

Compromised Data

From Social Media to Big Data

Edited by
Ganaele Langlois, Joanna Redden
and Greg Elmer

Bloomsbury Academic
An imprint of Bloomsbury Publishing Inc

B L O O M S B U R Y
NEW YORK • LONDON • NEW DELHI • SYDNEY

Big Data as System of Knowledge: Investigating Canadian Governance

Joanna Redden

Introduction

We are said to be entering a new era of governance through the collection and analysis of big data. The “data rush” is being compared to the gold rush and the oil boom as companies and governments spend billions to collect and analyze data (Hamm 2013). Utopian claims abound: big data is said to provide the means for governments to improve decision making, increase productivity, save money, become more efficient, improve health care, reduce fraud, get people to work sooner, provide more personalized services, prevent crises, and help society address a range of social and environmental problems, such as climate change and poverty. In short, it is said that big data will make us richer, healthier, safer, more civically engaged, improve government services, and lead to greater productivity (Manyika et al. 2011; Yiu 2012; Armah 2013; Microsoft 2013; TechAmerica Foundation 2012). There is often a dollar figure attached to big data promises. It has been argued that big data use could help the UK government save between £16 and 33 billion (Yiu 2012), that it could help the EU save between €150 and 300 billion a year, and help the US save \$300 billion in health-care costs (Manyika et al. 2011).

To date critical literature questioning the political and social impact of big data is limited (Crawford, Milner, and Gray 2014; Helles and Jensen 2013; Gitelman 2013). Recent revelations about the US National Security Agency’s PRISM surveillance program have drawn critique to various governments’ use of big data as related to privacy, rights, spying, and security (Bauman et al. 2014; Deibert 2014). Little attention has been directed to how governments around the world are using data analytics in their day-to-day operations to

inform decision making and policy development (Cook 2014; Bertot et al. 2014; Bhushan 2014; Margetts and Sutcliffe 2013). In the United Kingdom, the New Labour government oversaw the development of devices to connect data across services and departments (Ruppert 2012). More recently, big data is being tied to austerity measures as the coalition government investigates how data can be employed to reduce, target, and integrate programs and services. The American government is considered by many to be ahead of most other countries in its use of big data analytics. Big data is being used in security, law enforcement, education, and health care (Podesta et al. 2014). Like other G8 countries, the Canadian government has publicly endorsed the promise of governance through data. Data has been playing an ever-greater role as consumer marketing and political campaigning fuse. We do not know to what extent social media analysis and other forms of big data use are informing government decision making and policy development.

This chapter begins the process of detailing how the Government of Canada is using big data analysis now, and considers the political and social implications of this use. The specific research questions addressed are: How are the departments in the Government of Canada making use of data analytics? How is big data informing government decision making now? How might big data inform government decision making in the future? Research findings, I argue, raise important questions about big data and the computationalization of policymaking, the role of strategic ignorance, and the relationship of both to the intensification of neoliberal governance.

Method

This project employs semistructured interviews as the primary method of investigation. I also draw on government documents and research reports. Semistructured interviews with federal and provincial civil servants and private sector consultants working with government provide a means to track how and why government is using big data. To date, sixteen interviews have been conducted. Obtaining interviews with federal public servants has proven a major challenge. At the time of writing, interview requests had been sent to representatives from eighteen federal departments and three federal agencies; representatives from six federal bodies agreed to be interviewed and two of those representatives chose to engage in an email interview and provide

written responses to questions rather than engaging in an oral interview. It has proven much easier to obtain interviews with provincial public servants. To date, five interviews have been conducted, with more planned. Interviews were conducted with five consultants who work with the federal government. In addition, three government documents provide details about big data use within the federal government. These include two reports from the Office of the Privacy Commissioner, "Checks and Controls: Reinforcing Privacy Protection and Oversight for the Canadian Intelligence Community in an Era of Cyber-Surveillance" and "The Age of Predictive Analysis: From Patterns to Predictions," and a report by the Standing Committee on Access to Information, Privacy and Ethics entitled "Privacy and Social Media in the Age of Big Data" (Dusseault 2013).

A semistructured interview format was selected as the best method for research as this fluid approach abandons a focus on standardization and complete control, and instead promotes open-ended dialogue (Deacon et al. 1999). The semistructured approach provides the flexibility to follow up interesting leads as they emerge. As one of the goals of the project is to consider the thinking behind big data use, questions are designed to draw out how data analysis and tools shape working practices, decision making, and become embedded as common practice. Although work investigating elite actors on a micro and qualitative level is rare, Davis (2007) and Herbst (1998) present two excellent examples to follow. By using interviews, the goal is to inductively build "grounded theory" by engaging continually and reflexively in data collection, initial interview analysis, and theorizing (Strauss and Corbin 1990). Theories are developed incrementally as interviews are continually transcribed, analyzed, and considered in relation to other interviews and key documents. In this way, certain findings throughout the course of the study "become solidified and others fall by the wayside" (Herbst 1998, 194). A similar list of questions was provided to all interviewees before the scheduled interview. Interviewees were asked about how big data analysis is being integrated into government processes, how it is now and might in the future influence decision making and policy development. They were also asked about whether or not analysis was conducted in-house or contracted out, and about data analysts working in government. Questions were asked about the analysis process, such as how research questions are generated, who does the analysis, and how information is reported and shared. There were also questions about perceived benefits, challenges, and limitations of big data analysis. Finally, questions were asked about national and international influencers

Table 2.1 Interview list.

