



# The Benefits of Open Data



Royal Roads  
UNIVERSITY

Steve Reitano

[Beautifuldata.ca](http://Beautifuldata.ca)

May 24, 2013



## **An Organizational Management Project for public organizations**

# **The Benefits of Open Data**

**Royal Roads University**

Submitted to: Don Caplan

Steve Reitano, 323543

OMP

May 24, 2013

Word Count: 13,880

## **Acknowledgements**

The author acknowledges the participation of Treasury Board of Canada in this OMP. The organization's willingness to be a learning laboratory for the purpose of this project in the midst of a busy agenda is sincerely appreciated.

This research project was possible with the support of many individuals who offered their guidance and assistance which enriched the project. I would like to thank everyone that participated in the survey and interviewees and I would like to acknowledge and thank the following individuals:

Don Caplan, MBA Major Projects Manager and Deborah Wickins for their guidance throughout the OMP process;

The academic staff with the faculty of Business at Royal Roads University, especially the librarians for their assistance with secondary research and Lynnette Kissoon for her guidance and patience;

To my MBA cohort and teammates for sharing your thoughts, insights and ideas;

On a personal note I would also like to thank my family and friends for their support, and I am especially thankful to my daughter for her sacrifices which enabled my academic pursuits.

## Glossary of Terms

<b>Aggregating</b>	to gather two or more datasets together to form one; sometimes referred to as "combining" or "mashing up" data
<b>API</b>	Application Programming Interface – a middle tier component used to communicate with the data from a user interface or front end application
<b>Big data</b>	a dataset too large to process with traditional on-hand database management tools
<b>Data</b>	quantitative values represented in a structure used to create information (e.g., datasets)
<b>Dataset</b>	an identifiable collection of data
<b>Comma-delimited file</b>	a collection of data in a plain text file, presented in tabular form where fields are separated by commas
<b>Crowdsourcing</b>	to solicit contributions from a large group in order to obtain needed services, ideas, or content
<b>Dissemination</b>	the ability to access and use information
<b>Open Data</b>	structured electronic information in a machine-readable format that is accessible and available for use or re-use without any copyright restriction
<b>Open Source</b>	free software with access to source code developed in a public, collaborative manner
<b>Scraping</b>	removing or extracting data from web pages

## Table of Contents

Acknowledgements.....	4
Glossary of Terms.....	5
Table of Contents.....	6
Executive Summary .....	8
Introduction .....	10
Open Data .....	10
Organization .....	10
Industry.....	11
Project Purpose.....	11
Research Methodology .....	12
Primary Research.....	12
Literature Review and Secondary Research.....	12
Design .....	13
Structure .....	13
Samples .....	14
Data Analysis .....	15
Research Findings – Literature Review .....	17
Information Management .....	17
Value of Data.....	18
Culture of Openness .....	21
Research Findings – Secondary Research .....	23
Information Management .....	23
Standards .....	24
Recognized Benefits .....	24
Research Findings – Primary Research .....	26
Online Survey.....	26
Interviews .....	32
Focus Group Session.....	34
Analysis and Discussion .....	35
Availability of Data.....	36
Standards .....	39
Dynamic Data.....	41
Scenario Analysis.....	44

Externalities .....	48
Recommendations .....	50
Availability of Data .....	50
Standards .....	51
Dynamic Data .....	52
Implementation Steps .....	53
Conclusion .....	54
Bibliography .....	56
Appendices .....	62
Appendix 1: Survey Questions .....	62
Appendix 2: Interview Questions for Internal Stakeholders.....	65
Appendix 3: Focus Group Questions.....	65
Appendix 4: Interview Questions for External Stakeholders.....	66
Appendix 5: Initiatives and Projects Utilizing Open Data.....	67
Appendix 6: Statistical Data from Survey Participants .....	68

## Executive Summary

With a goal for creating transparency, publishing open data (OD) has become an increasing trend among governments over the past few years.

OD refers to information that can be used by anyone for any purpose and at no cost. It can take on many forms, but in order to be considered open it must be presented in a machine-readable format and available through the Internet with licensing agreements. This makes it adaptable and easier to analyze and combine with other data, which in turn provides greater usefulness and value. In addition, openly accessible data provides information, knowledge and wisdom that have the potential for a number of social, economic and environmental benefits. The major users of OD include the academic community, the private sector and governments.

The Government of Canada (GC) needs to extend its OD initiative beyond its initial three year commitment. Unnecessary cuts in the program could affect the availability and quality of Public Sector Information (PSI) while needless expenditures could adversely affect the government's perception with citizens. This paper aims to give public organizations research-based information and recommendations by determining the benefits of OD. More specifically, it aims to answer the question: What are the social, economic and environmental benefits of shared data from public organizations?

The analyses suggest that one of the great benefits of publishing OD lies within research and development and is highly likely to create and support innovation. For example, the no-cost dissemination of data by Statistics Canada since 2012 has provided social and economic benefits for the support of research initiatives and not-for-profit organizations. Additionally, private companies often re-publish OD in the form of web or mobile applications, which have strong benefits for consumers.

Benefits of OD are provided through several mechanisms including public service delivery that informs and allows citizens to connect with their governments and supports industries by reducing the cost of data and adding opportunities for new digital products and services. In addition, OD also provides insights into trends and markets that contributed to economic growth and jobs. Studies have shown that the value of PSI can provide considerable benefits to help government organizations manage internal operations more effectively. The re-use of information allows for internal efficiencies and the prevention of fraudulent operational activities.



Issues surrounding the publishing of OD include the proper management of information through its lifecycle, which allows for effective re-use of information. The GC has established new operating standards for their OD portal. However, compliancy with policies and directives are preventing the dissemination of data, and current mechanisms for managing PSI are causing some internal disruptions.

Research has demonstrated that there is a shortage in openly available data. This lack of OD has created and fostered a culture of scrapers, or individuals who take data from public web sites and openly re-use it. This re-use of data is technically not permitted as it occurs in the absence of appropriate licensing.

The potential benefits from combining data are dependent on standards. The value of data decreases if organizations utilize different standards. The efforts needed to amalgamate data rises when standards are missing or incompatible, which in turn prevents data from being re-used. In addition, tools and guidelines are also needed to effectively amalgamate OD. Internal practices of governments were never designed for the dissemination of openly available information.

Recommendations can be made to address the issues facing the GC. The need for publishing OD is rising and departments and agencies will need guidelines and standards with clear instructions and criteria for publishing data. A lack of consistency, defined standards and procedures are preventing the aggregation of datasets.

Furthermore, the GC needs to launch their Directive on Open Government with a strong mandate to publish all data that meets the criteria. In order to create and maintain a sense of urgency and compliancy, a top-down approach is needed. A communication campaign promoting the importance of the directive should be led by the highest level of senior official within the GC. Failure to create a sense of urgency will minimize the importance and credibility of OD. In addition, the directive must mandate them to publish all data that meets the criteria for publishing openly.

The benefit derived from OD starts with the release of all available data, included data from public web sites. The GC should set a concrete goal to convert data from public web sites into open and available formats. The Treasury Board of Canada Secretariat (TBS) needs to be the catalyst for establishing a mechanism and cycle for the continuous flow of information, thereby creating public accountability and a more cost-effective, transparent, efficient and responsive government.

## Introduction

The Internet has changed how organizations operate and people live. One of the most important changes that online connectivity has introduced is the advent of Open Data (OD), a new technological trend that is transforming access to information. Every day, countless people are able to find new information that is open, accessible and re-useable in ways that pre-digital systems simply could not facilitate.

## Open Data

The Open Data Institute defines OD as information that can be used by anyone for any purpose and at no cost.<sup>1</sup> OD is information that is available electronically and in a machine-readable format such as Extensible Markup Language (XML), Comma-Separated Values (CSV) and dataset. In most cases, OD is made available through the Internet and is free to be used and re-used without any copyright restriction. This is made possible through the use of license agreements that allow individuals to openly use and re-use data. For effective re-use and whenever possible, OD should be time stamped, available and accessible in an open format using a non-proprietary or open source software, accompanied by useful metadata and geospatial information, which provides innovative and interactive opportunities to aggregate data with maps.

## Organization

As part of the efforts for driving innovation and economic opportunities, the Government of Canada (GC) launched their online Open Data Portal in March 2011 to centralize freely available data. In April 2012 the GC joined the International Open Government Partnership (OGP) and endorsed the core principles of the multilateral initiative for Open Government Data (OGD). Today the TBS is responsible for the governance, including guidelines and policies, applicable to data, and since its launch the portal has increased its federal department participation and available datasets. The GC is expanding its Open Government initiatives along three main streams: Open Information for the release of information on government activities, Open Data for making information available in a machine-readable format, and Open Dialogue which gives citizens the opportunity to dialogue with its government about policies and

---

<sup>1</sup> Knowledge for everyone, Open Data Institute, Last modified November 30, 2012, <http://www.theodi.org/>.

priorities.<sup>2</sup> In addition, TB is working on the release of a new and common Open Government Licence for OD and developing a new Directive on Open Government.<sup>3</sup>

## Industry

Today's modern technology is providing users' newfound flexibility to access more information than ever before. Through efforts such as those undertaken by the GC, it is evident that the proliferation of content available on the Internet is empowering citizens and is changing governments. Open Government is part of an effort to use these technologies to make government more open and accessible.

Challenges surrounding the dissemination and release of OD can sometimes prevent or limit its benefits. Publishing OD may involve a number of labour- and time-intensive tasks, such as changing data formats, making sure that information is up-to-date, aligning datasets with existing licenses and meeting criteria for releasing information that could be sensitive. Despite these possible obstacles, the process of publishing data is critical for governments because it is the first step to engaging users and demonstrating transparency. In addition, it can foster internal changes to organizations, such as the implementation of new standards and technologies, and/or changes to organizational and cultural behaviors. It also begins the interactive process needed to validate and achieve a level of quality data.

The success of OD is dependent on the quality of the information. In order to assure that OD is of high quality, related information must be made available. Metadata about the datasets can provide users with information about the data and allow for ease of use by improving public understanding and incorporating other information, such as geographical information.

## Project Purpose

This paper will focus on examining the benefits and challenges of publishing OD for government organizations. It will attempt to identify advantages of OD published from the GC Open Data Portal. It is presumed that open and accessible data offers multiple benefits, including improved openness and accountability, as well as an increase in innovation and economic growth. This paper aims to help public organizations make sound and informed decisions for extending their OD initiatives by determining the social, economic and environmental benefits of shared data from public organizations, thereby creating a more cost-effective, transparent, efficient and responsive government.

---

<sup>2</sup> Open Data, Government of Canada, Last modified October 24, 2012, <http://www.data.gc.ca/>.

<sup>3</sup> Open Data, Government of Canada.

## **Research Methodology**

In order to identify the social, economic and environmental benefits of shared data from public organizations, an analysis of both primary and secondary research comprised mostly of quantitative information was conducted. The original intent of this research was to conduct a cost-benefit analysis, but after further assessment, the need to determine the return on investment for governments became a moot point since most organizations are mandated to share Public Sector Information (PSI). Identifying costs of publishing OD will not help identify benefits. The focus of the analysis will be to determine and categorize the various benefits that OD can provide. Similar to qualitative evaluation methods used to analyse human services programs, this research methodology will attempt to identify social benefits of public service programs.

Open and accessible data can sometimes include OD published by private organizations. For the purpose of this paper, references to OD or primary data will include data accessible from public organizations while secondary data originates from external parties. In addition, references to internal stakeholders consist of individuals from a public organization while external stakeholders refer to individuals representing academic, private, for-profit or not-for-profit organizations or associations.

## **Primary Research**

A survey targeted to individuals involved or working with OD was conducted to help establish economic and social benefits. Individual interviews with representatives from public organizations were conducted to identify challenges with publishing OD while interviews with external experts involved or working with OD were also conducted to help identify challenges and benefits of using OD. Furthermore, a focus group session was also conducted with external experts involved or working with OD.

## **Literature Review and Secondary Research**

An initial review of academic literature was carried out, focusing on research knowledge of public organizations, shared data information practices, technology trends and changes related to public sector information. A review of secondary research was also conducted on government transparency and public sector information including open data, open access, open source, open government and trends to open and accessible information.

## Design

Information gathered for this research is comprised mostly of qualitative information. This required a research method that was qualitative in nature and capable of understanding the phenomenon and the behaviors towards the OD movement.<sup>4</sup> This paper uses a scenario analysis to determine the likelihood of specific events occurring. In addition, this methodology can only demonstrate the benefits linked to the data being accessible and not the quality or value of the data itself.

There are limitations to the design employed in the current study because a scenario analysis attempts to determine plausible outcomes. Furthermore, sample restrictions from a limited scale of the population and few longitudinal measurements from different levels of governments can also affect the outcome. Despite its limitations, a scenario analysis was used because it analyses future events by considering possible alternatives, which determines the impact and benefits of OD.

## Structure

The three methods used for collecting primary data included: a survey of potential individuals involved with Open Data, a focus group session with external participants and interviews with both internal and external informants. The survey and interviews included quantitative questions that focused on dollar figures and performance indicators and qualitative questions focused on opinions, viewpoints and trends. Participants completed the survey within 20 minutes; the focus group session and most interviews lasted one hour.

Information was first collected and analysed with the survey. The outcomes of these analyses were then validated with the focus group session and interviews. The survey was presented in an online format using the "SurveyMonkey" platform; it included 22 questions, which are listed in Appendix 1. The survey was distributed to 42 specific groups or organizations working with OD, and invitations to participate in the survey were sent to all individuals involved with these groups. The survey was open for a period of seven weeks and 123 participants responded. The results were collected and analysed in order to inform further research.

The interviews included predefined, open-ended questions that allowed for improvised responses which generated additional questions. A list of the predefined questions can be found in Appendix 2 and Appendix 4. This unstructured format was ideal for qualitative questions and

---

<sup>4</sup> Pervez Ghauri and Kjell Gronhaug, *Research Methods in Business Studies*, Fourth Edition, Pearson Education Inc. (Essex: England), 2010, 105.

allowed further discovery of individual viewpoints.<sup>5</sup> Interviews were conducted face-to-face or by phone; each participant signed a consent form and hand written notes were taken. Five interviews were completed with internal stakeholders while five additional interviews were conducted with external stakeholders.

The objective of the focus group session was to generate a conversation and a diverse set of options derived from specific challenges with OD. The session was conducted face-to-face and included three participants. The process consisted of exploring ideas in a divergent manner consistent with the Creative Problem Solving (CPS) model for generating novel ideas to address specific challenges.<sup>6</sup> Two challenges were presented to the group in the form of a question; these questions can be found in Appendix 3. The informal setting and the divergent conversations allowed participants to share stories and engage in further discussions on the topic of OD.

## **Samples**

The topic of open and accessible government information is a profound issue for several advocacy groups, which included a wide range of external stakeholders.<sup>7</sup> Using qualitative research methodology, these non-random samples of the population were not randomly selected, and in turn provided useful insight into the phenomenon of OD.<sup>8</sup>

Specific groups of participants were targeted for each method of collecting primary data. Participants for the survey consisted of individuals involved or working with OD from any country and sector including academic, public, private, for-profit or not-for-profit organizations. Participants of the focus group also included individuals involved or working with OD but only from academic, private, for-profit and not-for-profit organizations. Interviews were conducted with two types of participants. The first included specific individuals from public organizations who were responsible for publishing OD. The second were individuals utilizing OD from private, for-profit and not-for-profit organizations as showed in the following diagram.

---

<sup>5</sup> Ghauri, *Research Methods in Business Studies*, 126.

<sup>6</sup> Gerard J. Puccio, Marie Mance and Mary C. Murdock, *Creative Leadership: Skills That Drive Change*, Second Edition, (Thousand Oaks, California: Sage Publications Inc., 2011), 171.

<sup>7</sup> Margaret Munro, "Information watchdog to probe feds' alleged muzzling of scientists," *Edmonton Journal*, April 2, 2013, <http://www2.canada.com/edmontonjournal/news/story.html?id=915b635d-e5b8-40fa-8902-dcd3a0217ffa&p=1>.

<sup>8</sup> Ghauri, *Research Methods in Business Studies*, 141.

Collection method	Responsibility	Sector	Level
Survey	Involved or working with Open Data	Academic, public, private, for-profit or not-for-profit organization or associations	Global
Focus group session	Involved or working with Open Data	Academic, private, for-profit or not-for-profit organization or associations	Federal, Provincial or Municipal
Interview (Internal)	Responsible for publishing Open Data	Public organizations	Federal, Provincial or Municipal
Interview (External)	Involved or working with Open Data	Academic, private, for-profit or not-for-profit organization or associations	Federal, Provincial or Municipal

## Data Analysis

In order to determine plausible outcomes derived from OD, a scenario analysis was conducted. A scenario analysis identifies the likelihood of a specific event occurring, and in turn determines its impact on the benefits of OD.<sup>9</sup> This provided a deeper examination of obvious issues and patterns related to OD – increasing the importance of social issues.<sup>10</sup>

Data source	Collection method	Technique	Approach
<ul style="list-style-type: none"> <li><u>Primary sources</u> Front-line stakeholders from public and private sector</li> <li><u>Secondary sources</u> Existing academic literature</li> </ul>	<ul style="list-style-type: none"> <li>Online survey</li> <li>Interviews</li> <li>Focus group session</li> <li>Review of relevant document literature and secondary research</li> </ul>	<ul style="list-style-type: none"> <li>Scenario analysis</li> </ul>	<ul style="list-style-type: none"> <li>Contingent valuation method</li> </ul>

<sup>9</sup> Constance Gustke, "All the Options." Conference Board Review 46, no. 1 (Jan, 2009): 18-23, <https://ezproxy.royalroads.ca/login?url=http://search.ebscohost.com.ezproxy.royalroads.ca/login.aspx?direct=true&db=buh&AN=36615374>.

<sup>10</sup> Jay Ogilvy and Peter Schwartz, "Plotting Your Scenarios," Global Business Network, 1998. 9 [http://www.gbn.com/articles/pdfs/gbn\\_Plotting%20Scenarios%20new.pdf](http://www.gbn.com/articles/pdfs/gbn_Plotting%20Scenarios%20new.pdf).

The analytic tools used to analyse the data collected from the survey included the online tool available from “SurveyMonkey” and Microsoft Excel. Data was transferred from the survey to Excel spreadsheets allowing for further analysis by aggregating data results from several sources.

