Open Civic Data Standards in Canada
Rachel Bloom and Renee Sieber
for Geothink (http://geothink.ca/)

Introduction

Municipal governments hold numerous goals in publishing data, among them, to encourage local economic development, improve service delivery, provide internal business intelligence, and increase transparency and public accountability (Sieber and Johnson, 2015). A precondition for any of these goals are datasets that are machine readable and interoperable. One way to ensure machine readability and interoperability are through open data standards. Our partners at Open North gave a clear example of why standards are important:

> When a technology is widespread, standardization provides better predictability for those who rely on the technology. Consider light bulbs, for example. All incandescent light bulbs have the same screw base, and you can reliably expect that when you buy a bulb, it will fit into your lamp's socket. Compact fluorescent and LED bulbs use the same screw base, so that your old light sockets still work with these new bulbs. You may have never even thought to worry about whether a bulb will fit your socket! All users of this technology – manufacturers, electricians, consumers – benefit from not having to think about a bulb's screw base (Guidoin and McKinney 2012).

Open data standards provide the semantic and schematic guidelines to ensure machine readability and interoperability so governments can actually succeed in opening up data to the public.

In Geothink, we identified predominant civic open data standards and created a set of quality and performance metrics by which to evaluate these standards. We then developed two spreadsheets to assist municipal governments in making choices about standardization of open data:

1. The first spreadsheet provides an inventory of 22 open data standards that local governments can apply to their datasets. This spreadsheet applies a set of metrics and evaluations to that inventory so potential adopters can easily compare across and within domains. An inventory of open civic data standards provides a resource for informing potential adopters about what standards exist for implementation. The hope is that civil society and municipal workers will become more knowledgeable about tools and strategies for standardization as they open and share their data.

2. The second spreadsheet evaluates ten high-value datasets of five Canadian municipal governments in order to observe discrepancies that exist between how the data is being published and how standards dictate the ways the data should be published by governments. The cities considered within the second spreadsheet consist of Vancouver, Toronto, Surrey, Edmonton, and Ottawa. We chose ten high-value datasets according to the spreadsheet created by Jury Konga for the Open Knowledge Foundation (OKF) (refer to http://bit.ly/1NCyPnh). This OKF spreadsheet recognizes the high-value dataset categories included in the G8's Open Data Charter. Our spreadsheet organizes the salient schematic and semantic information of each dataset; including the data's structuration, file formats, metadata, and any applied standards.

Spreadsheets 1 and 2
https://docs.google.com/spreadsheets/d/12wcUhE6waDz0RPT81E5aebcJf58AH92FstMZQKy5kRc/edit?usp=sharing.

These two spreadsheets reflect complementary approaches to show the value of open data standards. The first spreadsheet focuses on schematic and semantic guidelines for structuring the data for optimal interoperability. The second spreadsheet presents datasets at the civic level and allows municipalities to become aware and compare how other municipalities. Comparing Spreadsheets 1 and 2 demonstrates how closely published civic datasets resembles the guidelines of data standards in the first spreadsheet. Before providing further explanation for the two spreadsheets, I will first define an open data standard and present its advantages for government and civil society.

Defining an open data standard

At their most basic level, standards are the difference between technology and social agreements made about technology (Russell, 2014). A *data* standard refers to "agreements on representations, formats, and definitions of common data" (Public Health Standards Consortium, n.d.). These standards use communication and collaboration to maximize the *interoperability* and *compatibility* of data. In addition, data standards facilitate the measurability, performance, and comparability of data through establishing common file formats, schema, and unique identifiers for data elements (Russell, 2014). For example, data standards may allow one to determine whether drug treatment "a" is better than drug treatment "b" because one has standardized patient health outcomes.

This paper considers standards for open data. According to the Open Knowledge Foundation (2014), open data is raw digital data that should be freely available to anyone to use,

repurposable and re-publishable as users wish and absent mechanisms of control like restrictive licenses. For this project, the focus is on open data from governments and, particularly, municipalities. A standard for open data allows for greater understanding of the underlying raw data and enables the repurposability of that data. As a result, standards improve the utility of the data. For example, Huijboom and Van den Broek (2011) surveyed numerous countries' open data strategies, representatives of whom identified a "lack of open data standards between (levels of) government organisations … as a barrier to open data usage by citizens and businesses and subsequently new open data policy".

