EXAMINING THE VALUE OF SPATIAL VS. NON-SPATIAL OPEN DATA

by

Sarah Greene

Bachelor of Environmental Studies, University of Waterloo 2016

A Major Research Paper

Presented to Ryerson University

in partial fulfillment of the

requirements for the degree of

Master of Spatial Analysis

in the program of

Spatial Analysis

Toronto, Ontario, Canada, 2017

Author's Declaration

I hereby declare that I am the sole author of this MRP. This is a true copy of the MRP, including any required final revisions.

I authorize Ryerson University to lend this MRP to other institutions or individuals for the purpose of scholarly research.

I further authorize Ryerson University to reproduce this MRP by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

I understand that my MRP may be made electronically available to the public

Abstract

Examining the Value of Spatial vs. Non-Spatial Open Data

Sarah Greene

This research paper analyzes how the value of open data varies based on the goal of the open data program, and the format in which the data are provided. Four cities across Canada which make up the *G4* are examined to identify common themes of open data available, and assess the data formats found most often within these themes. Further, the City of Toronto is examined in a case study to evaluate their open data program and assess if spatial open data are more prevalent within the theme of innovation for economic development. Findings from this research indicate that there are some data themes which typically have mostly spatial and/or non-spatial data formats, while there is also a group of themes which have a wide variety of both formats available. This paper also finds that the City of Toronto has a high prevalence of spatial open data within the theme of innovation. The evaluation created for this study could be used in assessing the value of spatial open data within and between cities.

Acknowledgements

I'd like to thank Dr. Claus Rinner and Dr. Pamela Robinson for their time and efforts in helping me complete this paper. I'd also like to thank my parents, Mike and the Data Development team at Environics Analytics for all their support and encouragement along the way.

Table of Contents

List of Tables

List of Figures

List of Appendices

Chapter 1: Introduction

1.1 Research Context

Open data are provided by government organizations and can be freely used, re-used and redistributed by anyone (Sadiq & Indulska, 2017). They are available to private sector companies, Non-Governmental Organizations (NGOs), journalists, academics and residents through web-based portals (Johnson, 2016). These portals allow for the availability of data without delay, usually with no registration required (Johnson, 2016). The data are provided in a number of different formats depending on the resources available to the specific government organization (Johnson, 2016). These formats include pdf documents such as reports, Excel files such as spreadsheets of data, and geographic file formats including Keyhole Markup Language or Shapefiles which can be used with geographic information systems (Baculi, 2014). Open data is growing around the world with new portals being created and existing portals being enhanced to provide end-users with the data they want and/or need.

Open data programs provide significant benefits to Government organizations, along with their residents through a number of different avenues. By providing data openly, Governments are increasing their transparency to the public, enhancing their efficiency and are promoting economic development (Pereira et al., 2017; Zeleti & Ojo, 2017). These many benefits are what drive open data programs and are therefore components of how open data programs are perceived as successful (Sieber & Johnson, 2015; Thorsby et al., 2017). Open data programs tend to focus more heavily on economic development as a key benefit as it creates opportunities for local residents, private sector companies or NGOs to create innovative products and encourages them to be involved in entrepreneurial activities which can help the community as a whole (Sieber & Johnson, 2015). The economic benefits provided to these companies and/or organizations are then brought back to the government through taxes, efficiency improvements and overall job creation (Janssen et al., 2012). This creates a positive feedback loop, where both governments and end-users benefit from the open data provided.

The Open Data 150 is one current project which compiles a list of which private sector

companies and NGOs are using Canadian open data and how they are using it (Open Data 150 Canada, n.d.). Open Data 150 is part of an international project; The Open Data 500 Global Network which seeks to "study the use and impact of open data" (Open Data 500, n.d.). The focus of the overall international project and the specific Canadian study are on understanding the economic and social impacts of open government data (Open Data 500, n.d.). These programs are showcasing the importance of open data as a driver for economic development, encouraging the use of open data as a product for businesses and NGOs. Through the presence of open data in an array of different companies, it is clear that open data holds significant economic value to the community as a whole.

For a government to quantify if their open data program is successful, they may turn to some or all of the specified benefits to see if they are being realized within their community (Chan et al., 2016). To understand if an open data program is fulfilling its potential, there is a need to better understand what the main goals of the program are, and how they are, or can be realized (Zeleti & Ojo, 2017). A typical problem arises when those involved in an open data program focus only on the releasing of data, rather than understanding what data is valuable to the end-user, and in what format it will be most accessible to them (Johnson, 2016). Working towards the strategic release of open data will therefore provide greater benefits to all those involved.

1.2 Research Questions

Based on the relevance of geography and geospatial data for government operations and government-resident interactions, this research focuses on examining open data from the perspective of value related to geospatial data. The purpose of this study is to discover the frequency of spatial versus non-spatial open data, and further to discover if this varies by theme and towards particular program goals, specifically economic development. The following research questions will guide the study:

1. *What is the distribution of open data by general themes in Canadian open data*

catalogues

2. *Are spatial open data more prevalent within the theme of innovation which is used in achieving the goal of economic development*?

The aim of this paper is to provide evidence towards the prevalence of spatial open data towards the goal of economic development. The results of this study can therefore be used towards strategically planning for new, and assessing existing, open data programs in relation to their current and future goals of economic development and how it relates to the value provided to the end-user.

This paper focuses on examining existing Canadian open data catalogues at the municipal level to provide baseline statistics of what the general themes of data available are and which format the data are most commonly provided in. This baseline will then be used in a case study of the City of Toronto's open data catalogue to determine if the City's open data aligns with other catalogues and then further analyze the open data through examining additional statistics of the downloads, web clicks and general popularity of their open datasets. Using the baseline data coupled with a review of literature, an index was created and used on the City of Toronto's catalogue to assess spatial open data in relation to the goal of economic development.

1.3 Objectives

The first objective is to determine the general themes of data and their most commonly associated file type. This objective will focus on examining which data are published across different municipal open data programs, along with the file type they are most commonly associated with. This will also include gathering information on how many data files are published under each theme. The purpose of this objective is to gather summary statistics on the themes of open data and their associated file types to then be used as a guideline in further analysis undertaken in this study.

The second objective is to determine the prevalence of spatial open data within the theme of innovation towards the goal of economic development. This objective will also focus on providing a preliminary assessment of the City of Toronto's open data based on the documented access and use of spatial and/or non-spatial open data in their catalogue. The City of Toronto collects further statistics on website clicks and downloads associated with their open data, along with a catalogue of products creating using open data. By assessing which open data are viewed and/or downloaded most often, an assessment of which datasets are most popular to end-users can be assumed. Analyzing the products created using open data can help in understanding the value provided to the end-user, as well as how it relates back to the government itself, such as through greater transparency, economic development and/or improved efficiency.

The last objective is to provide an evaluation which can be used to assess spatial open data in relation to economic development. The purpose of this objective is to provide information related to what should be included when evaluating open data, based off of the results of this study. This can then be used towards the future evaluation of open data catalogues in relation to the goal of economic development.

1.4 Structure of the Research Paper

The paper begins with a review of current literature on the topic of open data, spatial vs. non-spatial data and different evaluation criteria. The literature review focuses on the topic of open data and the different initiatives behind it, while also developing the criteria for the evaluation to be used in this study. The study area, data, and methodology are then introduced. The index is developed, along with the analysis of Canadian open data catalogues. The results of the analysis are discussed in relation to the topics of strategic release of open data and tools for evaluating open data. The paper will conclude with final thoughts and recommendations for future research.

Chapter 2: Literature Review

2.1 Introduction

The discussion of open data in the scientific literature is recent due to its link to the evolution of e-governance and the Web 2.0. E-governance as defined by Sandoval-Almazan & Gil-Garcia (2012) is "the use of technology to enhance the access to and delivery of government services to benefit residents, business, partners and employees." This ties in with the Web 2.0 evolution which focuses on using web portals to create opportunities for better communication, collaboration and sharing within and between governments as well as with their residents (Sandoval-Almazan & Gil-Garcia, 2012). Open data programs fall within this category as they are found on web-based portals and/or webpages which focus on sharing data from governments to residents, or governments themselves free of charge (Gill & Corbett, 2017). Data catalogues are an increasingly popular format for the sharing of data, due to the end user's ability to download the data with ease, quickly, and at no cost (Borzacchiello & Craglia, 2012). Although there are no direct costs involved, there are secondary factors to be considered, such as the need for an internet connection to view and/or use the data that can exclude some users from being able to access the information (Sieber, R.E., & Johnson, P.A., 2015). Further, there are additional technical barriers towards accessing open data, further excluding some users due to their lack of knowledge on what certain datasets are, or how the dataset can be used, based on its format (Sieber & Johnson, 2015).

Open data may have evolved from e-governance through the idea of providing government information online, however, it has expanded past this to include various other key motivations and purposes including promoting transparency, efficiency, and economic development (Graves & Hendler, 2013). As open data continues to develop, the literature has expanded to focus on topics outside of examining solely the importance of open data programs in general, to further include evaluating existing programs based on program specific goals, successes and/or failures, along with the quality of datasets provided (Denham, 2013; WWW Foundation, 2016; Gill & Corbett, 2017).

This paper examines the value of spatial open data in achieving the goal of economic development for open data programs. To examine the value of spatial open data, there is a need to better understand what types of data are considered important to the end-user as well as which data are needed to meet the varying goals of open data programs. This chapter reviews the literature on the topic of evaluating open data and their associated programs. The first section focuses on understanding the evolution and importance of open data in its current state. The second section examines common data themes and file types found in open data catalogues, the varying value of open data and its relation to quality over quantity. The third section will examine the different motivations behind open data programs and how they relate to the wants and needs of end users. The fourth section will explore evaluation techniques and criteria used to evaluate open data catalogues and programs. This chapter will conclude with a summary of the literature.

2.2 Evolution and Importance of Open Data Programs across Canada

Open data programs are prominent around the world at varying levels of government such as at the federal, provincial, and municipal level. These levels of government may focus on releasing different quantities and types of datasets depending on their jurisdiction, and the department in which the open data program is located. As mentioned previously, much of the development of open data programs is due to the rising popularity of e-governance and the Web 2.0 evolution (Sandoval-Almazan & Gil-Garcia, 2012; Roy, 2014). Society has welcomed the internet as a platform and an array of increasingly advanced technology available within it, including online interactive tools and the expectancy of access to all of these new technologies on the go through mobile devices (Roy, 2014). This has created a shift towards an open and data-driven society (Roy, 2014). This includes two important factors that are found within open data programs; having information available openly by default, and providing statistics and other metrics on government related activities. These changes have been at the forefront behind the increasing popularity of open data, leading towards a steady increase of governments participating in creating and up keeping open data programs.

In Canada, open data became prominent following the success in countries such as the United States and the United Kingdom (Vogel, 2011). Prior to 2011, Canada was seen to be behind these countries, not yet fully engaged in the open data movement (Vogel, 2011). However, the momentum towards open data picked up with great speed, due to the federal government developing an *Open Government* strategy in March of 2011, and through developing an *Action Plan on Open Government* in 2012, which encourages and supports governments of all levels to provide data openly to their residents (Government of Canada, 2016; Gill & Corbett, 2017). Both of these documents outlined the importance of open data as a whole, including improving availability of information, encouraging resident participation in government, increasing professional and public integrity, improving public services, and improving efficiency throughout government operations (Government of Canada, 2016). As open data has continued to develop in Canada, there has been a focus at the municipal level in particular, which may be due to this level of government being closely connected to residents, therefore providing a great opportunity for governments to engage with these individuals. The pioneers of open data at the municipal level within Canada banded together after their inception to form a working group titled *the G4*, focused on exchanging current successes, ongoing problems, and overall providing support to one another, along with additional municipalities looking to develop their own open data programs (Giggey, 2012). The G4 cities are comprised of Vancouver, Edmonton, Toronto and Ottawa (Giggey, 2012). They have seen success through their involvement with one another, proving the need for local governments to collaborate and share not only with their residents, but with other government institutions as well (Giggey, 2012; Roy 2016). Through creating an atmosphere of collaboration and knowledge sharing, the G4 have helped share that open data programs are valuable and useful for residents, and have ample benefits for all parties involved.