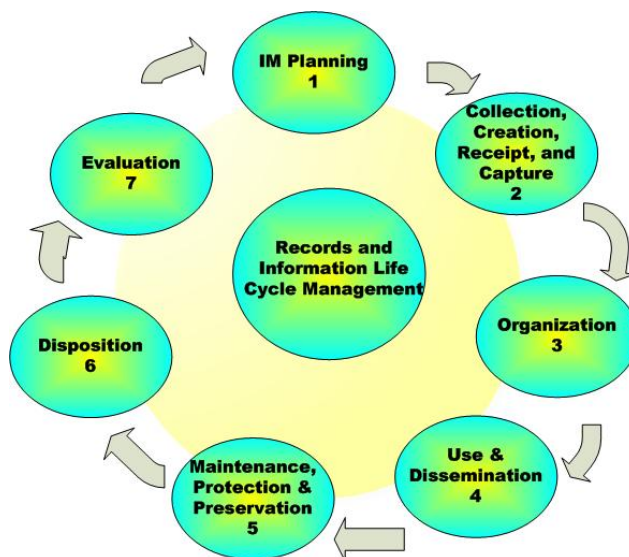


## Research Findings – Literature Review

The following is a brief summary of the literature review for this paper. In the context of the research question, the review touches on elements of effective Information Management (IM), data management and the value of data, organizational culture and the benefits of transparency from sharing Public Sector Information (PSI).

### Information Management

The principles of IM include a series of steps for managing information through a predefined lifecycle. These steps include the planning, capturing, organizing, dissemination, preserving, disposition and evaluating of information.<sup>11</sup> The Government of Canada (GC) has a well-defined IM framework which is part of a suite of policies and directives.<sup>12</sup> The IM lifecycle is used to implement and improve IM initiatives and best practices within each GC department and agency. Proper management of information through its lifecycle allows for effective access to information making it available for effective decision making and re-use. The following diagram outlines the IM lifecycle.<sup>13</sup>



<sup>11</sup> Records and Information Life Cycle Management, Library and Archives Canada. Last modified April 19, 2006, <http://www.collectionscanada.gc.ca/007/002/007002-2012-e.html>.

<sup>12</sup> Records and Information Life Cycle Management, Library and Archives Canada.

<sup>13</sup> Ibid,

With the emergence of OD, the GC has established new operating standards for their OD portal, influenced strongly by the dissemination step of the IM lifecycle. These standards include completeness of data, primary source data, timeliness, ease of access and machine-readable formats, non-discrimination, use of common standards, available licensing agreements, freedom of use, and permanence.<sup>14</sup> These principles allow government organizations to share information in ways that were not initially intended, such as dissemination to the public. To effectively share OD and create a new type of relationship with the public, GC departments and agencies need to consider these principles when managing information.<sup>15</sup>

With new guiding principles and outcomes for data use, the dissemination of OD requires a new cycle for information management.<sup>16</sup> The Center for Technology in Government at the University at Albany in New York has taken a heuristic approach to presenting the flow of data sources related to Open Government Data. It includes an iterative process that comprises of stakeholders with specific roles and both primary and secondary sources of data.<sup>17</sup> The center defines OD published on government portals as primary data sources, managed by primary data resources. Non-government data sources such as geo-coded data and third-party data are considered secondary data sources managed by secondary data resources.<sup>18</sup> The combination and consolidation of these sources is where the potential for innovation and economic growth lies; this potential is currently largely untapped.<sup>19</sup> Aggregating primary and secondary data sources brings in external stakeholders and in turn creates added value to data.

## Value of Data

Information and data can sometimes be used interchangeably, but there is a critical difference.<sup>20</sup> Data is considered raw and factual and it has little significance beyond itself.<sup>21</sup> Information is the

---

<sup>14</sup> Operating Principles for data.gc.ca, Government of Canada, Last modified October 24, 2012, <http://www.data.gc.ca/default.asp?lang=En&n=89BD0D43-1>.

<sup>15</sup> Natalie C. Helbig et al. "The Dynamics of Opening Government Data," Center for Technology in Government, (2012), 7, <http://www.ctg.albany.edu/>.

<sup>16</sup> Natalie C. Helbig et al. "The Dynamics of Opening Government Data, 12-13.

<sup>17</sup> Ibid., 12.

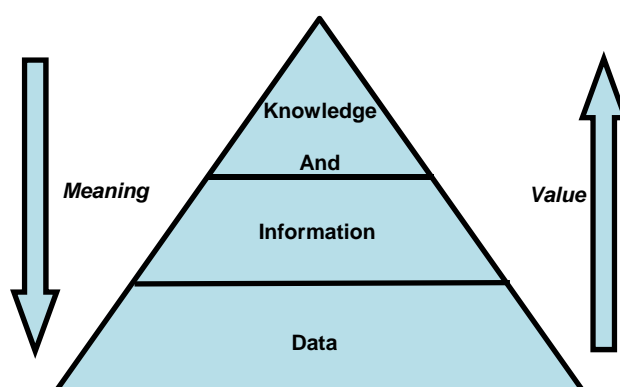
<sup>18</sup> Ibid.

<sup>19</sup> Ibid.

<sup>20</sup> Tim Davies, "Open data, democracy and public sector reform," A look at open government data use from data.gov.uk. (2010) 12 <http://practicalparticipation.co.uk/odi/report/wp-content/uploads/2010/08/How-is-open-governmentdata-being-used-in-practice.pdf>.

<sup>21</sup> Jennifer Rowley, "The wisdom hierarchy: representations of the DIKW hierarchy," Journal of Information Science 33, (2007), 166, <http://jis.sagepub.com.ezproxy.royalroads.ca/content/33/2/163.refs>.

value we extract from data, which then becomes knowledge and wisdom.<sup>22</sup> For example, data can be represented digitally as tabular rows and columns, which is a structured format that can take many forms. Information on the other hand can be derived from the value that was determined by the data; this requires human input and consideration, and brings forward knowledge and wisdom.<sup>23</sup> Furthermore, on the scale where data, information, knowledge and wisdom relate to each other, value is gained from the knowledge and wisdom that is initially taken from the meaning of data.<sup>24</sup> This paper discusses the benefits of data and more specifically openly access data that provide information, knowledge and wisdom. The relationship between these concepts is exemplified in the following diagram.<sup>25</sup>



### Public Sector Information (PSI)

Public Sector Information (PSI) includes all of the information collected by governments and can be “otherwise known as Open Government Data.”<sup>26</sup> PSI can encompass several domains of information including: business or administrative, geographic, legal, meteorological and transportation, plus social or statistical data. Due its breadth and scope, PSI has the potential for a number of economic benefits and advantages.<sup>27</sup>

---

<sup>22</sup> Rowley, “The wisdom hierarchy,” 166.

<sup>23</sup> Ibid..

<sup>24</sup> Ibid., 167.

<sup>25</sup> Ibid.

<sup>26</sup> Naomi Lillie, “The new PSI Directive – as good as it Seems?” Open Knowledge Foundation, Last modified April 19, 2013, <http://blog.okfn.org/2013/04/19/the-new-psi-directive-as-good-as-it-seems/>.

<sup>27</sup> Graham Vickery, “Review of recent studies on PSI re-use and related market developments.” Information Economics, Paris (2011). 3. [http://www.umic.pt/images/stories/publicacoes6/psi\\_final\\_version\\_formatted-1.pdf](http://www.umic.pt/images/stories/publicacoes6/psi_final_version_formatted-1.pdf).

A 2006 study measuring the European Union plus Norway (EU25) PSI market size estimated that it was worth EUR 27 billion.<sup>28</sup> This equates to approximately 0.25% of their total aggregated GDP. A subsequent review of the 2006 study was conducted in 2011 on the same basis with similar outcomes. This new study for the 27 members of the European Union (EU27) demonstrated a rapid growth of approximately 7% which equated to a PSI market size of EUR 28 billion for 2008 and EUR 32 billion for 2010. PSI can be used in a wide variety of applications to innovate several goods and services. The aggregation of PSI with secondary data adds “further economic and social benefit to the EU27 economy.”<sup>29</sup> In addition, this study also suggests that removing underlying barriers preventing access to data could lead to gains of 10-40% in the geospatial sector alone. Furthermore, if citizens can save two hours per year with more rapid and comprehensive access to public information it would be worth EUR 1.4 billion per year.<sup>30</sup> The European Commission believes that PSI holds a significant amount of potential and that the raw material from PSI can drive innovation and economic activities if open to the public sector.<sup>31</sup>

The benefits of PSI are not limited to primary data. Additional savings could be attained if public organizations were to leverage secondary data creatively with PSI. In the case of the US healthcare system, effective use of external and internal data could create efficiencies of more than \$300 billion each year, which the added value of combining data has the potential of reducing healthcare expenditures to the amount of 8%.<sup>32</sup>

The value chain for the re-use of PSI consists of capturing, organizing, packaging and disseminating information. For the purpose of this paper, only the costs and efforts of packaging and dissemination will be analysed since the costs and efforts needed to capture and organize are already incurred within public organizations’ existing IM functions. The value of PSI is derived by the use of the data from all stakeholders. This includes the ability for citizens to provide and contribute their “expertise and perspective to government decision making.”<sup>33</sup>

---

<sup>28</sup> Dekkers et al. “Measuring European Public Sector Information Resources,” 16.

<sup>29</sup> Ibid., 4.

<sup>30</sup> Ibid.

<sup>31</sup> Vickery, “Review of recent studies on PSI.” 16.

<sup>32</sup> James Manyika et al. “Big data: The next frontier for innovation, competition, and productivity,” McKinsey & Company (2011), 2, [http://www.mckinsey.com/insights/business\\_technology/big\\_data\\_the\\_next\\_frontier\\_for\\_innovation](http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation).

<sup>33</sup> Teresa Harrison et al. “Delivering Public Value Through Open Government,” Center for Technology in Government, (2012), [http://www.ctg.albany.edu/publications/issuebriefs/opengov\\_pubvalue](http://www.ctg.albany.edu/publications/issuebriefs/opengov_pubvalue).

Transparency, participation and collaboration are needed by governments in order to allow citizens to perform various roles.<sup>34</sup> The sheer volume of data or number of datasets published by a government is not a good indicator of value. Rather, the quality and accessibility of data determines whether value has been created.<sup>35</sup> Data with the most value will be centered on specific stakeholders and their interests as opposed to citizens in general.<sup>36</sup> Understanding who is being served and ensuring openness of data are both critical to creating value.

### **Adaptable Data**

OD can take on many forms, but to be open it must be in a machine-readable format and available through the Internet with licensing agreements. This makes it adaptable and easier to analyze, aggregate and process, which in turn provides greater service delivery to citizens.<sup>37</sup> On the other end of the spectrum, inert data includes printed reports, forms and machine-readable data that are not available through the means of the Internet. This form of data prevents ease of use. If data cannot be analyzed and aggregated dynamically with technology then most of its value is lost. To benefit from the dissemination of OD, governments need to embrace technology and minimize inert data.<sup>38</sup> Ease of access and assessment of adaptable data provides transparency allows for public scrutiny and accountability.

The act of releasing OD demonstrates basic government transparency, but the real benefit of accountability is only obtained with an added “degree of interaction.”<sup>39</sup> This interaction requires that data reaches its intended audience and that mechanisms are in place to allow citizens to react and governments to respond accordingly.<sup>40</sup> This two-directional flow is required in order to reap the full benefits of accountability as opposed to simply releasing data. This interaction can be easily achieved with today’s technology and the release of OD.

### **Culture of Openness**

The dissemination of information needs to be vetted through a series of specific criteria to protect the integrity of the organization. To mitigate this risk, many organizations have

---

<sup>34</sup> Harrison et al. “Delivering Public Value Through Open Government,”

<sup>35</sup> Ibid.

<sup>36</sup> Ibid.

<sup>37</sup> Harlan Yu and David Robinson, “The New Ambiguity of ‘Open Government’” Social Science Research Network (2012): 181, <http://dx.doi.org/10.2139/ssrn.2012489>

<sup>38</sup> Yu, “The New Ambiguity of ‘Open Government’” 181.

<sup>39</sup> Tiago Peixoto, “The Uncertain Relationship between Open Data and Accountability: A Response to Yu and Robinson’s ‘The New Ambiguity of Open Government,’” Social Science Research Network (2013): 203. <http://ssrn.com/abstract=2264369>

<sup>40</sup> Peixoto, “The Uncertain Relationship between Open Data and Accountability,” 203.

implemented a complex assessment process that is not well understood by internal staff because publishing OD is not yet a common practice. Several organizations will only release information that was requested through an access to information request, placing the onus of data transfer on external stakeholders. A government's ability to solve problems, meet challenges and be innovative is dependent on its ability to flow information to stakeholders. Knowledge is derived from information and to be effective it needs to reach the right person at the right time.<sup>41</sup>

Governments will need to change internal procedures and adapt to methods of disseminating information openly. In 2008, one of the top 10 disruptive technologies was open source software; today it is OD and surrounding technologies.<sup>42</sup> Public scrutiny and the demand from advocacy and special interest groups are triggering the disruptive nature surrounding OD.<sup>43</sup> In addition, OD's dependency on technology and external factors are likely to cause discontinuous or episodic changes within governments. If not managed properly these could create a culture of resistance to change.<sup>44</sup>

---

<sup>41</sup> Warren G Bennis, et al. *Transparency how Leaders Create a Culture of Candor*, 1st ed. San Francisco, CA. Jossey-Bass, (2008). 4. <http://voyager.royalroads.ca/vwebv/holdingsInfo?bibld=244171>.

<sup>42</sup> David Eaves, "Containers, Facebook, Baseball & the Dark Matter around Open Data," Eaves.ca, Last modified July 12, 2012, <http://eaves.ca/2012/07/12/containers-facebook-baseball-the-dark-matter-around-open-data-iogdc-keynote>.

<sup>43</sup> David Eaves, "Rules are no substitute for cultivating a culture of open government," Toronto Star Newspapers Ltd. April 5, 2013, <http://www.thestar.com>.

<sup>44</sup> Karl E. Weick and Robert E. Quinn, "Organizational Change and Development," *Annual Review of Psychology* 50, no. 1 (02, 1999): 7.

## Research Findings – Secondary Research

The following is a brief summary of the secondary research for this paper. In the context of the research question, the review touches on elements of Information Management (IM) practices within the GC, data standards and recognized benefits of sharing Public Sector Information (PSI).

### Information Management

In 2010, Library and Archives Canada (LAC) conducted a formative evaluation to determine government accountability regarding policy, standards and directives. Key informants were interviewed including 39 representatives from various GC departments and agencies.<sup>45</sup> The results of the evaluation highlighted resourcing as a main obstacle to implementing IM practices.

The Treasury Board of Canada Secretariat is leading the efforts to strengthen the management of electronic information and in 2011, the Secretariat conducted a government-wide audit of electronic recordkeeping practices in large departments and agencies.<sup>46</sup> The audit found that organizations are at risk of not effectively identifying and retrieving information needed for effective decision making. This risk is caused by exponential growth of electronic information outpacing internal resources for information management.<sup>47</sup>

In addition, the Canadian government has been criticized for its performance in responding to access to information requests.<sup>48,49,50</sup> Nearly half of the requests submitted to the Government of Canada exceed the thirty day limit prescribed by the Access to Information Act.<sup>51</sup> Furthermore, improvements to the dissemination of OD in response to these requests could lead to more cost-effective practices.<sup>52</sup>

---

<sup>45</sup> Formative Evaluation of LAC Strategic Outcome 1.0, Library and Archives Canada, Last modified December 21, 2012, <http://www.collectionscanada.gc.ca/evaluations-verifications/012014-380-e.html>.

<sup>46</sup> Horizontal Internal Audit of Electronic Recordkeeping in Large Departments and Agencies. Treasury Board of Canada Secretariat. Last modified December 9, 2012. <http://www.tbs-sct.gc.ca/report/orp/2011/itl-itg02-eng.asp>.

<sup>47</sup> Horizontal Internal Audit of Electronic Recordkeeping.

<sup>48</sup> "Letter on open government for the President of the Treasury Board," Office of the Information Commissioner of Canada, Last modified January 19, 2012, [http://www.oic-ci.gc.ca/eng/rr-sl-odi-adi\\_2012\\_1.aspx](http://www.oic-ci.gc.ca/eng/rr-sl-odi-adi_2012_1.aspx).

<sup>49</sup> "Feds 'not the most transparent,' says information commissioner," CBC News, February 9, 2013, <http://www.cbc.ca/>.

<sup>50</sup> "Open Government," Office of the Information Commissioner of Canada, Last modified September 1, 2010, [http://www.oic-ci.gc.ca/eng/rp-pr-ori-ari\\_2010\\_1.aspx](http://www.oic-ci.gc.ca/eng/rp-pr-ori-ari_2010_1.aspx).

<sup>51</sup> "Feds 'not the most transparent.'"

<sup>52</sup> David Eaves, Calculating the Value of Canada's Open Data Portal: A Mini-Case Study, Eaves.ca. Last modified March 8, 2012, <http://eaves.ca/2012/03/08/calculating-the-value-of-canadas-open-data-portal-a-mini-case-study/>.

## Standards

The European Commission is working on an initiative for a common standard that will facilitate the cross-referencing of data and interoperability in order to provide greater benefits to users.<sup>53</sup>

A working group has been assembled, consisting of data experts from public and private organizations across 20 countries including Australia and the United States (US). In addition, the Open Data Institute is actively crowdsourcing to develop criteria that will help organizations assess the value of possible datasets.<sup>54</sup> Meanwhile, Natural Resources Canada (NRCan) has implemented and adapted several geospatial standards, guidelines and best practices which allow applications and systems to effectively operate with each other.<sup>55,56</sup>

## Recognized Benefits

Since February 2012, Statistics Canada has been disseminating its data at no cost; information from their CANSIM database and census data is now OD.<sup>57</sup> This has important social and economic benefits, especially for small organizations, research initiatives and not-for-profit organizations which could not afford the cost of accessing these datasets in the past.<sup>58</sup>

OD can also be used to prevent fraudulent activities. A review of questionable expense claims and receipts released as OD has helped the Canada Revenue Agency save \$3.2 billion in tax receipts claims, which were disallowed.<sup>59</sup> In addition, the OD data provided by Environment Canada (weather, air and water quality data) helps inform citizens and identify areas with climate issues.<sup>60</sup>

---

<sup>53</sup> DCAT application profile for data portals in Europe, European Commission Joinup, Last modified March 8, 2013, [https://joinup.ec.europa.eu/asset/dcat\\_application\\_profile/description](https://joinup.ec.europa.eu/asset/dcat_application_profile/description).

<sup>54</sup> Knowledge for everyone, Open Data Institute, Last modified November 30, 2012, <http://www.theodi.org/>.

<sup>55</sup> Geospatial Standards and Operational Policies, Natural Resources Canada, Last modified April 12, 2012, <http://geoconnections.nrcan.gc.ca/1017>.