Standards in the open data world tend to work across a variety of levels. Davies outlined these in his 2015 presentation at the International Open Data Conference. From this presentation, Davies cited Palfrey and Gasser's (2012) book to highlight the difference between quality standards and interoperability standards for open data. The former refers to standards that intend to asses the quality of open data. These types of standards use metrics, rating systems, and filters to assess quality and assign benchmark accountability in how to publish the data. For example, the Open Knowledge Foundation relies on the Global Open Data Index to assess the state of publication of key datasets by governments around the world (http://index.okfn.org/methodology/). The Global Open Data Index applies quality metrics, such as whether the datasets are machine readable, free to use, and use open licenses, to assess the quality of the open datasets. The culmination of the evaluation through quality metrics result in a comparable rating of each government's performance in publishing open data. On the other hand, interoperability standards rely rely on different kinds of mechanisms, such as data schemas and controlled vocabularies, to make the open data more transferable across systems and places.

With this in mind, this project focuses on open data standards that intend to make open data more interoperable. In the most basic sense, interoperability among informational technologies refers to the power to transfer and render useful data across systems, applications, or components (Palfrey and Gasser, 2012). In other words, one of interoperability standards' greatest contributions to publishers and consumers of open data is their ability to make independently devised systems communicate with each other. This communication occurs in a way that is more lightweight and sophisticated than the complete integration of data. These standards utilize technical data elements such as data schemas and unique identifiers to relate these independent information systems. These technical underpinning are defined by the standard's schematic guidelines. These guidelines express the means of communications - specifically how the data should be structured and how the content should be stored (http://govex.jhu.edu/enabling-civic-data-standards/).

In addition to schematic guildines, standards must define their vocabulary and what they mean when they specify information such as aggregated expenditures or types of crime

(http://govex.jhu.edu/enabling-civic-data-standards/). These semantic specifications ensure that everyone is talking about the same thing when they are using common codes to encapsulate and express the data observations. While vocabularies should be controlled, they must also be flexible enough to accommodate the data at different localities and levels of government.

Beyond these core components, interoperable systems are complex and defined at multiple levels. As a result, a definition of interoperability is context-specific. Therefore, the term interoperability in the open data movement ought to remain flexible and evolving in order to acknowledge different levels of interoperability at work across different sectors and cases. To understand the meaning of interoperability, we must first ask *who* defines interoperability and *for what purposes* do they base this definition.