The creation of strategic documents and evaluation criteria has been another important element provided by the G4, encouraging each other to participate in active rather than passive involvement in the future of their open data programs (Giggey, 2012). The G4 have worked together in developing strategies, directives and other technical documents to be used towards

measuring general success of their respective programs (Giggey, 2012). These documents have played an important role in the future development and goals of open data programs across the country. These documents have included important factors such as considering the goals of governments, along with the needs of residents (Giggey, 2012). There is however, a gap between understanding how residents are fully utilizing open data, since most metrics currently focus on what is being accessed most often, such as through collecting the *number of downloads* of a dataset, but does not include details as to what the data were being used for (Roy, 2014; Johnson, 2016). Including further details alongside metrics such as number of downloads have been found to be necessary in understanding more details related to the wants and needs of open data end-users. (Roy, 2014; Johnson, 2016).

2.3 Data Themes, File Types and Quality over Quantity: Current Open Data Trends

Open data programs have continued to gain momentum across Canada to include over 30 participating municipalities (Roy, 2014). The amounts, themes and file formats can vary between these municipalities based on their "seniority", strategies and/or overall goal of their respective programs (Sieber & Johnson, 2015; Thorsby et al., 2017). There are however, also many similarities between municipalities, providing a perspective on current trends in open data catalogues across Canada. One such trend is towards the volume of data available in an open data catalogue. As programs begin to develop, they usually start with a handful of datasets provided on their catalogues, mainly including base data such as administrative boundaries, road networks, budgetary information and census data (Thorsby et al., 2017; Dong, et al., 2017). As an open data portal matures, larger quantities of datasets are found, covering a wider variety of themes (Roy, 2014; Robinson & Johnson, 2016).

Open data catalogues can cover large amounts of information under a variety of different themes. Overarching themes of data found in Canadian open data catalogues include infrastructure and transit, budgetary data including expenditures and other governmental information, geographic data, health and public safety, sport, leisure and cultural data, planning and development information, along with educational and social services data (Currie, 2013;

Thorsby et al., 2017). The literature suggests that there are many variations of themes available in an open data catalogue dependent on the goals and funds of the municipal government running it (Thorsby et al., 2017). While these themes are found most often, many catalogues include only some of these themes and/or include additional themes found less often such as statistical data, weather, or traffic counts to name a few (Janssen et al., 2012; Chan et al., 2016; Thorsby, et al., 2017).

Although data themes may vary between catalogues based on funds, goals and maturity of the data portal, the formats in which the open data are provided have been found to be quite consistent across Canadian municipalities. Open data are usually provided in one of eight formats as shown in Table 1 below. These eight formats can be broadly classified into two categories; spatial or non-spatial open data.

Table 1. Common Open Data File Formats

Data Format	Description	Spatial Categorization
CSV (Comma Separated Files)	Used for large datasets. Is compatible with multiple software including GIS and Microsoft Excel	Non-spatial
DWG (Drawing)	Binary file format, usually consisting of engineering drawings used with CAD software.	Spatial
JSON (Java Script Object Notation) or other API	Easy to read for any programming language. Can be used to create applications.	Non-spatial
KML (Keyhole Markup Language)	An XML notation for representing geographic information. Compatible with Google Earth.	Spatial
SHP (Shapefile)	Most widely used format for special data. Used specifically for Geographic Information System (GIS) software	Spatial
XLS (Microsoft Excel Spreadsheet)	One of most easily understood format, compatible with Microsoft Excel.	Non-spatial

XML (Extensible Mark-Up Language)	Most commonly used file format for data exchange. It keeps the integrity of the data while allowing for the end-user to modify the information in parts.	Non-spatial
PDF	Document and/or report that provides information, but is non-manipulative.	Non-spatial

(Modified from Dong et al., 2017).

For the purpose of this research, spatial open data are those available in a SHP, DWG, or KML file format. Sometimes geographic references can be found in an XLS or similar format, such as when latitude and longitude coordinates or addresses are provided in a spreadsheet (Currie, 2013). Although spatial data is not always limited to SHP and KML files, those datasets with geographic references in non-spatial formats require further manipulation to be visualized in a spatial setting. In general, spatial open data provides location-based information which can then be mapped using GIS related tools and software (Currie, 2013). The data can be provided as points such as school locations in a city, areas such as parks, and lines such as road networks (Currie, 2013). Non-spatial data fall within a number of different themes and does not have a geographic component to it. The data may be associated with the municipality as a whole, but the information provided is not tied to a specific geographic point, area of line. Examples of non-spatial data include budgetary information, annual reports, and schedules for upcoming events.

Throughout the literature, there are a number of varying opinions on the value of certain data formats over others. Some claim that data provided in a spatial format is considered less “open” as it requires a specific skillset and more expensive software to use it (Janssen et al., 2012; Chan et al., 2016; Throsby, et al., 2017). Others claim that spatial open data is a more valuable format, as it can be used to visualize data in an interesting and captivating way, leading to potential economic development through the creation of mobile and online applications (Janssen & Zuiderwijk, 2014; Robinson & Johnson, 2016). As for non-spatial open data, it has been found to serve an important role in increasing accountability and transparency between governments and residents, building trust between the two groups,

leading to greater participation and collaboration (Thorsby, et al., 2017). Since both spatial and non-spatial data have their own positives and potential pitfalls, there is a need to understand and consider both when releasing open data.

As open data continues to grow in popularity, there is a desire internally for governments to release as much data as possible to the public (Sadiq & Indulska, 2017). However, this is causing a disconnect between the data publishers and the end-users, as the focus is on releasing data, rather than spending time to discover and provide end-users with the data they want and/or desire (Johnson, 2016). This push for quality over quantity specifies how the value of open data varies. Providing a dataset in a portal for end-users to download does not guarantee that the data will be perceived as valuable (Johnson, 2016; Sadiq & Indulska, 2017). It is not until the data is utilized that its full potential will be reached. Depending on the goal of the open data program, coupled with the wants and needs of end-users, the value of open data will vary greatly (Chan et al., 2016; Machova & Lnenicka, 2016). This is of particular importance when considering the varying backgrounds of potential users, including both their technicality, and their motivations for using the data. Understanding the wants and needs of the end-user, however, will enhance the knowledge of open data program coordinators to better assess why certain datasets are being downloaded more than others, and what they may be doing with that data (Chan et al., 2016).

2.4 Goals and Motivations of Open Data Programs and their End-Users

Open data programs are considered valuable to a variety of different end-users, whether they are everyday residents, private sector companies, Non-governmental organizations, or governments themselves. The value that open data programs provide to these end-users differs, such as for increased transparency and accountability, innovation, collaboration, creating participatory governance, or enhancing efficiency within government itself (Currie, 2013). Along with the differing values provided to the end-users, municipal governments have varying goals and strategies behind the creation and upkeep of their open data programs. This can include similar values to the end-users, such as increasing transparency

and accountability, and/or efficiency within the municipal government (Graves & Hendler, 2013). Governments also have more advanced goals such as enhancing opportunities for economic development, where they believe open data programs can help create jobs and monetary gains for private sector companies and individuals (Janssen & Zuiderwijk, 2014; Chan et al., 2016). The overall strategies behind open data programs from a municipal perspective tend to focus on connecting the goals of the government with the motives of the end-users with.

Transparency and accountability are among the most widely discussed goals of open data (Robinson & Johnson, 2016; Martin & Begany, 2017). These goals focus on enhancing the relationship between government and citizens, including placing an emphasis on sharing information before being asked (Scassa, 2015). This is also referred to as transforming government information to "open by default." By having an open data catalogue, governments can release detailed information about their current projects, funding, and other budgetary information which citizens can then access (Roy, 2014). By having this detailed information available online, this is believed to help hold governments accountable, for example, in relation to what they are spending tax dollars on (Evans & Campos, 2013). This can further increase residents' feelings towards and relationship with their local government, allowing them to better understand if their governments are working towards their commitments. Transparency and accountability are considered to be at the forefront behind the reasoning for governments to create open data programs, and includes interest on the side of the end-users towards their motivations in using the data. The end-users however, do vary and include those who are looking to not only understand what their government is doing, but use government information towards innovative gains.

Many municipalities partake in open data programs with the promise of increased citizen engagement, which can then be combined with innovative activities to create new opportunities for residents and governments alike. A cyclical process is usually described, where governments hope that residents will engage with their open data, downloading it, to

further manipulate and reuse (Scassa, 2015). The data would then be coupled with varying technologies to create innovative applications which can be used to help solve everyday problems (Graves & Hendler, 2013; Scassa, 2015). These applications then lead back towards the government through two possible pathways, or a combination of the two; increasing efficiency and innovation at the local governmental level and/or creating monetary gains through the selling of the application which will return benefits to the government through taxes (Scassa, 2015). This broader theme of economic development therefore can provide a multitude of benefits for all users involved through creating job opportunities along with providing services for the public and governments themselves (Zeleti & Ojo, 2017). The broader theme of economic development, as discussed, focuses on driving innovation. This has been outlined as an important aspect of open data by both the Cities of Toronto and Edmonton (City of Edmonton, 2017; City of Toronto; 2017). Increasing opportunities for innovation is therefore seen as an important focus within municipal governments, while open data programs are seen as a way to achieve this goal.

Many municipalities have discovered that having open data portals does not necessarily mean that end-users will engage with the data in this fashion (Johnson, 2016). This is where civic hackathons have played an important role in connecting residents and private sector companies more closely to the open data, while also in some cases providing incentives through cash prizes (Johnson & Robinson, 2014). Civic hackathons are events run by governments which encourage the public to use open data to create different products, mainly under the category of mobile or web-based applications (Johnson & Robinson, 2014; Sieber & Johnson, 2015; Robinson & Johnson, 2016). These hackathons encourage residents to download and use open data, increasing the popularity of government open data portals, and further usually focus on solving everyday problems which can then be used towards improving resident's lives (Robinson & Johnson, 2016). Further, hackathons can result in residents and/or private sector companies creating applications which can provide them with monetary gains (Robinson & Johnson, 2016). Civic hackathons are quite common now, having taken place in all cities of the G4, and prove how knowledge sharing and encouragement through the G4 has led

to more resident engagement within their respective open data programs (Johnson & Robinson, 2014). The civic hackathons have seen a direct impact towards governments as the events usually revolve around one theme, focused on using particular datasets to solve specific problems (Sieber & Johnson, 2015; Sadiq & Indulska, 2017). Hackathons therefore continue to prove that economic development can take place through open data programs, and can create a collaborative environment where governments and residents are working together towards a common goal.