Oral Interview	Field	Identifier	Interview Date
	Federal and Provincial Public Service	Public Servant A	2013
		Public Servant B	2014
		Public Servant C	2013
		Public Servant D	2014
		Public Servant E	2014
		Public Servant F	2014
		Public Servant G	2014
		Public Servant H	2014
		Public Servant I	2014
	Consultant	Consultant A	2013
		David Eaves, Public Policy Expert	2014
		Consultant B	2014
		Tasha Truant, Consultant Manager,	
		Goss Gilroy Inc.	2014
		Consultant C	2013
Written Response	Public Service	Treasury Board Secretariat	2014
		Communications Security Establishment	2014

and potentials and barriers to data analysis. Interviews were transcribed and responses were grouped thematically and analyzed. They were grouped into the following categories: what is happening in the federal government now, perceived benefits of big data analysis, concerns and limitations, big data literacy, big data versus traditional methods, the future of policy development, and the role of big data in international relations. The discussion of findings below begins by detailing some of the examples of current big data uses. I also summarize comments concerning the benefits, challenges, and limitations of big data use within government. Further, findings are considered in relation to the existing big data literature.

Findings

The federal government does not at this point in time have any documents that provide an overview of how big data is being used across government (Executive Office of the President 2012). Understanding how big data analysis is informing decision making requires compiling an overview of how it is being used within government to gain a comprehensive picture of the range and variety of applications. One of the objectives of the “Big Data and Canadian Governance” project is to develop such an overview. The information provided below is the beginning of this work, and as an ongoing project it remains incomplete. Examples are provided from various departments to illustrate the range of big data applications.

Big data is commonly explained in the IT sector in relation to volume, velocity, and variety. Volume refers to the increasing amount of data that streams from social media, from sensors, from transactions, and from machine to machine. Velocity refers to the speed of information coming in that people are trying to collect and respond to, often quickly. Variety refers to the variety of formats in which data is presented, including structured numeric data, and unstructured formats like texts, video, audio, or financial transactions. Combining this data in ways that can enable analysis is a major issue for government. Usefully, SAS, a business analytics company, adds variability and complexity to its definition of big data. Variability draws attention to data peaks; data loads can speedup and slowdown in relation to daily, seasonal, topical, and event-triggered peaks. Complexity refers to the need to connect, correlate, and link data from multiple sources (SAS 2014). This very broad definition of big data is useful when thinking through the broad range of data analysis that spans government departments. Some departments, such as Environment Canada and Natural Resources Canada, work with large volumes of data generated by environmental and spatial sensors. Health Canada works with a lot of administrative data sets. Other departments may analyze online transactions and public communications on social media sites.

In general, many of those interviewed view the federal government as more of a producer than a consumer of data. The Treasury Board Secretariat indicates that much of the federal government’s focus at the moment is on open data, making more government data accessible to the public, than on developing the use of data analytics. Canada launched its open data site in 2011 and relaunched it with more publicity in June of 2013. Canada signed the G8 Open Data Charter

in 2013, pledging to release high-quality and comprehensive data as early as possible. The Open Data Portal (data.gc.ca) at present hosts nearly 200,000 data sets. Some see the site as a step toward more open government, while others argue that the Conservatives are providing access to some information and promoting the idea of open data and open government while simultaneously making access to other types of government information more difficult (Ligeti and Smith 2014; Larsen and Walby 2012).

The intelligence community, as might be expected, is widely cited as being ahead of the rest of government in its use of big data analytics (Deibert 2013, 2014). When discussing how big data analysis is being used by the feds, public servants and consultants often responded that how the federal government is using big data and how intelligence agencies are using big data are two separate things; the point being that the latter must be considered and discussed separately. It is largely agreed that the federal government is in the early stages of using big data, while the intelligence community is highly advanced.

In the sciences, some of the bigger users of big data are Natural Resources Canada, Environment Canada, and Agriculture Agri-Food Canada, which generate a lot of data through their environmental sensing devices. Natural Resources Canada's work leading the Federal Geospatial Platform was highlighted as noteworthy. The goal of this project is to share geospatial assets. Environment Canada generates a lot of data through its work with Meteorological Services of Canada. The science departments are also in the process of trying to consolidate their data so that data can be accessed collectively and more easily shared.

Health Canada is using big data, for example, to develop early warning systems for monitoring pandemics. The extent to which Health Canada can make use of big data is limited by the fact that health falls under provincial jurisdiction and many health records are held provincially. The provinces vary in the extent to which they are big data ready. Public servants from one Canadian province suggest that it is a fallacy at present to talk about big data in the health sector because, although there is plenty of data, it is not integrated and so cannot be "crawled through." However, there is interest in moving toward a more integrated data solution (Public Servant G 2014). Logistically, many departments are wrestling with how to link data sets in a common fashion. In some cases, this will require addressing privacy and security matters. In many cases, there are strict legislative structures in place to ensure data is collected and maintained in accordance with specific obligations. Any move toward big data analysis will require respecting these legislative frameworks (Public Servant A

2013). In addition, new technological infrastructure will be needed to integrate data that will come in a variety of shapes and sizes. Consultant A notes that tight budget constraints in areas such as health are pushing public servants to integrate more data analysis into their policy and program development as they look for ways to make programs more effective and efficient.

Other big data initiatives include the Canada Border Services Agency use of predictive analytics to manage and mitigate risks. Human Resources and Skills Development Canada (now Employment and Social Development Canada) is developing and implementing the use of predictive analytics to save money by reducing client errors and fraud (HRSDC 2013). The aim is to detect Employment Insurance fraud and abuse through the analysis of multiple databases to detect people who have been overpaid. Files that are flagged by the system are then investigated. The program is undergoing a shift to this type of automated fraud detection and risk management through the use of analytical tools (Office of the Privacy Commissioner 2012). The National Research Council is said to be making great use of data to support research and business decisions. Revenue Canada, Statistics Canada, and Elections Canada have and continue to generate large data sets. Canada Post presently sells its compilation of address data to businesses for marketing and business intelligence. More work is needed to investigate how these government bodies are incorporating data analysis into their processes. Citizenship and Immigration Canada has created an immigrant longitudinal database and is linking its Permanent Resident Data System records to annual income tax returns to generate more detailed information than that provided by longitudinal surveys (Spencer and McNutt 2014). The Canadian Research Data Centre Network (CRDCN) manages and analyzes confidential data, making this data available to researchers to inform policy decisions. The network of centers provide access to confidential Statistics Canada records and an increasing number of administrative files such as files from the Canadian cancer registry, some provincial health records, and vital statistics. The CRDCN is trying to acquire more administrative data to enable greater data analytics (Spencer and McNutt 2014).