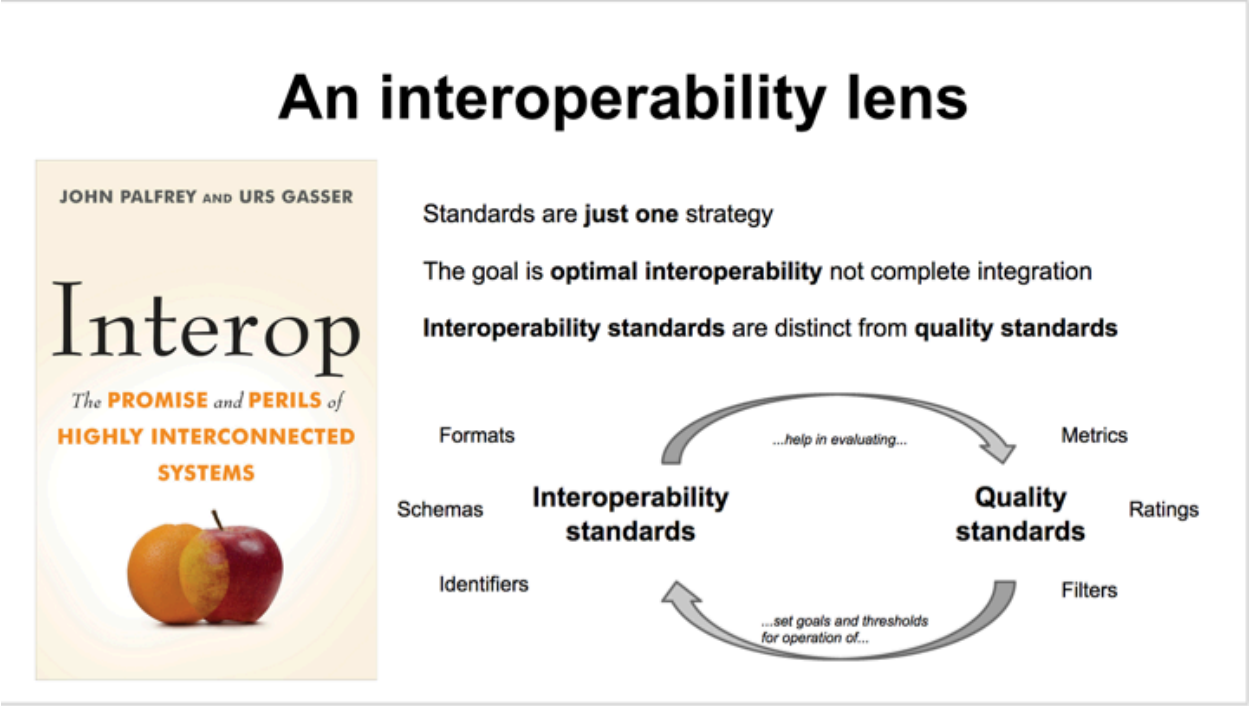

***Figure 1*:** Palfrey and Gasser's interoperability theory explained by Tim Davies

It should be noted that open data standards does not necessarily mean standards for open data. An *open* standard for data requires a collaborative, transparent, and consensus-driven process to maintain its development (Palfrey and Gasser, 2012). An open standard process ought to be made available and accessible to the general public without barriers or restrictions. In his book, *Open Standards and the Digital Age*, Russell (2014) asserts that open standards embody the vision of participatory democracy because they value fairness, transparency, due process, and rights of appeal.

However, there can be standards for open data that are not themselves completely open. The International Standards Organization, in collaboration with the World Council on City Data, created the performance standard ISO 37120 Sustainable Development of Communities: Indicators for City Services and Quality of Life (http://www.dataforcities.org/wccd/). ISO engages in an open process to create the standard but also requires users to pay for a patent license. This point emphasizes the subtle difference between talking about standards for open data and open standards for data. ISO reflects the case of the former. However, open standardization adheres most closely to the mandate of the open data movement and moves towards an optimal environment for opening up municipal data.

Consequently, this paper advocates open standardization for opening up civic datasets. An additional driver of opening the process is ensuring publishers and re-users of the data and technology do not get excluded from a standard's evolution. Open standards ensure that the open data movement's values of transparency, accessibility, and accountability are upheld throughout the process of standardization. Therefore, sustainable environments and optimal interoperability for open data and informational systems can be attributed to open agreements made about technology by multiple stakeholders.

Drivers and benefactors of open data standards

Standard processes, the standard's development and maintenance, include participatory stakeholders and benefactors. Stakeholders benefit from standards that appear "authoritative, objective, uncontroversial, and natural" (Russell 2014).

The standardization of the technological communication systems emphasize stakeholder organizational boundaries and any alliances that mobilize across those boundaries (Russell, 2014). Hybrid organizations and multiple stakeholder participation are a driver of the open standards processes. The types of publishers are not exclusively hierarchical or market driven; rather, they are a mixture of the two (*ibid*.). Involvement of multiple types of actors (whether public, private, or hybrid) reinforce the conceptual integrity of data standards.

Government bodies benefit from applying domain-specific standards to their openly published data. Standards help to ensure that government publishes the data in a way that is more understandable and usable for the public. Consistent schematic and semantic data, equipped with metadata, encourages a benchmark accountability by government in publishing open data. It also improves the transparency and comparability of public information and government services.

In addition, data standards may enable positive economic feedback cycles by easing accessibility to and implementation of the data. In a neoliberal system, standards that remain open in process may remove barriers to innovation and provide a more diverse array of goods and services at

lower prices (*ibid*.). Lower prices in turn allow for more consumption of goods and services. Therefore, standards applied to local datasets facilitate third party applications and services for the public.

For example, the General Transit Feed Specification (GTFS), developed through a partnership between Google and the public transit agency in Portland, Oregon, exemplifies a data standard that has been successfully applied to open transit data for hundreds of cities around the world. GTFS was started because a government employee realized the public value of making transit data more interoperable and repurposable.The standard publishes the municipal transit data in simple CSV files so that smaller agencies and less technically skilled data reusers can enter into transit technology. The implementation of GTFS by transit agencies has made accessing public transit directions on mapping platforms such as Google Maps a reality (McHugh 2013).

GTFS, as a standard for open data, has succeeded in easing transit data has opened up the data by providing interoperability across of variety of transit services and urban areas. A crucial component of the success of GTFS comes from its commitment to the principles of open source and open data. GTFS exemplifies how open standards for open data are capable of benefiting publishers and consumers of civic data on a large-scale.

Methodology

Our project consists of two spreadsheets. The first is an inventory of domain-specific open data standards that are evaluated by a set of metrics. The second is a survey of 10 high-value datasets for each of the cities of Vancouver, Toronto, Surrey, Edmonton and Ottawa.

Selection of open data standards

There are numerous potential open data standards to study. Therefore, for the purposes of this project, we have identified a set of high value data domains. Domains for high-value datasets are taken from Jury Konga's spreadsheet for the Open Knowledge Foundation (OKF) of top twenty high-value datasets (refer to http://bit.ly/1NCyPnh). The OKF table applies G8 categorization, description of dataset, and rationale for why the dataset is considered key or not. We selected 12 data domains reflect datasets that offer the great value to the public and should therefore be published openly. In addition, this table covers the domains presented for the second spreadsheet, which evaluates the adoption of these domain-specific standards by city. … The data domains chosen for the survey are annual budgets, building permits, crime, elections results, public facilities, road construction, service requests, transit, zoning, expenditure, procurement contracts, and food safety.