Government open data can also be used towards improving efficiency. This includes both within government operations and in residents' everyday lives. This overlaps with both of the categories described above. Efficiency improvements can be found through increasing transparency, adopting an "open by default" mandate (Scassa, 2015; Martin & Begany, 2017). This could include a decrease in time spent on tasks related to *freedom of information* requests, as much of the data will already be provided online. This could also include easier and quicker movement of data within governmental organizations themselves (Martin & Begany, 2017). Efficiency improvements may also be seen through the economic development lens, through private sector companies, residents, or governments themselves using new and evolving technologies to visualize, update, and share data (Scassa, 2015). This can also include improvements of existing processes through encouraging residents to create products to solve everyday problems, and/or encouraging them to create products to display and share information not previously thought of. Efficiency may also be improved more indirectly through the category of accountability such as through focusing on engaging with residents, collaborating with them and listening to their concerns and thoughts towards current and future processes (Johnson, 2016). Efficiency improvements are therefore an important motive behind open data programs, but are of a secondary nature, with many of the improvements being provided through the broader categories of transparency/ accountability and/or economic development.

2.5 Techniques and Criteria used to Evaluate Open Data

Although the motives and strategies behind the creation and upkeep of an open data catalogue can vary by municipality, there is a need to understand how to assess if the goals of the open data program are being met, along with if the end-users find the data provided useful and valuable. There are many different types of techniques used to evaluate open data portals and/or programs. Two of the main ones fall under self-evaluation, usually in the form of informal discussions by governmental bodies, and external evaluations completed by NGOs and/or academics, which focus on comparing governmental programs to one-another, assessing a number of different characteristics (Evans & Campos, 2013). Further, dependent on the goal of the program, the evaluation will focus on different types of criteria, whether it be the number of datasets available, the "openness" of the datasets, or other factors (Johnson, 2016; WWW Foundation, 2016). The literature did not discuss a preferred evaluation technique, but instead focused on highlighting a number of techniques and criteria that can be used, dependent on the goal of evaluation.

Some municipal governments participate in self-assessment, through examining and altering their strategic documents on a regular to semi-regular basis (Evans & Campos, 2013; Johnson, 2016). Many municipal governments band together, such as the G4, to write these documents, and meet to discuss progress, along with successes and failures to consider while moving forward (Giggey, 2012; Evans & Campos, 2013). These evaluation techniques are very important towards proving that the open data program is of value to all those involved, and therefore should be a project that is continually funded by the government. As self-evaluations tend to focus on the overall success of the program, the statistics captured are usually broad. This could include assessing the number of datasets added to a portal over time, the number of online visits to the portal, the number of datasets downloaded, along with other more detailed information if applicable, such as if any products were created with the data (Denham, 2013; Johnson, 2016). Self-evaluation is helpful for open data officers, or other individuals in a similar role, as it gives them the opportunity to share the importance and necessity of the program to other departments in the government, as well as to other governments, helping them develop,

or further enhance their own open data programs.

NGOs and academics are a significant group who contribute to evaluating open data programs. These groups focus on evaluating open data with the purpose of providing insights, trends and results relevant to a number of different open data goals. The results of these evaluations are usually more specific and focus on a specific element of an open data program, or work towards evaluating a specific theme (Denham, 2013). A number of different methods are used to evaluate open data from a researcher's perspective. A popular method is through comparing an open data portal to standards created by external governments or other entities (Johnson, 2016; Chan et al., 2016). An example of this type of method was completed in Denham (2013)'s report, where she evaluates the government of British Columbia's open data initiative. This report focused on providing a list of recommendations for future improvement within four specific areas based on comparing the program at its current state with a directive created by a number of top open data countries around the world (Denham, 2013). The directive outlined a number of suggestions and guidelines on a number of elements of open data catalogues (Denham, 2013). This evaluation method is quite simple, but can provide concrete steps and/or goals to focus on moving forward.

An additional method used to evaluate open data programs is through creating open data portal assessments. These assessments include rating a program on a number of specific characteristics. Gill and Corbett (2017) used a heuristic evaluation model to examine how user-friendly and accessible open data portals are. This evaluation used a score sheet approach, where they rated each open data portal on a scale from zero to five based on a number of characteristics (Gill & Corbett, 2017). This type of evaluation is useful for examining specific characteristics of an open data portal and further allows for the comparison of multiple portals at once, to show how they rank against one another.

Another similar method is the evaluation created by the World Wide Web Foundation called the Open Data Barometer. The Open Data Barometer is a well-known evaluation that

includes government self-assessments, peer-reviewed expert survey responses, detailed dataset assessments and also includes some additional secondary data (WWW Foundation, 2016). Their evaluation examines the open data programs in the perspective of governments, residents and civil society and lastly entrepreneurs and business (WWW Foundation, 2016). They use a rating system, applying weights to yes or no questions related to data specifics (WWW Foundation, 2016). They also provide standards as to the meaning and naming conventions of specific types of datasets, including general themes and data formats (WWW Foundation, 2016). This robust evaluation provides in depth analysis on the current state of open data portals, outlining where programs align against one-another and how they can improve. The Open Data Barometer is used internationally to evaluate data and is seen as a valuable source of information regarding where countries stand in their current state of open data.

2.6 Summary

Overall, open data is a new and still evolving phenomenon, with momentum building in Canada around 2011 due to increased interest in e-governance and the Web 2.0 movement (Roy 2016; Gill & Corbett, 2017). There are many different themes of data provided openly, in an array of formats. These mostly include information that can be connected back to increasing transparency, accountability, efficiency and economic development for the local government involved (Currie, 2013). Economic development in particular is considered an important component to open data programs, connecting entrepreneurs to data, encouraging them to create innovative products with the goal of helping society (Robinson & Johnson, 2016). As open data portals continue to increase in popularity and governments turn to focusing on meeting evolving goals, such as the ones described above, a number of different tools and criteria are used to evaluate open data portals. These evaluations focus on providing suggestions and recommendations for governments to consider when moving forward towards enhancing their open data programs (Denham, 2013; WWW Foundation, 2016; Gill & Corbett, 2017). Using these evaluations as guidelines, governments can better understand if their current goals are being met and/ or determine how they can reach them. The methodology

section presented next, focuses on outlining how this paper evaluated spatial and non-spatial open data towards their economic development benefits, using methods and definitions found within the literature as a guide.

Chapter 3: Methodology

As described throughout the literature, there is a need to continually collect information on and evaluate open data catalogues based on a number of different factors, including current trends, and the value of the data provided. A number of evaluation techniques were found, most of which focused on evaluating a sole municipality or a small group of government programs to compare and contrast current successes and outline areas for improvement (Denham, 2013; WWW Foundation, 2016; Gill & Corbett, 2017). To address the first objective of this paper, four Canadian municipalities were examined, as outlined in Table 2, all of which are considered members of the *G4*. As previously described, the G4, which consists of the City of Vancouver, Edmonton, Toronto and Ottawa, are municipalities with established open data programs who were some of the first to adopt programs within the country of Canada (Giggey, 2012). They are considered pioneers within Canada for open data, and continually provide feedback to other developing programs towards how to ensure or improve their success.

Table 2. Open Data Catalogues

City Name	Link to Catalogue	Year Started
Vancouver	http://vancouver.ca/your-government/open-data-catalogue.aspx	2009
Edmonton	https://data.edmonton.ca/	2010
Toronto	http://toronto.ca/open	2009
Ottawa	http://data.ottawa.ca/	2010

The open data catalogues of the G4 municipalities were reviewed by visiting their individual web portals and collecting two distinct statistics. Firstly, the number of datasets found throughout each of the catalogues was documented. This grand total of datasets available was collected from each municipality on the same day, as updates and new datasets are not uploaded similarly between municipalities. The number of datasets was also documented within one of nine themed categories, as shown in Table 3 below. The categories were chosen from the literature as commonly occurring themes by which to summarize open

data availability. A dataset may fall under one or more theme, but was placed within the theme that best describes it, meaning each dataset was only represented once across all nine themes.

Table 3. Data Themes

Theme	**Example Datasets**
Business	Food truck locations, business permits
Planning and Development	Ward boundaries, neighbourhood names, building permits
Parks, Recreation and Culture	Parks, recreation schedules, outdoor pool locations
Health, Public Safety and Legal	Crime statistics, paring tickets, food safety
Educational, Community and Social Services	School locations, 311 data, homeless shelters
Governmental Data	Meeting minutes, budgetary data, census information
Environmental	Garbage and recycling schedules, watermain breaks
Transportation	Bus schedules
Infrastructure	Street network

Lastly, the file format of each open dataset was collected. To better understand what general themes of data are provided in which file formats, a summary of the number of different file formats for the datasets within each theme described in Table 3 was collected. Datasets refers to individual catalogue items. Data files encompass the different file formats that a dataset can be downloaded as. A dataset will have at least one data file, with the possibility of having multiple data files associated with it. It was further documented if the file format was spatial or non-spatial. The spatial data files are those which are a file type that is GIS-ready. These data do not have to be further manipulated before being inserted into a spatial-based software. Non-spatial data files are those which are not spatially ready. They could have coordinate data within them, but would have to be further manipulated before they can be used in GIS. These statistics were then combined to answer the first research question, providing details on the number of datasets and their file formats by general theme.

To further examine open data and the presence of spatial versus non-spatial data files, a case study on the City of Toronto was undertaken. An evaluation was created that focused on answering a number of questions related to the program goal of economic development, specifically focusing on examining if spatial open data are more prevalent within the theme of innovation and therefore if it may be more valuable towards reaching the goal of economic development. Spatial open data are open data provided in file formats that are spatially ready, meaning they are file formats that can be used directly with spatial software, such as ESRI's ArcGIS and Google Earth. The value of the open data was assessed in relation to the goal of economic development. Value in the context of this paper is defined as the usefulness of the data provided. The goal of economic development is defined, as outlined in the literature, as a cyclical process, where governments hope that residents and private sector companies will engage with their open data, downloading it, to further manipulate and reuse it. As mentioned in the literature, increasing innovative activities is a way in which the goal of economic development can be achieved within a city. Innovation in the realm of economic development focuses on encouraging job growth through supporting entrepreneurship, start-ups and other activities focused on creating new jobs and income sources for residents. In many cases in relation to open data, innovation is found through end-users using the data to produce applications and other products that can then be used to help solve everyday problems, creating job opportunities, which then provide benefits to the government through taxes and improved efficiency, and further can provide the creator can profit from. The end-users of innovation-based data mainly consist of private sector companies, entrepreneurs, and residents who have a technical background.

Therefore, the open data provided by the City of Toronto were evaluated to determine if they are useful for technical residents and/or private sector companies to create products and applications. The City of Toronto provided a number of statistics which were used towards creating the evaluation, including website visits and number of downloads of each data file. Further, the City's website provides a list of products created by third parties using their open

data. The proposed evaluation was created to illustrate how the City of Toronto can be evaluated looking specifically at the goal of economic development and its relationship to spatial open data. The results will therefore be provided to outline how the evaluation can be used and additionally to provide some general insights into the value of spatial open data.

Table 4. Additional Information Used to Evaluate City of Toronto Open Data

Data	Year	Format	Source
Web/ Accessibility Statistics	2017	XLS	City of Toronto
Gallery of Products Created using Open Data	2017	Webpage	City of Toronto

To create the evaluation, using the information provided in Table 4, the methodology used towards creating the Open Data Barometer was examined. The Open Data Barometer's (ODB) methodology included survey questions and detailed catalogue investigations. The survey questions were peer reviewed and scored based on a scale from 0 to 10. Detailed dataset assessments were undertaken by investigating the availability of different types of data, additionally answering a ten-point checklist regarding the quality of the datasets available within the specific catalogue. The ten-point checklist was further weighted to provide a score from 0 to 100 regarding the quality of data found in a particular data catalogue. The survey questions and dataset investigations were focused on providing a broad overview of the data catalogues examined, mainly relating to the overall "openness" of the data provided. The ODB is an internationally accepted document used for reviewing open data and therefore is a great source to follow in terms of evaluating an open data catalogue.