As of November 2013, the federal government did not have a program to do social media analysis, but public works had put out a tender for social media monitoring. At this point, it is unclear what this would entail. Through correspondence between the privacy commissioner and Tony Clement, the president of the Treasury Board, we know that some departments are collecting social media data (Wingrove 2014). Further, interviews indicate social media

monitoring and analysis in relation to specific initiatives, and experimental uses by some departments to test its research value. Communications staff members are doing basic social media monitoring. For example, staff at one federal agency are monitoring and trying to influence social media outcomes in relation to specific campaigns.

The federal government is in the process of consolidating 1,500 websites to one portal over the next four years. When these are consolidated, one of the goals is to collect more data about how Canadians are accessing their information about government in order to provide better information and to also integrate Web and social media presence. The Government of Canada has also created Shared Services Canada with the aim of consolidating some of their basic systems. Their goal is to free up capacity by eliminating some of the redundancy happening in relation to IT delivery and maintenance, so that there would be more resources available to make use of big data analytics (Public Servant A 2013).

Interviewees repeatedly stress that the use of big data analysis in relation to policy development and decision making is in early stages. Consultants note that there is increasing interest among public servants and that more are becoming open to the idea of making use of big data, but there is a long way to go before use of this method of analysis reaches anything close to common practice. The consultant manager of Goss Gilroy Inc., Tasha Truant, is actively going to government departments and giving presentations about the uses of big data analysis. Truant says the goal is to try and get the word out about what these technologies and methodologies can do. For many, big data is new. Truant's company taps into public content already published on the Web. She argues that this type of big data analysis provides a list of benefits such as access to a large quantity of data, that it is a relatively less expensive method of research, and that this method does not present some of the biases introduced through other methods. One of the benefits of big data analysis, she argues, is the ability to access unprompted conversations by groups around certain topics. As an example, she notes that bringing immigrants into immigrant serving organizations for a focus group could introduce bias depending on the circumstances. Truant argues that it is possible to get a better read on immigrant attitudes and experiences by analyzing what is said online where huge diaspora communities are already talking about their experiences without any guidance. Further, she argues that big data analysis is one way to deal with decreasing participation rates in surveys. Others argue that monitoring unprompted social

media comments and conversations on public sites does not plant questions and therefore introduce bias in the way that surveys do (Consultant B 2014). Others see potential benefits in big data as enabling public servants to make policies and services more responsive to real-time events, demands, and feedback. Eaves (Interview 2014) sees the potential for big data analytics to enable public servants to provide more customized services, provide better services according to citizen needs, and eventually enable real-time policy responses to health or food inspection reports. Analysts argue that big data can enable evaluations of the impact of projects and policies almost immediately, leading to more time-sensitive policy development.

At present, big data analysis is done by government employees and also contracted out. Basic social media analysis is being conducted and compiled by communications staff through social media platforms and services such as Hootsuite and Heartbeat. It is likely that the number of automated dashboard services will increase. These dashboard devices provide an easy and accessible way for communications staff to compile social media analytics. An analysis of the limitations and blind spots of these dashboard services would be useful. More detailed big data analysis is being conducted for a range of purposes. For example, it is being used to track how discourse around certain topics evolve online, who the key influencers are, how organizations are and are not changing debates, the impact of initiatives and programs, or how to meet increasing health needs in a better way. Approaches vary; public servants are sometimes going to analysts with specific questions or analysts are part of the process of formulating the analytical questions asked. Some are cautioning that there will need to be oversight the more big data analysis drives decision making.

One concern is the quality of the data itself (Busch 2014). There are issues with the very processes involved in trying to figure out how reliable the data is. As Leavey notes, "establishing information about samples, such as their demographic profiles and the provenance, reliability and timeliness of data is a challenge in itself. This makes correcting the data problematic" (2013, 16). Others have drawn attention to the question of authenticity online. The central question here is to what extent the information gleaned from social network analysis and other forms of online communication can be treated as "true," "accurate," and "authentic." The argument made by Manovich (2011) is that we know people behave differently in different environments and with different people, and in this respect, peoples' online behaviors are no different than their offline behaviors.

Given this, it would be a mistake to treat people's posts, comments, and other forms of information sharing as "transparent windows into their selves" (ibid.). The Canadian Privacy Commissioner has publicly warned that any social media data being used to inform government policy must be proven accurate, and that it is not clear that this obligation is being met (Wingrove 2014). Further, there is unease about the lack of information people have about the information that is being generated and/or collected about them and the ability of governments to protect people from harm if their personal data is released. More transparent debate is needed about what kinds of information governments plan to combine, analyze, and use (Eynon 2013).

There is a danger that an emphasis on data mining, data analysis, and data-informed decision making will transform the nature of government. The danger is that the public will be viewed as a set of subpopulations defined by risk profiles instead of a social body (Milakovich 2012). Another related danger is that, as data analysis approaches are largely being adopted from commercial uses, it is likely that citizens will be viewed and even treated as consumers and/or in connection to the particular subpopulation they are identified with rather than as citizens. One analyst noted that banks, for example, are restricted from targeting and offering special offers to people based on religion, age, or ethnicity, but that in a social media setting it is very easy for an analyst to figure out someone's age, religion, and ethnicity. Through social media analysis, it is possible to target groups and give some people special offers over others. He notes that this type of targeting and segmenting of groups of people could lead to unequal access to programs and services if incorporated into government and business practices (Consultant A 2013).