We then selected 22 open data standards. These standards, situated within their domains, explicitly emphasize openness in the standard's conceptualization. Therefore, these 22 chosen standards provide enough documentation in their specification to indicate a strategy for data standardization that values transparency, multi-stakeholder participation, and that is consensus driven.

Quality metrics of domain-specific open data standards

Metrics of the project's first spreadsheet represent a measured evaluation of the standard's performance in interoperability and openness. These metrics, instead of evaluating what is easy to measure, focus on important evaluations of each standard. Therefore, the metrics and standards applied to each civic data standard are meant to provide background about the who, what, where, and how of the open data standards process across domains. The metrics are inspired by the Dublin Core metadata element set.

The inventory equips potential adopters of open data standards with information about the publishers, schema, the extensibility of components, and the level of use for each of the standards specified in the spreadsheet. It also provides information regarding whether the open standard process includes stakeholder participation and whether its conception is consensus-based. The inventory also offers a resource for learning how to contribute to the development of a standard.

In addition, the inventory highlights any overlaps and gaps in the open data standards initiative. For example, zoning data, because it is sensitive to geographic idiosyncrasies, is difficult to standardize across government jurisdictions. On the other hand, initiatives to establish standards for budget data have been popular to the extent that publishers of these standards would benefit from collaborating with each other to avoid any further overlap.

I chose the evaluation metrics assuming that publishers adopt standards to improve the interoperability and quality of data for data re-users. Inspiration for these metrics came from the demands and requirements voiced over forums, articles, and texts by *both* publishers and re-users of open municipal data. For example, users of real time transit data require standards that publish the data in real time. Therefore, metrics provide practical and relevant information about how best to implement the standards in order to make the data more useful to re-users. Moreover, open data publishers require information regarding the processes of standard adoption and implementation. Such relevant information includes background information about the standard's publisher and whether the standard is still active or in its draft stage.

The process was iterative and developed across domains rather than within them. Therefore, the metrics did not serve the interests of a particular domain, but aimed to observe strategies for standardization consistent among all types of data. In conclusion, the applied evaluations of Spreadsheet 1 provide information about the potential for achieving optimal interoperability through open data standards.

For the first spreadsheet, columns are organized into a set of metrics to assess each observation's quality and performance. Rows are organized by standards specific to the data domains mentioned above. When possible, column variables include categorical options of yes, no, or unsure. These categorical columns have an accompanying column that provides the rationale.

Table 1. Background information about the standard and metrics for interoperability (Spreadsheet 1)

Name of Metric/Characteristic	Description	Reasoning for Choice	Range
Name	Name signifies what publishers label their standard	This metric is for identifying the standard being evaluated.	- -
Domain	These domains are considered to be high-value to data publishers and data re-users. Government prioritizes publishing these datasets and these domains are considered a priority for publishing in the "G8 Open Data Charter".	This project is concerned with open data standards that are domain-specific. Therefore, chosen standards take into account vocabularies that are rooted within a certain domain.	**Name**
Publisher	Actors in charge of the standard's conception and further development.	This metric identifies the coordinating body and implies the degree of partnerships and collaboration associated with the standard's conception.	**Name**
Publisher Reputation	Publisher reputation provides background information about an organization or company and their level of expertise regarding standards.	Open standards process distinguishes itself from unilateral approaches by frequently requiring collaboration and coordination among civil society, state actors, and the private sector. Municipal publishers may prefer to partner with certain kinds of publishers over others.	- -

License Information	This column provides for information about intellectual property rights of the standard.	These institutional reinforcements help to protect the interests of parties and instill confidence in stakeholders who create the technology.	**License name**
Description and Purpose	This background information indicates the publisher's objective for the standard. It includes a brief statement about the required content and schema of the standard.	Gives a brief idea about why publishers conceived the standard and provides a justification for its proposed implementation.	- -
Transferability to Other Jurisdictions	The metric for whether a standard is easy to adopt across jurisdictions is split up into two columns. The first column defines whether the standard is easy to adopt across jurisdictions on a yes, no, or unsure basis. The second columns gives the rationale for the first column. There is a hierarchy regarding the ease of implementation for a standard. For example, CSV format of a standard requires a minimal degree of resources and technical knowledge. On the other hand, more complex and sophisticated ways of formatting standards, such as RDF and SOAP, are not as easy for municipal bodies to implement. More often than not, sophisticated formats tend not to be manageable for municipal actors that lack resources and technical background. Standards that handle dynamic data and cURL APIs exemplify more complex ways of publishing city datasets.	The ease by which standards can be implemented depends on the domain and required relevancy of the data, in addition to other factors. For example, an ideal standard for zoning data would facilitate apps that map a city's most important regulatory characteristics across jurisdictional boundaries. However, devising zoning standards is particularly difficult to do because codes are tailored to jurisdictions (https://groups.google.com/forum/#!topic/ospt-ecosystem/1GUkwcTjBBE). Initiatives to create standards for budget data, on the other hand, have made great strides because budget data tends to utilize common vocabulary and schemas for aggregated expenses and incomes. In addition, agreeing on vocabulary for a domain can impede adoption of a standard. For election data, definitions for what constitutes an electoral process can be flexible and depend on context. Therefore, contested narratives and	**"Yes"/ "No"/ "Unsure"; Rationale**

