50(1): 181–201.

Miller, H.J. and Goodchild, M.F. (2015) 'Data-driven geography', *GeoJournal*, 80: 449–461.

Phillips, G.W. (2017) 'Indigenous data sovereignty and reconciliation', opening key-note address, Data Power Conference, Ottawa, Canada.

Pollock, R. (2006) 'The value of the public domain'. Institute of Public Policy Research, London. Available online: https://www.ippr.org/files/images/media/files/publication/2011/05/value_of_public_domain_1526.pdf (accessed 1 February 2018).

Poovey, M. (1998) *A History of the Modern Fact: Problems of Knowledge in the Sciences of Wealth and Society*. Chicago: University Chicago Press.

Rosenberg, D. (2013) 'Data before the fact', in L. Gitelman (ed.), *'Raw Data' Is an Oxymoron*. Cambridge, MA: MIT Press. pp 15–40.

Ruppert, E. (2012) 'The governmental topologies of database devices', *Theory, Culture & Society*, 29: 116–36.

Schuurman, N. (2000) Critical GIS: Theorizing an emerging science. PhD dissertation, University of British Columbia. Available online: https://open.library.ubc.ca/cIRcle/collections/ubctheses/831/items/1.0089782 (accessed 1 February 2018).

Shelton, T. (2017) 'The urban geographical imagination in the age of Big Data', *Big Data & Society*, 4(1). DOI: 10.1177/2053951716665129.

Taylor, D.R.F. and Lauriault, T.P. (2014) *Developments in the Theory and Practice of Cybercartography*. Amsterdam: Elsevier Science.

Vale, S. (2015) 'International collaboration to understand the relevance of Big Data for official statistics', *Statistical Journal of the IAOS*, 31, 158–163. Available online: https://statswiki.unece.org/display/bigdata/International+collaboration+to+understand+the+relevance+of+Big+Data+for+Official+Statistics+2015 (accessed 1 February 2018).