DJIM Dalhousie Journal of Interdisciplinary Management

Volume 12 – Spring 2016
djim.management.dal.ca
doi: 10.5931/djim.v12.i1.6449

Do We Need Data Literacy?

Public Perceptions Regarding Canada's Open Data Initiative

Abstract: This research report is a product of the course "Research Methods" (INFO 5520), taught by Dr. Mike Smit in the School of Information Management, Dalhousie University during the 2015 winter semester. The four researchers credited with authorship chose the topic, created research questions to address the topic, wrote and circulated a survey in order to receive public perceptions of the research questions, and finally wrote this paper to present their analysis of the survey results.

Author(s): Michelle Boychuk (mc443366@dal.ca), Mark Cousins (mr796112@dal.ca), Amanda Lloyd (amanda.lloyd@dal.ca), Charlotte MacKeigan (charlotte.mackeigan@dal.ca)
School of Information Management, Dalhousie University

Introduction

Open Government data is an emerging concept where government departments make their datasets freely available for public use in a machine-readable format. Members of the public can use Open Government data by downloading, reading, and manipulating the datasets to better understand how the Canadian Government operates and to engage more critically and knowledgeably in the government decision-making process. For example, there is a data set entitled "Air Pollutant Emission Inventory (APEI) Historical Trends" published by Environment Canada hosted on Canada's Open Data Portal, open.canada.ca. It covers data from 1990 through 2014 and could be used by citizens to inform environmental policy and building regulations. Open Government data can also be used to help inform citizen decisions like "Who should I vote for?", encourage collaboration between government, citizen, and private sector, and help to identify service needs or improve services based on demographic data (Davies, 2010).

In 2011, the Canadian Government introduced Canada's Action Plan on Open Government, which positions Open Data as one of its primary elements. The Action Plan launched the Open Data Portal, where the public can access raw data released by various government departments like Environment Canada and the Department of Finance in many formats such as .XML and GeoTIFF. Open Government strives to use Open Data to promote civic engagement and economic growth, and to support government transparency and accountability.

In order to obtain, understand, and manipulate raw data, a certain skill level--data literacy--is important (Schield, 2005). It is therefore important to assess the impact of the Open Data initiative for the Canadian public and whether they possess the necessary skills and motivation to use Open Data. Alternatively, the raw data can be curated, meaning an intermediary works to transform the most salient pieces of information out of raw data and organize them into an easily understandable presentation. However, what the Canadian public perceives to be the most effective method for interacting with Open Data is unknown, as the public has not been adequately consulted. Our study investigates the unknown variable that is public perceptions of Open Data in terms of its relevance and importance to the lives of Canadian individuals and to Canadian society. The following research questions guided our study: To what extent has the Canadian Open Government initiative, specifically the publishing of Open Data, impacted the general public? What importance do Canadian citizens place on data literacy to fully exercise their democratic rights in utilizing Open Data? How important is teaching the public to use

Open Data to realizing the principles of Open Data: transparency, accountability, and collaboration? How would teaching data literacy skills best be achieved? Does the public prefer to interact with raw Open Data, or have third parties translate that data into usable applications?

We conducted a literature review on Open Government Data strategies, problems of access, the role of educators and librarians, and data literacy to identify gaps in current research and to develop our research questions. The methods section outlines the qualitative and quantitative data collection and analysis approaches we took to answer our research questions. The results and discussion sections include the most significant findings and emergent themes from our data analysis. The conclusion includes suggestions for improving data literacy education, Canada's Open Data Portal usability, and how Open Data could potentially be curated for public consumption.

Literature Review

Open Government Data Strategies and Principles

Numerous governments, such as the United States and Canada, have developed initiatives geared toward making government data publicly accessible as a part of the Open Government Partnership (Open Government Partnership, n.d.). Open Government aims to promote transparency, accountability, and civic engagement in order to strengthen the overall democratic process (Government of Canada, 2014a; Open Government Partnership, 2014; Jaeger & Carlo Bertot, 2010). Greater government transparency provides public citizens with "greater control over their lives and improv[es] both their material and social conditions" (Johnson, 2014, p.263). As part of Open Government, both the Canadian and US Governments developed websites (www.open.canada.ca; www.data.gov) where they release raw government data for free public use (Jaeger & Carlo Bertot, 2010; Bertot et al., 2014; Government of Canada, 2014a). Although the government initiatives addressed providing free public access to their data, initially they did not consider the ability of citizens to effectively read and use the data.

In addition to releasing government data online to the public, Canada's Open Data Charter attempts to address the public's lack of data literacy by employing developers to "unlock the value of Open Data" because not all citizens are able to read and use Open Data (Government of Canada, 2014b, p. 29). "Appathons" like the Canadian Open Data Experience (CODE) invite developers to utilize Open Data

by creating widely usable applications that address this lack of comprehension (Government of Canada, 2014a).

Problems of access

Open government initiatives have the potential to stimulate all sectors of the economy and foster democratic participation; however, there is little evidence to suggest that these initiatives have made a transformative impact as of yet (Martin, 2014). Approximately 70% of surveyed members of the United Kingdom's Open Government Data community generally agreed that civil society does not possess adequate skills and resources to use Open Data (Martin, 2014). That being said, they also expressed conflicting perceptions concerning whether or not data literacy skills are confined to a small community of educated users. Respondents were not an adequate representation of the public, and therefore their perceptions of barriers to access Open Data may differ from how the general public perceives their ability to utilize datasets. Tools to assess and measure open government data initiatives tend to focus on the data's quality, accessibility, authenticity, and availability (Veljković et al., 2014; Ren & Glissmann, 2012). These tools concentrate on the back-end usability of the datasets and do not incorporate or evaluate the general public's ability to use or manipulate them. Andreae & Anderson (2012) propose a new theoretical model in which access extends beyond the physical availability of technological equipment, instead focusing on an individual's ability to engage with the information itself, which could be achieved once skills in critical thinking and analysis are obtained.