		alternative vocabularies complicate the process of devising standards for election data results.	
Stage in Development	Standards in the dashboard exist at different stages of development. Standards are either in their draft or completed stage.	Timelines help potential adopters observe the evolution and future development of the standard.	**“Completed”/ “Draft”**
Version	Version types are noted for completed standards. Draft standards include any known timeline information about future plans for development.	Information about timelines help potential adopters observe the evolution and future development of the standard.	**Version number; Date**
Date Last Updated by Publishers	The last time publishers updated the standard.	Lets potential adopters know if the standards is still active and being improved by publishers.	**Date**
Level of Use by Governments	Level of use can be reflected in the quantity of governments that have adopted or have proposed to adopt the standard.	These indicators specify whether a standard is well known and embraced by stakeholders. Level of use does not necessarily indicate the quality and effectiveness of a standard. Instead, this metric intends to give an idea about the scope of the standard’s adoption.	**City names**
Extensions	The first column indicates whether the standards has extensions. The second column gives rationale for the first column. Extension indicate the flexibility for a standard’s implementation. In addition, extensibility of a standard provides insight into how a standard is being implemented and enhanced for specific purposes.	An open standard may not prohibit extensions except to protect against predatory practices. In these cases, the standard may use license terms to require publishing reference information and provide software that is compatible with any occurring extensions. Such limitations keep the standard open for potential implementers.	**“Yes” / “No” ; Rationale**

Example	This column exemplifies an actual implementation for each standard.	Real world examples of the standard allow potential users to understand how the standard has been adopted and applied to the data.	**Link or text sample**
Example Information	This variable gives background on the example of the standard's implementation.	Gives context to the example provided in the Example column.	**Text**
Machine Readable	Acceptable machine readable structures include XML, RSS feed, CSV, RDF, JSON, TXT, XLS(X), and KML formats. Formats that are *not* machine readable include PDF, HTML, DOC(X), anything scanned, anything faxed, and anything typed in an email (Suszan, 2014). Standard's ought to compliment techniques to provide human and machine readable structures for the data. Publishing data as machine readable includes the following: (1) established standard vocabularies, (2) enriching the HTML resources with metadata, semantics, and identifiers, (3) and implementing simple, manageable, and stable URIs (Bennett and Harvey, 2009). Data tables, according to the standard's specification, should be normalized so to be incorporated into a relational database.	Open standards should publish the data in a machine readable structure so that data can be parsed through automated processes.	**"Yes"/ "No" ; Rationale**
Human Readable	Human readable requires a medium of data or information that can easily be understood by people. Therefore, the standard should encode the data by using easily identifiable text. Of course, there are semantic consideration for human	Human readable standards keep the data open by allowing consumers of the data to read the standardized datasets for themselves.	**"Yes"/ "No" ; Rationale**

	readable standards. For example, there could be a variety of interpretive meanings associated with encoding the data through text.		
Specifies Up-to-Date Data	This metric varies depending on the domain of the data. Some domains require formats that handle data in real time. However, other domains may require that the standard specify that data be updated quarterly or annually. For example, standards that handle transit and road construction data would require a web feed format to deliver updates about developments as they occur. However, budget datasets only requires a quarterly or yearly update. In practice, many municipal publishers still publish data in static files.	This metric checks that the data standard is relevant and appropriate for potential users. The adoption of GTFS real-time and the Open511 API by municipal data publishers show a progression towards dynamic schemas that make data more usable and relevant to users.	**“Yes” / “No” ; Rationale**
Takes into Account Associated Metadata for the Dataset	This metric checks whether the standard schema requires metadata. A “yes” for this metric indicates a presence of both descriptive and structural metadata for the primary data. Each standard should readily make available the time and date of the data’s creation, the author, location of the data on the computer networks, and information about any standard applied to the raw data. Metadata should have embedded permanent and/or discoverable URIs and should utilize electronic citations of the data in the form of hyperlinks (Bennett and Harvey, 2009).	An optimal standard includes information about the content and structure of the raw data.	**“Yes”/ “No”/ “Unsure” ; Rationale**