<sup>56</sup> Best practices for sharing sensitive environmental geospatial data, Natural Resources Canada, Last modified February 12, 2012. [http://ftp2.cits.nrcan.gc.ca/pub/geott/ess\\_pubs/288/288863/cgdi\\_ip\\_15\\_e.pdf](http://ftp2.cits.nrcan.gc.ca/pub/geott/ess_pubs/288/288863/cgdi_ip_15_e.pdf)

<sup>57</sup> Mehta, Diana, "Free census data a big perk but public site experiencing growing pains," *Winnipeg Free Press*, April 18, 2012, <http://www.winnipegfreepress.com>.

<sup>58</sup> Diana, "Free census data a big perk but public site experiencing growing pains."

<sup>59</sup> Kevin Donovan, "Charity rules beefed up," Toronto Star Newspapers Ltd, December 21, 2007, [http://www.thestar.com/news/investigations/2007/12/21/charity\\_rules\\_beefed\\_up.html](http://www.thestar.com/news/investigations/2007/12/21/charity_rules_beefed_up.html).

<sup>60</sup> Apps 4 Climate Action, British Columbia, Last modified April 24, 2013, <http://www.livesmartbc.ca/A4CA/gallery.html>



The US government has also embarked on similar initiatives where the release of weather data has “benefited the American people and contributed to economic growth and jobs.”<sup>61</sup>

Additionally, the government has created mobile and web applications which provides a directory for community clinics to help citizens gain easier access to locations near them.<sup>62</sup>

Meanwhile, the Danish government estimates that making basic data open and freely accessible will save the public sector \$45 million per year. Likewise, the Danish private sector is estimated to save \$87 million by reducing the cost of acquiring data, improving public services and adding opportunities for new digital products and services.<sup>63</sup>

The British Government has reduced costs through the prevention of fraudulent activities and internal operating efficiencies with dissemination of internal PSI. A recent case study demonstrates how data sharing, both within and between departments, could save \$65 billion through the internal and effective use of PSI.<sup>64</sup>

Other benefits can be achieved when governments or private companies re-publish OD in the form of web or mobile applications. Recollect, a Vancouver company, developed and implemented a standard for managing garbage and recycling related data to remind citizens about their local collection days; this system has proven successful in several cities.<sup>65</sup>

Applications that can provide useful and beneficial information to citizens can range from services that include border crossing waiting time, duty calculators, environmental and land conditions, transportation and bus information, tourist and recreational information, vehicle recalls, currency converters, property taxes, vaccine clinics, health products and locations of external defibrillators.<sup>66,67,68</sup> Other initiatives and projects utilizing open data are provided in Appendix 5.

---

<sup>61</sup> Eric Braverman, and Michael Chui, “Unleashing government’s ‘innovation mojo,” McKinsey Quarterly, 2012, [http://www.mckinsey.com/insights/public\\_sector/unleashing\\_governments\\_innovation\\_mojos\\_an\\_interview\\_with\\_the\\_us\\_chief\\_technology\\_officer](http://www.mckinsey.com/insights/public_sector/unleashing_governments_innovation_mojos_an_interview_with_the_us_chief_technology_officer).

<sup>62</sup> Braverman, “Unleashing government’s ‘innovation mojo,”

<sup>63</sup> Denmark’s open data initiative estimated to save government over \$45 million per year and generate returns of \$87 million for the private sector, Between the Poles, Last modified October 14, 2012, <http://geospatial.blogs.com/geospatial/>.

<sup>64</sup> Chris Yiu, “A Right to Data: Fulfilling the promise of open public data in the UK,” Policy Exchange, 2012, <http://www.policyexchange.org.uk/digital-government>.

<sup>65</sup> How Recollect works in 5 easy steps, Recollect, Accessed on April 20, 2013, <https://recollect.net>.

<sup>66</sup> Open Data Applications, Government of Canada, Last modified October 24, 2012, <http://www.data.gc.ca/default.asp?lang=En&n=CF8AAC8B-1>.

<sup>67</sup> Mobile App Gallery, City of Ottawa, Last modified April 24, 2013, <http://ottawa.ca/en/mobile-apps-and-open-data>.

## Research Findings – Primary Research

The following is a summary of the primary research for this paper. In the context of the research question, the review touches on the characteristics of the participants that used OD, their motivations and issues surrounding the use of OD in project delivery.

### Online Survey

The survey included three series of questions. The first referred to information about the participants, the second examined their involvement with OD, and the third asked questions related to specific OD projects. Requests to complete the survey were sent to 42 separate organizations or groups known to work with open data. A total of 123 participants responded to the survey but only 99 completed it in its entirety. The most common geographic location of participants was Canada (44%), followed by the European Union (22%) and the United States (18%). The geographic locations of participants are provided in the following table.

Nation	Requests	Response	Percentage
Canada	14	44	44.44%
United States	12	18	18.18%
EU27	12	22	22.22%
Others	4	15	15.15%
	<b>42</b>	<b>99</b>	

In addition, nearly 70% of respondents were between the ages of 25 to 44 and more than 80% of all participants had a university degree. Specific percentages related to age groups and educations are provided within Appendix 6.

### Users of Open Data

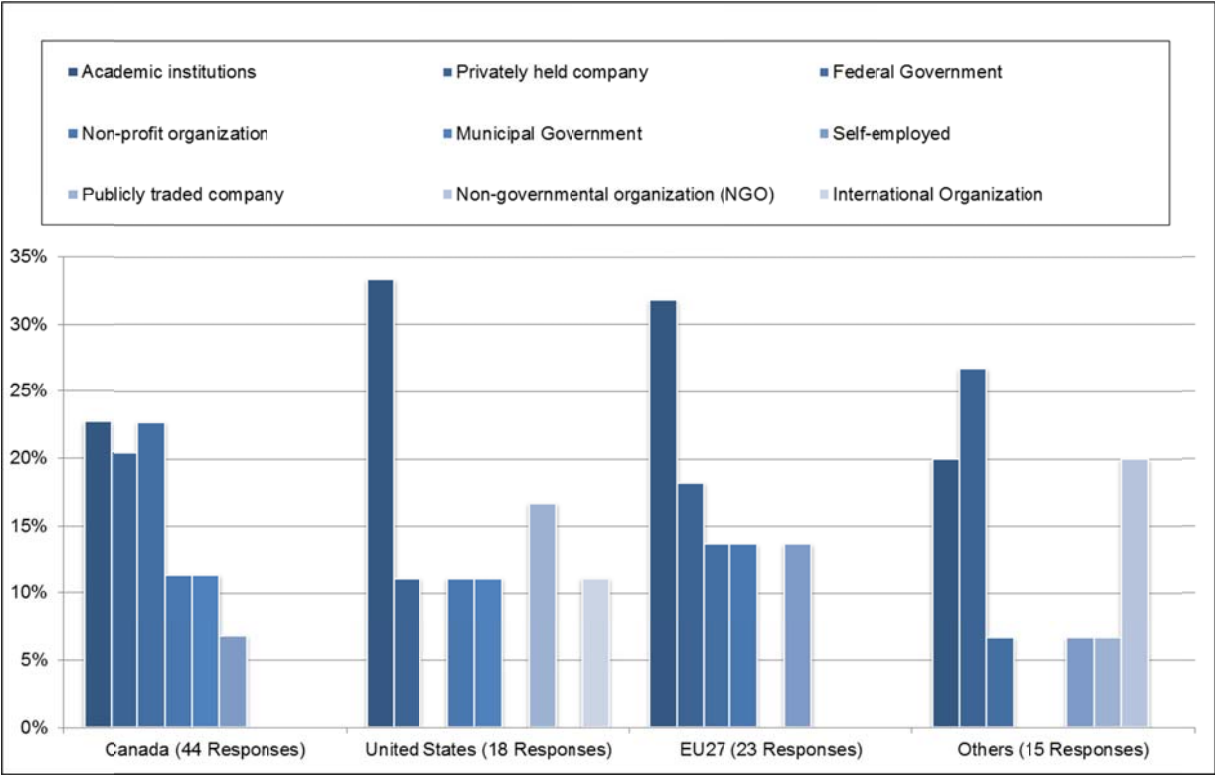
The users of PSI are generally multi-stakeholders and represent groups that are loosely connected by community interest. The users of OD can be categorized within three separate groups. The first group consists of analysts who prefer data in a raw format with the least amount of filtering applied. This group includes academics, economists and journalists who need to analyse and interpret data in order to contribute to their work-related deliverables. In addition, this group can also include other government organizations or internal departments

---

<sup>68</sup> Amazing Mobile Apps, City of Toronto, Last modified April 24, 2013, <http://www1.toronto.ca/wps/portal/contentonly?vgnextoid=7e57e03bb8d1e310VgnVCM10000071d60f89RCRD>.

within the same organization. The second group includes individuals who would prefer dynamic or automated access to data. These are the software developers who need access to clean and reliable data through automated methods such as an Application Programming Interface (API). For the most part their objective is to render the data in a visual interface for a specific audience, which necessitates the need for some level of data analytics. In most cases, this involves aggregating multiple datasets. The last group is comprised of regular citizens or special interest groups who prefer accessing data through an interface with a visual interpretation of the data. Most have limited or no skills for analysing and aggregating datasets and believe that governments are obligated to provide visual interfaces to OD. Organizations working with OD can be found in the following table.

### Organizations working with Open Data



### Motivation for using Open Data

Interest in OD is developing in a bottom-up direction. Special interest groups and not-for-profit organizations are realising the potential benefits of OD and how it can help them achieve strategic objectives. For most countries, the movement is still just beginning and a lot of people are very interested in learning more about the topic. Notwithstanding, a very small portion of people are using OD in the hope to build applications and/or web services. The large users of OD remain academic institutions and the private sectors.

## Micro and Macro data

The benefits of OD information are segregated into two levels of information, micro and macro data. Micro data refers to information that is relevant to individuals and their daily activities, such as bus schedules, road closures and recreational activities. Not surprisingly, smaller governments like municipalities publish micro data. Macro data, on the other hand, consists of data with a wider geographical span such as national and international level information. Macro data can impact larger subsets of the population and be specific to an area or topic. This level of data is dependent on statistical, population and geographical data. In some cases, several micro datasets can be aggregated with macro datasets. Although both are important to citizens, macro data tends to impact a greater number of people. The majority of the OD users that responded to the survey commonly use macro data, including statistical data closely followed by research, population and geographical data. Specific percentages related to the type of data used by survey participants are provided within Appendix 6.

## The Source of Open Data

Surprisingly, the source for data most commonly used by survey participants was not open. Significant amounts of data are published as web content on public organizations' web sites but are not available in an open format. This lack of availability has created a culture of scrapers – individuals who take data from web sites and openly re-use it.<sup>69</sup> That data is then transferred into structured datasets or databases where it can be re-used. In some cases this process is automated, and if web content changes the scraping process is triggered again. It is a complicated method of extracting data and an even more complex way to maintain and update data. In addition, scraping web content falls short of OD license agreements since the data is not available through an OD portal. An interesting outcome of the survey was that 8% of respondents that were working on a specific OD project and working for a federal government organizations were also scraping data. It seems that in some cases it is more cost-effective for governments to scrape themselves rather than to extract the data from its original source.<sup>70</sup> Percentage of macro and micro sources of OD from specific organizations can be found in the

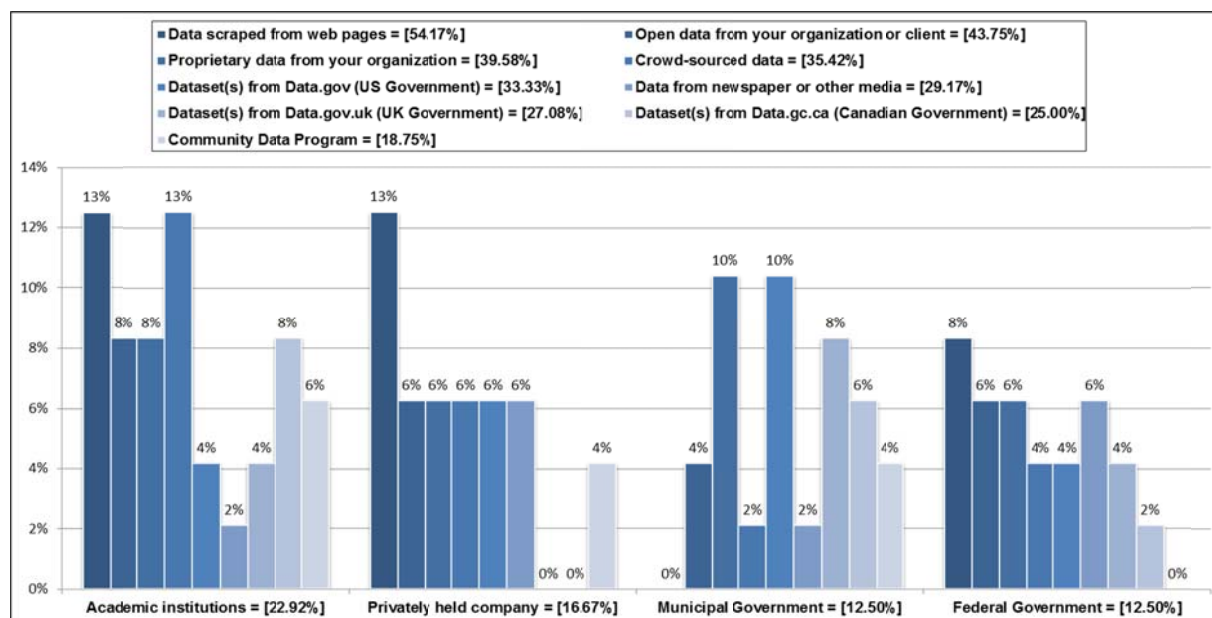
---

<sup>69</sup> Steven Marjanovic, "Screen Scraping' Lawsuit may Clarify Liability Issue," American Banker 165, no. 18 (01/27, 2000): 7, <https://ezproxy.royalroads.ca/login?url=http://search.ebscohost.com.ezproxy.royalroads.ca/login.aspx?direct=true&db=buh&AN=2756797>.

<sup>70</sup> Francis Irving, "Why the Government scraped itself," ScaperWiki Data. June 13, 2011. <http://blog.scrapewiki.com/2011/06/13/why-the-government-scraped-itself/>

following diagram while information related to sources used by separate nations is provided within Appendix 6.

### Sources of data by Organization



### Preferred methods for accessing Open Data

The preferred methods for accessing OD remain with tools and software that are commonly used or openly available. Formats such as comma-delimited files, which are not proprietary to specific software, are preferred for accessing and managing OD. In addition, the preferred tools used to manage and manipulate OD are also not proprietary; open source software is commonly used but in most cases personal preference outweighs the decision and individuals will use tools that are familiar to them. Specific percentages related to the preferred method for accessing the data by survey participants are provided within Appendix 6.

When asked what benefits are derived from OD initiatives the majority of the participants indicated that OD will support innovation and citizens have the right to public data and that public information should be openly accessible and available online. In addition, releasing new data was more important for participants than improving the quality of existing data. Specific percentages of related questions to the importance and benefits of OD initiatives are provided within Appendix 6.

### Open Data Projects

Included in the survey were specific questions related to OD projects. A total of 48 participants responded to these questions. Projects included local initiatives for housing, charities, education

and health-related data. International projects were also included, which looked at opportunities with open source software, data visualizations and aggregation with social media and big data. A few projects also included government initiatives for publishing OD and others included specific work toward new standards.

Several participants working on projects belong to different advocacy or special interest groups, which create personal projects in the attempt of making governments accountable for their actions. These groups also lobby governments with an aim to convince them to adopt OD policies. Other personal initiatives included the aggregation of statistical data in an attempt to identify trends to help charities and not-for-profit organizations determine their strategic direction.

From the participants working on specific projects, 23% identified federal governments as their primary client and nearly 84% stated that they are likely to use OD again. The average of primary clients from participants working on specific projects can be found in the following table.

Clients	Responses
Federal government	23%
Non-governmental organization (NGO)	23%
Non-profit organization	23%
Clubs or association	19%
Municipal government	17%
Publicly funded company	17%
Academic institutions	15%
Provincial government	15%
Community-based organizations	15%
My organization	15%
Privately held company	13%
Publicly traded company	10%
I am a stakeholder	6%

In addition, the following table defines the participants working on specific projects that agreed with each comment.

## Comments related to Open Data Projects

Thinking about your project(s), do you agree or disagree	Strongly Disagree	Disagree	Neither	Agree	Strongly Agree
I am more likely to use open data	2%	0%	14%	48%	36%
They could be repeated in another country/area if the data was available	7%	0%	18%	39%	36%
They will provide social benefits	5%	0%	14%	52%	30%
They will have an impact on people's lives	5%	0%	27%	36%	30%
They were shaped by the data that was available	7%	11%	2%	55%	25%
They would have been different if additional datasets were available	7%	9%	14%	45%	25%
They were experiments which I will continue to develop	5%	16%	14%	43%	23%
I have been waiting for open data	5%	9%	30%	36%	18%
I will need to update my project(s) when datasets get updated	7%	7%	25%	41%	18%
They will provide economic benefits	11%	7%	30%	36%	14%
I would have started them anyway without open data	18%	23%	14%	34%	11%
They will generate an anticipated profit for me or my organization	20%	34%	27%	9%	7%
They will generate a need for additional staff	14%	16%	43%	18%	7%

## Critical issues in government today

The survey offered an open-ended question allowing participants the opportunity to contribute what they believed were the top three issues facing government organizations today. In response to this question, 35% of the participants referred to the relevance of available OD, specifically expressing concern regarding its accuracy and availability. Financial restrictions were also identified as an issue for governments to effectively capture and publish OD. More than 25% of the participants believed that effective dissemination of OD could help governments cut costs and mitigate risks caused by cutbacks.

Furthermore, participants expressed that governments have an embedded organizational culture that does not share information. One quarter of participants stated that governments need to take action and make top-down changes:

The governments will need to build a culture of open data engagement, which will require a change in mindset, and culture. (Participant 97)

There were also several comments regarding the quality of OD. Relevant information about datasets, including the method of collection, metadata and instructions for re-use, is often missing. Inconsistency was also a relevant concern, which relates to a lack of existing standards for information management. In some cases, participants indicated that information about the datasets are simply insufficient to effectively aggregate with other datasets; this was especially true for geographic information. Real transparency was also a common theme. Several participants believed that members of government do not understand the significant importance of OD and the impact it can have for citizens:

Increased visibility of data tends to increase its accuracy and quality through feedback from users. (Participant 74)

The issue of enabling and facilitating public access to open data with visual interface was also a concern.

## Interviews

The interviews included 10 participants, 5 from public organizations and 5 from private organizations.

### Public Organizations

The majority of the interview questions for public officials consisted of quantitative questions to identify direct and indirect costs for the preparation and dissemination of OD. However, most governments' operating costs attributed to OD are small or unknown because they are absorbed within existing information technology operations and sections. Extracting the cost was complex and in most cases organizations were not aware of the actual funds used for OD initiatives. In other cases, some GC departments and agencies were simply not willing to share these costs.