For the purposes of this study, the ODB methodology was adapted, through using a survey-style recording sheet and dataset investigation approach to ask specific questions relating to spatial open data and its value to the goal of economic development for the City of Toronto. To help with creating the questions in the evaluation, information from the literature review was used to help determine general themes of open data and how they relate to the

different goals of an open data program. The ODB was also used, as it provides details related to which open data themes cluster together to create one of three *impact* groups. These three groups; Accountability, Social Policy and Innovation correspond to the different goals of an open data program. This study will focus on the goal of economic development, which aligns with the impact group of Innovation. The ODB groups five of the original nine themes outlined in Table 3 together to form the innovation impact group, including:

- Planning and development
- Transportation
- Infrastructure
- Crime related data
- Business data

The survey-style and dataset investigation questions used in evaluating the City of Toronto's open data program can be found in Table 5, where a higher score outlines a higher prevalence of spatial-related information.

Table 5. Criteria for Evaluating the City of Toronto's Spatial Open Data

Question	Weighted Response
1. What are the percentage spatial data files in the category of Innovation?	Less than 25% - 0 Between 25-50% - 3 Between 50-75% - 7 Greater than 75% -10
2. Are innovation-related spatial open data files downloaded more often on average?	Below Average - 0 Average - 5 Above Average - 10
3. Are the webpages of innovation-related spatial open data files visited more often on average?	Below Average - 0 Average - 5 Above Average - 10
4. Of the products created by third party users, how many created a spatial product?	Less than 25% - 0 Between 25-50% - 3 Between 50-75% - 7 Greater than 75% -10
Total Score	/40

Question 1 of the evaluation was answered by calculating the total number of spatial data files and dividing that by the total number of data files available within the City of Toronto's open data portal. The second and third questions were answered by using the web/ accessibility data provided by the City of Toronto. The City of Toronto provided their top 100 downloaded data files from the time period between January 1st 2017 and May 28th 2017. These data files were reduced to 75 data files, which included only those which fall under the category of Innovation, as specified by the Open Data Barometer. The data were provided in a spreadsheet which included each data file name, a description where possible and both the number of times it was downloaded and visited, as a data file could be viewed without being downloaded. The final question was answered by examining the gallery of products created by third party users, available through the City of Toronto's website. The products were provided as URL links to webpages which either shared the product in question, or linked to a page in which the product could be downloaded. The products were examined to see if they included a spatial component to them. Those that did, were classified as a spatial product, while those which had no spatial component were classified as non-spatial products. An example of how a spatial and non-spatial product were assessed can be found in Appendix A. Each answer was given a score, as outlined in Table 5, which was then used towards answering the final research question, including the prevalence of spatial open data within the theme of innovation, and further the introduction of a tool to assess the value of spatial open data in the context of economic development.

Chapter 4: Results

4.1 Introduction

The analysis of the G4 open data portals was undertaken and the results were captured in a series of tables and graphs outlining the most common data themes and their associated file formats. A summary of the overall frequency of spatial and non-spatial open data was also documented for each city. Further, the City of Toronto's open data catalogue was then examined, using the evaluation created, providing details as to whether spatial open data is more valuable for the goal of economic development. The results of the evaluation were presented in a series of tables, sharing the results to each question. This section ends with a summary of the results, answering the two research questions, as outlined in the introduction.

4.2 G4 Open Data Catalogue Review

The open data portals of the G4 cities were examined and the results from each city were documented and compared to one another. Table 6 outlines the total number of datasets along with data files provided within each open data catalogue.

Table 6. Dataset and Data File Results for the G4

Name of City	Number of Datasets	Number of Data Files
Vancouver	284	597
Edmonton	774	4487
Toronto	248	312
Ottawa	129	460

Across all four cities, most datasets had at least two data files associated with them. The City of Vancouver provided the majority of their datasets in at least two different file formats. If a dataset was provided in a spreadsheet format, it was usually provided in both .CSV and .XLS format, while their spatial data was provided in .DWG, .KML and .SHP format. Other common data formats included XML, JSON and Google Maps. The City of Toronto and the City of

Ottawa's datasets varied significantly on the number of data files provided, ranging from only one, such as an .XLS spreadsheet, to more than five different data files including both spatial and non-spatial data formats. The City of Edmonton provided multiple data formats for each dataset uniformly throughout their catalogue. All non-spatial data were provided in eight different formats, which included spreadsheets and web-based services, while their spatial open data was provided in five geospatial-ready formats as well as spreadsheets.

4.2.1 G4 Datasets by Theme

Figure 1 provides an overview of the number of data files found within the nine identified data themes described within the methodology. The data are further broken down in Appendix B, providing details on the number of datasets by theme for each individual city. Overall, the category with the largest number of datasets was *Government Data*. This included an array of data ranging from census data, budgetary information and government staff related details. The theme of *Education, Community and Social Service*s had the second largest amount of datasets. These datasets included community-based surveys, social services provided by the city and school related information. The category of *Business* had the smallest number of datasets, with business-related data being quite sparse across the G4 cities.

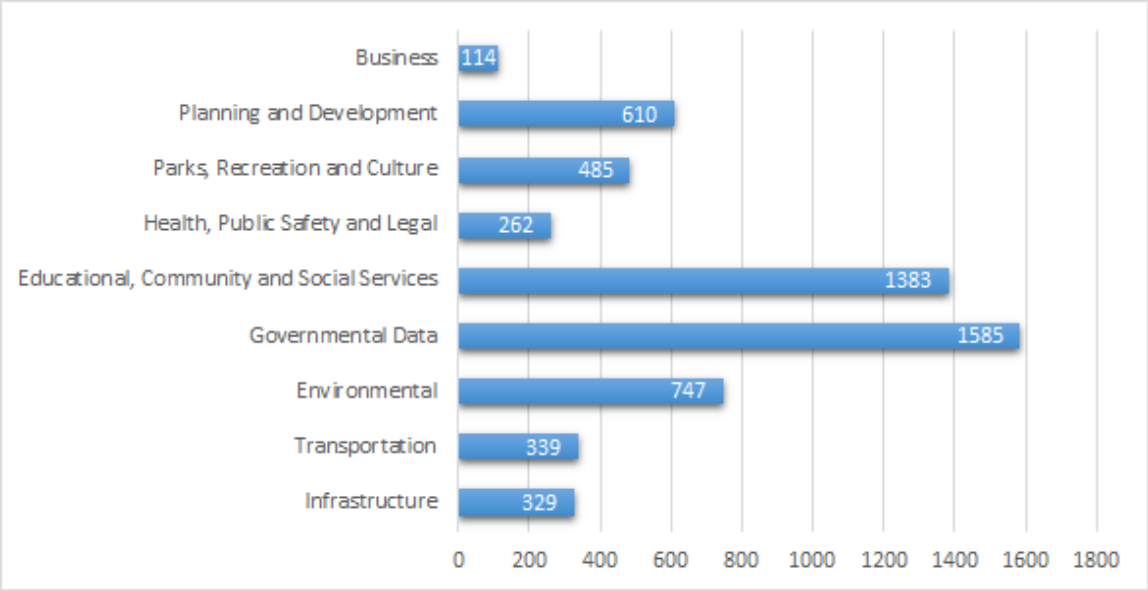


Figure 1. Total Number of Data files by Theme

4.2.2 G4 Data Files by Theme

Figure 2 outlines the total number of spatial versus non-spatial data files by theme for the G4 cities. Non-spatial open data were found more frequently throughout the four catalogues, reaching a total of 4698, almost 80 percent of all data files. There were 1175 spatial data files throughout the G4 cities, making up about 20 percent of the total number of data files. Across the G4 cities, non-spatial open data were therefore more prominent, having more data files present within each category, except for two; *Planning and Development* and *Infrastructure*. Three themes in particular; *Parks, Recreation and Culture, Educational, Community and Social Services,* and *Environmental* had non-negligible amounts of spatial data files. There were however both spatial and non-spatial data files throughout all the nine themes. Based on the results of analyzing the G4 catalogues, there is consistently a lower frequency of spatial data over non-spatial open data.

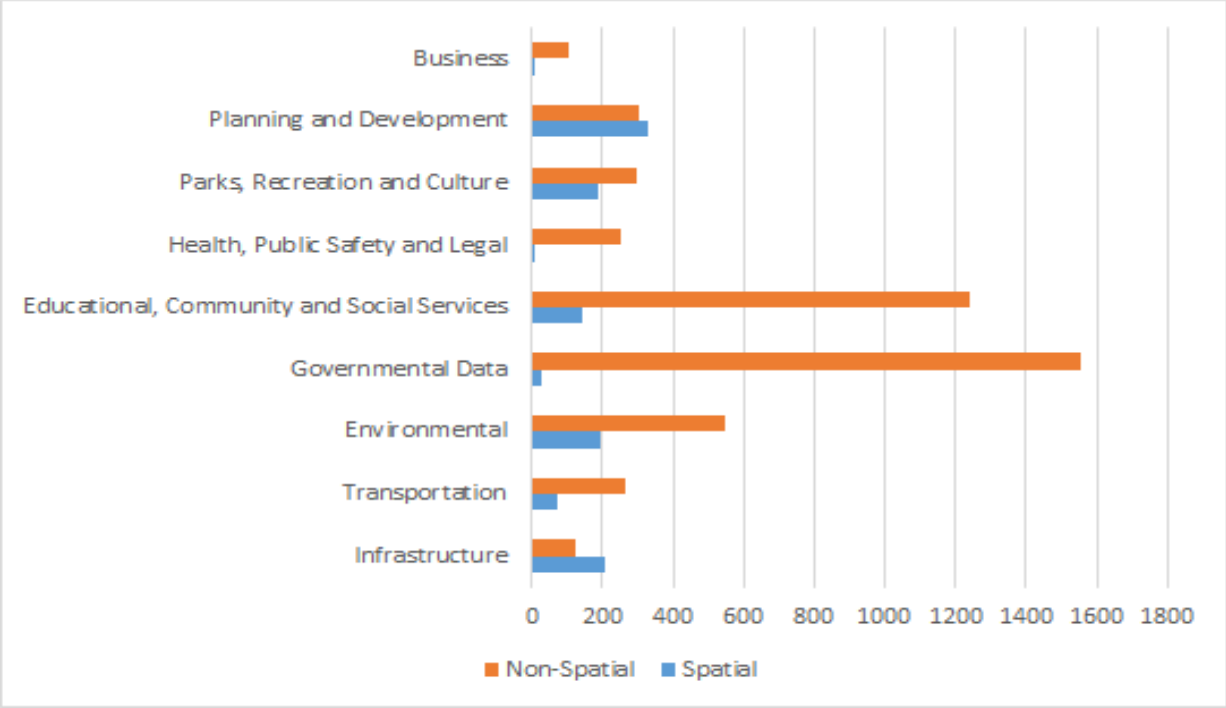


Figure 2. Spatial vs Non-Spatial Data Files by Theme

4.3 City of Toronto Evaluation

The City of Toronto was evaluated to assess if their spatial open data are more valuable toward the goal of economic development. An evaluation was created, as outlined in the methodology, which presents a series of questions used to discover these results. This section outlines the answer to each individual question, including the score given based on the result. A summary of the evaluation will also be provided, assessing if spatial open data are more prominent within the theme of innovation. Additionally, the results will outline if the evaluation is a useful tool to compare the role of spatial open data between cities, and within cities over time.