As noted by boyd and Crawford (2012), there is an assumption that the size of big data sets will lead to new and better information, and thus, insights. In part, this is related to the myth of objectivity and authenticity that surrounds big data analytics. However, in reality, subjective practices of selection continue. One of the dangers of big data analysis is becoming obsessed with the thing you know how to search for (Eaves 2014). Further, using big data for decision making and policy requires making sense of data and making sense of the data requires often complex models that few are able to design and understand. These models can be used to develop information that informs decision making, yet there are assumptions and biases involved, even at the model development stage, that are difficult for those outside of the process to question. In addition, these models are designed to work with the data that is available, not necessarily the

question that is being asked. This presents a limitation on what can be done with the information generated. The potential for people to challenge decisions being made based on data and complex models is limited to the extent that they can actually challenge the models used to generate information and therefore trace the reasoning behind information generation. In this new big data world data scientists wield considerable power and there is a growing data analysis divide between those with the expertise to use and understand data analytics (Manovich 2011). Controllers of the information generated, owners of the vast data archives, and data miners who select what information to observe and ignore possess considerable political power (Gehl 2011).

How will data analysts be trained given the wide variety of skills they will need to possess? As discussed by Meyer (2013), data analysts working in policy development will need to be diplomatic. They will also need to have a good understanding of the areas they are working in so that they can grasp what seems to make sense and is plausible and not simply respond to what the data appears to indicate. It will be crucial that faulty conclusions based on false correlations are avoided. Further, these data analysts will need to be very skilled in asking questions of data and assessing the value of the data they have.

Of concern in relation to the development of social policy is the extent to which big data analytics may replace other forms of knowledge. One of the central claims about big data is that it provides us with a better picture of things because rather than relying on samples we can now have access to “more” and even “all” of the data in some cases (Mayer-Schönberger and Cukier 2013). Such claims mask how the data available limits the kinds of questions we can ask. Since we do not have data for everything, the data we do have can shape the kinds of questions we ask and the type of research conducted (Eynon 2013). Given the increasing pressure on governments to reduce costs we must be conscious of how the use of big data in this regard can sideline other types of research or the questions people ask. The importance of this is underlined in the Canadian context as the turn to big data is happening while other bodies of knowledge are eliminated or silenced. In the last seven years, the Canadian Conservative government has systematically cut agencies and institutions generating information that counter the government’s neoliberal policy and legislative agenda. A few of the agencies and institutions that have been cut include: the Courts Challenges Program, the Canadian Council for International Co-operation, the Status of Women Canada, the National Council of Welfare, Statistics Canada’s long-form census, and a wide range of environmental agencies including the Canadian Foundation for

Climate and Atmospheric Science. Also, scientists, academics, and bureaucrats are being systematically silenced. In a recent survey by the Public Service Alliance of Canada, 90 percent of scientists said they could not speak freely. One public servant described the chill across the civil service: “The current government is re-creating federal departments to serve the interests of its industry and business supporters and subverting the science. . . . Public servants with a conscience live in fear [of opening] their mouths to the media or the public” (PIPSC 2013). The turn to big data within government should be contextualized in relation to the turn away from other modes of generating information.

Big data analysis may facilitate neoliberal approaches by effectively erasing poverty. The effect could further be compounded by the elimination of the mandatory long-form census in Canada. The information gathered from this census provides detailed information about inequality and poverty in Canada. This information was used to justify funding for social programs and to also ensure money was directed where needed. The elimination of the mandatory long-form census has the effect of making the poor invisible and “easier to ignore” (McQuaig 2010). The use of big data for government decision making could contribute to a further erasure of poverty and even more unequal treatment of individuals (Meyer 2013; Lerman 2013). For a start, not everyone is online. Big data may both reinforce and even exacerbate existing social and educational inequalities. The people who leave data traces to be analyzed are likely to be those people of higher income and education who are online more and participate more on Twitter, the blogosphere, Google, etc. Also, it is a challenge to correctly weigh data for the wider population that does not have any kind of social media presence (Leavey 2013).

Three of the consultants interviewed and one public servant worry about the potential for big data analysis to exclude and marginalize particular groups of people. For Truant, potential exclusion and marginalization must be addressed in any research. She cautions that big data analysis should not be about replacing other forms of research, but about complementing other methods of investigation:

The more we turn to online technologies to develop our policies and our programs the more we risk leaving out groups of people who are already marginalized, and we’re already silencing people whose voices may not be all that loud in the first place. . . . To me, that’s the thing I hope at the end of all of this I haven’t contributed to. What it comes down to is we can’t be lazy, we can’t do it all online. Again it’s one line of evidence amongst many. You still have to

go out there into the rural communities where people aren't online and talk to that person. . . . My answer is always "Look, it's not one or the other. It's about bringing this into what we're already doing and using it as one line of evidence.

The extent to which big data analytics will be added to other modes of investigation, and not simply replace them, will require continual monitoring particularly in this budgetary climate. Many IT companies are clearly marketing big data analysis as a cost-saving practice.

Some of the examples provided of uses of social media monitoring demonstrate how analysts are attempting to ensure their online samples are representative of offline realities. One public servant set out, through a pilot project, to test the extent to which online conversations were representative of offline dynamics (Public Servant B 2014). Public Servant B notes that online conversations, and the data gathered from social media monitoring, are only relevant when representative of real-world dynamics (Public Servant B 2014). Consultant B argues that analysts should take potential blind spots into account. This can involve identifying how and where the groups of interest are communicating and if their communication channels are accessible for analysis.

One of the major benefits of big data from a governmental perspective is that it enables probabilistic policymaking. This type of policymaking relies on developing rules and approaches based on ideas about what small groups of people will probably do rather than what they have done. A good example of this is the predictive policing in use in a number of cities across the United States. At this stage, it is unclear if predictive analytics are being used by the Canadian justice department, although it is known that cities across Canada are trying it. Toronto has tried predictive policing and the Vancouver Police Department is using predictive analysis (Silva 2012; Allen 2013). One of the dangers of predictive analysis and probabilistic policymaking is that it can lead to a "feedback loop of injustice" as groups identified as "at risk" or as likely to reoffend are increasingly watched and targeted (Margetts 2013). In this way, a reliance on future predictions to develop policies builds upon and reinforces present power dynamics. The effect of this is compounded by a turn away from causality to an emphasis on correlations. The big data argument is that we do not necessarily need to know why something is happening, just that it is happening (Mayer-Schönberger and Cukier 2013). This raises serious questions from a policy perspective, particularly when you consider it in relation to an issue like poverty, which is tied to a range of political, economic, and social factors. Eliminating causality from the equation can only contribute

to the idea that poverty is a product of individual failing and an individual's responsibility.