Role of educators, data literate librarians, and developers

Stephenson and Carvello (2007) conducted a study where UCLA students were taught data literacy skills that employ critical thinking and analysis. Two different teaching approaches were used – one in a standalone lab, and the other through the incorporation of data literacy concepts into an introductory Sociology course. Students found that both approaches were useful and that they helped them grasp the data with a deeper understanding, but the researchers felt the integration of the data literacy concepts into the curriculum was more useful because they could be related to course content and learned alongside other critical thinking skills (Stephenson & Caravello, 2007). Data literacy is presented as imperative for understanding complex datasets, but many open government data initiatives do not address this skill, nor do they clearly lay out plans to teach this skill to the public. Instead, teaching data literacy is often deferred to educators, librarians, or developers.

Library patrons rely on libraries for access to e-government and on the expertise of information literate librarians to assist them with using e-government resources. This is due to a lack of literacy skills needed to accurately navigate and make use of online data, which points to issues with using e-government resources (Jaegar & Bertot, 2010). Libraries have addressed this issue by creating positions for data librarians who are dedicated to understanding, curating, and assisting in the use of e-government resources and complex datasets (Hyams, Martinez-Uribe & Macdonald, 2008). In terms of both student and citizen populations, the onus is often put on libraries to teach these skills (Shield, 2005; Bertot et al., 2014). Beyond educators teaching data literacy skills and librarians assisting in the use of Open Data, other professionals, such as developers, also curate data for public consumption. Using data literacy skills, these professionals combine government data with other files to create usable websites or applications.

Kassen (2013) suggests that independent developers could use open government datasets to create applications that could promote civic engagement. Citizens would not be directly accessing or using the open government datasets, but could provide the developers with additional data to improve their applications, for example, by ranking them (Kassen, 2013). Instead of citizens accessing the unrefined datasets, this symbiotic relationship places the application developer between them. It would be the role of the independent application developer to translate open government data into usable information for the public. Possessing data literacy skills are relevant to more than simply the conversion of Open Data—they can be critical for various infrastructure improvements as well.

Data literacy is counted as an obstacle in the implementation process of using web analytics to meet social aims of innovators or organizations (Fotopoulou & Couldry, 2014). A study of a civil society organization explicates how harnessing web analytics aids in creative and innovative website development, allowing it to better align with the vision and objectives of the organization. A lack of data literacy skills presents a challenge in using web analytics to effectively "translat[e] measurements into concrete changes" (Fotopoulou & Couldry, 2014, p.243). This case study demonstrates the importance of collective data literacy skills in achieving the goals of an organization.

The Importance of Data Literacy as it Relates to Open Data

Gurstein (2011) posits that because of the digital divide Open Data serves only to make those who have the skills and resources to use the data more powerful, and

further marginalize those who do not. In order to prove this point, he cites the Zanesville lawsuit, wherein publicly accessible data was used to overthrow discriminatory water provision practices. The Zanesville lawsuit reveals the bias of Open Data towards those who already possess necessary skills. Although the citizens of Zanesville were the motivating force behind the lawsuit, the data mashup was created by a firm that specializes in GIS manipulation and the lawyer representing the case is a graduate of Harvard law (Gurstein, 2011). What Gurstein classifies as empowered effective use can be observed where community advocates used data literacy skills they learned in free and and accessible workshops provided by the UCLA Centre for Health Policy Research (CHPR). By using raw data from a survey conducted by the CHPR to visualize a map, community advocates were able to show that an unusually high concentration of asthma cases were present in the area of a proposed truck stop. The map successfully aided the advocates in their argument against the proposed truck stop, and it was never built (Gurstein, 2011). Although both Zanesville and Solano County supported their cause through the use of data literacy skills, the people of Zanesville had to recruit legal and data specialists who possessed the necessary skills, while the people in Solano County were themselves empowered. When citizens are empowered not only with easily accessible information, but with the skills to use that information, there is a possibility to affect real change.

Summary

One of the primary benefits of data literacy is to support and exercise principles of democracy. Accessibility is essential to the use of open government data, but it is ineffective if those who have access do not have sufficient skills in analysis and critical thinking to comprehend and manipulate it. Data literacy of all public citizens is essential to full democratic use of open government. Hence, there is a need for citizens to be data literate. Solutions to this problem thus far include: a focus of teaching data literacy skills to students of higher education; having data literate librarians act as a medium for presenting data in an understandable way; and having data literate individuals curate raw data into easily understandable forms, for example, through the development of applications or APIs. While governments and existing research assert the need for data literacy in Open Data, particularly from governments, there is no consideration of public opinion about data literacy, which calls into question the validity of open government's basic democratic principles.

Methodology

After reviewing the literature, we determined that there was not an existing theoretical framework currently in place to effectively interpret the data we collected; thus, modified grounded theory provided the basis for our analysis. We administered an online survey that ran for twenty-eight days, and included quantitative and qualitative questions derived from the main research questions. We used the content of the responses to develop a theoretical framework. Using a mixed-methods online survey allowed our research team to obtain an understanding of the general public's overarching opinions more easily than could be obtained from other, more concentrated methods. Anonymous online surveys also allow for a larger sample size and are not limited to a specific geographic region. Participants' anonymity may also encourage them to provide more truthful answers.