4.3.1 Frequency of Spatial Open Data Files

Table 7 outlines the total number of data files found during the examination of the City of Toronto's open data portal, including the number of spatial and non-spatial data files.

Table 7. City of Toronto Portal Examination Results

	Total Data Files	**Non-Spatial Data**	**Spatial Data**
All	312	226	86
Innovation Data Only	123	82	41

Less than half of all data files were classified within the category of Innovation. Of the data under this category, non-spatial data files were more frequent, having more than double the amount of files available, at a total of 82, compared to only 41 spatial data files. Specifically, 33 percent of the data files under the category of Innovation were spatial. This led to a score of 3 for the first question of the evaluation, as shown in Table 8. Of note, there was a greater percentage of spatial data files within the category of innovation compared to the presence of spatial open data within the total amount of data files found in the City of Toronto's portal. This was complemented by a reduction in the frequency of non-spatial data files from 72 percent of the total to 66 percent of the innovation themed data.

Table 8. Spatial Data Evaluation - Question 1

Question	Weighted Response	Result
1. What is the percentage of spatial data files in the category of Innovation?	Less than 25% - 0 Between 25-50% - 3 Between 50-75% - 7 Greater than 75% -10	41/123 data files = 33% **Between 25 - 50% = 3 points**

4.3.2 Popularity of Spatial Open Data Files

To answer questions 2 and 3 of the evaluation, the popularity of the spatial data files within the Innovation category was assessed. The number of webpage visits and downloads were provided by the City of Toronto and used to examine if spatial data files were downloaded, and/or visited more often than the non-spatial open data files. As outlined in the methodology, the City of Toronto provided their top 100 data files which had the highest combination of webpage visits and downloads. 75 of the 100 original data files were classified under the category of Innovation. 30 of these top 75 were non-spatial, while the remaining 45 data files were classified as spatial, making up 60 percent of the top data files. The data used to answer these two questions can be found in Appendix C. Table 9 outlines the average downloads and webpage visits for the top data files, broken down further by spatial versus non-spatial data files.

Table 9. Average Webpage visits and Downloads per Data File over 6 months within the Innovation Category

	Average Webpage Visits	Average Downloads
Total Data Files	1992	387
Spatial Data Files	2238	436
Non-Spatial Data Files	1621	318

The average number of downloads during the study period throughout the 75 data files was

387 times. The average amount of downloads for spatial data files was higher than the total average, at 436 times. The non-spatial data files were downloaded less, with an average of 318 times. This led to a score of 10 for the second question of the evaluation, as shown in Table 9. The top data files ranged in number of times downloaded from 137 to 3369. The top five most downloaded data files were all classified as spatial, with eight of the top 10 also under the category of spatial.

Table 10. Spatial Data Evaluation - Question 2

Question	Weighted Response	Result
2. Are innovation-related spatial open data files downloaded more often on average?	Below Average - 0 Average - 5 Above Average - 10	Average of total data files = 387 Average of Spatial Data Files = 436 Average of Non-Spatial Data Files = 318 **Above Average =10 points**

Table 8 also outlines the average webpage visits for each data file. The data files ranged from being viewed 230 to 9785 times. The average webpage visits was 1992, with spatial data files having a higher average of webpage views; 2238 versus a lower average for non-spatial data files; 1621. This led to a score of 10 for the third question of the evaluation, as shown in Table 10 below.

Table 11. Spatial Data Evaluation - Question 3

Question	Weighted Response	Result
3. Are the webpages of innovation-related spatial open data files visited more often on average?	Below Average - 0 Average - 5 Above Average - 10	Average of total data files =1992 Average of Spatial Data Files = 2238 Average of Non-Spatial Data Files = 1621 **Above Average =10 points**

The data provided by the City of Toronto included a spreadsheet which contained both the downloads and webpage visits for each data file. When comparing the number of downloads to page visits, the highest ratio was 158 downloads to 7117 webpage visits, while the lowest ratio was 225 downloads to 341 page visits. The highest ratio belonged to a spatial data file, while the lowest ratio belonged to a non-spatial data file. The top five data files with the highest ratio were all non-spatial data files, while seven of the top ten were also non-spatial data files.

4.3.3 Spatial vs. Non-Spatial Products Created using Open Data

Some of the products created using open data have been collected by the City of Toronto and presented in an online gallery. This gallery contained 51 different products created. Each product was examined to determine if it included a spatial component to it. This included applications that used web maps, and/or static maps to show information based on the dataset. Table 12 outlines the total number of products, along with their spatial versus non-spatial classification.

Table 12. Spatial vs. Non-Spatial Products Created using Open Data

	Number of Products
Total	51
Spatial	33
Non-Spatial	18

Spatial products made up over 50 percent of all products presented within the gallery, with 33 products being documented as spatial. These products included a number of mobile applications which focused on transportation. Additional products included a game, along with online webpages that focused on providing analysis on a number of different topics. The non-spatial products totaled 18, and mainly consisted of applications related to garbage and recycling schedules and/or reminders, along with applications highlighting upcoming events. The spatial products outweighed the non-spatial products, making up 65 percent of all those

presented in the gallery. This led to a score of 7 for the final question of the evaluation, as shown in Table 13 below.

Table 13. Spatial Data Evaluation - Question 4

Question	Weighted Response	Result
4. Of the products created by third party users, how many created a spatial product?	Less than 25% - 0 Between 25-50% - 3 Between 50-75% - 7 Greater than 75% -10	33/51= 65% **Between 50-75%** **=7 points**

4.4 Value of Spatial Open Data

Based on the results of the individual questions of the evaluation, spatial open data are more prevalent within the theme of innovation. The spatial open data have a higher frequency of data files within the theme of innovation compared to its presence among all data files. Further, the spatial open data were downloaded and viewed at a higher rate than their non-spatial counterparts. Additionally, through the gallery presented by the City of Toronto, the majority of the products included a spatial component to them.

The evaluation used to derive these answers is an example of how the results can be combined and compared to another city and/or future years of collection to assess the role of spatial open data towards the goal of economic development. This proposed evaluation can report a final index, which can then be used across and within cities to assess the role of spatial data compares to non-spatial data, further assessing if spatial open data is more valuable than non-spatial data towards the goal of economic development.

Chapter 5: Discussion and Conclusion

5.1 Open Data in the G4 Cities

Throughout the G4 cities, there were significant differences in terms of the number of datasets and their associated data files. However, there also were consistent patterns in the types of data files released with certain themes, and the overall frequency of spatial versus non-spatial open data. Overall, the City of Edmonton had significantly more datasets and data files found within their portal, potentially skewing the results of the G4 portal survey. It was interesting to find that within the total number of datasets of the G4 catalogues, the categories of *Government Data* and *Education* and *Community and Social Services* were most populated, and also consistently had more non-spatial data files. This may be related to the program-based goal of transparency, which many cities cite as their original purpose in creating an open data portal. These themes are geared towards informing residents about their government's actions or future plans along with the services provided to residents by the city, leading to a focus on engaging residents and holding the government accountable. The audience for these datasets will therefore more likely be non-technical users, simply looking to view the information rather than manipulate it. The most common themes for spatial open data were *Planning and Development* along with *Infrastructure*. These datasets revolve around more technical information related more closely with geography, such as political boundaries, roads, and building permits. Those who are accessing these files will most likely be viewing them in a spatial environment to better understand the data. Further, many of these datasets can be seen as 'baseline' data, used with additional information to provide a bigger picture, such as by sharing the location of child care centres, while using roads as a point of reference to better understand where the centres are located. Through discovering the clear distinctions between the types of data files found in different themes, it is clear that the value of spatial versus non-spatial data is heavily dependent on the end-user, which is directly related to the purpose of the data and how the user may want to use them.

Further discussing the higher frequency of non-spatial data files in the G4 cities, it is important to note that many of these data files may have included spatial information, such as

latitude and longitude coordinates, but were not provided in a GIS-ready format. This may have been the case for many of the .CSV and .JSON files present in the G4 data portals. These files can still be used to produce spatial products, and in some cases can be quite valuable to the end-user looking to create a product that is not necessarily compatible with a particular spatial-ready data format. As this study focused on a clear distinction between spatial and non-spatial data files based on the format they were provided in, those data files that included spatial information were not captured. Although it is important to consider that a non-spatial data file as defined by this paper may include spatial information, it is provided in a format which may need to be further manipulated to be used by a technical user in the appropriate software. A trade-off therefore occurs in the case of non-spatial data files including spatial information, by potentially increasing the audience of a data file due to it being a less technical file type, while also potentially deterring technical users from gaining value from the data based on the additional work needed to use the data as desired.

This trade-off ties back to the availability of different data files across datasets and between portals. There does not seem to be a standard in terms of the number or type of data files that should be released with a certain type of dataset. As the four open data programs examined in this study began at roughly the same time, the stage of development of an open data portal did not seem to have a large, if any, impact on the data files available. The City of Edmonton's approach of uniformly releasing datasets in a set number and type of data files may help in terms of providing open data to meet the needs of all potential end-users, but does bring up the potential issue of wasted resources in terms of producing and uploading the various data files as well as storage space and bandwidth for accessing the portal. By considering which data themes are more valuable in a spatial versus non-spatial format, along with considering which data files may complement one another, such as by providing .CSV and .SHP files for the same dataset, the open data may be more strategically released, focusing on meeting the needs of all the end-users.

5.2 The City of Toronto's Spatial Open Data

Based on the results of the evaluation created, spatial open data were found to have a higher prevalence within the theme of innovation for the City of Toronto's goal of economic development. The evaluation showed that spatial data files were found less frequently in the catalogue, yet were visited and downloaded at a higher rate. Interestingly, among the top data files, many of the non-spatial data files found in the list were accompanying files such as "readme" files or other metadata documents related to the spatial data files. This further demonstrates the popularity of spatial data files amongst open data end-users. Additionally, the products created with the open data also tended to feature a spatial component to them, further proving the higher prevalence of the spatial open data. As the goal of economic development focuses on encouraging residents and private sector companies to use the open data to create applications and other products, the data that are available to these end users must focus on including valuable content to be used in an application, and also should be in a format that allows for these products to be created with ease.

As the City of Toronto has taken the time to create a gallery of products, there is a connection being made between the open data and how they are being used. Connecting with the end-user once they have interacted with the open data can allow for governments to better understand which data are being used, and most importantly, how they use it. The city may expect a dataset to be most valuable in one format, while the end-users most inclined to interact with that dataset find it valuable in a variety of formats not previously thought of. By having an open and on-going discussion with the public relating to their wants and needs, an open data program can focus on providing the right type of data, most importantly in the right type of format.

Using the results of this study however can help in providing general guidelines towards releasing data related to specific themes in certain types of formats. As well, for the case of the City of Toronto, where spatial data files seem to be more popular, some effort may be taken in releasing existing data files in spatial-ready format. This may be of interest, but mostly importantly, a focus should be placed on the future release of datasets. This may include, for

example, releasing datasets within similar data themes, a more uniform approach may be taken, releasing a set amount of related data files which complement each other, but most importantly focus on providing the end-user with the data most valuable to them. Generally, the evaluation can be used to help the city create a strategic plan towards releasing open data, including which types of data to focus on, and further which data formats to release it in.