The common view among interviewees is that government is still "testing the waters," but that the government will be moving toward more, and not less, big data analysis. As one analyst suggests, "it's hard to say what all of this will mean for the future of policy development, we're really only a couple of years into this" (Consultant B 2014). The cultures of policymaking, and whether or not these cultures change, will influence how big data analysis is and is not integrated into the policy process. High-ranking bureaucrats, as some note, have often achieved their position of authority through their political experience and instincts. There will need to be a change in mindset and a shift in policymaking practices by this group for big data analysis to be further integrated into government processes. "It's [going to involve] getting those folks to change their mindset, that they have to take policy advice from the pimple faced kid with the earring" (Consultant A 2013). One consultant argues that big data analysis may be threatening. "The notion that there is a whole bunch of data that might actually reveal some interesting insights threatens the group of people who don't have the skill sets to actually use and manage and play with that data" (Eaves 2014). Others note that there is always the question of whether or not higher-level decision makers care about the evidence when making policy decisions, and that this impacts the take-up and use of data analysis.

One interviewee argues that public servants and political staff worry about public perceptions. The fear is that the public is not making a distinction between data consultants and the analysis of already published content online, versus the more covert surveillance conducted by security agencies gaining much recent attention. One of the issues raised by consultants and analysts is that they are operating in a "murky, grey area" right now, as there are no federal guidelines about how the government should and should not be using social media analysis. Companies in Canada that do predictive analysis or use personal information have to follow the Personal Information Protection and Electronics and Documents Act. The Act does not provide the necessary oversight or guidelines for action in relation to the types of social media monitoring being done. In response to recent revelations about American, British, and Canadian intelligence surveillance practices, the Privacy Commissioner published a special report to parliament with a number of recommendations regarding government access and use of online communications (2014). The report recommends government develop guidelines for the collection, use, and dissemination of

intelligence products using online sources and social network sites. The Privacy Commissioner also takes the position that just because personal information is available on the Internet, does not mean that this information is not personal. The commissioner cautioned that departments should not access personal information on social media unless they can demonstrate a direct correlation to legitimate government business. The report has had a chilling effect on users, and potential users, of social media data (Consultant B 2014). Further, interim Privacy Commissioner Chantal Bernier warned Treasury Board President Tony Clement in February 2014 that the government's collection of social media data might violate the Privacy Act. Minister Clement has said the government is conducting an internal review of these practices.

Cost is proving another barrier to big data integration. While consultants argue that big data analysis will save money, bureaucrats view big data as a new tool requiring new people to make use of it. Public Servant A describes the budget issue as one of capacity:

As you know we are in a period of fiscal restraint and our budgets from an IT perspective do not grow on regular basis, nor do budgets for program delivery grow, so everybody is under the buzz word of constraint. Doing new things on that scale requires capacity, requires financial resources for computing capabilities and software and also the skillsets and the teams ready to focus on those, and that involves in today's environment having to make some tough decisions as to what will I stop doing if instead I want to work on this. I would say in the next five years we will see a growing move towards more big data.

She says the government's aim in integrating big data analysis will be to become more productive and efficient and provide better program and service delivery. She notes that departments are well aware of the potential to use big data but are limited by what they can do. Another public servant argues that it is difficult for government to do anything new in IT because of the increasing top-down control over expenditures. He argues that the increasing bureaucratization and control over decision making is dysfunctional and slows down and in some cases prevents innovation. Public Servant C is particularly troubled by how cost recovery models within various departments can be prohibitive to the sharing of ideas and information. His examples include one government body charging another to access its information, or IT staff being required to charge other departments for time spent in meetings. In combination, there is consensus that the future of big data – driven policy development depends largely upon

how high-ranking policymakers and political staff respond to it as a tool. Their response is expected to be influenced most by how useful it seems, the outcomes of privacy debates, and if they conclude that the benefits exceed the costs.

Conclusion

This chapter presents an early sketch of big data approaches within the Canadian federal government. We are dealing with broad strokes and anecdotes of micro details. The ability to construct a systematic illustration of how big data analysis is being used by the federal government at present is limited by the sheer size of government operations and also by a lack of access to the information required. Nevertheless, this initial sketch raises some significant questions, concerns, and areas for future research. The lack of access to federal civil servants is an important methodological finding, and a problem that will need to be considered and addressed in future research if we are to better understand government practices and how they are changing. Interviews are a crucial social science research method, particularly when researching the present in order to consider future political and social implications. Interviews are also an important way to better understand how digital technologies are being used by elite actors and influencing the development of policy. What strategies can be used to unblock interview access? What other methods might we use to better understand the thinking and reasoning influencing the ways digital technologies are being integrated into government processes?

This research contributes to some of the epistemological concerns about the uses of big data. With the turn to big data, as raised by boyd and Crawford (2012), there is a concern about the computational turn in thought and research. As noted by boyd and Crawford, big data shapes the reality it measures by staking out new methods of knowing (2011, 665). Because of the sheer amount of information available, subjective decisions are made about what is measured and what is ignored. The “subtractive methods of understanding reality”—that is, the reduction of information flows into numbers that can be stored and then mined—produce very particular forms of information and computational knowledge (Berry 2011, 2). At issue is how this computationalization of reality, the “datafication” of reality, can lead to a computationalization of reason (Berry 2011), and moreover the congruence between the computationalization of reason and the intensification of neoliberalism. Epistemological concerns draw

attention to how big data shapes the reality it measures through the subjective and value-laden decisions made about what to measure, the significance of results, and the value placed on big data findings as opposed to other methods of information generation. The process of converting our messy and unruly social world into numbers to be stored and mined, into data that can be collectively “crawled through” and compared, fits some frameworks of meaning better than others. In this way, the “datafication” of reality can privilege some lines of thought over others. Most importantly the datafication of reality may reinforce neoliberal frameworks of meaning over social justice frameworks.