We used the social media platform Reddit to post the survey link to two relevant, high-traffic subreddits; /r/canada which has over 170,000 Canadian subscribers and /r/samplesize, which is a dedicated subreddit for surveys. We also posted our survey link numerous times to both Twitter--using the hashtags #opendata, #opengov, and #canadianopengovernment--and Facebook, where it was then shared by various Facebook friends. Posters advertising the survey were displayed at all fourteen Halifax Regional Municipality public libraries in an attempt to reach populations who do not use social media. We also circulated the survey through three different company listservs in Alberta and New Brunswick. We encouraged snowball sampling through sharing, retweets and word of mouth. We used multiple methods of promotion in order to reach the largest and most diverse number of participants given the time constraints.

Adhering to the grounded theory method, we analyzed our qualitative data by assigning codes to each response and iteratively comparing them to ascertain common themes, beliefs, and opinions. These comparisons were completed as a group so that each researcher agreed upon the coding choices. The quantitative data was used to supplement the qualitative data by providing demographic and educational context.

Results

Forty-two people completed the survey. Respondents' ages range from 16 to 64 with a median age of 26.5 and 36% possess at least one post-secondary degree. Fifty-five percent have visited Canada's Open Data Portal, as illustrated in Figure 1, and of these respondents 89% have downloaded and used datasets from the website. This

high statistic is due to the high respondent rate of students from a Research Methods course available through Dalhousie.

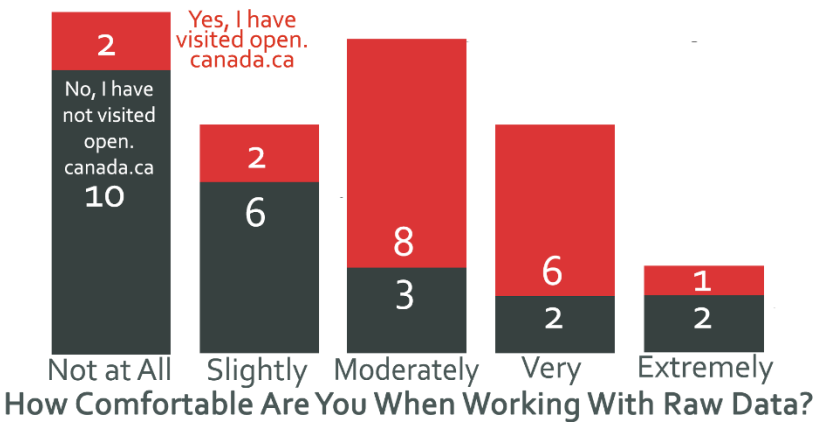

Figure 1: How comfortable Participants are when working with raw data compared to if they have visited Canada's Open Data portal

Forty-three percent of respondents received formal training to use raw data sets, while 40% received no formal training, and 17% taught themselves. The 17% that taught themselves, as well as the 43% that were formally taught, have obtained secondary education at some level as you can see below in Figure 2.

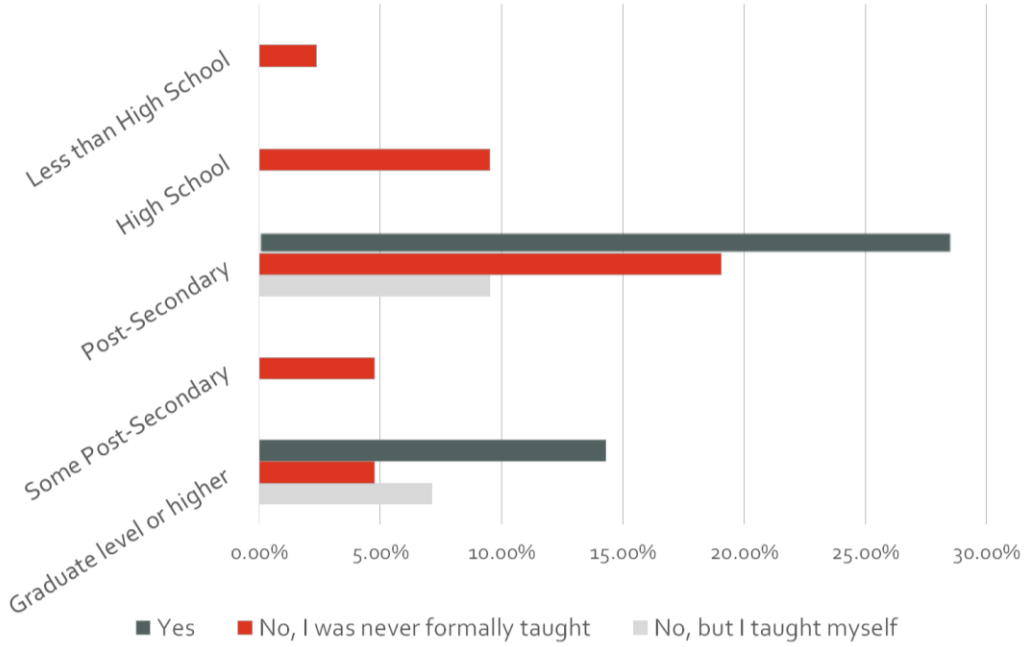

Figure 2: Formal data literacy training by level of education

When asked if Canadians should be taught the skills to manipulate the type of data presented on open.canada.ca, 58% of respondents indicate a belief that teaching is critical to fulfill the Open Government initiative's three principles; 28% of the respondents for this question express that data literacy should not be taught because it is not an important skill or that it is not relevant to many people. Additionally, 14% feel that teaching data literacy, like the Open Data itself, should be optional, yet accessible to all.