With the exception of large GC departments, manual intervention by staff members is required for the dissemination of new datasets or updates to existing ones. Three public organizations interviewed provided approximate costs to human resource and operating or maintenance costs, including direct and indirect costs. The average cost for the dissemination of OD within these organizations averaged \$130 thousand CDN per year. The following table contains the average costs from each organization.

Organization	Operating Costs per year	Human Resource per year	Total
Public organization 1	\$ 50,000	\$ 85,000	\$ 135,000
Public organization 2	\$ 65,000	\$ 75,000	\$ 140,000
Public organization 3	\$ 15,000	\$ 100,000	\$ 115,000
<b>Average</b>	<b>\$ 43,333</b>	<b>\$ 86,667</b>	<b>\$ 130,000</b>

Two organizations spent time developing a cost recovery model for the preparation and dissemination of OD by individual datasets. They believed that estimated time spent to publishing each new dataset is approximately 275 hours, from the analysis to the deployment phase. What was not clear is the indirect costs and time spent on publishing OD from other



sections within organizations. What was very clear and voiced from all of the interview participants from public organizations was that the cost of publishing OD would definitely increase.

Furthermore, interview participants commented on the lack of internal policies and standards that in turn reflect on the poor quality of the OD published:

Internal IM practices are lacking, we are building data from the outside in. The internal practices of governments were never designed to manage information, which informs the community in this open format. Governments need to re-think their frameworks.

(Interview Participant 4)

### **Private Organizations**

A common concept among interview participants from the private sector was the lack of available data and the difficulty in locating data. These participants work with data on a daily basis and frequently need to locate new and interesting datasets for their projects or initiatives.

Participants indicated that in many cases information was simply not available in an open format but is publicly available online from public web sites. As such, individuals and companies frequently resort to scraping data off public web pages.<sup>71</sup> Although scraping raises many concerns, the consensus among participants is that it is a common practice, even within government organizations.

In the case of one participant that was scraping public web sites, the information he was collecting was simply not available in any other format. By aggregating the scraped data, he was able to create a valuable and appealing dataset, since this information was not available anywhere else. By offering an interface to access this data, he created value by allowing regular citizens to search and query the data without analytical skills. Unfortunately, existing license agreements do not include the scraping of public websites; the web content is publicly available but the right to re-use the data is technically not permitted. Scraped data is simply not OD. Interestingly, the added value of the information that he created by aggregating this data is so significant that members of the GC are subscribed users to his service.

---

<sup>71</sup> Marjanovic, "'Screen Scraping' Lawsuit may Clarify Liability Issue," 7.

While publishing public information on web sites governments should make the effort to publish data in an open format also. Any data available publicly should also be available in an open format also. (Interview Participant 2)

Another issue brought forward by interview participants was common standards. Standards provide a common method for aggregating OD that can protect privacy. This was especially relevant for geographic standards. The aggregation of OD is dependent on standards across datasets and if separate departments utilize different standards it will take a considerable amount of effort to amalgamate data. As an example, geographical frameworks used within OD vary between GC departments and this lack of a common framework poses challenges to analysts who need to aggregate data across departments. Existing geographies used in datasets can include postal code zones, federal electoral districts, census geographies or health districts. For example, health and environmental issues can span many jurisdictions and geographic boundaries.

### **Focus Group Session**

The focus group session was held during a conference event and included three candidates from the private sector. The goal of the session was to explore ideas in a divergent manner and identify possible solutions for known issues preventing OD from achieving benefits. The group identified transparency as an issue and related the problem to the organizational culture found in government organizations. Experiences from participants were shared among the group; the discussion focused on examples where information was not released in re-usable formats.

When requesting information that was not publicly available online it was provided to me in a compact disk or in a paper format. (Session Participant 2)

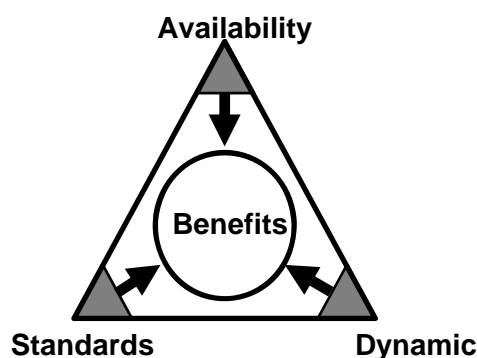
Other issues were identified which related to the efforts required to find information. There is still much information that is not yet available in open format, such as information related to grants and contributions. Participants that work on specific research initiatives indicated a need for a large amount of statistical data to identify trends. According to the group, a considerable amount of effort is used to try to determine if the information actually exists and is available.

## Analysis and Discussion

The results of the current research show that benefits of OD can be driven through six separate mechanisms that can directly or indirectly provide benefits to society through social, economic and environmental means. The first provides social and environmental benefits through the mechanism of public service delivery, which enables citizens to stay informed and interact with their governments. The second provides economic benefits with the support of industries and commerce through the innovation and creation of products and services. Further economic advantages can be obtained by re-using knowledge and wisdom drawn from OD across internal government organizations, leading to efficiency gains and operational cost savings. Finally, the last three mechanisms have the potential to provide social, economic and environmental benefits with the support of academic research, evidence-based policy decisions, and public accountability.

1. Public service delivery
2. Support of industries and commerce
3. Re-using of information internally
4. Support for academic research
5. Support for evidence based policy decisions
6. Public accountability

The findings of this research indicate three factors that can either help or hinder the effectiveness of OD: availability of data, standards, and dynamism. The following diagram illustrates how benefits can be obtained when governments make data readily available, that is dynamic, and that adheres to established standards.



## Availability of Data

Data needs to be made available in an open format while rigorously protecting privacy. This includes data that can be extracted from PSI and that has not yet been made public. It can also include public data that has not yet been made available in a machine-readable format.

## Information Management Practices

The ability to publish data starts with effectively capturing and organizing PSI. For the most part, existing IM practices within the GC are not effective and not compliant with existing policy and directives. This makes the work of publishing OD even more complex for the civil servants managing the data. Salient obstacles to implementing IM practices identified in Library and Archive Canada's (LAC) evaluation include lack of resourcing and lack of familiarity with policies. Nearly one third of participating departments and agencies felt that IM was simply not receiving sufficient attention and priority within their organization.<sup>72</sup> Furthermore, only 20% believed that their organization was capable of managing information effectively.<sup>73</sup> The GC's approach to managing information has not proven to be effective across departments and agencies and these issues will impact their ability to make OD available.

The role of government needs to include a mandate to manage data. Within their IM policy and directives, the GC has defined information of business value as an asset.<sup>74</sup> These assets help to support and facilitate decision making, operations and the delivery of programs and services. Unfortunately, the link between IM policy and directives and the benefits of publishing OD is still missing. The GC is attempting to address this issue with the development and implementation of a Directive on Open Government, which should provide guidance to departments and agencies on how to identify information, and how to determine formats and standards for publishing Open Government Data.<sup>75</sup>

Other countries have taken different approaches to the same issue. President Barack Obama has released an executive order committing all federal government data to be open and

---

<sup>72</sup> Formative Evaluation of LAC Strategic Outcome 1.0.

<sup>73</sup> Ibid.

<sup>74</sup> Directive on Recordkeeping, Government of Canada, Last modified July 18, 2011, <http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=16552>.

<sup>75</sup> Canada's Action Plan on Open Government. Government of Canada, Last modified April 11, 2013, <http://open.gc.ca/open-ouvert/ap-pa04-eng.asp#toc6>.

machine-readable.<sup>76</sup> The British Prime Minister has attempted to mitigate the issue with a public letter addressed to all senior civil servants.<sup>77</sup>

New policies and directives from the GC will help minimize challenges around effective management of information. But given other priorities, the lack of resources and understanding of policy, the GC still needs to mitigate the issue of compliance. Stronger action is needed to prevent issues with compliance or the publishing of datasets will suffer from cutbacks. In addition, GC departments and agencies need guidelines and tools to help them identify and manage OD.

Furthermore, GC compliance with the access to information legislation has been in steady decline for the past several years.<sup>78</sup> A special report to Parliament in 2010 revealed issues related to a lack of commitment from senior management within departments and agencies, mismanagement of time extensions, and an insufficient pool of resources and training devoted to access functions.<sup>79</sup> With an increase in the volume of electronic records and a lack of sound IM practices, this will lead to additional costs and delays. OD has the potential to help the GC resolve issues surrounding access to information requests.<sup>80</sup> Opportunities can be realized where requests for information, which would have normally gone through the lengthy process of a formal access to information request, can be accomplished with access to OD. Not all requests could be satisfied with OD, but the potential for operational savings could be significant. This would require collaboration among OD teams and people responsible for access to information requests.

## **The Value of Data**

According to the current program of research, the most prominent users of OD are members of the research community. These individuals have the skills required for in-depth analysis and prefer data in its original state untouched and unfiltered. They are also users of macro data

---

<sup>76</sup> Executive Order -- Making Open and Machine Readable the New Default for Government Information, The White House. Last modified May 9, 2013, <http://www.whitehouse.gov/the-press-office/2013/05/09/executive-order-making-open-and-machine-readable-new-default-government->

<sup>77</sup> Letter to Government Departments on Opening Up Data, The Official Site of the British Prime Minister's Office, Last modified April 24, 2013, <http://webarchive.nationalarchives.gov.uk/20130109092234/http://number10.gov.uk/news/letter-to-government-departments-on-opening-up-data/>

<sup>78</sup> Letter on open government for the President of the Treasury Board, Office of the Information Commissioner of Canada, Last modified January 19, 2012, [http://www.oic-ci.gc.ca/eng/tr-si-odi-adi\\_2012\\_1.aspx](http://www.oic-ci.gc.ca/eng/tr-si-odi-adi_2012_1.aspx).

<sup>79</sup> Letter on open government for the President of the Treasury Board.

<sup>80</sup> Eaves, Calculating the Value of Canada's Open Data Portal.

sources that help social causes that impact a larger portion of the population. The long-standing social benefits of research can be significant. Research outcomes have the potential to foster collaboration with the community and promote sharing of information and innovation. The GC needs to support academic research, foster further collaboration with the academic community, and promote sharing of information.

The possible economic advantage that can be obtained from combining data is creating a demand and encouraging private companies and individuals to develop, implement and maintain mobile and web applications. This is especially true with micro data and mobile applications that provide a public service. These types of initiatives can provide the government with insight into the type of data citizens are interested in seeing.

Currently, the development of applications is a relatively small outcome of OD. When asked if OD projects would create a profit for them or their company, only 16% of survey participants agreed, 34% said no and 27% did not know. Notwithstanding, when asked if they are likely to use OD in the future 84% said yes. Furthermore, 95% of survey participants believe that OD is important for the support of innovation. Making more data available openly will only increase the potential for such applications, as well as support industries and commerce which can in turn create further economic benefits.

### **Adaptable Data**

When governments provide data they have two options. They can publish the underlying data in a machine-readable format, such as OD, or they can provide access to the data through a visual and interactive user interface, such as web sites. Underlying data is considered adaptable and can easily be re-used and aggregated with other adaptable data. It is the preferred format for the skilled user who prefers to analyse unfiltered data. However, such data is useless to citizens with limited data analysis skills.

Although web sites are useful to users with limited skills, they are considered inert data and cannot be easily re-used. In the absence of OD, inert data on public web sites creates dependence on scraping, which carries a number of ethical and economic concerns. The re-use of extracted data through scraping is occurring without proper license agreements. The GC OD license agreements do not take into account the re-use of extracted data from public web sites.

With nearly 50% of surveyed participants scraping public web sites this is an issue that governments need to address. When inert data is published on public web sites, the motivation for departments and agencies to release the data in an adaptable format is considerably low

since the task of filtering and protecting sensitive data has already been completed. Unfortunately, this added level of effort is not being accomplished, despite the true value of OD being found in the knowledge and wisdom that is extracted from combining datasets.

To accomplish this, PSI needs to be made available in adaptable formats. It is the ability to aggregate machine-readable data that adds value to data. This is especially true when combining primary data from public sectors with secondary data sources from external sectors. Such combinations significantly increase the potential for further knowledge and wisdom. This is an advantage that cannot be achieved without technology and, although potentially disruptive for governments, it is beneficial to society. If OD is not available, citizens would be dependent on governments to render and display information and the knowledge and wisdom drawn from it. OD should be made more accessible in machine-readable formats so that everyone can benefit from its knowledge and wisdom.

Inert data can pose a number of challenges as well. Government organizations struggle to meet an overwhelming amount of end-user requirements when they develop and deploy applications. Adherence to Common-Look and Feel (CLF) and language requirements drives the cost of developing applications and in many cases the public's need for data is not met. There are also the costs and resources associated with the ongoing support and maintenance of systems; governments have a tendency of keeping systems online past their usefulness, which brings additional unnecessary costs that they will need to incur. The private sectors have already demonstrated how they can develop and implement applications faster and more efficiently. In addition, private parties will address the market need and meet their requirements if adaptable data is made available by governments. Because of this the GC should spend their limited resources publishing adaptable data instead of deploying applications.

Inert data has its place; it is useful for users with limited skills to analyse data. In some cases, governments have a mandate to make information available in a visual interface for the public, including such users. Nevertheless, the GC should prioritize OD over inert data. Publishing OD will allow private companies to develop and deploy applications and in turn provide faster access for citizens.

## **Standards**

It has been established that combining data is an important beneficial process. It is necessary, then, to establish standards for OD. Available data allows us to access information and provides openness and transparency for governments but the need to aggregate OD is key to extracting its full potential. Standards can make this process easier for users of OD and in turn provide

greater breadth and quality of information. Without standards to validate and identify data, GC departments and agencies will be reluctant to release OD.

There are several existing initiatives within the GC that could help to establish standards for OD. For example, Natural Resource Canada has conducted extensive assessments for validating the dissemination of sensitive information.<sup>81</sup> Decision trees with detailed step-by-step instructions have been developed and could be re-used to help develop OD assessment tools across multiple departments. A similar tool was developed and implemented in the province of British Columbia, where such tools are a requirement of their OD policy.<sup>82</sup> Outside of Canada, the World Bank is using an innovative method to seek suggestions for the development of their Open Data Readiness Assessment Tool through crowdsourcing.<sup>83</sup>

### **Geographical Standard**

The formalities involved in extracting data from several government departments can be unwieldy, preventing effective outcomes and innovations. The survey revealed how individuals obtained data from multiple sources, and how individuals are using data by aggregating OD. One important benefit of OD is the ability to aggregate it with other data; from which value, information and knowledge is derived. One of the challenges of aggregating data is the lack of using common standards. Standards determine how data is structured within the packaged dataset; when data sources use different standards, it can be difficult or even impossible to aggregate data.

This is particularly true with geospatial standards. The ability to aggregate OD with maps requires that geographical data be part of the dataset. Two issues with geographical data were identified in the findings: the first was that geographical standards need to effectively protect personal information and privacy, and the second was that the GC departments and agencies are not leveraging any one standard when collecting and organizing data. Standards used within datasets today can include simple address and postal codes, federal electoral districts and the census geography from Statistics Canada. In addition, Natural Resources Canada has

---

<sup>81</sup> Best practices for sharing sensitive environmental geospatial data, Natural Resources Canada, Last modified February 12, 2012,

[http://geoscan.ess.nrcan.gc.ca/starweb/geoscan/servlet.starweb?path=geoscan/geoscanfastlink\\_e.web&search1=R=288863](http://geoscan.ess.nrcan.gc.ca/starweb/geoscan/servlet.starweb?path=geoscan/geoscanfastlink_e.web&search1=R=288863).

<sup>82</sup> Open Information and Open Data Policy, DataBC, Last modified April 12, 2013, [http://www.cio.gov.bc.ca/local/cio/kis/pdfs/open\\_data.pdf](http://www.cio.gov.bc.ca/local/cio/kis/pdfs/open_data.pdf).

<sup>83</sup> Readiness Assessment Tool, The World Bank, Last modified April 12, 2013, <http://data.worldbank.org/about/open-government-data-toolkit/readiness-assessment-tool>.



implemented geospatial standards and operational policies.<sup>84</sup> These can remove barriers and enable individuals and organizations to exchange location information efficiently. This type of situation is best described by the following hypothetical situation:

A researcher or a government department wishes to study respiratory illness in children between the ages of 0-5 years old. One objective is to create a geographical representation to determine areas of increased risk for respiratory issues using Canada's geographic standard.

The data required to complete this assessment would necessitate looking at several federal and provincial government sources. Respiratory illness data may reside with the respective governments of the 10 provinces and 3 territories or with the Canadian Institute for Health Information (CIHI) which is a federal cost recovery institution. In addition, a list of deaths caused by respiratory diseases can be acquired from Statistics Canada; this data is only aggregated into provinces and territories making it difficult to identify issues in urban areas or smaller, rural areas. Statistics Canada also produces population data by age. Depending on the type of geography, a custom request for information might be required; if so, this would lead to additional costs and processing time. Air quality data would need to be acquired from Environment Canada. This is point source data collected from air quality sensors.

None of these datasets use the same standards for geography. If the analysis were to include income and country of origin to the mix, data from Canada Revenue Agency and Citizenship and Immigration Canada would be required, which would add new formats for geographies. The overall analysis is impeded by costs, time delays, multiple jurisdictions, multiple licenses and a lack of a geographic aggregation standard which reduces its benefits.

## Dynamic Data

Governments face certain challenges when attempting to publish OD; these are due to the new actors involved in OD and the flow that data takes on.<sup>85</sup> OD has the potential to travel through a dynamic loop and return to its origin.<sup>86</sup> The added value derived from OD initiatives comes from combining primary data sources with other data sources and connecting multiple providers and stakeholders.<sup>87</sup> For example, health care data related to a specific illness that is combined with geographic locations could benefit a series of stakeholders including government, health care

---

<sup>84</sup> Geospatial Standards and Operational Policies, Natural Resources Canada, Last modified April 12, 2012, <http://geoconnections.nrcan.gc.ca/1017>.

<sup>85</sup> Helbig et al. "The Dynamics of Opening Government Data," 21.

<sup>86</sup> Ibid., 22.

<sup>87</sup> Ibid., 12.

workers, researchers, not-for-profit organizations, and individual citizens. In addition, the data could also be enhanced while circulating through data sources and providers. This flow of information provides benefits and is what governments need to think about when publishing data.

A government which is accountable is a more cost-effective, transparent, efficient and responsive government. The release of OD can account for transparency but this is only one step to obtaining accountability.<sup>88</sup> Accountability needs information to cycle back to governments. After data reaches its intended audience, mechanisms need to be in place that allows stakeholders to react, and governments need to respond accordingly.<sup>89</sup> These series of events are required to achieve the full benefits of public accountability.