Computational, or big data, decision making may reinforce neoliberal rationality, and therefore reinforce neoliberal calls for cuts and further marketization of government programs and services. This is particularly dangerous given the intensification of neoliberal governance since the financial crisis of 2007–08 (Mirowski 2013; Peck 2010; Hall, Massey, and Rustin 2013). The transformation of information about the social world in all its complexity into numbers, models, and calculations complements and reinforces the instrumental and market-based thinking of neoliberal rationality (Foucault 2008; Davies 2013). The problem, and cause for alarm, is the further dissemination of market values to all spheres of life given the role that neoliberal approaches have played in not simply increasing poverty and inequality (Coburn 2000; Navarro 2007; Ruckert and Labonté 2014; Jacobs and Lindsey 2014; Wacquant 2009) but encouraging attacks on the poor (Briant, Watson, and Philo 2013; Mooney 2011; Kendall 2011). The idea, as argued by Foucault and others, is that neoliberalism is not just a political and economic project, but also a project to change the way we think; neoliberalism becomes a method of thought, a grid of economic and sociological analysis (Foucault 2008, 218). The goal is for the rationalities of the market and its schemas of analysis, its decision-making criteria, to be extended to all facets of life (Couldry 2010; Brown 2005). These new tools complement ongoing attempts to quantify life, to make all decision making based on calculative reasoning. Information systems are shaped by value systems, and economic values, as opposed to social and personal values, are driving big data analysis (Baym 2013). Big data analysis supports and furthers governmental emphasis on efficiency, cost-benefit, productivity, quantification, and targets, continuing the neoliberal colonization of government practices since the advent of new public management strategies in the 1980s (Boltanski and Chiappello 2005; Lorenz 2012; Power 1999).

As the Canadian government increasingly integrates big data use in policy development processes, big data analysis must be placed within its wider

informational context. In the Canadian case, this context is full of seeming contradiction: increasing government use of big data analysis, more social media monitoring, increasing efforts to make more data open to the public, increasing cuts to significant statistical services such as cutting the long-form census, cuts to key information bodies such as the National Council of Welfare, greater control of access to information, limits on journalistic investigation, and barriers to public servants speaking publicly. This context is important because while some sources of information are being eliminated or silenced, others are being pursued. As an example, information, such as the long-form census, provides government authorities with the high-quality socioeconomic data necessary to justify redistributive programs designed to increase social and economic equality. The census renders the poor “visible” in ways that social media monitoring cannot. The sheer fact of its elimination in the face of massive protest brings forth a reminder about the role of strategic ignorance in furthering the neoliberal project (Davies and McGoey 2012; Mirowsky 2013). As argued by Hayek, ignorance can be desirable and necessary to diffuse the authority of central planners by taking away the authority and reasons for action (Davies and McGoey 2012). The widespread claims of ignorance in the face of the 2007–8 financial crisis demonstrate the effectiveness of ignorance as strategy, and also its value to the neoliberal project (Davies and McGoey 2012). Going forward, it will be crucial to evaluate how government further integrates big data analysis, to consider its relationship to strategic ignorance, and to evaluate the extent to which big data analysis is used to supplement and complement research or to replace and reduce the expense of research endeavors.

References

- Allen, Mary. 2013. “‘Intelligence-led Policing’ in Reach with Analytics.” *InsightaaS.com*. September 24. Accessed July 9, 2014. <http://www.insightaas.com/intelligence-led-policing-in-reach-with-analytics/>.
- Armah, Nii Ayi. 2013. “Big Data Analysis: The Next Frontier.” *Bank of Canada Review*. Summer. Accessed July 9, 2014. <http://www.bankofcanada.ca/wp-content/uploads/2013/08/boc-review-summer13-armah.pdf>.
- Bauman, Z., D. Bigo, P. Esteves, E. Guild, V. Jabri, D. Lyon, and R. B. J. Walker. 2014. “After Snowden: Rethinking the Impact of Surveillance.” *International Political Sociology* 8(2): 121–44.

- Baym, Nancy K. 2013. "Data Not Seen: The Uses and Shortcomings of Social Media Metrics." *First Monday* 18(10). Accessed June 10, 2014. <http://firstmonday.org/ojs/index.php/fm/article/view/4873>.
- Berry, David. 2011. "The Computational Turn: Thinking About the Digital Humanities." *Culture Machine* 12: 1–22. Accessed July 10, 2013. <http://www.culturemachine.net/index.php/cm/article/viewDownloadInterstitial/440/470>.
- Bertot, J. C., U. Gorham, P. T. Jaeger, L. C. Sarin, and H. Choi. 2014. "Big Data: Open Government and E-Government: Issues, Policies and Recommendations." *Information Polity* 19: 5–16.
- Bhushan, Aniket. 2014. "Fast Data, Slow Policy: Making the Most of Disruptive Innovation." *SAIS Review of International Affairs* 34(1): 93–107.
- Boltanski, L. and E. Chiapello. 2005. *The New Spirit of Capitalism*. New York: Verso.
- boyd, d. and K. Crawford. 2012. "Critical Questions for Big Data." *Information, Communication & Society* 15(5): 662–79.
- Briant, Emma, Nick Watson, and Gregory Philo. 2013. "Reporting Disability in the Age of Austerity: The Changing Face of Media Representation of Disability and Disabled People in the United Kingdom and the Creation of New 'Folk Devils.'" *Disability & Society* 28(6): 874–89.
- Brown, Wendy. 2005. "Neoliberalism and the End of Liberal Democracy." In *Edgework: Critical Essays on Knowledge and Politics*. Princeton: Princeton University Press, 37–59.
- Busch, Lawrence. 2014. "A Dozen Ways to Get Lost in Translation: Inherent Challenges in Large-Scale Data Sets." *International Journal of Communication* 8: 1727–44.
- Coburn, David. 2000. "Income Inequality, Social Cohesion and the Health Status of Populations: The Role of Neo-liberalism." *Social Science & Medicine* 51: 135–46.
- Cook, Thomas D. 2014. "'Big Data' in Research on Social Policy." *Journal of Policy Analysis and Management* 33(2): 544–47.
- Couldry, Nick. 2010. *Why Voice Matters: Culture and Politics after Neoliberalism*. London: Sage.
- Crawford, K., K. Milner, and M. L. Gray. 2014. Special Section Introduction: "Critiquing Big Data: Politics, Ethics, Epistemology." *International Journal of Communication* 8: 1663–72.
- Davies, William. 2013. "Neoliberalism and the Revenge of the 'Social.'" *openDemocracy*. July 16. Accessed September 29, 2014. <http://www.opendemocracy.net/william-davies/neoliberalism-and-revenge-of-%E2%80%99social%E2%80%99D>.
- Davies, William and McGoey, Linsey. 2012. "Rationalities of Ignorance: On Financial Crisis and the Ambivalence of Neo-Liberal Epistemology." *Economy and Society* 41(1): 64–83.
- Davis, Aeron. 2007. *The Mediation of Power*. London: Routledge.
- Deacon, David, Michael Pickering, Peter Golding, and Graham Murdoch. 1999. *Researching Communications: A Practical Guide to Methods in Media and Cultural Analysis*. London: Arnold.