In one of the survey questions, we suggested six venues where data literacy skills could be taught. Figure 3, displayed below, shows respondents' preferences about the most appropriate venues for this type of education. The question prompted participants to choose all that applied. Those who chose 'other' suggested community workshops,

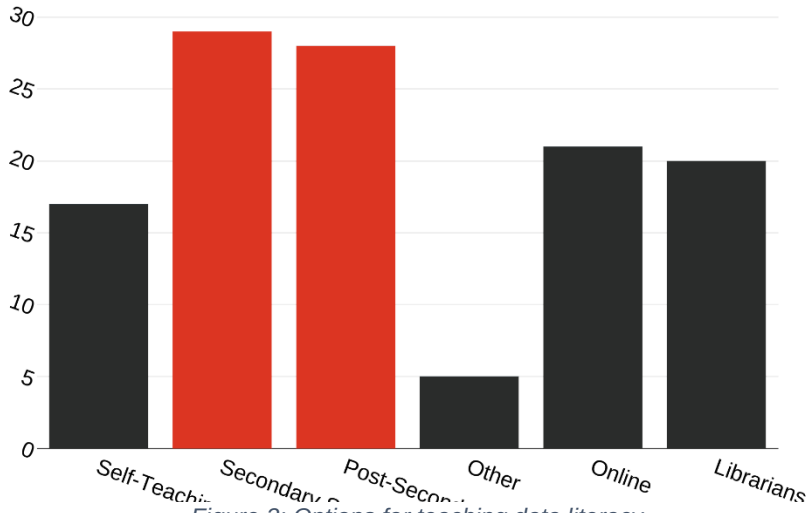

Figure 3: Options for teaching data literacy

friends, family, and government programs.

The majority of respondents indicate that the principles of Canada's Open Data initiative are not entirely fulfilled if members of the Canadian public are not able to understand raw Open Data. Some also state that the public need not be educated in data literacy skills, but that an intermediary should translate raw data into widely comprehensible information. Over half of the responses agreed with the use of professional curation, like the CODE events, as a substitute for understanding the raw data. However, some raise issues with the potential for developers to misrepresent the data. Others recognize the importance of understanding raw data but they understand that people may want to take the path of least resistance when it comes to receiving information.

Discussion

We identified seven major themes from the responses that suggest trends in present public perceptions of Canada's Open Data initiative and the usefulness of Open Data itself with respect to the Canadian public. These themes are outlined below in Table 1 and will be further explained below.

Themes	Concepts	Number of Responses
Impact on individuals/society	-No impact at present -Civic engagement -Personal motivation -Privacy Concerns	39
Impact on government relations/operations	-Government transparency -Government accountability -Tax-related significance -Accessibility	19

Mistrust	-Government mistrust -Data validity	13
Mismanagement	-Open government data initiative mismanagement	10
Data Literacy Divide	n/a	10
Education	-Teaching to fulfill ideals of open data -Affirmative for teaching -Negative for teaching -Optional for teaching	26
Curation	-Positive curation -Negative curation -Both	26

Table 1. Category and concepts coding chart. As discussed in the Methodology, these code were selected based on trends of perceptions found within the qualitative responses. We then used the code to organize our paper.

Impact on Individuals and Society

One of the primary goals of our research study was to determine how Canadians perceive the impact of Canada's Open Data initiative on the lives of individuals and on society. We found there is a variety of attitudes about how Open Data impacts their personal lives, society as a whole, and the relationship between Canadian citizens and their government. Just over half of respondents are optimistic that Open Data can foster a sense of civic engagement by enabling voters to make more informed and responsible decisions and become better educated about how the government functions and allocates resources. Open Data is cited as another means of communication between the public and the Canadian Government and can assist

individuals in becoming more actively engaged in policy decisions. Belying the respondents' perceptions of how Open Data affects civic engagement is the belief that "an informed populace is the foundation of a powerful society" (survey respondent, personal communication, 2015).

In contrast, just over a quarter of respondents feel that Open Data has little to no impact on their personal lives or on society. Among them, six think that Open Data serves no practical value for the general population of Canada, as one wrote "it affects me the way a mild wispy breeze would - I don't really have any feeling of where it came from or where it is going and don't find myself feeling the need to respond to it in any way." Two respondents express that the Open Data initiative has yet to have a remarkable impact because it has not been widely discussed by Canadians. Another indicates that they are unaware of how Open Data could benefit them personally because they already access data from Statistics Canada and from Nova Scotia's Community Counts website. Conversely, Open Data may only impact those who are personally motivated by either a specific interest or are required to use it through their line of work. One respondent states "I believe that Open Data itself won't affect society, but the people who decide to do something with it."

Generally, respondents who have worked with raw datasets are more optimistic about how they can be used to foster civic engagement. Those who have not downloaded or used datasets are more likely to believe that Open Data does not serve much value to the public, perhaps because they are unaware of what datasets are available and what kind of information they contain. This reveals a lack of personal connection between the public and Open Data because they do not see how it can serve them. People also appear to idealize how the tool can help improve civic engagement, but are often unclear about concrete ways in which Open Data can actualize engagement. Whether or not there is a clear idea of what is released onto the Open Data portal may contribute to the spectrum of perceived impacts of Open Data on individuals and society.