The lean approach used by entrepreneurship could benefit government operations. A major reason for entrepreneurship success with OD relates to the speed with which they can deploy solutions. In most cases, these are interdisciplinary teams that develop the simplest solutions and go to market to learn from actual experiences and feedback.<sup>90</sup> OD contests that are hosted by different governments are a good example of how fast the private sector can deploy applications. The city of Ottawa's last contest received more than 60 applications within weeks as opposed to months.<sup>91</sup> This rapid and iterative approach to developing solutions could benefit governments when it comes to publishing OD. In most cases, the traditional approach to government development is cumbersome and time consuming. With the release of OD, governments will need to make changes within internal processes to allow for dynamic data. GC departments and agencies could benefit from interdisciplinary teams to deal with OD initiatives.

### **Change-Friendly Culture**

Change is much more than the analytical side of a business case. There needs to be a broader approach to change that includes what people think and how they feel.<sup>92</sup> Lewin conceived the three stages of a transition process as unfreeze, change and refreeze.<sup>93</sup> It is in the unfreeze stage where planned changes are likely to fail because of resistance. The major reasons are

---

<sup>88</sup> Peixoto, "The Uncertain Relationship between Open Data and Accountability," 203.

<sup>89</sup> Ibid.

<sup>90</sup> Braverman "Unleashing government's 'innovation mojo,'" 35.

<sup>91</sup> Ottawa Open Data App Contest, City of Ottawa, Last modified May 13, 2013, <http://www.apps4ottawa.ca/>.

<sup>92</sup> John P. Kotter, "Developing a Change-Friendly Culture: an interview with John P. Kotter." *Leader to Leader* 48 (2008): 35. <http://onlinelibrary.wiley.com.ezproxy.royalroads.ca/doi/10.1002/ltl.278/pdf>.

<sup>93</sup> Kotter, "Developing a Change-Friendly Culture," 35.

linked to a lack of strategic direction, clear vision, communication and the failure to create a sense of urgency.<sup>94</sup> If a sense of urgency and/or importance regarding OD is not understood then motivation for change within public organizations will remain low and plans to release OD will progress slowly or not at all.

President Obama has understood this; by releasing an executive order committing all federal government data to be open and machine-readable, he is fostering an environment of urgency and importance that is needed for systematic change.<sup>95</sup> Furthermore, to minimize complacency the US government hired a new Chief Technology Officer (CTO) from the private sector with OD experience.<sup>96</sup> An individual that was not previously in government can bring innovative perspectives and ideas to the government system and underscore the environment of urgency. This sense of urgency from senior officials is required to help people transition from the old to the new. People have a tendency to revert to familiar routines and systems, particularly if anxiety about the change sets in.<sup>97</sup> Persuading people to leave their comfort zone requires ongoing effort and dedicated leadership. Managers can sometimes become impatient, overestimate the success of the change, and stop creating a sense of urgency.<sup>98</sup> Failure to communicate a clear vision and a strategic direction will prevent governments from sustaining a sense of urgency and inadvertently encourage employees to go back to their old ways.

Additional resistance from organizational culture can include an episodic process of change. Public scrutiny, advocacy and special interest groups trigger the disruptive nature surrounding OD; this external environment is shifting and precipitating internal pressures, which move organizations away from their equilibrium.<sup>99</sup> Triggered by external environments, internal

---

<sup>94</sup> John J. Lucey, "Why is the Failure Rate for Organisation Change so High?" *Management Services* 52, no. 4 (2008): 14. <https://ezproxy.royalroads.ca/login?url=http://search.proquest.com/docview/234257649?accountid=8056>.

<sup>95</sup> Executive Order -- Making Open and Machine Readable the New Default for Government Information, The White House. Last modified May 9, 2013, <http://www.whitehouse.gov/the-press-office/2013/05/09/executive-order-making-open-and-machine-readable-new-default-government->.

<sup>96</sup> Philip Yam, "How to Kick-Start Innovation with Free Data," *Scientific American*, March 23, 2013, <http://www.scientificamerican.com/article.cfm?id=how-to-kick-start-innovaton>

<sup>97</sup> Chip Heath, "Making the Emotional Case for Change." *McKinsey Quarterly* no. 2 (06, 2010): 88. <https://ezproxy.royalroads.ca/login?url=http://search.ebscohost.com.ezproxy.royalroads.ca/login.aspx?direct=true&db=buh&AN=49469781&site=ehost-live>.

<sup>98</sup> John P. Kotter, "Developing a Change-Friendly Culture: an interview with John P. Kotter." 35.

<sup>99</sup> Karl E. Weick and Robert E. Quinn, "Organizational Change and Development," *Annual Review of Psychology* 50, no. 1 (02, 1999): 7. <https://ezproxy.royalroads.ca/login?url=http://search.ebscohost.com.ezproxy.royalroads.ca/login.aspx?direct=true&db=buh&AN=1776889&site=ehost-live>.

performance and new strategic directions, an episodic change is created and will prove to be more disruptive and slower to implement with less chance of success.<sup>100</sup>

Resistance should be minimized however possible, as it can clash with the strategic direction of OD initiatives. Deeply embedded organizational culture can prevent the success of planned change. Trying to change culture is like changing the belief of individuals because behaviors that are driven from organizational culture are embedded in members' values. These values inform cultural actions and behaviors. Motivating and utilizing behaviors are the only way to prosper from organizational culture.<sup>101</sup> Aligning the organization's objectives with its culture will help focus the strategic direction; such an alignment can accelerate organizational change.<sup>102</sup> If management ignores the organizational culture of government, they will risk losing positive attitudes and behaviors that can make OD a success. In addition, a critical success factor will be selecting key individuals that can propel change and lean operations.

## Scenario Analysis

Scenario planning is a business analysis tool which helps determine if a scenario is plausible. Indicators are identified to help determine the likelihood of specific events occurring.<sup>103</sup> A scenario analysis requires a focal question that describes our scenario. For the purpose of this analysis and in the context of the research question, the focal question is: What are the benefits of OD?

---

<sup>100</sup> Weick, "Organizational Change and Development," 11.

<sup>101</sup> Jon Katzenbach and Ashley Harshak, "Stop Blaming Your Culture," *Strategy+Business* 62 (2011): 1-10.  
[http://www.strategy-business.com/media/file/sb62\\_11108.pdf](http://www.strategy-business.com/media/file/sb62_11108.pdf).

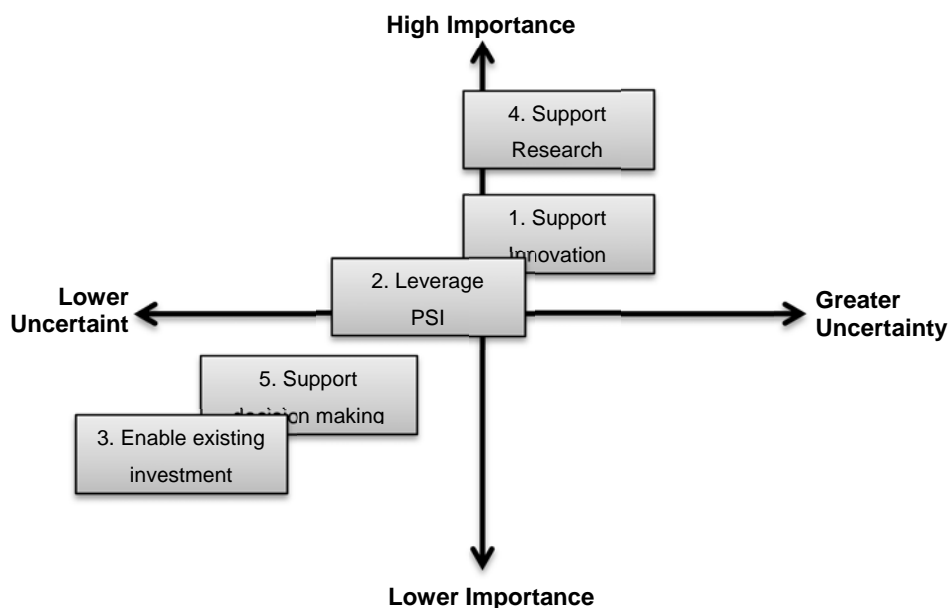
<sup>102</sup> Katzenbach, " Stop Blaming Your Culture," 113.

<sup>103</sup> Constance Gustke, "All the Options." *Conference Board Review* 46, no. 1 (Jan, 2009): 18-23,  
<https://ezproxy.royalroads.ca/login?url=http://search.ebscohost.com.ezproxy.royalroads.ca/login.aspx?direct=true&db=buh&AN=36615374>.

## Uncertainties

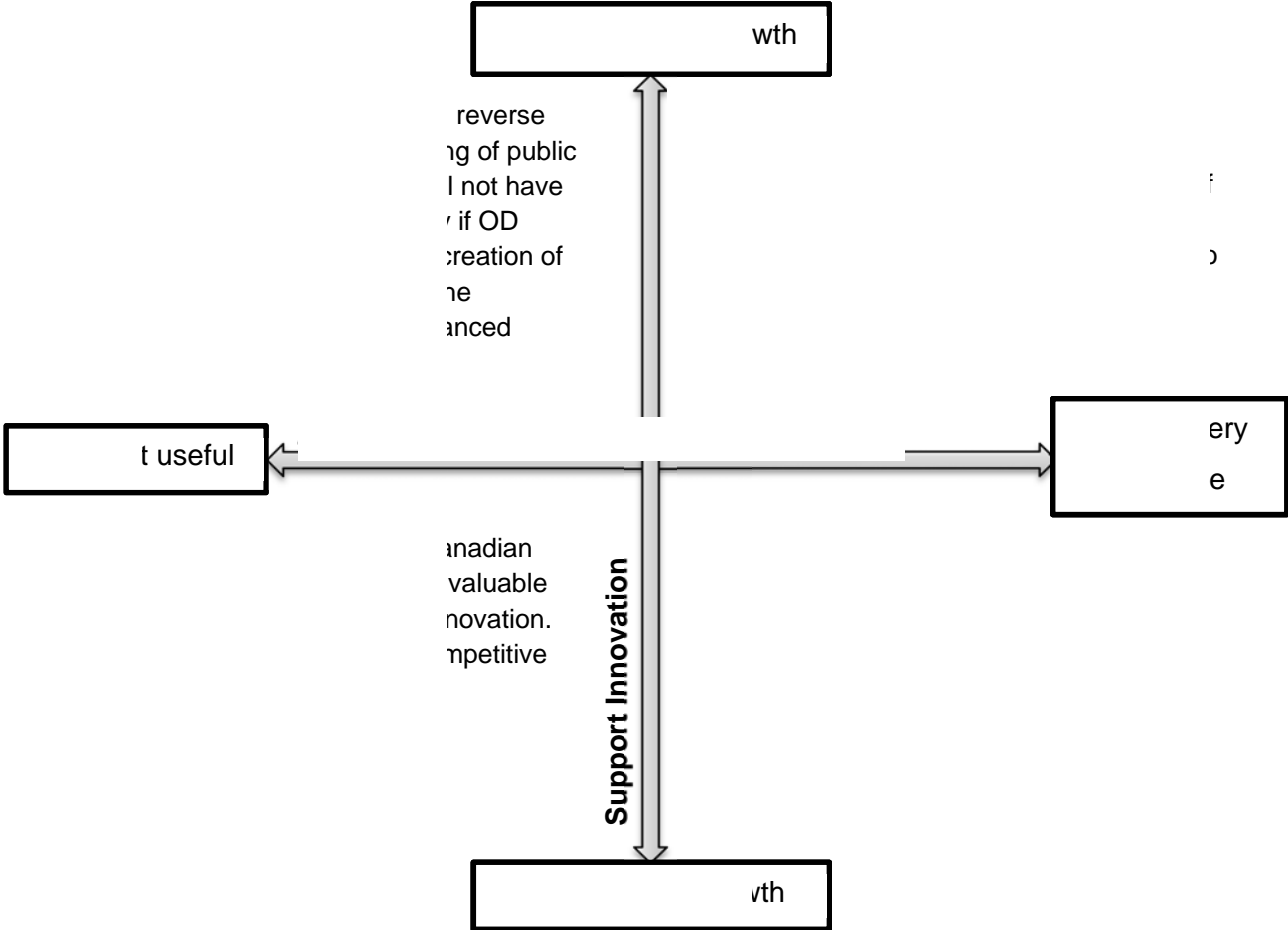
The following is a list of uncertainties about future outcomes and potential value of OD, as reflected by the presumed benefits of OD identified by the GC.<sup>104</sup> These are uncertainties of how OD might benefit Canadians:

1. Supports innovation
2. Leverages public sector information to develop consumer and commercial products
3. Enables better use of existing investment in broadband and community information infrastructure
4. Supports research
5. Supports informed decisions for consumers



It is likely that OD will support innovation and research but it is difficult to measure. The mere process of releasing OD will in turn help leverage public sector information; private sectors and governments will benefit from the release of OD. The two most uncertain and important excerpts remain the support of innovation and research. This is aligned with the survey result where 96% of participants agreed that innovation was very important for Canada. In addition, there are still uncertainties to how OD might support research and innovation. The previous diagram represents the importance and uncertainty of each presumed benefit.

<sup>104</sup> Benefits of Open Data, Government of Canada, Last modified October 10, 2012, <http://www.data.gc.ca/default.asp?lang=En&n=1E4722A8-1>



**Scenario Narratives**

The previous diagram provides a graphical representation of four possible outcomes within this scenario framework. By taking the two most important but uncertain excerpts, the support of innovation and the research, we can build a plausible story narrative that will explain how the support of innovation and the research will provide benefit from the current state of affairs to a future described by each of the following quadrants.

**Scenario A: Spur economic growth and value to research**

In this scenario OD informs and supports research and private sectors with valuable data and in turn helps research and development to increase Canada’s competitive ability in the world market. New markets for open data are created through the development of solutions and novel ideas. Collaboration is heightened and governments are viewed as transparent. Innovation provides knowledge that informs strategic directions for governments and industries. Research and development is enhanced, rapid advancements and developments are created, and

research outcomes become a catalyst of growth in Canada. OD enables the creation of new markets and services. A new demand for knowledge is created which creates a need for new enterprises. The sharing of data is in high demand and new companies are created to meet the demand for new products and services.

### **Scenario B: Spur economic growth but no value to research**

In this scenario, OD enables the creation of new markets and services. A new demand for knowledge is created which drives new enterprises. The sharing of data is in high demand and companies are created to meet the demand for new products and services. Meanwhile, OD does not support research and newfound knowledge is not obtained. If research and development efforts do not enhance the advancement of new and innovative products and services, research will not become a catalyst of growth in Canada and the economy will eventually suffer. Even if the sharing of public data heightens the economy, it will not have a long-term effect on the economy if OD does not reinforce research. The creation of new markets will be temporary if the economy is not supported by enhanced research and development.

### **Scenario C: Value to research but no economic growth**

In this scenario OD informs and provides the research community and private sector with valuable data but does not help increase Canada's competitive ability in the world market. OD does not support innovation and economic growth and new products and services are not created. The development of solutions and novel ideas do not enable the creation of new markets and services. Research and development does not enhance advancements and innovation does not become a catalyst of growth in Canada.

### **Scenario D: No economic growth and no value to research**

This scenario is highly unlikely given the value of shared information. In this scenario OD does not support research and private sectors with valuable data. It does not help economic growth, nor does it increase Canada's competitive ability in the world market. New markets are not created through the development of solutions and novel ideas. Furthermore, collaboration would not improve and governments would not be viewed as transparent. Innovation does not become a catalyst of growth in Canada.

### **Summary**

The plausible outcome of this analysis is determined by the probability of a specific scenario occurring. The likelihood that OD creates and supports innovation and research is very high. We

have already identified the academic community as primary users of OD and the private sector is increasingly using OD to bolster innovation. These are the users that have the skills and tools to aggregate data and create added value from their in-depth analysis of the data, which will in turn provide heightened information, knowledge and wisdom to society. The release of OD is likely to create benefits for these groups and in turn provide them with additional resources to complete other relevant tasks.

## Externalities

Externalities refer to a scenario when unexpected outcomes occur which impact the marginal costs or benefits of a specific effort.<sup>105</sup> In the case of OD, the effort of publishing data results in marginal benefits. Positive externalities are created when further benefits are obtained from the aggregation of data, a series of complex reactions are ignited, and data is shared across multiple jurisdictions with other stakeholders. Marginal benefits combined with positive externalities result in marginal social benefits. The challenge lies with attempting to measure the amount of positive externalities.

An evaluation of qualitative data is required to measure the outcome and performance of publishing OD. We have already identified the marginal benefits as:

1. Public service delivery
2. Support of industries and commerce
3. Re-using of information internally
4. Support for academic research
5. Support for evidence based policy decisions
6. Public accountability

We can also determine some marginal costs from publishing OD which include time spent analysing and publishing data. Some citizen groups and media outlets believe that the release of government information could be misinterpreted and the time spent on unwanted data can be wasteful on government resources.<sup>106</sup> This perspective is influenced by the inherent difficulty of determining the value of data prior to its publication and aggregation.

In order to demonstrate benefits of publishing OD, positive externalities must outweigh any negative externalities. There are few negative externalities from publishing OD that could create

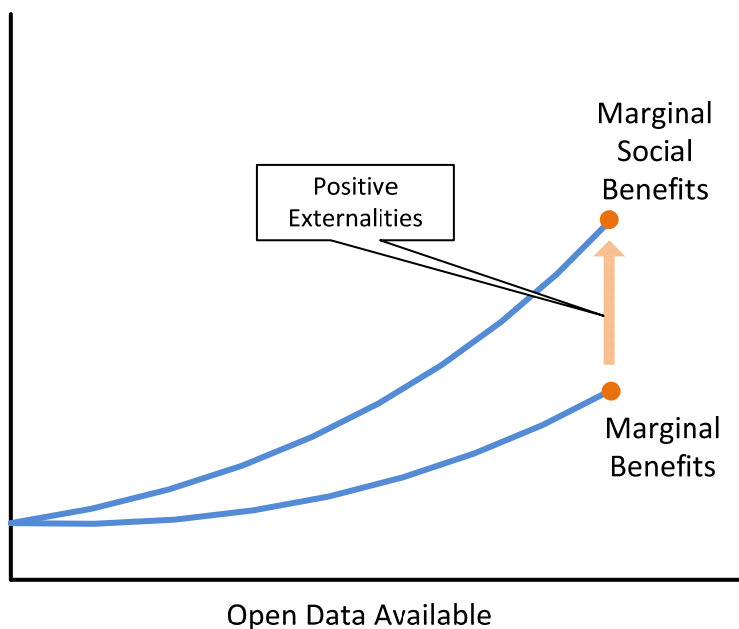
---

<sup>105</sup> Laurence Booth and Sean W. Cleary, *Introduction to Corporate Finance* (Missauga: Wiley, 2010), 40.