- Deibert, Ronald. 2013. *Black Code: Inside the Battle for Cyberspace*. Toronto: McClelland and Stewart.
- Deibert, Ronald. 2014. "Now we know Ottawa can snoop on any Canadian. What are we going to do?" *Globe and Mail*. January 31. Accessed May 3, 2014. <http://www.theglobeandmail.com/globe-debate/now-we-know-ottawa-can-snoop-on-any-canadian-what-are-we-going-to-do/article16625310/>.
- Dusseault, Pierre-Luc, M.P. 2013. "Privacy and Social Media in the Age of Big Data." Report of the Standing Committee on Access to Information, Privacy and Ethics. Ottawa: Government of Canada. 41st Parliament, First Session, April. Accessed June 10, 2014. <http://www.parl.gc.ca/content/hoc/Committee/411/ETHI/Reports/RP6094136/ethirp05/ethirp05-e.pdf>.
- Executive Office of the President. 2012. "Big Data Across the Federal Government." Washington: The White House. Accessed June 4, 2014. http://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_fact_sheet.pdf.
- Eynon, Rebecca. 2013. "The Rise of Big Data: What Does it Mean for Education, Technology, and Media Research?" Editorial. *Learning, Media and Technology* 38(3): 237–40.
- Foucault, Michel. 2008. *The Birth of Biopolitics: Lectures at the College De France 1978–1979*. New York: Palgrave Macmillan.
- Francoli, Mary. 2014. Independent Reporting Mechanism Canada: Progress Report 2012–2013. Ottawa: Government of Canada. Accessed June 10, 2014. http://www.opengovpartnership.org/sites/default/files/Canada_final_2012_Eng.pdf.
- Gehl, Robert. W. 2011. "The Archive and the Processor: The Internal Logic of Web 2.0." *New Media & Society* 13(8): 1228–44.
- Gitelman, Lisa. 2013. *Raw Data is an Oxymoron*. Cambridge, MA: The MIT Press.
- Hall, S., D. Massey, and M. Rustin. 2013. "After Neoliberalism? The Kilburn Manifesto." *Soundings*. Accessed June 5, 2014. <http://www.lwbooks.co.uk/journals/soundings/manifesto.html>.
- Hamm, Dax. 2013. "How Data-Driven Society Can Benefit Us." *Huffington Post Business Canada* (blog). May 6. Accessed July 31, 2013. http://www.huffingtonpost.ca/dax-hamman/big-data_b_3215520.html.
- Helles, Rasmus and Klaus Bruhn Jensen. 2013. Introduction to the special issue: "Making data – Big Data and Beyond." *First Monday* 18(10). Accessed June 10, 2014. <http://firstmonday.org/article/view/4860/3748>.
- Herbst, Susan. 1998. *Reading Public Opinion: How Political Actors View the Democratic Process*. London: University of Chicago Press.
- Human Resources and Skills Development Canada. 2013. 2013–2014 Report on Plans and Priorities. Ottawa: Government of Canada. Accessed July 7, 2014. <http://www12.hrsdc.gc.ca>.
- Jacobs, David and Lindsey Myers. 2014. "Union Strength, Neoliberalism, and Inequality: Contingent Political Analyses of U.S. Income Differences Since 1950."

- American Sociological Review*. June 9: 1–23. Accessed July 2, 2014. <http://asr.sagepub.com/content/early/2014/05/29/0003122414536392>.
- Keen, Justin, R. Calinescu, R. Paige, and J. Rooksby. 2012. "Big Health Data: Institutional and Technological Challenges." Conference presentation at Internet, Politics, Policy 2012: Big Data, Big Challenges, Oxford Internet Institute.
- Kendall, Diana. 2011. *Framing Class: Media Representations of Wealth and Poverty in American*. Lanham, MD: Rowman & Littlefield.
- Larsen, Mike and Kevin Walby. 2012. *Brokering Access: Power, Politics, and Freedom of Information Process in Canada*. Toronto: University of British Columbia Press.
- Leavey, Jason. 2013. *Social Media and Public Policy: What is the Evidence*. Discussion Paper. Alliance for Useful Evidence.
- Lerman, Jonas. 2013. "Big Data and Its Exclusions." *Stanford Law Review Online* 66: 55–63.
- Ligeti, Arik and Marie-Danielle Smith. 2014. "Open Data, Restricted Access." *Capital News Online*, March 21. Accessed July 9, 2014. <http://www.capitalnews.ca/index.php/news/open-data-restricted-access>.
- Lorenz, Chris. 2012. "If You're So Smart, Why Are You Under Surveillance? Universities, Neoliberalism, and New Public Management." *Critical Inquiry* 38(3): 599–629.
- Manovich, Lev. 2011. "Trending: the Promises and the Challenges of Big Social Data." In Matthew K. Gold, ed. *Debates in the Digital Humanities*. Minneapolis: The University of Minnesota Press, 460–75.
- Manyika, J., M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. Hung Byers. 2011. "Big Data: The Next Frontier for Innovation, Competition, and Productivity." McKinsey Global Institute.
- Margetts, Helen. 2013. "The Promises and Threats of Big Data for Public Policy-Making." *The Policy and Internet Blog*. Oxford Internet Institute, University of Oxford. Accessed October 28, 2013. <http://blogs.oii.ox.ac.uk/policy/promises-threats-big-data-for-public-policy-making/>.
- Margetts, Helen and David Sutcliffe. 2013. "Special Issue: Potentials and Challenges of Big Data." *Policy & Internet* 5(2): 139–46.
- Mayer-Schönberger, Viktor and Kenneth Cukier. 2013. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. New York: Houghton Mifflin Harcourt.
- McQuaig, Linda. 2010. "Making it Easier to Ignore the Poor." *Rabble.ca*. Accessed August 20, 2011. <http://rabble.ca/columnists/2010/07/harpers-attack-census-bad-news-poor>.
- Microsoft. 2013. "Seizing the Power of Public Data to Transform Government." Microsoft and Socrata. Accessed September 25, 2013. <http://www.microsoft.com/government/ww/public-services/initiatives/pages/openness-interopability.aspx>.
- Milakovich, Michael E. 2012. "Anticipatory Government: Integrating Big Data for Smaller Government." Conference presentation for Internet, Politics Policy 2012: Big Data, Big Challenges. St. Anne's College, Oxford University, UK. September.