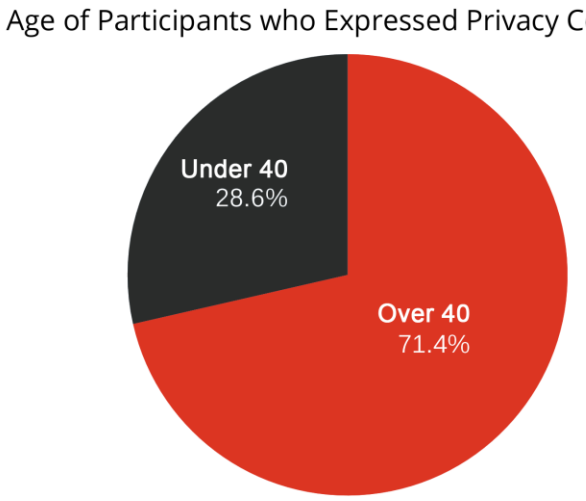

Figure 4: Age of participants who expressed privacy concerns

Seven respondents question how much of their private information is made publicly available on the Open Data Portal, revealing that they are not familiar with what datasets are published and how they can be used. Interestingly, of those who expressed privacy concerns, the majority were over the age of 40, which is visible in Figure 4. One respondent mentions the publicized incident where Jeb Bush publicly disclosed emails containing citizens' private information (Viebeck, 2015), and how beliefs surrounding personal privacy tie into the broader debate over how citizens' personal information can be released without their consent or knowledge. It is possible that the number of survey responses that reveal this anxiety is indicative of this popular social issue. However, most of the respondents who express concerns about their privacy also accept the release of the data providing that it is properly anonymized and redacted. This demonstrates a certain level of support surrounding the Open Government Data initiative and some level of trust in the government's motives with releasing Open Data, despite concerns of privacy.

Impact on Government Relations/Operations

Respondents express their appreciation that the government provides public access to government data in order to further exercise democratic principles, instead of hiding data that could better inform the public. Approximately half of respondents believe in the importance of government transparency and that Open Data is a way to ensure that the government is giving citizens insight into its functions and operations. One respondent says "I find it helpful and comforting to know my country provides this information", while another states "transparency indicates a trustworthy government," suggesting how open government affects the level of trust given by citizens to the

government and its regular operations. This increase in government trust through the release of Open Data is further demonstrated by several expressions of the potential for Open Data to hold the government accountable for their actions.

Those who wrote about government accountability see potential for Open Data to allow the public to monitor the distribution of government funds and, more specifically, tax dollars. One respondent states, "I would . . . use this data to comprehend information about the government and use it to establish if I felt the government were spending money wisely," and another says, "as a tax-payer[sic], I appreciate that the government is making more information publicly available, particularly when the data is in relation to how tax dollars are spent." Responses such as the latter suggest a way that Open Data can assist concerned citizens in ensuring that tax money is being spent in what they believe to be an appropriate and effective manner. Similarly, several respondents believe that if their tax money was being spent to collect the data, then that data should be readily available to citizens. Therefore, Open Data is perceived to be a tool that can be used for citizens to hold the government accountable, but also can potentially motivate the government to hold themselves accountable because they are required to release their raw data.

Mistrust

Some question the motives behind releasing Open Data and the subsequent validity of the data. A subset of these believe that the release of Open Data may be motivated by a desire for the government to improve their reputation. They suggest that the government does not consider whether the public will make use of the data, like one respondent who said "[Open Data] gives the government a pat on the shoulder for apparent 'transparency' but leaves Canadians wondering how this Open Data is being used, manipulated and applied to policies." One respondent also expressed concerns about whether Open Data could be used by terrorist organizations or by corporations to exploit Canadians.

Respondents were also wary about the way that data was collected and how it could affect the validity of published datasets. The eradication of the long-form census calls into question the legitimacy of the published data, as one participant states "without the long-form census, I have tremendous doubts as to the future validity of any other data sets." It was also noted that the government in power could be selective about the type of data it releases as a means to "sway the public to current ideologies." One respondent points out that some of the datasets they had downloaded only included aggregated sums instead of the raw data, which does not align with the Open Data initiative's commitment to completeness of data and the principle of transparency.

These examples of mistrust demonstrate how Open Data's potential can be undermined as the public doubts the integrity and authenticity of the government's intentions.

Open Data Portal Usability and Data Mismanagement

Respondents who have both visited and downloaded datasets from Canada's Open Data Portal mention usability issues with both the website and the available datasets. One highlights the website's poor searching and browsing functions as an impediment to using the datasets, which decreases the likelihood that they would use Open Data in the future. Another respondent elaborates on the usability issue of Open Data by mentioning that the datasets contain very little metadata that is needed to contextualize the meaning of datasets. Searchability and browsing issues are handled more effectively on the American Open Data portal, data.gov, where datasets are sorted into overarching topics to help guide users towards what they are interested in, for example, health, education, and climate-related datasets (Bertot et al, 2014). Within those topics, datasets with high-interest value are presented as highlights, with recently updated datasets located below. A possible solution to these usability problems could be to remodel the website, drawing from more successful portals like data.gov. This may increase the number of citizens who use open government data, making the initiative more likely to fulfill its objectives.

One respondent states "I also think that it will require the contribution of a much wider scope of data that is actually relevant to the average Canadian. That is not to say that the data that is currently available from the GoC [Government of Canada] is not important, but that due to design weaknesses, a lack of promotion, and a seemingly high volume of isolated studies, the Open Data portal is not yet ready to be a success in Canadian society." These perceptions reveal how Canada's Open Data initiative is not likely to impact Canadians when mismanagement occurs through limited portal usability and inconsistent presentations of data.