<sup>106</sup> Steve Paikin, "Time for Open Government," *The Agenda*, April 28, 2010, <http://www.youtube.com/watch?v=KCJooWIHLyC>.



marginal social costs. In addition, the action of not publishing OD is not a negative externality derived from the effort of publishing OD. Conversely, when governments don't publish OD it prevents benefits but it does not reduce the marginal social benefits from the effort of publishing OD. Because of this we can determine that there will always be more positive externalities than negative has showed in the following diagram.



Nevertheless based on the sample of the survey, not publishing OD would have an impact on the most common users, the academic community. They would lose the most from limited access to data. These users possess skills and knowledge for analysis and combining data; not having access to OD would hinder research and development, which in turn would negatively impact future innovation, the release of new products and services, and potential economic benefits. In the 1980s, Canada witnessed such negative impact when Statistics Canada charged the public a fee for data.<sup>107</sup> At the time, the prevailing opinion was that the cost of producing powerful data had to be recovered, and that this was best accomplished by charging the user. Canadian research questions went unanswered as some researchers were forced to turn to US or European while and other researchers gave up using these types of data altogether "causing a decade of lost capacity in quantitative expertise."<sup>108</sup>

<sup>107</sup> Ernie Boyko and Wendy Watkins, "Canada's Data Liberation Initiative," International Household Survey Network, 2011, <http://ihsn.org/home/node/102>.

<sup>108</sup> Boyko, "Canada's Data Liberation Initiative,"

## Recommendations

The analysis has demonstrated how OD can provide social, economic and environmental benefits to society. Several challenges surrounding the dissemination of OD are still preventing these benefits from being achieved. This section provides recommendations for obtaining further benefits with the dissemination of OD.

The five recommendations outlined in this section stem from the analyses, with the objective of achieving the most benefits to society.

### Availability of Data

Stronger action is needed by senior officials to prevent issues with compliancy or else availability of datasets will suffer from economic cutbacks. GC departments and agencies need to understand that they are obligated to publish datasets and OD practices need to be part of their IM practices. Legislation surrounding OD along with guidelines and tools needs to be in place to help departments and agencies manage PSI and publish OD.

**Recommendation #1:** The GC needs to launch the Directive on Open Government to help departments and agencies publish more datasets. As part of the requirement of the directive, there must be standard tools and guidelines such as criteria for publishing that will help departments and agencies identify and publish OD.

In addition, the directive should not allow departments and agencies to pick and choose the data that will be published. Instead, the directive must mandate them to publish all data that meets the criteria for publishing openly. This must also include data from access to information requests and public web sites.

Similar to the Directive on Recordkeeping, the new Directive on Open Government must also contain deadlines and consequences for not meeting requirements. To mitigate the issue of compliancy, requirements of the directive must be key objectives within the performance appraisals for all Information Management Senior Officers (IMSO) and Chief Information Officers (CIO) from each department and agencies.

Furthermore, a communication campaign promoting the importance of the directive should be led by the highest level of senior officials within the GC. Departments and agencies need to understand the urgency and benefits that can be obtained from publishing OD. With effective communication and a proper mandate, departments and agencies will have the ability to achieve compliance and make more data openly available.

**Recommendation #2:** The GC should set a concrete goal to convert inert data available on their websites into a dynamic open format within the next year. This includes data that is available on public websites from all departments and agencies. If information is publicly available on government websites then it must also be available openly. Data that is only available on physical disks or in a printed format must be converted.

A large amount of OD users are scraping public websites without appropriate licenses. This is an obvious indication that information published on public websites is both needed by OD users and currently unavailable in open format. With a growing number of users wanting to use OD in subsequent projects and the benefits that can be achieved from making more data available, this is a cost-effective approach for the GC to publish data that users are requesting. In addition, this would help support the academic community which are the biggest users of OD and would improve transparency with advocacy and special interest groups which lobby governments for more data to be published.

## Standards

There are several initiatives from other jurisdictions that could be leveraged to help implement standards, guidelines, and lessons learned from other government departments.

**Recommendation #3:** The GC needs to identify guidelines and standards for the publishing of OD. A lack of consistency of common formats across departments was identified as an issue for combining datasets which was caused by the lack of defined standards and procedures. This lack of standards only increases costs and time delays for groups needing to aggregate datasets. Additionally, since members of the academic community are the biggest users of macro data available from the GC, the lack of standards is obstructing future innovation from research and development within Canada.

The GC needs to work in collaboration with other jurisdictions to establish standards for metadata and geographic information. Furthermore, tools and procedures are required to help departments and agencies to manage OD. Existing work within other departments like Statistics Canada and Natural Resources Canada could be leveraged for the process of identifying potential standards for publishing geospatial datasets. In addition, the GC should participate in international initiatives for the development of standards and best practices. The European Commission is working on an initiative for a common standard, as is the Open Data Institute in the UK. Providing common standards will help aggregated data and gain added benefits from OD.

## Dynamic Data

The GC needs to provide mechanisms that will allow information to flow to all stakeholders. External stakeholders should be seen as partners in the effort to publish OD.

**Recommendation #4:** The GC needs to collaborate with external stakeholders and all departments and agencies. TBS needs to be the catalyst for establishing a degree of interaction for the continuous flow of information. Provide the ability to change the patterns for interaction among existing and new stakeholders. This interaction requires that data reaches its intended audience and that mechanisms are in place to allow users to contribute. This two-directional flow will allow for transparency from the benefit of public accountability.

The first step of the circular process could be to leverage external stakeholders to validate the need for new datasets. A series of stakeholders that are subject matter experts could be identified to provide feedback. This would give valuable insight into the value of the datasets and enhance the quality of OD published by the GC with relatively no cost.

Furthermore, an advisory board needs to be created to allow stakeholders to collaborate with the publication of OD. External parties should be seen as partners in the effort to publish OD; the skills and efforts for in depth analysis that they offer could help governments reduce internal costs for identifying and publishing OD. External participation will also help create a sense of urgency for OD and remove complacency. In addition, this level of collaboration with external stakeholders would allow for the interaction needed for public accountability. Information will reach targeted audiences, provide a mechanism to react, and allow GC to respond.

**Recommendation #5:** Changes within GC departments and agencies need to include a lean operational process for publishing OD. An interdisciplinary team within each organization will need to be identified and implemented which will use a rapid and iterative approach to publishing OD. The team will work with internal divisions to identify efficiencies and develop processes for publishing needed data in days and weeks instead of months and years. It will be important for organizations to identify key individuals with the experience and knowledge to propel lean operational changes. In addition, the team will help implement internal procedures and best practices that will meet legislative obligations for publishing OD.

## Implementation Steps

Several of the recommendations that have been made will require the involvement and participation of all departments and agencies, extending beyond the responsibility and reach of the Treasury Board of Canada.

The first step to this high-level plan is to develop and implement a communication strategy at the GC level that will be able to reach out across all departments and agencies. IMSOs and CIOs from each organization need to be made aware of how OD practices will affect them. This strategic plan will provide horizontal communication between government departments and agencies and help to share the requirements of the Directive on Open Government and maintain a sense of urgency for OD. Furthermore, the communication plan needs to be championed and governed by the President of the Treasury Board. The plan should also be used to communicate the goal of converting inert data available from their websites into an open format within the next year.

The next step will be to create an Open Data advisory board (ODAB) with external stakeholders and all departments and agencies. This board needs to be a working group with subject matter experts responsible for the development and implementation of best practices and guidelines. The advisory board would report to the Advisory Panel on Open Government. A representative from each granting council would need to participate and help with the creation of a partnership with institutions and research initiatives. The board would advise the Treasury Board on the implementation of best practices and key principles including the development of an OD framework for standards and guidelines, as well as a checklist to help assess and validate information for dissemination. In addition, several tools and policies could be re-used from other departments. A review of the operational policies and best practices developed and implemented by Natural Resource Canada would provide significant insight. The advisory board would also be responsible for helping government organizations establish a lean operational process with an interdisciplinary team.

The third step needs to include the implementation of the Directive on Open Government across departments and agencies. Objectives related to the requirements of the directive needs to be part of the performance appraisal of all IMSO's and CIO's for the fiscal year 2014-2015. Tools and guidelines for the implementation of lean operational teams should accompany the directive.

## Conclusion

Many government organizations are currently seeking effective methods for publishing OD and are implementing strategic objectives for creating transparency and openness while attempting to support innovation and economic growth. However, governments will need to review their internal practices and procedures related to information management. The ability to effectively re-use information is dependent on how well an organization is able to manage information effectively. The cycle of collecting and organizing information will prove beneficial if changes are made that align with the principles of OD.

Internal changes will build trust with communities, and governments will need to be as open as possible by publishing all their data. The change will also need to be driven from the top down in order to maintain a sense of urgency. Legislation alone will not remove public scrutiny, a new organizational context that imposes new roles and responsibilities will be needed to mitigate the issue of complacency.

A closer look at the stakeholders will provide insight into what data is needed. External stakeholders are lobbying for the release of more data to every level of government. The data that provides the most benefit consists of national or international level data which includes statistical, population and geographical data. These are the types of data that have the most benefit and impact for research and innovation. Economic benefits are expected to increase with the use of OD, but the growth is predicted to increase slowly on a long-term basis. Making OD available to external stakeholders will help direct the publication of OD and will lead to the development of applications servicing citizens in new and innovative ways. The value created by the market for the development of applications is only just beginning, but has a strong potential for growth. Research communities with the skills and tools to create value with OD will profit from innovation and will in turn improve the Canadian economy.

Standards will also emerge and provide better tools for sharing. The worldwide movement of OD will create sharing practices across nations with new standards, best practices and guidelines that will make sharing of information easier than ever. Knowledge and wisdom is derived from information and be effective it needs to reach the right audience at the right time.

OD can achieve many benefits, and several governments have already demonstrated the ways in which transparency and openness can help multiple stakeholders. These governments have realized that a change-friendly culture of openness is required. Governments need to learn to share the data they use to inform themselves and the people they serve. In addition,

governments need to understand that data is more valuable when it is aggregated with other data sources. It is information from aggregated data that provides the benefits for supporting research, innovation and economic growth. OD is not about return on investment as much as it is about creating value.

## Bibliography

Amazing Mobile Apps. City of Toronto. Last modified April 24, 2013.

<http://www1.toronto.ca/wps/portal/contentonly?vnextoid=7e57e03bb8d1e310VqnVCM10000071d60f89RCRD>.

Apps 4 Climate Action. British Columbia. Last modified April 24, 2013.

<http://www.livesmartbc.ca/A4CA/gallery.html>.

Benefits of Open Data. Government of Canada. Last modified October 10, 2012.

<http://www.data.gc.ca/default.asp?lang=En&n=1E4722A8-1>.

Bennis, Warren G., Daniel Goleman, James O'Toole, and Inc ebrary. Transparency how Leaders Create a Culture of Candor /. 1st ed. San Francisco, CA. Jossey-Bass. (2008).

<http://voyager.royalroads.ca/vwebv/holdingsInfo?bibId=244171>.

Best practices for sharing sensitive environmental geospatial data. Natural Resources Canada. Last modified February 6, 2012.

[http://ftp2.cits.rncan.gc.ca/pub/geott/ess\\_pubs/288/288863/cgdi\\_ip\\_15\\_e.pdf](http://ftp2.cits.rncan.gc.ca/pub/geott/ess_pubs/288/288863/cgdi_ip_15_e.pdf).

Braverman, Eric and Michael Chui. "Unleashing government's 'innovation mojo.'" McKinsey Quarterly. 2012.

[http://www.mckinsey.com/insights/public\\_sector/unleashing\\_governments\\_innovation\\_moj\\_o\\_an\\_interview\\_with\\_the\\_us\\_chief\\_technology\\_officer](http://www.mckinsey.com/insights/public_sector/unleashing_governments_innovation_moj_o_an_interview_with_the_us_chief_technology_officer).

Booth, Laurence and Sean W. Cleary. Introduction to Corporate Finance, 2nd ed. (Missasauga: Wiley, 2010).

Boyko, Ernie and Wendy Watkins. "Canada's Data Liberation Initiative." International Household Survey Network. 2011. <http://ihsn.org/home/node/102>.

Canada's Action Plan on Open Government. Government of Canada. Last modified April 11, 2013. <http://open.gc.ca/open-ouvert/ap-pa04-eng.asp#toc6>.

CBC News. "Feds 'not the most transparent,' says information commissioner." February 9, 2013. <http://www.cbc.ca/news/canada/story/2013/02/09/access-information-legault.html>.

Containers, Facebook, Baseball & the Dark Matter around Open Data. Eaves.ca. Last modified July 12, 2012. <http://eaves.ca/2012/07/12/containers-facebook-baseball-the-dark-matter-around-open-data-iogdc-keynote>.



Davies, Tim. "Open Data, Democracy and Public Sector Reform." A Look at Open Government Data use from Data.Gov.Uk. (2010) <http://practicalparticipation.Co.uk/odi/report/wp-content/uploads/2010/08/How-is-Open-Governmentdata-being-used-in-Practice.Pdf>.

DCAT application profile for data portals in Europe. European Commission Joinup. Last modified March 8, 2013. [https://joinup.ec.europa.eu/asset/dcat\\_application\\_profile/description](https://joinup.ec.europa.eu/asset/dcat_application_profile/description).

Dekkers, Makx and Femke Polman, Robbin te Velde, Marc de Vries. "Measuring European Public Sector Information Resources." Paris (2006). [http://ec.europa.eu/information\\_society/policy/psi/docs/pdfs/mepsir/executive\\_summary.pdf](http://ec.europa.eu/information_society/policy/psi/docs/pdfs/mepsir/executive_summary.pdf).

Denmark's open data initiative estimated to save government over \$45 million per year and generate returns of \$87 million for the private sector. Between the Poles. Last modified October 14, 2012. [http://geospatial.blogs.com/geospatial/2012/10/denmarks-open-data-initiative-estimated-to-save-government-over-45-million-per-year-and-generate-ret.html?utm\\_source=feedburner&utm\\_medium=feed&utm\\_campaign=Feed%3A+typepad%2Fzeissg%2Fgeospatial+%28Between+the+Poles%29](http://geospatial.blogs.com/geospatial/2012/10/denmarks-open-data-initiative-estimated-to-save-government-over-45-million-per-year-and-generate-ret.html?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed%3A+typepad%2Fzeissg%2Fgeospatial+%28Between+the+Poles%29).

Directive on Recordkeeping. Government of Canada. Last modified July 18, 2011. <http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=16552>.

Donovan, Kevin. "Charity rules beefed up." Toronto Star Newspapers Ltd. December 21, 2007. [http://www.thestar.com/news/investigations/2007/12/21/charity\\_rules\\_beefed\\_up.html](http://www.thestar.com/news/investigations/2007/12/21/charity_rules_beefed_up.html).

Eaves, David. "Rules are no substitute for cultivating a culture of open government." Toronto Star Newspapers Ltd. April 5, 2013. <http://www.thestar.com>.

Eaves, David. Calculating the Value of Canada's Open Data Portal: A Mini-Case Study. Eaves.ca. Last modified March 8, 2012. <http://eaves.ca/2012/03/08/calculating-the-value-of-canadas-open-data-portal-a-mini-case-study/>.

Executive Order -- Making Open and Machine Readable the New Default for Government Information. The White House. Last modified May 9, 2013. <http://www.whitehouse.gov/the-press-office/2013/05/09/executive-order-making-open-and-machine-readable-new-default-government>.

- Formative Evaluation of LAC Strategic Outcomes 1.0. Library and Archives Canada. Last modified December 21, 2012. <http://www.collectionscanada.gc.ca/evaluations-verification/012014-380-e.html>.
- Geospatial Standards and Operational Policies. Natural Resources Canada. Last modified April 12, 2012. <http://geoconnections.nrcan.gc.ca/1017>.
- Ghuri, Pervez, and Kjell Gronhaug. Research Methods in Business Studies Fourth Edition. Essex England: Pearson Education Inc., 2010.
- Gustke, Constance. "All the Options." Conference Board Review 46, no. 1 (Jan, 2009): 18-23. <https://ezproxy.royalroads.ca/login?url=http://search.ebscohost.com.ezproxy.royalroads.ca/login.aspx?direct=true&db=buh&AN=36615374>.
- Harrison, Teresa, Theresa Pardo, Anthony Cresswell and Meghan Cook. "Delivering Public Value Through Open Government." Center for Technology in Government. (2012). [http://www.ctg.albany.edu/publications/issuebriefs/opengov\\_pubvalue](http://www.ctg.albany.edu/publications/issuebriefs/opengov_pubvalue).
- Heath, Chip. "Making the Emotional Case for Change." McKinsey Quarterly no. 2 (06, 2010): 88-97. <https://ezproxy.royalroads.ca/login?url=http://search.ebscohost.com.ezproxy.royalroads.ca/login.aspx?direct=true&db=buh&AN=49469781&site=ehost-live>.
- Helbig, Natalie C. Cresswell, Anthony M. Burke, G. Brian and Luis Luna-Reyes. "The Dynamics of Opening Government Data." Center for Technology in Government. (2012). <http://www.ctg.albany.edu/publications/reports/opendata>.
- Horizontal Internal Audit of Electronic Recordkeeping in Large Departments and Agencies. Treasury Board of Canada Secretariat. Last modified December 9, 2012. <http://www.tbs-sct.gc.ca/report/orp/2011/itl-itg02-eng.asp>.
- How Recollect works in 5 easy steps. Recollect. Accessed on April 20, 2013. <https://recollect.net>.
- Information and Communications Technologies (ICT). Industry Canada. Last modified October 2012. [http://www.ic.gc.ca/eic/site/ict-tic.nsf/eng/h\\_it05838.html](http://www.ic.gc.ca/eic/site/ict-tic.nsf/eng/h_it05838.html).
- Irving, Francis. "Why the Government scraped itself." ScraperWiki Data. June 13, 2011. <http://blog.scrapewiki.com/2011/06/13/why-the-government-scraped-itself/>.

Knowledge for everyone. Open Data Institute. Last modified November 30, 2012.

<http://www.theodi.org/>.

Kotter, John P. "Developing a Change-Friendly Culture: an interview with John P. Kotter." Leader to Leader 48 (2008): 33-38.

<http://onlinelibrary.wiley.com.ezproxy.royalroads.ca/doi/10.1002/ltl.278/pdf>.

Letter to Government Departments on Opening Up Data. The Official Site of the British Prime Minister's Office. Last modified April 24, 2013.

<http://webarchive.nationalarchives.gov.uk/20130109092234/http://number10.gov.uk/news/letter-to-government-departments-on-opening-up-data/>.