Applications Built from Standard	Indicates if third parties have developed applications that use the standard.	Open data standards facilitate third party applications. This reflects how end users, whether it be public or private sector or civil society, engage with the standard. More applications that use the standard indicate that the data is popular and repurposable.	**Names of third party applications**
Standard's Documentation	Consists of a URL that points to the publisher's documentation of the standard.	Source websites usually host forums for participating in the standard's development. These forums include GitHub issue trackers and Google groups. They also contain technical information about the formatting of the standard, publishing information, and any timeline for future developments to the standard.	**Hyperlink**
References	Contains electronic citations in the form of hyperlinks for information documented in the spreadsheet. Related sources include blog posts, documentation manuals, news articles, and government websites.	Provides transparency for the origins of information collected in the table.	**Hyperlink(s)**
Date Last Updated in Dashboard	Provides the date of the last time that information about the standard was updated on the dashboard.	This indicator notes how relevant information about the standard is for the user and potential adopter.	**Date**

Table 2: Additional metrics for openness of standards process (Spreadsheet 1)

Name of Metric	Description	Reasoning for Choice	Range
Open License	What qualifies a standard as being "open" is debated. However, openness may be inferred when the standard is published under an open license. Open licenses iterate	Openness of the standards process ensures that consumers and producers of the technology do not get left behind during the standard's evolution.	**"Yes"/"No"**

	that anyone has the right to repurpose and share the material without restriction. Examples of open licenses include public domain licenses, the UK Open Government License v3, creative commons licenses, and open data common licenses (World Bank, Open Data Essentials).		
Multi-Stakeholder Participation	Stakeholders for a standard include civil society, government, and the private sector. An open standard should aim to include all types of stakeholders in its conception and maintenance. Types of stakeholder participation can be inferred based on the types of publisher reputations.	Stakeholder participation sets up a dialogue between those who publish the data and those who use it.	**“Yes”/ “No”/ “Unsure” ; Rationale**
Consensus-Based Governance	Standardization implies an ongoing dialogue between producers and consumers of data. It is important to note that consensus-based governance does not mean that all inputs are accepted if the majority agrees. Instead, consensus-based indicates a process *willing* to address any request pertaining to the standard’s statement of purpose. A charter providing transparency of decisions about the standard’s evolution support a consensus-based approach. Consensus-based governance can be inferred by the presence of a mailing list or active working group for the standard.	Consensus based metrics reflects the nature of critique as a productive act. This dimension of open standards acknowledges any possible concerns and critic associated with the standard’s evolution. An approach based on the consensus of stakeholders offers resistance to authorities that intend to control, censor, or ignore alternative voices and perspectives. Russell states, “the “open systems” created in the late twentieth century, and the “open standards” described by the title of this book, thus constitute critiques and rejections of ideologies of centralized control“ (Russell, 6).	**“Yes”/ “No” / “Unsure”; Rationale**

Evaluation of Adoption of Standards by City

Ten high-value datasets and their formats were compiled into a spreadsheet for the Cities of Vancouver, Toronto, Surrey, Edmonton and Ottawa. These cities were chosen for the project because they are leaders in the Canadian open data movement and have well-established open data catalogues. Just recently, Toronto, Surrey, Edmonton and Ottawa were listed as the top 4 and Vancouver was rated sixth (https://www.publicsectordigest.com/articles/view/1547) for their open data initiatives. The ten dataset domains were chosen from a list of high value datasets by Jury Konga for the OKF (refer to http://bit.ly/1NCyPnh). High value datasets are also cited in the G8 Open Data Charter as datasets that the government is committed to publishing openly.

This spreadsheet of surveyed datasets records the types of file formats, structuration, metadata/description of the data, and any open data civic schema applied to each of the datasets and recorded by each city. This spreadsheet gives an idea of the current state of these open datasets. Comparing Spreadsheet 2 with Spreadsheet 1 gives an idea of how much the published data deviates from the semantic and schematic guidelines of each standard.