Data Literacy Divide

Ten respondents indicated that the use of Open Data is potentially limited to those with adequate data literacy skills. One respondent describes a key problem with Canada's Open Data initiative: "Open Data provides the opportunity for Canadians to become better-acquainted with how their country functions. Unfortunately, this knowledge is limited to those who possess the resources to access and interpret the Open Data." Some respondents believe that the majority of citizens will not be able to use the raw data because they do not have the required data literacy skills; therefore,

the public will not likely use the data and the Open Data initiative will only affect a small, skilled portion of the public. If only a small portion of the public can use Open Data, then the initiative's intention of promoting citizen engagement, collaboration, and transparency to a wider range of Canadian citizens will not be achieved--"If the public is unable to understand how to use a tool the government is offering then it is yet to be a tool." The success of the initiative is sorely weakened if citizens are unable to independently interpret Open Data.

Education

Some responses suggest that having the option to access the data via the Open Data Portal is sufficient; thus, teaching data literacy is not required. Others are concerned about the data illiterate portion of the population and how this impacts the success of Open Government. One respondent indicates that "if Canadian society is to benefit from having access to Open Data, the members of that society must be able to actually use it. Knowing that it is there is not enough," and this is echoed by another who states, "it is vital to the success of the program to have teaching the public available." A few of these participants articulate that it is necessary for the public to learn data literacy skills in order to draw their own conclusions from the data and to form unbiased opinions. It was also noted that teaching data literacy skills also could benefit citizens in other areas as they are highly transferable skills. When asked hypothetically about where data literacy should be taught, most respondents specified in high school or post-secondary education.

Not all respondents agree that it is necessary to teach data literacy skills because it is either not a national priority, because citizens would not be interested, or because data literacy is not an essential skill. One respondent expresses that "the average Canadian would never be able to deal with data in this format. Nor should they have to", and six others respond similarly, which suggests that they believe data literacy is not essential to Canadian citizens and it is not the responsibility of the public to obtain these skills. Two respondents indicate that there are more important skills that deserve attention before introducing data literacy into the education curriculum as these other skills would benefit more of the population.

A small segment of responses express the importance of data literacy skills in effectively using open government data but added that they do not believe it to be a feasible expectation for the public, since many people would not be motivated to learn. As a solution, some suggested making optional learning opportunities available for citizens who want to learn how to use Open Data. One respondent suggests creating free college courses to teach citizens about how to use Open Data, stating

that "to effectively use Open Data requires a certain skill and comfort level . . . so there need to be ways, with wide availability, to get those skills."

Curation

The majority of participants are concerned with the general population's ability to understand the raw data. They feel that without an intermediary party to transform raw datasets into a format that is comprehensible to the general public, the data will be too difficult to understand and will discourage people from engaging with it. They believe that curating the data--thus creating easier-to-understand mediums to display the data--ensures that the general public is able to interact with and make optimal use of it. One respondent identifies that they "hope that researchers, journalists, etc. are using Open Data to disseminate information to the broader public about national trends. It is not in a format to be used by the general public so it does need interpretation for use." Another believes that Open Data could be better worked with by researchers who already possess the skills to effectively analyze and synthesize raw data for public consumption.

Eight participants perceive curating data for public use as a detraction from the stated principles of Open Government. One expresses this perception by stating, ". . . this removes the transparency that the government was trying to achieve. Once citizens are interacting with data through an application, it is no longer raw." Other respondents concede this criticism, conveying that this is a form of data curation and that it does not provide full transparency to Canadians. Some participants believe curation is problematic because raw data is not viewed by each individual and the person or organization that curates the Open Data could present a biased interpretation.

Still other responses reveal a more tentative opinion concerning curation. Although curation extends the reach of Open Data and makes it accessible to more citizens, they are wary of curation misrepresenting or subverting the raw data: "the unfortunate truth is that the easier something is to access, the more people will do it. I suppose as long as it doesn't water down the information, I am OK with it."

Limitations

The time that we had to complete our research process was restricted to only three months, which limited our ability to obtain a large sample size that would better represent the Canadian public. We also could not ensure that all citizens were Canadian. We attempted to allow only Canadian citizens to complete the survey by

sharing the survey in the specific Canadian subreddit and placing a disclaimer stating Canadians only, but we cannot be certain that only Canadian citizens completed the survey. The manner in which we circulated our survey lowered the external validity of our results, likely producing a bias toward students in a Master's program at a Canadian university on the East Coast. In particular, we realize that several of the respondents have completed Masters-level work that involved being taught how to use raw datasets from the Open Data Portal to complete an assignment, therefore the number of respondents who are familiar with Open Data and had used raw datasets is most likely disproportionate to the Canadian public. By circulating the survey largely online our respondents may have been more likely to be familiar with many of the concepts we ask about in our survey. We attempted to mitigate this by hanging posters in public libraries.

Additionally, data literacy skills exist along a spectrum. A person can be thought of as data literate if they are able to manipulate CSV files in Microsoft Excel, but one can have an even more complex understanding of data literacy enabling them to manipulate more difficult data like JSON or Geographic Information System (GIS) files. Our study did not address the public's perception of using and understanding these complex file formats even though they are also available on the Open Data Portal because we perceived these file types to require more specialized skills and resources.

Conclusion

Although many of the respondents feel that appathon events like CODE are an innovative and useful way to disseminate raw data without teaching data literacy skills, they also feel that teaching these skills could benefit Canadians and the Open Data initiative. Instead of reassigning resources to teach data literacy in schools as a mandatory skill, workshops that teach citizens how to work with Open Data could be offered as an option alongside CODE events and also be hosted on Canada's Open Data Portal. This would foster more interest in CODE by allowing a wider range of citizens to feel included in the event's goal of encouraging engagement with Open Data. The usability of Canada's Open Data Portal could also be improved by remodeling the website after more successful Open Data portals to make searchability and discoverability easier for users. Additionally, the usefulness of datasets would be enhanced with the addition of more robust metadata that explain how the data were collected and what is included in the datasets.