Letter on open government for the President of the Treasury Board. Office of the Information Commissioner of Canada. Last modified January 19, 2012. [http://www.oic-](http://www.oic-ci.gc.ca/eng/rr-sl-odi-adi_2012_1.aspx)

[ci.gc.ca/eng/rr-sl-odi-adi\\_2012\\_1.aspx](http://www.oic-ci.gc.ca/eng/rr-sl-odi-adi_2012_1.aspx).

Lillie, Naomi. "The new PSI Directive – as good as it Seems?" Open Knowledge Foundation. Last modified April 19, 2013. [http://blog.okfn.org/2013/04/19/the-new-psi-directive-as-](http://blog.okfn.org/2013/04/19/the-new-psi-directive-as-good-as-it-seems/)

[good-as-it-seems/](http://blog.okfn.org/2013/04/19/the-new-psi-directive-as-good-as-it-seems/).

Lucey, John J. "Why is the Failure Rate for Organisation Change so High?" Management Services 52, no. 4 (2008): 10-18.

<https://ezproxy.royalroads.ca/login?url=http://search.proquest.com/docview/234257649?accountid=8056>.

Manyika, James and Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, Angela Hung Byers. "Big data: The next frontier for innovation, competition, and productivity." McKinsey & Company (2011).

[http://www.mckinsey.com/insights/business\\_technology/big\\_data\\_the\\_next\\_frontier\\_for\\_innovation](http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation).

Marjanovic, Steven. "'Screen Scraping' Lawsuit may Clarify Liability Issue." American Banker 165, no. 18 (01/27, 2000): 7.

<https://ezproxy.royalroads.ca/login?url=http://search.ebscohost.com.ezproxy.royalroads.ca/login.aspx?direct=true&db=buh&AN=2756797>.

Mehta, Diana. "Free census data a big perk but public site experiencing growing pains." Winnipeg Free Press. April 18, 2012.

<http://www.winnipegfreepress.com/fpnewstopstory/free-census-data-a-big-perk-but-public-site-experiencing-growing-pains-147879915.html>.

Mobile App Gallery. City of Ottawa. Last modified April 24, 2013. <http://ottawa.ca/en/mobile-apps-and-open-data>.

Munro, Margaret. "Information watchdog to probe feds' alleged muzzling of scientists." Edmonton Journal, April 2, 2013. <http://www2.canada.com/edmontonjournal/news/story.html?id=915b635d-e5b8-40fa-8902-dcd3a0217ffa&p=1>.

Ogilvy, Jay and Schwartz, Peter. "Plotting Your Scenarios." Global Business Network. 1998. [http://www.gbn.com/articles/pdfs/gbn\\_Plotting%20Scenarios%20new.pdf](http://www.gbn.com/articles/pdfs/gbn_Plotting%20Scenarios%20new.pdf).

Open Data. Government of Canada. Last modified October 24, 2012. <http://www.data.gc.ca/>.

Open Data Applications. Government of Canada. Last modified October 24, 2012. <http://www.data.gc.ca/default.asp?lang=En&n=CF8AAC8B-1>.

Open Government. Office of the Information Commissioner of Canada. Last modified September 1, 2010. [http://www.oic-ci.gc.ca/eng/rp-pr-ori-ari\\_2010\\_1.aspx](http://www.oic-ci.gc.ca/eng/rp-pr-ori-ari_2010_1.aspx).

Open Government Initiative. The White House President Barack Obama. Last modified December 8, 2009. <http://www.whitehouse.gov/>.

Open Information and Open Data Policy. DataBC. Last modified April 12, 2013. [http://www.cio.gov.bc.ca/local/cio/kis/pdfs/open\\_data.pdf](http://www.cio.gov.bc.ca/local/cio/kis/pdfs/open_data.pdf).

Operating Principles for data.gc.ca. Government of Canada. Last modified October 24, 2012. <http://www.data.gc.ca/default.asp?lang=En&n=89BD0D43-1>.

Ottawa Open Data App Contest. City of Ottawa. Last modified May 13, 2013. <http://www.apps4ottawa.ca/>.

Paikin, Steve. "Time for Open Government." The Agenda. April 28, 2010. <http://www.youtube.com/watch?v=KCJooWIHLyc>.

Peixoto, Tiago. "The Uncertain Relationship between Open Data and Accountability: A Response to Yu and Robinson's 'The New Ambiguity of Open Government'" Social Science Research Network (2013): 202-213. <http://ssrn.com/abstract=2264369>.

- Puccio, Gerard J., Marie Mance, and Mary C. Murdock. *Creative Leadership: Skills that drive change* (2nd ed.). Thousand Oaks, CA: Sage Publications, 2011.
- Records and Information Life Cycle Management. Library and Archives Canada. Last modified April 19, 2006. <http://www.collectionscanada.gc.ca/007/002/007002-2012-e.html>.
- Readiness Assessment Tool. The World Bank. Last modified April 12, 2013. <http://data.worldbank.org/about/open-government-data-toolkit/readiness-assessment-tool>.
- Rowley, Jennifer. "The wisdom hierarchy: representations of the DIKW hierarchy." *Journal of Information Science* 33, (2007). <http://jis.sagepub.com.ezproxy.royalroads.ca/content/33/2/163.refs>.
- Vickery, Graham. "Review of recent studies on PSI re-use and related market developments." *Information Economics*. Paris (2011). [http://www.unic.pt/images/stories/publicacoes6/psi\\_final\\_version\\_formatted-1.pdf](http://www.unic.pt/images/stories/publicacoes6/psi_final_version_formatted-1.pdf).
- Weick, Karl E. and Robert E. Quinn. "Organizational Change and Development." *Annual Review of Psychology* 50, no. 1 (02, 1999): 361. <https://ezproxy.royalroads.ca/login?url=http://search.ebscohost.com.ezproxy.royalroads.ca/login.aspx?direct=true&db=buh&AN=1776889&site=ehost-live>.
- Yam, Philip. "How to Kick-Start Innovation with Free Data." *Scientific American*. March 23, 2013. <http://www.scientificamerican.com/article.cfm?id=how-to-kick-start-innovaton>.
- Yiu, Chris. "A Right to Data: Fulfilling the promise of open public data in the UK." *Policy Exchange*. 2012. <http://www.policyexchange.org.uk/digital-government>.
- Yu, Harlan and Robinson, David G. "The New Ambiguity of 'Open Government'" *Social Science Research Network* (2012): 180-208. <http://dx.doi.org/10.2139/ssrn.2012489>.

## Appendices

### Appendix 1: Survey Questions

Question 1 – Are you a Canadian citizen?

Yes                       Other Country: \_\_\_\_\_

Question 2 – Your Age Group

Under 18       18 – 30       31 – 45       46 – 65       65 and over

Question 3 – What is the level of education you have completed?

- |  |                                      |
|--|--------------------------------------|
| <input type="radio"/> Secondary School     | <input type="radio"/> Masters Degree |
| <input type="radio"/> Further Education    | <input type="radio"/> Doctorate      |
| <input type="radio"/> Diploma, Certificate | <input type="radio"/> Other: _____   |
| <input type="radio"/> Undergraduate Degree |                                      |

Question 4 – What was your field of education?

- |   |   |
|---|---|
| <input type="checkbox"/> Arts and Humanities              | <input type="checkbox"/> Mathematics                    |
| <input type="checkbox"/> Computer Science                 | <input type="checkbox"/> Politics, Public Policy or Law |
| <input type="checkbox"/> Engineering and Physical Science | <input type="checkbox"/> Social Sciences                |
| <input type="checkbox"/> Life Sciences                    | <input type="checkbox"/> Statistics                     |
|   | <input type="checkbox"/> Other: _____                   |

Question 5 –What is your employment status?

- |                                      |                                     |
|--------------------------------------|-------------------------------------|
| <input type="radio"/> Student        | <input type="radio"/> Employed      |
| <input type="radio"/> Volunteer      | <input type="radio"/> Unemployed    |
| <input type="radio"/> Self Employed  | <input type="radio"/> Retired       |
| <input type="radio"/> Business owner | <input type="radio"/> Other:: _____ |

Question 6 – Do you associate yourself with the following labels?

- |  |  |   |
|--|--|---|
| <input type="checkbox"/> Web Developer | <input type="checkbox"/> IM Specialist | <input type="checkbox"/> Consultant         |
| <input type="checkbox"/> Researcher    | <input type="checkbox"/> Activist      | <input type="checkbox"/> Politician         |
| <input type="checkbox"/> Journalist    | <input type="checkbox"/> Entrepreneur  | <input type="checkbox"/> Data specialist    |
| <input type="checkbox"/> Blogger       | <input type="checkbox"/> Citizen       | <input type="checkbox"/> Open data advocate |
| <input type="checkbox"/> Policy Maker  | <input type="checkbox"/> Civil servant | <input type="checkbox"/> Other: _____       |
| <input type="checkbox"/> IT Specialist | <input type="checkbox"/> Designer      |   |

Question 7 – Which best describes your organization?

- |  |  |  |
|--|--|--|
| <input type="checkbox"/> Academic institutions         | <input type="checkbox"/> Non-governmental organization (NGO) | <input type="checkbox"/> Publicly funded company |
| <input type="checkbox"/> Clubs or association          | <input type="checkbox"/> Non-profit organization             | <input type="checkbox"/> Publicly traded company |
| <input type="checkbox"/> Community-based organizations | <input type="checkbox"/> Privately held company              | <input type="checkbox"/> Self-employed           |
| <input type="checkbox"/> Federal Government            | <input type="checkbox"/> Provincial Government               | <input type="checkbox"/> N/A                     |
| <input type="checkbox"/> Municipal Government          |  | <input type="checkbox"/> Other: _____            |

Question 8 – What size is the organization?

- |  |  |  |
|--|--|--|
| <input type="radio"/> Just me                  | <input type="radio"/> Between 51 and 100 people  | <input type="radio"/> Between 501 and 1000 people  |
| <input type="radio"/> Less than 10 people      | <input type="radio"/> Between 101 and 200 people | <input type="radio"/> Between 1001 and 5000 people |
| <input type="radio"/> Between 10 and 50 people | <input type="radio"/> Between 201 and 500 people | <input type="radio"/> More than 5000 people        |
|  |  | <input type="radio"/> N/A                          |

Question 9 – Which of the following sources of open data have you explored or made use of?

Question 10 – What is your motivation for working with open data?

- |   |  |
|---|--|
| <input type="checkbox"/> Curiosity  | <input type="checkbox"/> Being recognized as the creator of something useful or insightful |
| <input type="checkbox"/> Getting a better understanding of government     | <input type="checkbox"/> Providing a better service to citizens or customers               |
| <input type="checkbox"/> Building an innovative website /service          | <input type="checkbox"/> Providing a platform for other people to build upon               |
| <input type="checkbox"/> Building an innovative mobile application        | <input type="checkbox"/> Solving a specific problem  |
| <input type="checkbox"/> Building a report for an article or presentation | <input type="checkbox"/> Democratic deliberation   |
| <input type="checkbox"/> Creating a mapping mash-up                       | <input type="checkbox"/> Inform citizen about elected officials                            |
| <input type="checkbox"/> Making a profit                                  | <input type="checkbox"/> Evidence based policy decision making                             |
| <input type="checkbox"/> Meeting the requests of a manager / client       | <input type="checkbox"/> To improve access to recreational information                     |
| <input type="checkbox"/> Providing insight to social programs             | <input type="checkbox"/> It is just fun to play with data                                  |
| <input type="checkbox"/> Making a difference to my local community        | <input type="checkbox"/> I like making apps  |
| <input type="checkbox"/> Making government more efficient                 | <input type="checkbox"/> Other: _____  |
| <input type="checkbox"/> Learning new skills                              |  |
| <input type="checkbox"/> Making government more accountable               |  |

Question 11 – Which categories of data are you most interested in obtaining?

- |   |   |   |
|---|---|---|
| <input type="checkbox"/> Working with Statistical data          | <input type="checkbox"/> Elevation and Derived Products data    | <input type="checkbox"/> Language and Linguistics data        |
| <input type="checkbox"/> Working with Location data             | <input type="checkbox"/> Environment and Conservation data      | <input type="checkbox"/> Legal and Administrative Data data   |
| <input type="checkbox"/> Working with Service data              | <input type="checkbox"/> Environment and Weather Data           | <input type="checkbox"/> Locations and Geodetic Networks data |
| <input type="checkbox"/> Working with Real-time data            | <input type="checkbox"/> Geographical data                      | <input type="checkbox"/> Military data                        |
| <input type="checkbox"/> Working with Research data             | <input type="checkbox"/> Geological and Geophysical data        | <input type="checkbox"/> Nature and Environment data          |
| <input type="checkbox"/> Administrative and Political data      | <input type="checkbox"/> Government and Politics data           | <input type="checkbox"/> Oceans and Coasts data               |
| <input type="checkbox"/> Agriculture and Farming data           | <input type="checkbox"/> Government Income and Spending data    | <input type="checkbox"/> Persons data                         |
| <input type="checkbox"/> Arts, Music, Literature data           | <input type="checkbox"/> Health and Safety data                 | <input type="checkbox"/> Population data                      |
| <input type="checkbox"/> Atmosphere and Climate data            | <input type="checkbox"/> History and Archaeology data           | <input type="checkbox"/> Processes data                       |
| <input type="checkbox"/> Biology and Ecology data               | <input type="checkbox"/> Housing Data                           | <input type="checkbox"/> Public Health and Disease data       |
| <input type="checkbox"/> Business Data                          | <input type="checkbox"/> Inland Water Resources data            | <input type="checkbox"/> Public Safety and Security data      |
| <input type="checkbox"/> Cadastral data                         | <input type="checkbox"/> Imagery and Basemaps data              | <input type="checkbox"/> Science and Technology data          |
| <input type="checkbox"/> Crime and Justice Data                 | <input type="checkbox"/> Information and Communications data    | <input type="checkbox"/> Society and Culture data             |
| <input type="checkbox"/> Cultural, Society and Demographic data | <input type="checkbox"/> International Aid and Development Data | <input type="checkbox"/> Trade data                           |
| <input type="checkbox"/> Economics and Industry data            | <input type="checkbox"/> Labour data                            | <input type="checkbox"/> Transport data                       |
| <input type="checkbox"/> Education and Training data            |   | <input type="checkbox"/> Web Analytics Data                   |
| <input type="checkbox"/>  |   | <input type="checkbox"/> None of the above                    |
|   |   | <input type="checkbox"/> Others: _____                        |

Question 12 – In your opinion how important are the following (Don't know, Not Important, Important, Very important, Extremely important)

- |   |   |
|---|---|
| <ul style="list-style-type: none"> <li>• Open data can be leveraged to develop consumer and commercial products</li> <li>• Open data can support innovation</li> <li>• All government data should be openly accessible online</li> <li>• Data should only be made open when there is a clear demand for it</li> <li>• Open data can be used to keep governments accountable</li> <li>• Open data can support informed decisions for consumers</li> <li>• Government data was paid for by citizens, so it should be accessible to citizens</li> <li>• Open data can be used to develop the semantic web / link web data</li> </ul> | <ul style="list-style-type: none"> <li>• Open data can be used to help reform public services</li> <li>• Companies can build services with open data</li> <li>• Open data can be used to inform the electorate</li> <li>• Government should provide tools that make it easier for citizens to view datasets</li> <li>• Government are not very good at building online services</li> <li>• Improving the quality of available data is more important than releasing new open government data</li> <li>• Communities will be able to use government data to solve local problems</li> <li>• Innovators from outside government will use</li> </ul> |
|---|---|

Question 13 – In relation to Open Data, what are the three most important issues facing government today?

1. \_\_\_\_\_ 2. \_\_\_\_\_ 3. \_\_\_\_\_

Question 14 – For each of the issues you have listed above, please indicate how important you think increased access to open government data will be in addressing these issues. (Don't know, Not Important, Important, Very important, Extremely important)

Question 15 – Please rank your preferred method of accessing data (Don't know, Not Important, Important, Very important, Extremely important)

- |  |  |   |
|--|--|---|
| <input type="checkbox"/> CSV - Comma Separated Values      | <input type="checkbox"/> RDF - Resource Descriptor Framework     | <input type="checkbox"/> Through an API (Application Programming Interface) |
| <input type="checkbox"/> Excel Spreadsheet                 | <input type="checkbox"/> KML - Keyhole Markup Language (geodata) | <input type="checkbox"/> Through an online data browser                     |
| <input type="checkbox"/> PDF File                          | <input type="checkbox"/> SQL Database Dump                       | <input type="checkbox"/> Through a SPARQL endpoint (for querying RDF data)  |
| <input type="checkbox"/> Web pages                         | <input type="checkbox"/> RDFa - RDF Embedded in web pages        | <input type="checkbox"/> Other: _____                                       |
| <input type="checkbox"/> Webservice or API                 | <input type="checkbox"/> Bulk data download                      |   |
| <input type="checkbox"/> XML - extensible Markup Language  |  |   |
| <input type="checkbox"/> JSON - JavaScript Object Notation |  |   |

Question 16 – Please select the tools and technologies you made use of for managing open data.

- |   |  |
|---|--|
| Spreadsheet (e.g. Excel)                              | Online mapping mash-up tools (free)    |
| Word processor (e.g. Word; Open Office Writer)        | Online data visualisation tools (free) |
| Web scripting language (e.g. PHP, Python, Ruby)       | Commercial mapping tools               |
| JSON  | Commercial data visualisation tools    |
| XML   | Statistical software                   |
| SPARQL  | Web mapping                            |
| Google Spreadsheets, Yahoo Pipes or similar platforms | Google Maps                            |
| Flash   | Geographic information system (GIS)    |
|   | Other: _____                           |

Question 17 – Please provide a short description of your open data project(s) or initiative(s).

Projects: \_\_\_\_\_

Question 18 – Who were the stakeholders of your project(s)?