5.3 Limitations of the Study

As the results of this study focused on a select group of open data catalogues, there are some possible associated limitations. Firstly, this study only focused on the G4 cities. These cities have highly developed catalogues and therefore may provide different results related to newer portals. Additionally, the results may have varied if a greater number of catalogues were assessed. In particular, the City of Edmonton had significantly more datasets and associated data files than the other three catalogues. The summary provided on the data may have been affected by the greater number of Edmonton data files. Examining the individual results in Appendix B, however, can help in providing more details on the variation of data files within certain themes. Further, different municipalities may have varying responsibilities towards the types of open data they can release. This may vary based off of the responsibilities of other levels of government, such as the province. Additionally, the department in which the open data program is currently run within can impact the focus of the program and therefore the types of data the program focuses on releasing. This further shows the variation that can exist between catalogues even within one country, or even province.

Further limitations may be associated with the choice of criteria to assess the value of open data. The criteria for the evaluation included only four questions. This was due to the lack of control over the data provided by the City of Toronto. Further related to the data provided by the City of Toronto, the top 100 data files included information relating to metadata files. In many cases, these files are usually not accessed independently, and are mostly downloaded with their accompanying data. Having separately tracked some metadata files, there may have been an increased amount of non-spatial open data found. Overall, the criteria for the

evaluation provided robust questions for the limited data provided, and therefore can be enhanced with the availability of more in-depth data sources. For example, there is a lack of knowledge towards who exactly is downloading or viewing the data, and additionally what they are doing with the data in the end. The City of Toronto has made some progress on the latter through providing some products in their online gallery, but this is only a small portion of the products created with the data. This study is therefore a preliminary analysis of spatial open data in the context of economic development and has the potential to be further enhanced with additional data and criteria.

5.4 Conclusion

This research has outlined the many examined different aspects that surround open data, from starting a program to enhancing a portal to meet the needs of their growing end users. This further included understanding the varying formats that open data are provided in, and how these impact the goal(s) of the open data program, and in particular, the end-users who are interacting with the data. Spatial open data has been found to be less frequent than non-spatial data, yet through examining the City of Toronto, are some of the most popular data files within their portal. The spatial data are quite heavily used in creating applications and other products by residents and private sector companies, potentially leading towards economic development in the city. The proposed evaluation index can be used to compare the role of spatial data within and between cities, further creating potential for the analysis of spatial open data in the context of economic development.

Within the City of Toronto, there were a higher prevalence of spatial data within the category of innovation, suggesting a greater value of the spatial open data provided. However, there are many benefits to including non-spatial data files in accompaniment with these spatial files, providing more opportunities for a wider audience to engage with data that was previous limited to technical users. Therefore, there is a need to strategically release open data, where the considerations of the government and its current and future goals are considered, along with the end-user's wants and needs. Simply releasing datasets to reach a higher total number within a catalogue may be a wasted effort, while spending time to discover which datasets are

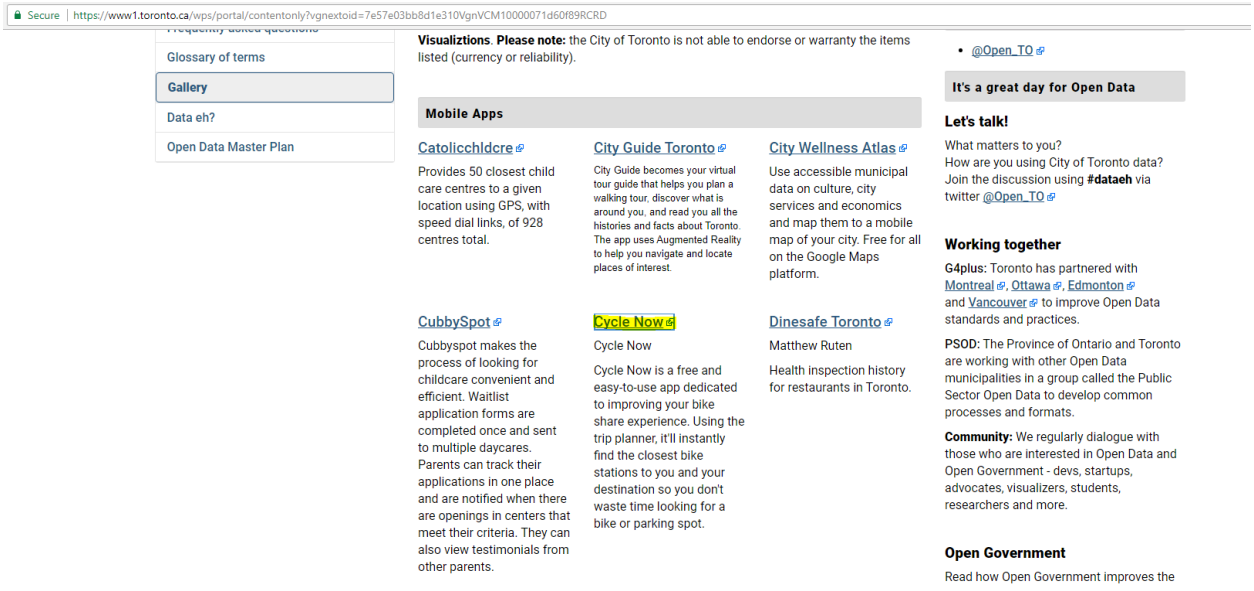
wanted by the community may improve engagement. Further, understanding how a dataset may be used will help in releasing it in a format that makes more sense to all those involved. Based on the results of this study, this may include releasing data formats more uniformly across datasets, using the theme of the data as a guideline.

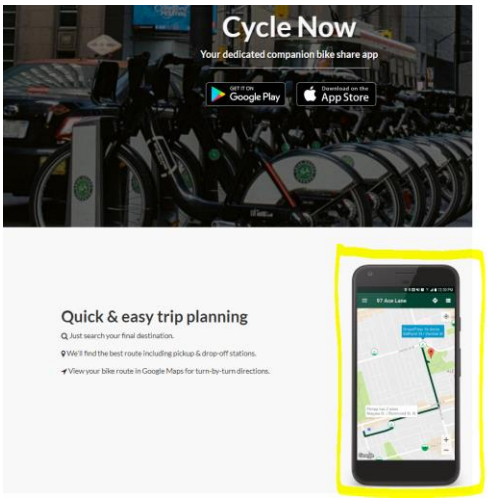
The consistent and ongoing evaluation of open data programs are a key component to the program's current and future success. Further analysis on the varying value provided by spatial versus non-spatial data files should be undertaken at the municipal level. This can be done by using the evaluation created in this study to compare the role of spatial open data within and between cities, further focusing on connecting innovation within government as well. Including the involvement of both governments and their end-users would provide a great opportunity to further understand which users are engaging in what types of data and how all those involved find value in the data provided. Spatial data in particular can be further analyzed to include not just spatial-ready file formats, but additionally non-spatial file formats which include spatial information such as latitude and longitude coordinates. A focused study on all those files that include spatial information may provide even further insight into the types of data which are most valuable, including the specific information which is most sought after by end-users.

Appendix A - Spatial vs. Non-spatial Product Classification

Spatial Example:

The gallery provides a list of products created with open data. The title is a link to a webpage showing either the product or where to download it.





This webpage highlights a map provided in the map, clearly outlining it has a spatial component to it.

Non-spatial Example:

Toronto Cycling

Brisk Synergies

The app will allow cyclists to record their cycling routes and provide this data to the City. The data will be part of the toolkit used to help inform the City's data collection and analysis when developing new cycling network plans.

Toronto Events

LunarFrog

Searchable current and upcoming city events and festivals.

Toronto Events & Festivals

Philip Vlach

Use Toronto Events & Festivals to search by 'what', 'where' and 'when'.

Toronto Eventz

SysAlto Corporation

Find events in Toronto including art exhibits, theatre and music performances, family events and celebrations.

Toronto Public Bikes

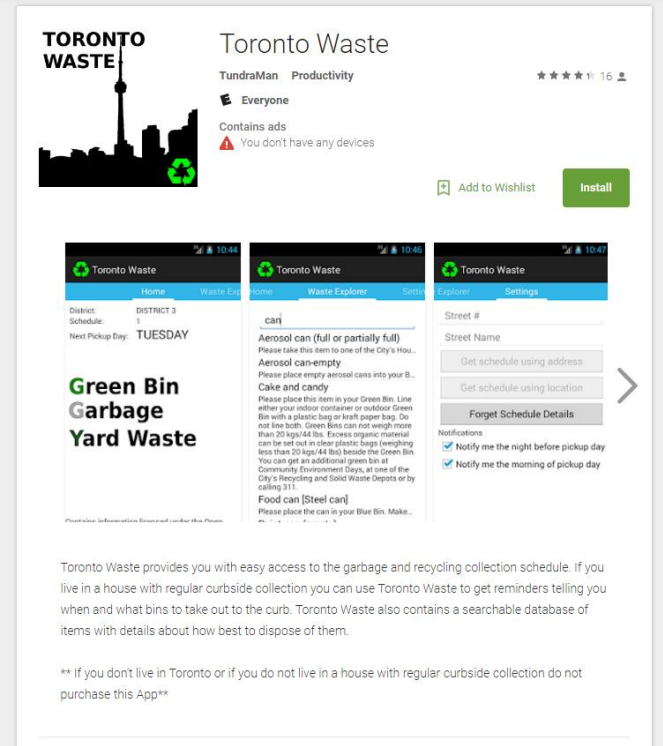
People & Code Inc.

The current bike/dock availability of Toronto Bixi Bike locations.

Toronto Waste

TundraMan

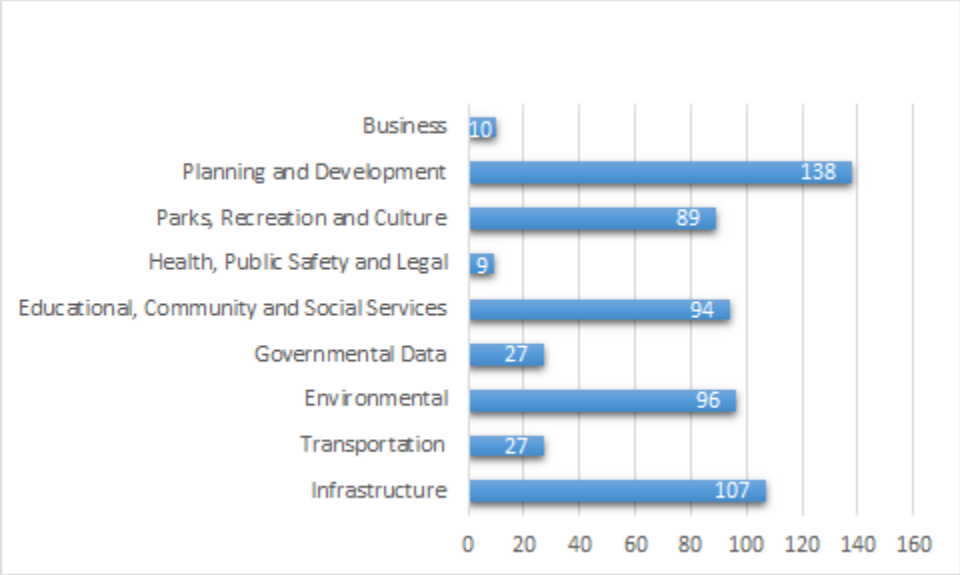
Pickup schedule, reminders and searchable item database.