This research study reveals a variety of differing opinions regarding the impact of Canada's Open Data initiative on Canadian individuals, society, and government

operations. Respondents generally perceived Open Data to have a positive effect on government transparency, civic engagement, and accountability. Others believe that Open Data has not yet made any personal impact nor has it impacted the Canadian public. There are also opposing beliefs concerning whether data literacy skills should be taught to the public. Respondents express that teaching data literacy to citizens should be available if they are motivated to learn. However, the majority feels it is more feasible to have Open Data curated for easier understandability as an alternative to teaching these skills. By conducting this study, we have provided some insight into how Canadians citizens perceive open government data and data literacy, and will help inform future decisions regarding the integration of the Open Data initiative into Canadian society.

References

Andreae, J., & Anderson, E. (2012). Re-conceptualizing Access: The new role of information literacy in post-secondary education. *Communications In Information Literacy, 5*(2). Retrieved from http://www.comminfolit.org/index.php?journal=cil&page=article&op=view&path%5B 5=v5i2p74

Bertot, J. C., Gorham, U., Jaeger, P. T., Sarin, L. C., & Choi, H. (2014). Big data, open government and e-government: Issues, policies and recommendations. *Information Polity: The International Journal Of Government & Democracy In The Information Age, 19*(1/2), 5-16. doi:10.3233/IP-140328

Brito, J. (2008). Hack, mash, & peer: Crowdsourcing government transparency. *Columbia Science and Technology Law Review*, 9. Retrieved from http://surprisinglyfree.com/wpcontent/research/brito-hack-mash-peer.pdf

Couldry, N., & Fotopoulou, A. (2014). Telling the story of the stories. *Information, Communication & Society, 18*(2), 235-249. DOI: 10.1080/1369118X.2014.952317

Davies, T. (2010). *Open data, democracy and public sector reform. A look at open government data use from data.gov.uk* [Partial MSc dissertation]. Retrieved from http://www.opendataimpacts.net/report/wp-content/uploads/2010/08/How-is-open-government-data-being-used-in-practice.pdf

Government of Canada. (2014a). *Canada's action plan on open government 2014-16*. Retrieved from http://open.canada.ca/en/content/canadas-action-plan-open-government-2014-16

Government of Canada. (2014b). *G8 Open Data Charter*. Retrieved from http://www.international.gc.ca/g8/open_data_charter charte_du_g8_sur_les_donnees_ouvertes.aspx?lang=eng

Gurstein, M. (2011). Open Data: Empowering the empowered or effective data use for everyone? *First Monday, 16*(2). doi:10.5210/fm.v16i2.3316

Hyams, E., Martinez-Uribe, L., & Macdonald, S. (2008). Data librarianship: a gap in the market. *CILIP Update*, *7*(6), 20-21. Retrieved from *https://www.era.lib.ed.ac.uk/handle/1842/2499*

Jaeger, P.T., & Bertot, J.C. (2010) Transparency and technological change: ensuring equal and sustained public access to government information. *Government Information Quarterly, 27*(4), p. 371-376. doi:10.1016/j.giq.2010.05.003

Johnson, J.A. (2014). From Open Data to information justice. *Ethics and Information Policy, 16*(4), p. 263-274. doi:10.1007/s10676-014-9351-8

Kassen, M. (2013). A promising phenomenon of Open Data: A case study of the Chicago Open Data project. *Government Information Quarterly, 30*(4), 508-513. doi.org/10.1016/j.giq.2013.05.012.

Martin, C. (2014). Barriers to the Open Government Data agenda: Taking a multi-level perspective. *Policy & Internet, 6*(3), 217-240.

Open Government Partnership. (2014). *About: What is the open government partnership?* Retrieved from http://www.opengovpartnership.org/about

Open Government Partnership. (n.d.). *Participating countries*. Retrieved from http://www.opengovpartnership.org/countries

Ren, G. J., & Glissmann, S. (2012, September 9-11). *Identifying information assets for Open Data: The role of business architecture and information quality*. Paper presented at the 14th International Conference on Commerce and Enterprise Computing, Hangzhou, China. doi:10.1109/CEC.2012.23

Schield, M. (2004). Information literacy, statistical literacy and data literacy. *IASSIST Quarterly, 28*(2/3), 6-11.

Stephenson, E., Caravello, P.S. (2007). Incorporating data literacy into undergraduate information literacy programs in the social sciences: A pilot project. *Reference Services Review, 35*(4), 525 - 540.

Viebeck, E. (2015). Jeb Bush email dump include social security numbers. *The Hills.*

Retrieved from http://thehill.com/policy/cybersecurity/232324-jeb-bush-email-dump-includes-social-security-numbers

Veljković, N., Bogdanović-Dinić, S., & Stoimenov, L. (2014). Benchmarking open government: An Open Data perspective. *Government Information Quarterly, 31*(2), 278-290. doi.org/10.1016/j.giq.2013.10.011

Appendix A

1. Please select your age from the dropdown list.
2. Please select the type of location that best describes where you live.
    a. Urban
    b. Suburban
    c. Rural
3. How often do you use the internet?
    a. Very frequently
    b. Frequently
    c. Occasionally
    d. Rarely
    e. Very rarely
4. What is the highest level of education you have completed? Please check one.
    a. Less than High School
    b. High School
    c. Some Post-Secondary
    d. Post-Secondary
    e. Graduate level or higher
5. How familiar are you with the concept of Open Data?
    a. Very Familiar
    b. Somewhat familiar
    c. Not very familiar
    d. Not at all familiar
6. How familiar are you with Canada's Open Government Initiative?
    a. Very Familiar
    b. Somewhat familiar
    c. Not very familiar
    d. Not at all familiar
7. Have you ever visited open.canada.ca?
    a. Yes
    b. No

8. If yes, have you ever dowloaded any of the datasets provided by open.canada.ca?
    a. Yes
    b. No
9. Have you ever attempted to use datasets provided by open.canada.ca?
    a. Yes
    b. No

Canada's commitment to open government is part of the federal government's efforts to foster greater openness and accountability, to provide Canadians with more opportunities to learn about and participate in government, to drive innovation and economic opportunities for all Canadians and, at the same time, create a more cost effective, efficient and responsive government.