- |  |  |  |
|--|--|--|
| <input type="checkbox"/> Academic institutions         | <input type="checkbox"/> Non-governmental organization (NGO) | <input type="checkbox"/> Publicly funded company |
| <input type="checkbox"/> Clubs or association          | <input type="checkbox"/> Non-profit organization             | <input type="checkbox"/> Publicly traded company |
| <input type="checkbox"/> Community-based organizations | <input type="checkbox"/> Privately held company              | <input type="checkbox"/> Self-employed           |
| <input type="checkbox"/> Federal Government            | <input type="checkbox"/> Provincial Government               | <input type="checkbox"/> N/A                     |
| <input type="checkbox"/> Municipal Government          |  | <input type="checkbox"/> Other: _____            |

Question 19 – Thinking about your projects, do you agree or disagree with the following: ( Strongly Agree, Agree, Neither, Disagree, Strongly Disagree)



## Appendix 2: Interview Questions for Internal Stakeholders

1. Most people believe that open data fosters innovation and helps society. Can you provide any examples of social or economic benefits from the release of open data?
2. In my survey, nearly 60% of participants agreed that governments are not very good at building online services. Should governments get out of the business of developing interfaces and stick to publishing raw data only?
3. Has your organization implemented (or in the process) policies and/or guidelines to publishing open data?
4. Has your organization implemented (or in the process) standards for metadata and/or data-sets?
5. What are your organization's direct or indirect costs of publishing open data? Please note that the cost of capturing data should not be included, these are costs and activities that would be incurred if the data is published or not.
  - a. Direct costs include the support and maintenance of preparing and hosting the data which would not exist if you were not publishing open data. These should also include the cost of human resources.
  - b. Indirect costs if available might include costs incurred by other divisions or sections

## Appendix 3: Focus Group Questions

### Purpose of Focus Group

To generate a diverse set of options to open data challenges. Explore ideas in a divergent manner and identify possible methods for public organization's to make sound and informative decisions for extending their Open Data initiatives with the hope of creating a more cost-effective, transparent, efficient and responsive government.

### Challenges:

1. In my survey, nearly 60% of participants agreed that governments are not very good at building online services. A challenge that GC faces is releasing data in machine-readable formats that enable's citizens, the private sector, and non-government organizations to leverage information in an innovative and value-added way. Should governments get out of the business of developing interfaces and stick to publishing raw data only?
2. The Quebec Minister Bernard Drainville said that an open and transparent government is required to fight corruptions, promote civic engagement and restore public trust. Knowing this, the issue of a government with an embedded organizational culture that does not share

## Appendix 4: Interview Questions for External Stakeholders

1. Most people believe that open data fosters innovation and helps society. Can you provide any examples of social or economic benefits from the release of open data?
2. In my survey, nearly 60% of participants agreed that governments are not very good at building online services. Should governments get out of the business of developing interfaces and stick to publishing raw data only?
3. Has your organization implemented (or in the process) policies and/or guidelines to publishing open data?
4. Has your organization implemented (or in the process) standards for metadata and/or data-sets?
5. What are your organization's direct or indirect costs of publishing open data? Please note that the cost of capturing data should not be included, these are costs and activities that would be incurred if the data is published or not.
  - a. Direct costs include the support and maintenance of preparing and hosting the data which would not exist if you were not publishing open data. These should also include the cost of human resources.
  - b. Indirect costs if available might include costs incurred by other divisions or sections
6. Customer Segments - For whom are you creating value?
7. Customer Segments - For whom are you creating value?
8. Value proposition - What value do you deliver to customers?
9. Customer Relation - What type of relationship does each of your Customer Segments expect you to establish and maintain with them?
10. Key resources - What Key Resources do your Value Propositions require?
11. Key Activities - What Key Activities do your Value Propositions require?
12. Cost structure – What is your cost model?
13. Key Partners - Who are your Key Partners?

## Appendix 5: Initiatives and Projects Utilizing Open Data

Initiatives and projects utilizing open data:

Find the safest way to walk your children to school <http://schoolzone.moappi.com/default.html>

Trees of the city of Montreal <http://www.quebio.ca/en/arbresmtl>

Transit application, works in 37 cities <http://thetransitapp.com/>

BC Property Tax Comparatron <http://www.opendatabc.ca/bc-property-tax-comparatron.html>

Free flu vaccine clinics <http://flushot.newurbanmechanics.org/>

Claim responsibility for shoveling out a fire hydrant <http://adoptahydrant.org/>

Adopt-a-Sidewalk <http://www.adoptasidewalk.org/>

Ottawa Snow Day <http://snowdayshutdowns.com/>

Interactive info graphics <http://infogr.am/>

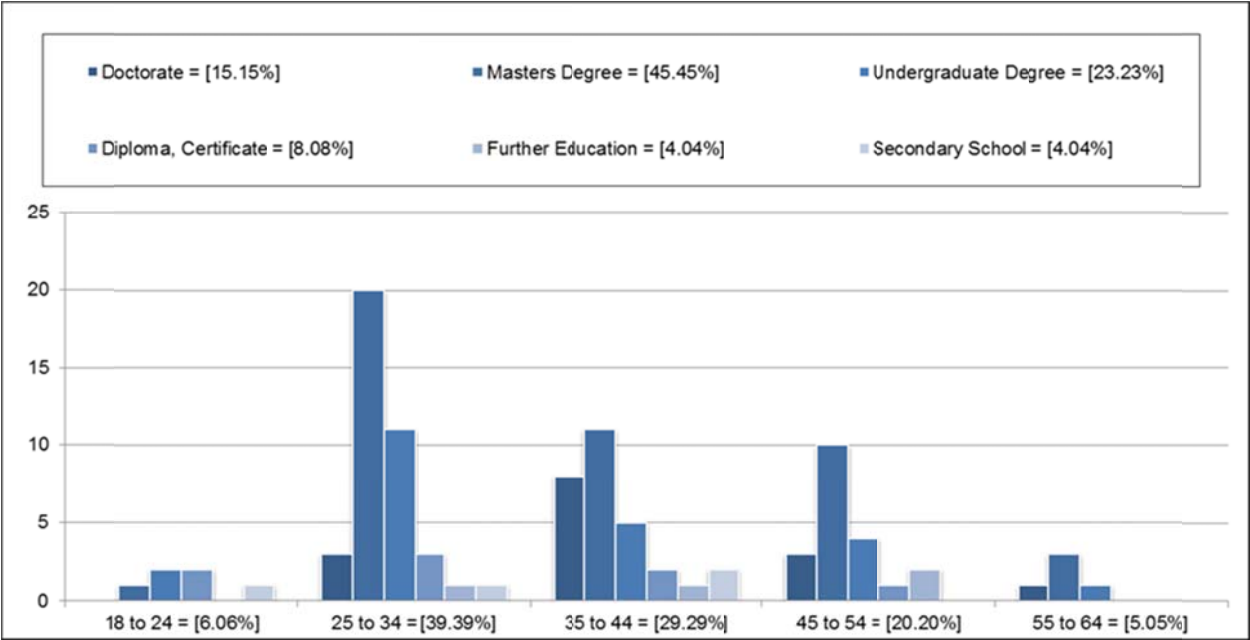
Water and environmental data <http://waterandenvironmentalhub.ca/>

# Appendix 6: Statistical Data from Survey Participants

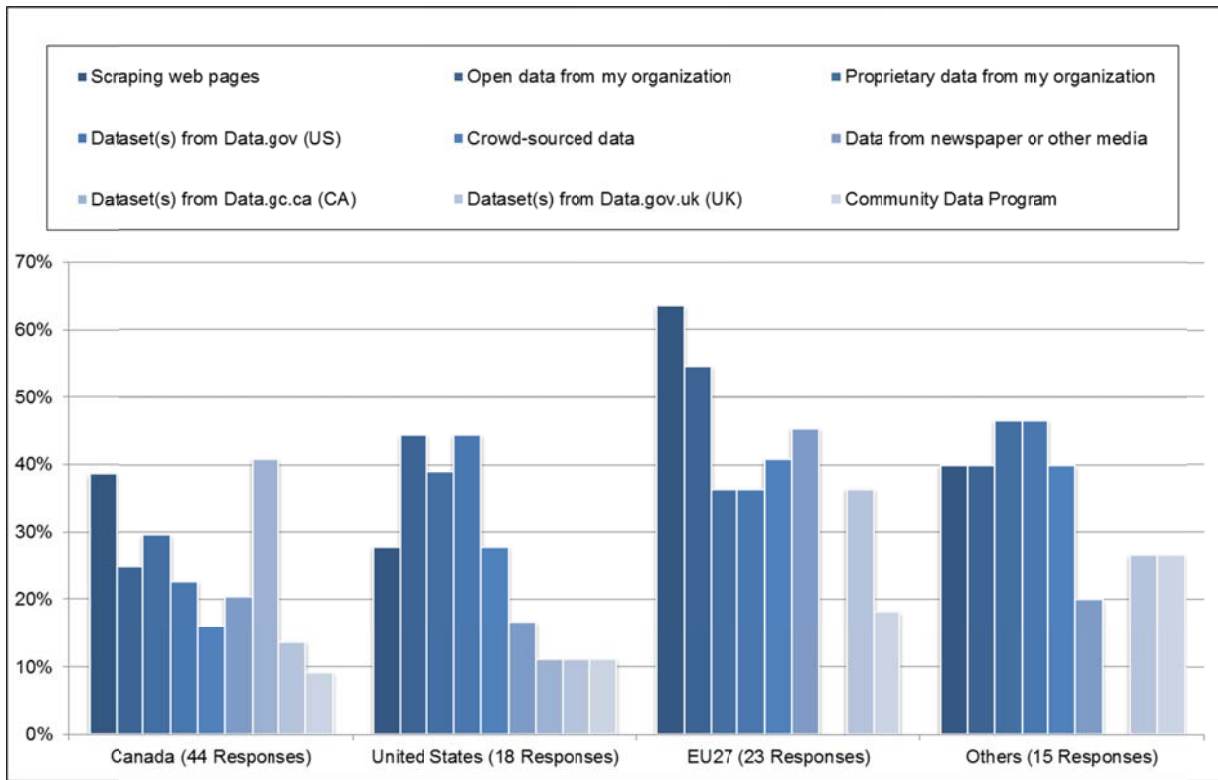
## a. Country Participants

Country	Requests	Responses	Percentage
1 Canada	14	44	44.44%
2 United States	12	18	18.18%
3 Spain	1	7	7.07%
4 Great Britain	7	5	5.05%
5 India	0	3	3.03%
6 France	0	2	2.02%
7 Italy	1	2	2.02%
8 Australia	1	2	2.02%
9 Greece	0	2	2.02%
10 Netherlands	3	2	2.02%
11 Norway	0	2	2.02%
12 South Africa	0	1	1.01%
13 Chile	0	1	1.01%
14 Peru	0	1	1.01%
15 Uruguay	0	1	1.01%
16 New Zealand	0	1	1.01%
17 Mexico	0	1	1.01%
18 Macedonia	0	1	1.01%
19 China	0	1	1.01%
20 Hungary	0	1	1.01%
21 Germany	0	1	1.01%
22 Singapore	1	0	0.00%
23 Switzerland	2	0	0.00%
	<b>42</b>	<b>99</b>	

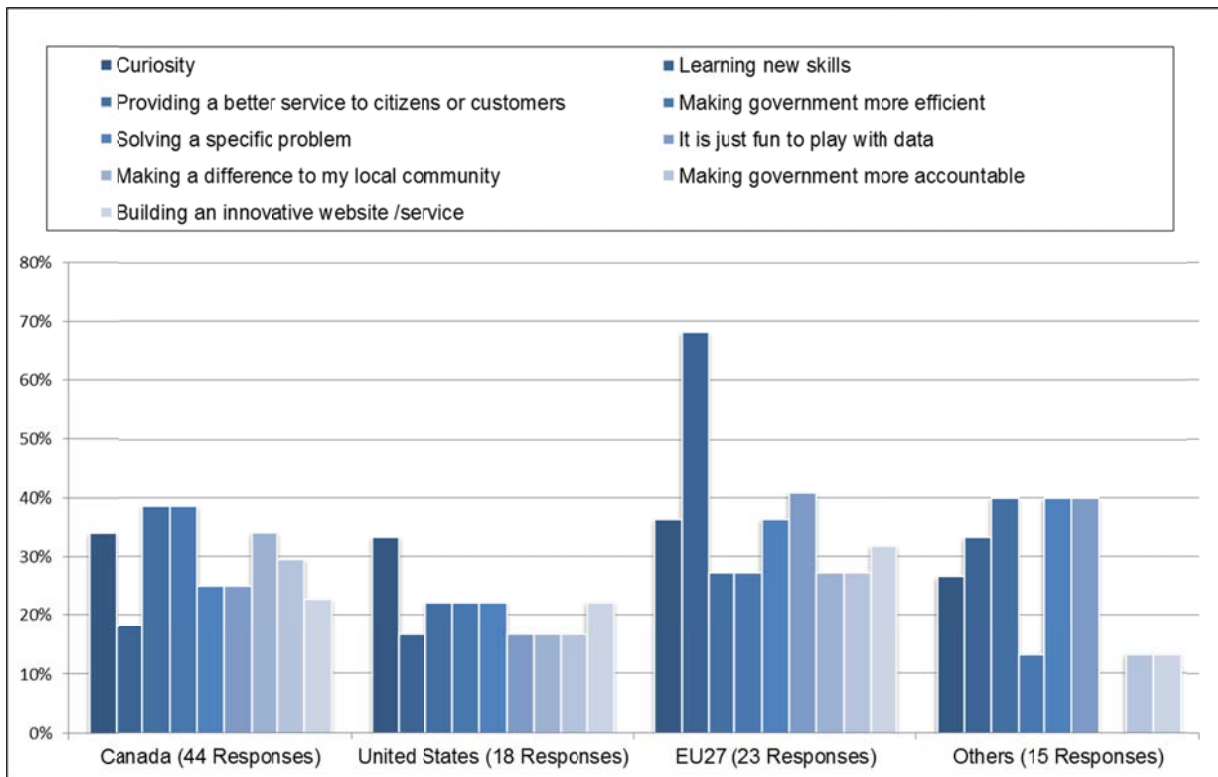
## b. Age group and education



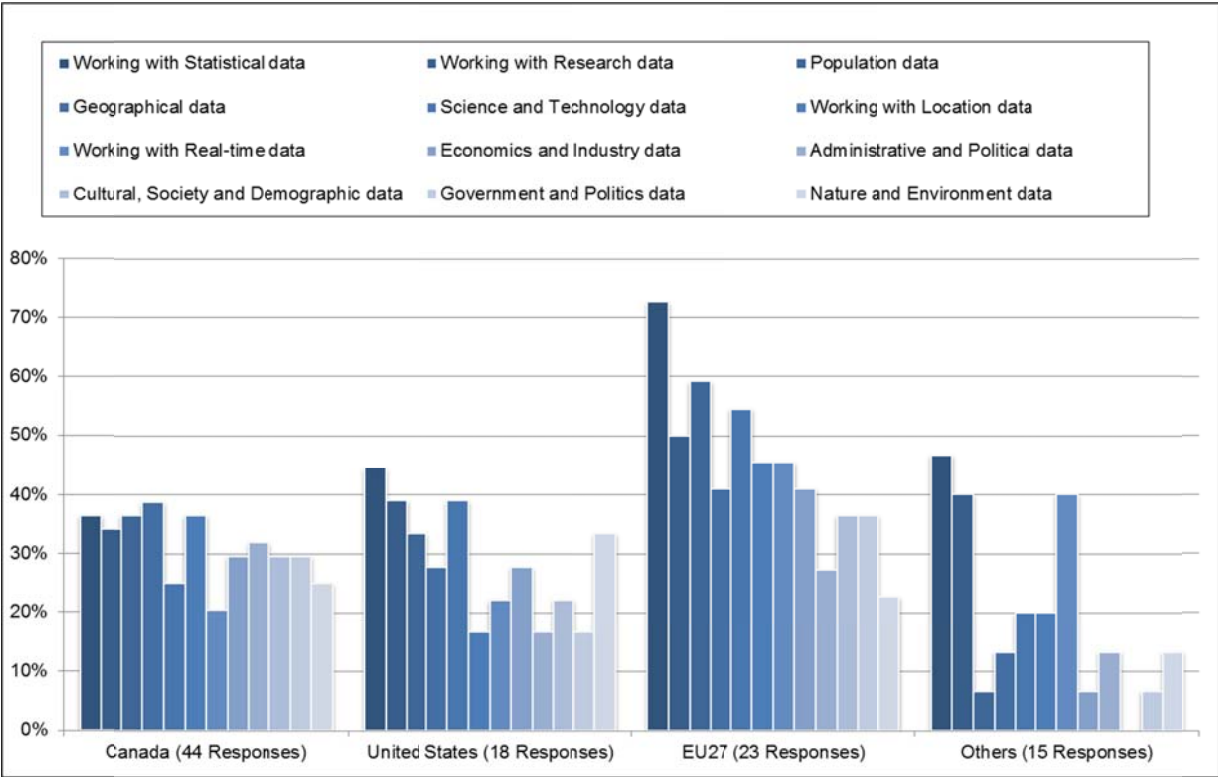
### c. Sources of data by Nations



### d. What motivates people to work with Open Data?



**e. What type of Open Data interests them?**



**f. Preferred methods for accessing the data**

Method	Don't Know	Not Important	Important	Very Important	Very Important	Extremely Important
CSV - Comma Separated Values	5%	9%	27%	21%	21%	38%
Webservice or API	10%	3%	30%	23%	23%	31%
XML - extensible Markup Language	6%	13%	19%	34%	34%	23%
Through an API (Application Programming Interface)	13%	9%	21%	30%	30%	22%
JSON - java script Object Notation	25%	10%	26%	16%	16%	18%
Bulk data download	13%	14%	36%	18%	18%	17%
RDF - Resource Descriptor Framework	34%	17%	22%	9%	9%	13%
SQL Database Dump	14%	19%	30%	22%	22%	13%
Web pages	4%	13%	49%	21%	21%	10%
Through an online data browser	5%	13%	38%	31%	31%	9%
KML - Keyhole Markup Language (geodata)	23%	10%	27%	29%	29%	8%
Through a SPARQL endpoint (for querying RDF data)	44%	16%	21%	12%	12%	5%
Excel Spreadsheet	5%	29%	35%	25%	25%	4%
PDF File	6%	55%	26%	8%	8%	4%
RDFa - RDF Embedded in web pages	42%	19%	27%	5%	5%	4%

## g. Important for Open Data initiatives

How important are the following	Not Important	Important	Very Important
Open data can support innovation	0%	25%	70%
Government data was paid for by citizens, it should be open to citizens	0%	24%	61%
All government data should be openly accessible online	4%	18%	60%
Open data can be used to inform the electorate	1%	42%	49%
Open data can be used to keep governments accountable	0%	52%	43%
Innovators can't use open data to build better online services	1%	43%	43%
Open data can help identify and correct errors within the data	4%	42%	43%
Open data can support informed decisions for consumers	1%	51%	40%
Government should provide tools that make it easier for citizens to view datasets	4%	31%	40%
Open data can be used to help reform public services	1%	45%	39%
Open data can be leveraged to develop consumer and commercial products	1%	49%	34%
Open data can be used to develop the semantic web / link web data	1%	36%	34%
Communities will be able to use government data to solve local problems	1%	52%	31%
Companies can build services with open data	0%	64%	30%
Government are not very good at building online services	10%	31%	28%
Improving the quality is more important than releasing new data	16%	31%	15%
There are many good reasons for restricting access to government datasets	28%	19%	9%
Data should only be made open when there is a clear demand for it	31%	18%	6%