Table 3. Metrics that assess state of municipal open datasets and their compliance to standardization

Name	Description	Rating?
Name	Refers to how the dataset is titled by the municipality publishing the data.	
File format	Indicates the standard way of storing the data in a computer file. Examples of file formats include both machine readable formats, such as CSV and JSON, and non-machine readable formats, such as PDFs and Web Documents. File formats can also be categorized by whether they are saved as a common interchange format (e.g., CSV) or a proprietary file format (e.g., XLS). An interchange format reflects a more open way of publishing the data, considering that XLS requires a proprietary application to read the data file.	
Structuration	Indicates how data is structured within the file format. For example, structuration field gives insight to	

	whether a database is normalized and can be related to other databases in a system.	
Metadata/description of data	Notes any applied metadata standard or background information about the data. In addition, this field notes how and where metadata is stored in the open data catalogue.	
Open data (OD) civic schema	Refers to any municipally applied open standard to the dataset. This category is limited to *domain-specific* standards. As seen in Spreadsheet 1, common examples include GTFS and Open311: GeoReport API	

Results

When we initially reviewed the datasets, we realized there existed no clear and solidified definition of an open data standard. Before conducting the research, we were aware of GTFS as a model standard for open transit data. Thus, by learning how GTFS works and why it works well, we were able to refine what we meant by an open data standard. With this in mind, we focused on standards that use specified data values and vocabularies, data schemas, and unique identifiers to make the data more interoperable. Therefore, we excluded standards for measuring the quality and availability of open datasets from the research results. In addition, metadata standards that are not rooted in a specific domain, such as RDF and Dublin Core Metadata Element Set, are excluded from this project's definition of an open data standard.

Spreadsheet 1

Results of the first spreadsheet show that some domains have more existing open data standards than other domains. Annual budget and election data domains have multiple available open data standards. However, domains for zoning data and public facilities data lack cohesive and widely applied open data standards. The figure below, visualized by Julia Conzon, organizes standards that exist for open data.

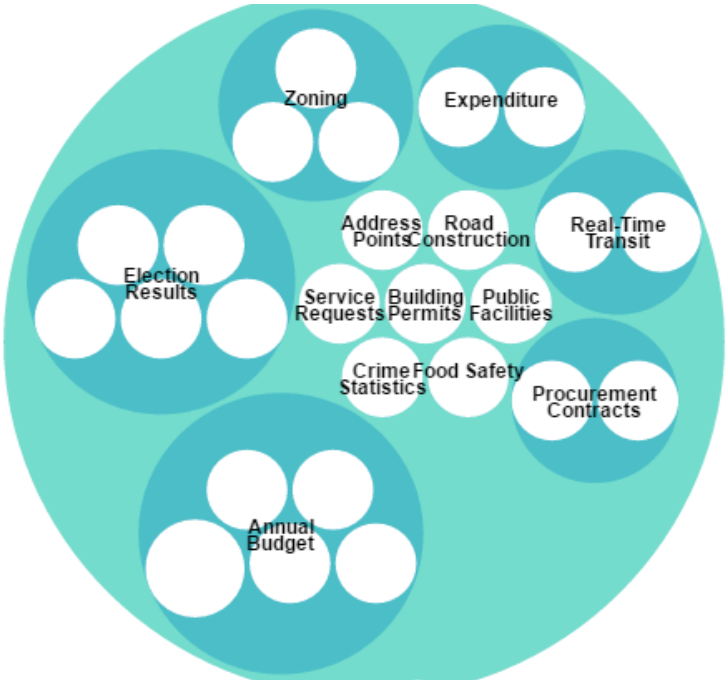

Figure 2: Visualizing standards for open data compiled in Spreadsheet 1

Types of publishers and reputations

Publishers of standards for open data include standard bodies, private technology companies, city governments and non-governmental organizations. Socrata, a company that offers technological services for opening up government data, has contributed to creating standards for open datasets. Socrata developed two open budget schemas, and also helped to create the BLDS specification. Socrata's Operating Budget Schema and Capital Budget Schema standards makes annual budget data more compatible with their budget open data platform and visualization software. Yelp and Google are two other private technology companies involved in the creation of standards. These companies perceive the value of utilizing standards to better reuse public data for consumer use by means of their products and services.

In addition, the majority of standards involved input from city governments for their conception. Standard bodies, such as the World Wide Web Consortium (W3C), European Committee for Standardization (CEN), and the National Institute for Standards Technology (NIST), have contributed to publishing standards. Non profit organizations and networks, such as the Open Knowledge Foundation and the Open Data Institute, have also contributed to creating standards for open data.

The publisher's column indicates that there are many influential players in the process to standardize open data. These actors intentions for creating standards vary. For instance, private technology firms may be more interested in making standards that are compatible with products

and services. In contrast, non-profit organizations and standard bodies may be interested in creating standards that maximize widespread proliferation and accessibility to the data.

Description and purpose of standards

Standards publishers claim that their standards improve the accessibility and widespread dissemination of data. Documentation for standards such as Spotcrime specify data values and acceptable file formats. Moreover, standards specify geospatial data so that it may be geocoded more seamlessly by developers. For instance, OpenAddresses is a specification that parses and imports address data into a database that can be geocoded more seamlessly.