- Mirowski, Philip. 2013. *Never Let a Serious Crisis Go to Waste: How Neoliberalism Survived the Financial Crisis*. London: Verso.
- Meyer, E. T. 2013. "Big Data and Public Policy Workshop." Oxford Internet Institute. Accessed September 30, 2013. <http://people.oii.ox.ac.uk/meyer/2013/09/15/big-data-and-public-policy-workshop/>.
- Mooney, Gerry. 2011. *Stigmatising Poverty? The "Broken Society" and Reflections on Anti-Welfarism in the UK Today*. Oxford: Oxfam. Accessed September 20, 2013. <http://oro.open.ac.uk/29714/>.
- Navarro, Vincente. 2007. *Neoliberalism, Globalization, and Inequalities*. Amityville, NY: Baywood Press.
- Office of the Privacy Commissioner of Canada. 2012. "The Age of Predictive Analytics: From Patterns to Predictions." Research Group at the Office of the Privacy Commissioner of Canada. Ottawa: Government of Canada. Accessed July 1, 2014. https://www.priv.gc.ca/information/research-recherche/2012/pa_201208_e.asp#_ftn47.
- Office of the Privacy Commissioner of Canada. 2014. "Checks and Controls: Reinforcing Privacy Protection and Oversight for the Canadian Intelligence Community in an Era of Cyber-Surveillance." Special Report to Parliament. Ottawa: Government of Canada, January 28. Accessed July 1, 2014. https://www.priv.gc.ca/information/sr-rs/201314/sr_cic_e.pdf.
- Peck, Jamie. 2010. *Constructions of Neoliberal Reason*. Oxford: Oxford University Press.
- Podesta, John, Penny Pritzker, Ernest J. Moniz, John Holdren, and Jeffrey Zients. 2014. *Big Data: Seizing Opportunities, Preserving Values*. Executive Office of the President. Washington, DC: The White House. Accessed June 10, 2014. http://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf.
- Power, Michael. 1999. *The Audit Society*. Oxford: Oxford University Press.
- Professional Institute of the Public Service of Canada. 2013. "The Big Chill, Silencing Public Interest Science, A Survey." Ottawa, Canada. Accessed October 15, 2013. <http://www.pipsc.ca/portal/page/portal/website/issues/science/bigchill>.
- Ruckert, Arne and Ronald Labonté. 2014. "The Global Financial Crisis and Health Equity: Early Experiences from Canada." *Globalization and Health* 10(2): 1–10.
- Ruppert, Evelyn. 2012. "The Governmental Topologies of Database Devices." *Theory, Culture & Society* 29(4/5): 116–36.
- SAS. 2014. "Big Data: What it Is and Why it Matters." SAS. Accessed July 9, 2014. http://www.sas.com/en_us/insights/big-data/what-is-big-data.html.
- Silva, Luis. 2012. "High-Tech Crime-Fighting." *Toronto Star*, July 29. Accessed July 2, 2014. http://www.thestar.com/opinion/editorialopinion/2012/07/29/hightech_crimefighting.html.
- Spencer, Byron G. and Robert McNutt. 2014. Science and Technology Consultation by Industry Canada, Canadian Research Data Centre Network, February. Accessed May 8, 2014. http://www.rdc-cdr.ca/sites/default/files/submission_to_industry_canada.pdf.

- Strauss, Anselm and Juliet M. Corbin. 1990. *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*. Newbury Park, CA: Sage.
- Swanson, Jean. 2001. *Poor-Bashing: The Politics of Exclusion*. Toronto: Between the Lines.
- TechAmerica Foundation. 2012. "Demystifying Big Data: A Practical Guide to Transforming the Business of Government." Washington, DC. Accessed September 20, 2013. <http://breakinggov.sites.breakingmedia.com/wp-content/uploads/sites/4/2012/10/TechAmericaBigDataReport.pdf>.
- Wacquant, Loic. 2009. *Punishing the Poor: The Neoliberal Government of Social Insecurity*. Durham: Duke University Press.
- Wingrove, Josh. 2014. "Ottawa Launches Data Collection Review." *Globe and Mail*. May 8. Accessed June 9, 2014. <http://www.theglobeandmail.com/news/politics/ottawa-launches-data-collection-review/article18578578>.
- Yiu, Chris. 2012. "The Big Data Opportunity: Making Government Faster, Smarter and More Personal." *Policy Exchange*. <http://www.policyexchange.org.uk/media-centre/blogs/category/item/the-big-data-opportunity>.