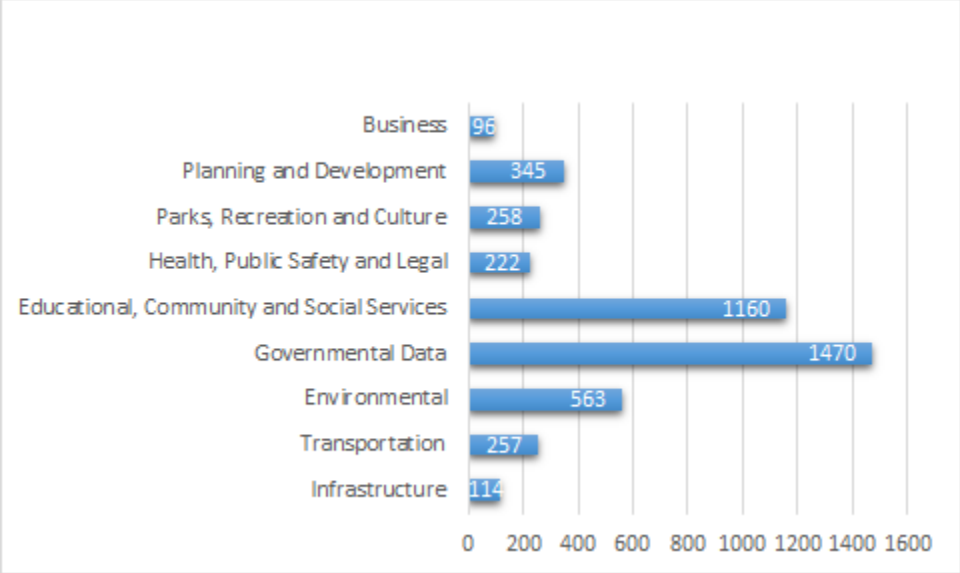
This application inputs your address, yet only provides information related to how to properly dispose of waste, and reminders of the location collection schedule.

Appendix B – Number of datasets found within the nine identified data themes

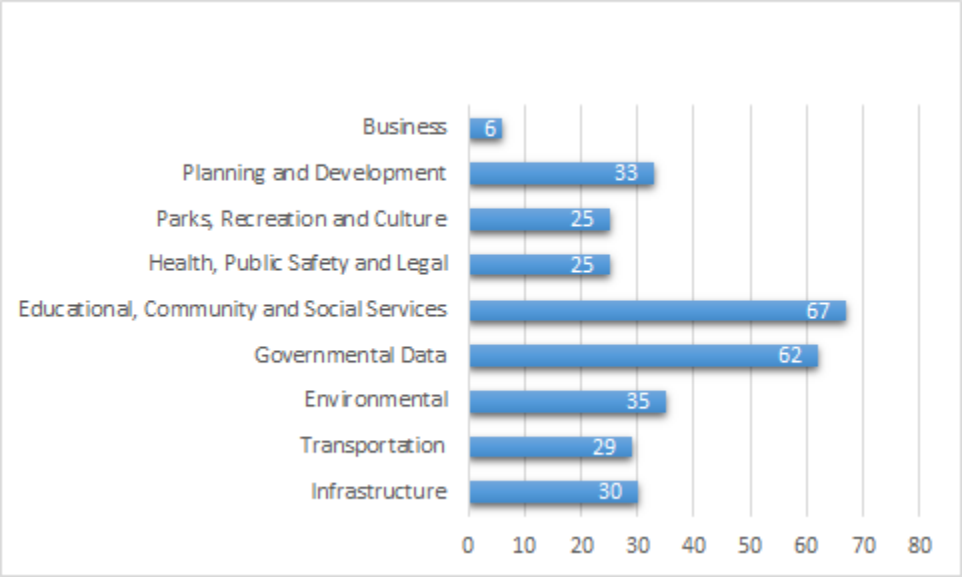
City of Vancouver Datasets by Theme



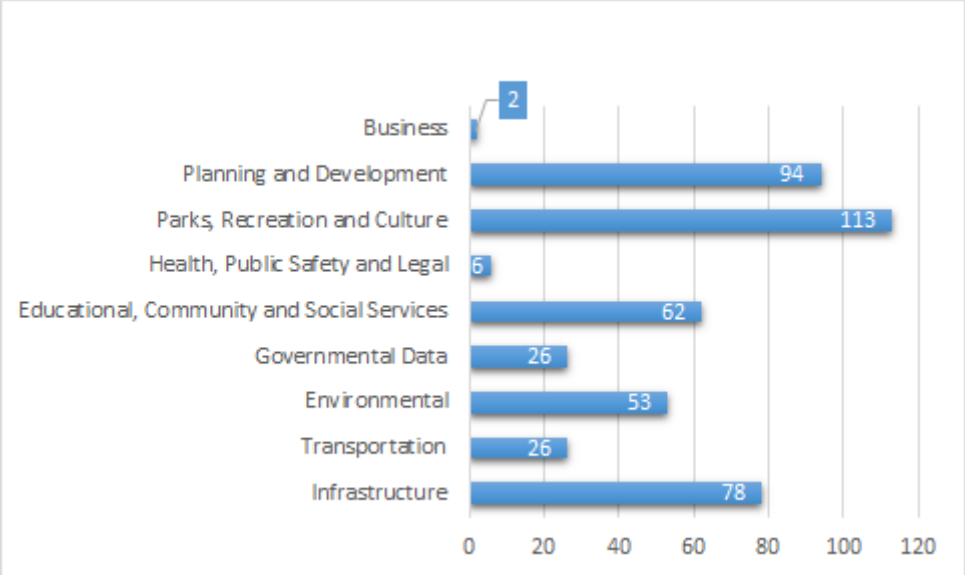
City of Edmonton Datasets by Theme



City of Toronto Datasets by Theme



City of Ottawa Datasets by Theme



Appendix C – City of Toronto Web/ Accessibility Statistics

Red highlighted data files = non-spatial data file type, ranked by downloads

Rank	Event Source Page	Specific file in the dataset	Visi	Download	Downloads to Visits Ratio
1	On-Street Permit Parking Area Maps	http://opendata.toronto.ca/gcc/parking_permit_areas_wgs84.zip	9785	3369	0.34
2	Property Boundaries	http://opendata.toronto.ca/gcc/property_bnds_gcc_mtm3.zip	4801	1537	0.32
3	On-Street Permit Parking Area Maps	http://opendata.toronto.ca/gcc/parking_permit_areas_mtm3.zip	9785	1530	0.16
4	Building Permits - Active Permits	http://opendata.toronto.ca/building/active.permits/activepermits_csv.zip	3010	1005	0.33
5	Signalized Intersection Traffic and Pedestrian Volume	http://www1.toronto.ca/City Of Toronto/Information & Technology/Open Data/Data Sets/Assets/Files/Signalized Intersection Traffic and Pedestrian Volume.xlsx	1907	957	0.50
6	Neighbourhoods	http://opendata.toronto.ca/gcc/neighbourhoods_planning_areas_wgs84.zip	2157	842	0.39
7	TTC Routes and Schedules	http://opendata.toronto.ca/TTC/routes/OpenData_TTC_Schedules.zip	1632	780	0.48
8	Property Boundaries	http://opendata.toronto.ca/gcc/property_bnds_gcc_wgs84.zip	4801	775	0.16
9	Economic Indicators	http://opendata.toronto.ca/it/com/economic_indicators.xlsx	1291	768	0.59
10	Building Permits - Active Permits	http://opendata.toronto.ca/building/active.permits/BuildingPermitsActiveReadme.xls	3010	728	0.24
11	TTC Ridership - Ridership Numbers and Revenues Summary	http://opendata.toronto.ca/city.manager/performance.management/PM_TTC.xls	1910	709	0.37
12	Toronto Centreline (TCL)	http://opendata.toronto.ca/gcc/centreline_wgs84.zip	1789	578	0.32
13	Toronto Centreline (TCL)	http://opendata.toronto.ca/gcc/centreline_mtm3.zip	1789	558	0.31
14	TTC Ridership Analysis	http://www1.toronto.ca/City Of Toronto/Information & Technology/Open Data/Data Sets/Assets/Files/Analysis of ridership 1985-2016.xlsx	1150	518	0.45
15	3D Massing	http://cot-planning.maps.arcgis.com/apps/webappviewer/	7117	499	0.07
16	TTC Ridership - Ridership Numbers and Revenues Summary	http://opendata.toronto.ca/city.manager/performance.management/PM_TTCRidershipNosRevenueSummaryReadme.xls	1910	483	0.25
17	TTC Real-Time Next Vehicle Arrival (NVAS)	http://www.nextbus.com/xmlFeedDocs/NextBusXMLFeed.pdf	1367	474	0.35
18	3D Massing	http://opendata.toronto.ca/planning/3dmassing/3DMassingShapefile_2016_MTM3.zip	7117	458	0.06
19	Signalized Intersection Traffic and Pedestrian Volume	http://www1.toronto.ca/City_Of_Toronto/Information_&_Technology/Open_Data/Data_Sets/Assets/Files/signalizedTrafficPedestrianVolumesReadme.xls	1907	447	0.23
20	Parks	http://opendata.toronto.ca/gcc/city_green_space_wgs84.zip	867	419	0.48
21	Building Permits - Active Permits	http://opendata.toronto.ca/building/active.permits/activepermits.zip	3010	409	0.14

22	3D Massing	http://opendata.toronto.ca/planning/3dmassing/3DMassingShapefile_2016_WGS84.zip	7117	397	0.06
23	Address Points (Municipal) - Toronto One Address Repository	http://opendata.toronto.ca/gcc/address_points_wgs84.zip	1779	309	0.17
24	Green P Parking	http://www1.toronto.ca/City Of Toronto/Information & Technology/Open Data/Data Sets/Assets/Files/greenPParking2015.json	5261	307	0.06
25	City Wards	http://www1.toronto.ca/City_Of_Toronto/Information_Technology/Open_Data/Data_Sets/Assets/Files/wards_may2010_wgs84.zip	854	306	0.36
26	Ambulance Station Locations	http://opendata.toronto.ca/gcc/ambulance_facility_wgs84.zip	673	305	0.45
27	Zoning By-law	http://opendata.toronto.ca/gcc/zoning_wgs84.zip	1065	282	0.26
28	TTC Ridership - All Day Weekday for Surface Routes	http://www1.toronto.ca/City Of Toronto/Information & Technology/Open Data/Data Sets/Assets/Files/Ranking surface routes 2016 Open Data Toronto.xlsx	797	281	0.35
29	TTC Ridership - Ridership Numbers and Revenues Summary	http://opendata.toronto.ca/city.manager/performance.management/TTC-Ridership.csv	1910	276	0.14
30	Neighbourhoods	http://opendata.toronto.ca/gcc/neighbourhood_planning_areas_utm6.zip	2157	263	0.12
31	TTC Subway Shapefiles	http://opendata.toronto.ca/gcc/TTC_subway lines_wgs84.zip	591	263	0.45
32	Bicycle Stations (Bike Share Toronto)	http://www1.toronto.ca/City_Of_Toronto/Information_&_Technology/Open_Data/Data_Sets/Assets/Files/bicycle_stations_bixi_readme.txt	1661	255	0.15
33	Tenant Notification for Rent Reduction	http://www1.toronto.ca/City Of Toronto/Information & Technology/Open Data/Data Sets/Assets/Files/tenantnotification2016.xlsx	615	250	0.41
34	Neighbourhoods	http://opendata.toronto.ca/gcc/neighbourhoods_planning_areas_mtm3.zip	2157	239	0.11
35	Traffic Cameras	http://opendata.toronto.ca/transportation/tmc/rescucameraimages/Data/tmcearthcameras.csv	846	231	0.27
36	Bicycle Stations (Bike Share Toronto)	http://www.bikesharetoronto.com/stations/json	1661	227	0.14
37	Wellbeing Toronto - Safety	http://opendata.toronto.ca/social.development/wellbeing/WB-Safety.xlsx	341	225	0.66
38	Neighbourhood Improvement Areas	http://opendata.toronto.ca/gcc/neighbourhood_improvement_areas_mtm3.zip	682	223	0.33
39	Dinesafe	http://opendata.toronto.ca/public.health/dinesafe/dinesafe.zip	875	222	0.25
40	TTC Ridership - Subway/Scarborough RT Station Usage	http://www1.toronto.ca/City Of Toronto/Information & Technology/Open Data/Data Sets/Assets/Files/2015 Subway Platform Usage Open Data Toronto.xlsx	571	220	0.39
41	Building Permits - Cleared Permits Prior Years	http://opendata.toronto.ca/building/cleared.permits/TBPCleared2001_2014_csv.zip	1211	217	0.18