We define [Open Data] as making raw data available in machine-readable formats to citizens, governments, not-for-profit and private sector organizations to leverage it in innovative and value-added ways. There are two commitments in this stream. (http://open.canada.ca/en/canadas-action-plan-open-government)

1. Based on this definition of Open Data, how do you think it affects you personally? Please explain your thoughts below.
2. How do you think Open Data affects Canadian society? Please explain your thoughts below.
3. How comfortable are you with working with sets of raw data like those displayed below?
    a. Very Comfortable
    b. Somewhat Comfortable
    c. Not Very Comfortable
    d. Not at All Comfortable

```
PRESTATIONS D'INVALIDITÉ SELON LA CAUSE D'INVALIDITÉ ET L'ÂGE,,,,,,,,,,
,,,,,Age of Beneficiary / Âge du bénéficiaire,,,,,
Period / Période,Classes of Principal Diagnosis / Causes principales d'invalidité,,-30,30-34,35-39,40-44
,,,,,,,,,,
Dec. / déc. 2011,Infectious and Parasitic Diseases / Maladies infectieuses et parasitaires,,10,30,87,269
Dec. / déc. 2011,Neoplasms / Tumeurs,,148,309,620,"1,111","2,307","3,799","5,598","6,506"
Dec. / déc. 2011,"Endocrine, Nutritional and Metabolic Diseases, and Immunity Disorders / Maladies endoc
Dec. / déc. 2011,Diseases of the Blood and Blood-Forming Organs / Maladies du sang et des organes hématc
Dec. / déc. 2011,Mental Disorders / Troubles mentaux,,724,"2,051","4,014","7,448","14,234","19,731","23,
Dec. / déc. 2011,Diseases of the Nervous System and Sense Organs / Maladies du système nerveux et des or
Dec. / déc. 2011,Diseases of the Circulatory System / Maladies de l'appareil circulatoire,,49,154,324,77
Dec. / déc. 2011,Diseases of the Respiratory System / Maladies de l'appareil respiratoire,,X,X,54,131,35
Dec. / déc. 2011,Diseases of the Digestive System / Maladies de l'appareil digestif,,39,128,262,535,970,
Dec. / déc. 2011,Diseases of the Genitourinary System/Maladies des organes  génito-urinaires,,29,103,217
Dec. / déc. 2011,"Complications of Pregnancy, Childbirth and Puerperium / Complications de la grossesse,
Dec. / déc. 2011,Diseases of the Skin and Subcutaneous Tissue / Maladies de la peau et du tissu cellulai
```

		Age of Beneficiary / Âge du bénéficiaire						
Period / Période	Classes of Principal Diagnosis / Causes principales d'invali	-30	30-34	35-39	40-44	45-49	50-54	55-59
Dec. / déc. 2011	Infectious and Parasitic Diseases / Maladies infectieuses e	10	30	87	269	775	952	1,222
Dec. / déc. 2011	Neoplasms / Tumeurs	148	309	620	1,111	2,307	3,799	5,598
Dec. / déc. 2011	Endocrine, Nutritional and Metabolic Diseases, and Immu	54	125	237	488	916	1,540	2,321
Dec. / déc. 2011	Diseases of the Blood and Blood-Forming Organs / Maladi	X	X	30	33	65	100	134
Dec. / déc. 2011	Mental Disorders / Troubles mentaux	724	2,051	4,014	7,448	14,234	19,731	23,946
Dec. / déc. 2011	Diseases of the Nervous System and Sense Organs / Malac	236	632	1,442	2,853	5,284	7,683	9,916
Dec. / déc. 2011	Diseases of the Circulatory System / Maladies de l'apparei	49	154	324	775	1,785	3,868	7,262
Dec. / déc. 2011	Diseases of the Respiratory System / Maladies de l'appare	X	X	54	131	359	820	1,642
Dec. / déc. 2011	Diseases of the Digestive System / Maladies de l'appareil	39	128	262	535	970	1,470	1,829
Dec. / déc. 2011	Diseases of the Genitourinary System/Maladies des organ	29	103	217	336	575	859	1,032

4. Were you ever formally taught how to work with raw data sets like those pictured above?
    a. No, I was never formally taught
    b. Yes
    c. No, I taught myself
5. In your opinion, do you think that that all Canadians should be taught how to work with Open Data? Please explain why or why not below:
6. Who do you think should teach these skills: Choose all that apply:
    a. Self teaching
    b. Librarians
    c. Secondary schools (Junior High School or High School)
    d. Post secondary
    e. Other: _____
    f. I don't think these skills should be taught
7. The Government of Canada publishes Open Data to increase government transparency, accountability to voters, and encourage collaboration between

users of the data. How important is teaching the public to use Open Data to accomplish these goals? Please explain your thoughts below:

8. The Canadian Government hosts a "hackathon" event where developers come together, using Open Data to create applications that are easier to work with. These applications are promoted to citizens to use instead of raw data. As a consumer of this information, please share your opinion about this strategy.