42	Bicycle Counts	http://opendata.toronto.ca/transportation/bicycle.counts/bicycle.counts.zip	540	212	0.39
43	Building Permits - Cleared Permits Prior Years	http://opendata.toronto.ca/building/cleared.permits/clearedpermits2016_csv.zip	1211	211	0.17
44	Green P Parking	http://www1.toronto.ca/City_Of_Toronto/Information_&_Technology/Open_Data/Data_Sets/Assets/Files/GreenPParkingReadme.xls	5261	210	0.04
45	Building Permits - Cleared Permits Prior Years	http://opendata.toronto.ca/building/cleared.permits/clearedpermits2015_csv.zip	1211	208	0.17
46	Address Points (Municipal) - Toronto One Address Repository	http://opendata.toronto.ca/gcc/address_points_mtm3.zip	1779	208	0.12
47	Building Permits - Cleared Permits Current Year	http://opendata.toronto.ca/building/cleared.permits/clearedpermits2017_csv.zip	592	208	0.35
48	Parking Tickets	http://opendata.toronto.ca/revenue/parking/ticket/parking_tickets_2015.zip	1148	207	0.18
49	Former Municipality Boundaries	http://opendata.toronto.ca/gcc/formerMunicipalityBoundaries_mtm3.zip	475	204	0.43
50	3D Massing	http://opendata.toronto.ca/planning/3dmassing/3DContextMassing_Tile_Locator.pdf	7117	202	0.03
51	Parking Tickets	http://opendata.toronto.ca/revenue/parking/ticket/parking_tickets_2016.zip	1148	196	0.17
52	Zoning By-law	http://opendata.toronto.ca/gcc/zoning_mtm3.zip	1065	193	0.18
53	Municipal Licensing and Standards - Business Licences and Permits	http://opendata.toronto.ca/mls/business.licences/business.licences.csv	563	184	0.33
54	Wellbeing Toronto - Transportation	http://opendata.toronto.ca/social.development/wellbeing/WB-Transportation.xlsx	312	184	0.59
55	City Wards	http://www1.toronto.ca/City_Of_Toronto/Information_Technology/Open_Data/Data_Sets/Assets/Files/wards_may2010.zip	854	182	0.21
56	Hotel Association Member List for the Greater Toronto Area	http://opendata.toronto.ca/gcc/hotel_gtha_membership_list.xls	381	178	0.47
57	Bicycle Stations (Bike Share Toronto)	http://www.bikesharetoronto.com/data/stations/bikeStations.xml	1661	176	0.11
58	Bikeways	http://opendata.toronto.ca/gcc/bikeways_wgs84.zip	605	173	0.29
59	Fire Station Locations	http://www1.toronto.ca/City_Of_Toronto/Information_Technology/Open_Data/Data_Sets/Assets/Files/fire_stns.zip	533	173	0.32
60	Wellbeing Toronto - Health	http://opendata.toronto.ca/social.development/wellbeing/WB-Health.xlsx	282	172	0.61
61	Regional Municipal Boundary	http://opendata.toronto.ca/gcc/torontoBoundary_wgs84.zip	416	164	0.39
62	3D Massing	http://opendata.toronto.ca/planning/3dmassing/3DMassingMultipatch_2016_WGS84.gdb.zip	7117	158	0.02

63	Address Points (Municipal) - Toronto One Address Repository	http://www1.toronto.ca/City Of Toronto/Information & Technology/Open Data/Data Sets/Assets/Files/readme_address_points_Jan2013.txt	1779	154	0.09
64	Intersection File - City of Toronto	http://opendata.toronto.ca/gcc/centreline_intersection_wgs84.zip	469	154	0.33
65	Building Permits - Cleared Permits Prior Years	http://opendata.toronto.ca/building/cleared.permits/clearedpermits2014_csv.zip	1211	149	0.12
66	Parking Tickets	http://opendata.toronto.ca/revenue/parking/ticket/parking_tickets_readme.xls	1148	149	0.13
67	Intersection File - City of Toronto	http://opendata.toronto.ca/gcc/centreline_intersection_mtm3.zip	469	149	0.32
68	Parking Ticket Trials	http://opendata.toronto.ca/revenue/parking/trial/Parking_Ticket_Trials.zip	268	149	0.56
69	Bicycle Stations (Bike Share Toronto)	http://tor.publicbikesystem.net/ube/gbfs/v1/	1661	146	0.09
70	Wellbeing Toronto - Economics	http://opendata.toronto.ca/social.development/wellbeing/WB-Economics.xlsx	230	146	0.63
71	Ward Profiles	http://opendata.toronto.ca/it/com/Ward Profiles - WardAreas.xlsx	954	144	0.15
72	Traffic Signals Tabular	http://app.toronto.ca/opendata/cart/traffic_signals.csv	465	143	0.31
73	Bikeways	http://opendata.toronto.ca/gcc/bikeways_mtm3.zip	605	142	0.23
74	Red Light Cameras	http://www1.toronto.ca/City Of Toronto/Information & Technology/Open Data/Data Sets/Assets/Files/RLC List.xlsx	379	142	0.37
76	One Way Streets	http://opendata.toronto.ca/gcc/oneways_mtm3.zip	519	137	0.26

References

Baculi, E. (2014). *Content and growth of municipal open data catalogues.* GoGeomatics Canada. Retrieved from http://www.gogeomatics.ca/magazine/content-and-growth-of-municipal-open-data-catalogues.htm#

Borzacchiello, M.T. & Craglia, M. (2012). The impact of innovation of open access to spatial environmental information: a research strategy. *Int. J. Technology Management* 60(1/2): 114-129.

Chan, M., Johnson, P.A., & Shookner, M. (2016). Assessing the use of government open data and the role of data infomediaries: The case of Nova Scotia's community counts program. JeDEM 8(1): 1-27.

City of Edmonton. (2017). *What is open data*? Retrieved from https://www.edmonton.ca/city_government/initiatives_innovation/open-data.aspx

City of Toronto. (2017). *Open data master plan – update report.* Retrieved from http://www.toronto.ca/legdocs/mmis/2017/ex/bgrd/backgroundfile-102582.pdf

Currie, L.J. (2013). *The role of Canadian municipal open data initiatives: A multi-city evaluation*. Unpublished master's thesis, Queen's University, Ontario, Canada.

Denham, E. Government of British Columbia. (2013). *Evaluating the government of British Columbia's open data initiative.* (Report No. F13-03).

Dong, H., Singh, G., Attri, A., & El Saddik, A. (2017). Open data-set of seven Canadian cities. *IEEE Access 5:* 530-543

Evans, A.M., & Campos, A. (2013). Open government initiatives: Challenges of citizen participation. *Journal of Policy Analysis and Management* 32(1): 172-203.

Giggey, R. (2012, May 7) *The G4: Setting city data free.* Retrieved from http://canadiangovernmentexecutive.ca/the-g4-setting-city-data-free/

Gill, M., & Corbett, J. (2017). Downscaling: Understanding the influence of opendata initiatives

in smaller and mid-sized cities in British Columbia, Canada. *The Canadian Geographer DOI: 10.1111*

Government of Canada (2016). *Canada's action plan on open government 2012-2014.* Retrieved from http://open.canada.ca/en/canadas-action-plan-open-government

Graves, A., & Hendler, J. (2013). Visualization tools for open government data. *Dg.o'13 Proceedings of the 14th Annual International Conference on Digital Government Research* (pp. 136-145). New York, NY, USA: ACM.

Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, adoption barriers and myths of open data and open government. *Information Systems Management* 29(4): 258-268.

Janssen, M. & Zuiderwijk, A. (2014). Infomediary business models for connecting open data providers and users. *Social Science Computer Review* 32(5): 694-711.

Johnson, P. (2016). Reflecting on the success of open data: How municipal government evaluates their open data programs. *International Journal of E-Planning Research* 5(3). DOI: 10.4018/IJEPR.2016070101.

Johnson, P. and Robinson, P. (2014). Civic hackathons: Innovation, procurement, or civic engagement? *Review of Policy Research* 31(4): 349-357.

Machova, R., & Lnenicka, M. (2016). Evaluating the quality of open data portals on the national level. *Journal of Theoretical and Applied Electronic Commerce Research* 12(1): 21-41.

Martin, E.G., & Begany, G.M. (2017). Opening government health data to the public: benefits, challenges, and lessons learned from early innovators. *Journal of the American Medical Informatics Association* 24(2): 345-351.

Open Data 150 Canada (n.d.). *About this project*. Retrieved from http://canada.opendata500.com/index.html

Open Data 500 (n.d.). *The OD500 global network.* Retrieved from http://www.opendata500.com/

Pereira, G.V., Macadar, M.A., Luciano, E.M., & Testa, M.G. (2017). Delivering public value through open government data initiatives in a smart city context. *Information Systems Frontiers* 19(2): 213-229.

Robinson, P.J., & Johnson, P.A. (2016). Civic hackathons: New terrain for local government-citizen interaction? *Urban Planning* 1(2): 65-74.

Roy, J. (2016). Open data and open governance in Canada: A critical examination of new opportunities and old tension. *Future Internet* 6(3): 414-432.

Sadiq, S., & Indulska, M. (2017). Open data: Quality over quantity. *International Journal of Information Management.* 37(3): 150-154.

Sandoval-Almazan, R., & Gil-Garcia, J.R. (2012). Are government internet portals evolving towards more interaction, participation, and collaboration? Revisiting the rhetoric of e-government among municipalities. *Government Information Quarterly* 29(S1): S72-S81.

Scassa, T. (2015). Public transit data through an intellectual property lens: Lessons about open data. *Fordham Urban Law Journal* 41(5): 1759-1810.

Sieber, R.E., & Johnson, P.A. (2015). Civic open data at a crossroads: Dominant models and current challenges. *Government Information Quarterly* 32(3): 308-315.

Thorsby, J., Stowers, G.N.L., Wolslegel, K., & Tumbuan, E. (2017). Understanding the content and features of open data portals in American cities. *Government Information Quarterly* 34(1): 53-61.

Vogel, L. (2011). The secret's in: Open data is a foreign concept in Canada. *Canadian Medical Association. Journal.* 183(7): E375-376.

(WWW Foundation) World Wide Web Foundation (2016). *Open Data Barometer: ODB methodology - v1.0*

Zeleti, F.A., & Ojo, A. (2017). Open data value capability architecture. *Information Systems Frontiers* 19(2): 337-360.