Transferability to other jurisdictions

18 of the 27 standards in the inventory are transferable (*see figure 3*). Transferability of the standard among jurisdictions ensures that the vocabulary is flexible enough to be applied to other cities. In addition, standards that utilize less complex file formats enable cities with less resources to enter the technology. Standards that require more complex formats for the data may exclude smaller municipalities to adopt. Transferable standards may also be more extensible and flexible so that they may be implemented more easily by a variety of governments.

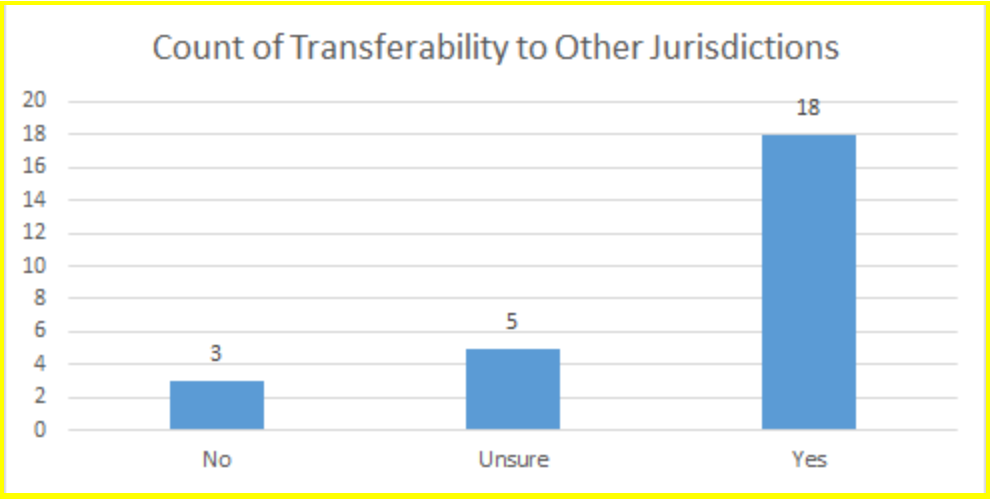

Figure 3: Count of Transferability (Spreadsheet 1)

Timeline information

The spreadsheet indicates four of the 27 standards are in their draft stage (*see figure 4*). Most of the standards had been updated since 2014, with the exception of the Election Markup Language, which was last updated in 2011. Timeline information represented by the Stage in Development, Version, and Last Updated variables indicate that open data standards are a relatively new and evolving technology among these high value domains.

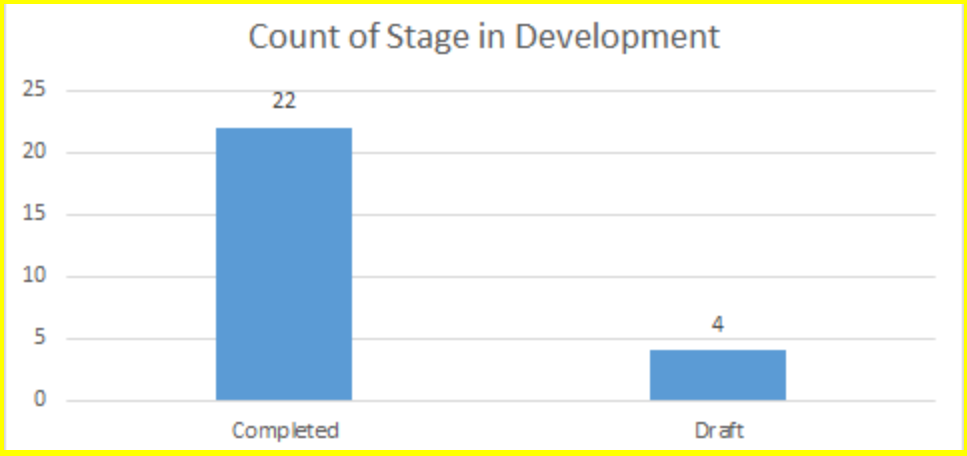

***Figure 4*: Count of Version (Spreadsheet 1)**

Level of adoption by governments

Most standards are implemented locally. Only Open311 GeoReport API, GTFS, SIRI, Open Contracting Data Standard, LIVES, TCIP, and OpenAdresses have been implemented at the international level. These standards are implemented mainly in North America and Europe. GTFS and Open311 Georeport API are the most common standards used for open data.

Extensions

The inventory indicates that only 13 of the 27 standards have extensions (*see figure 5*). Extensions refer to the re-invention of a standard to fit changes or a particular context of the data. For example, extensible standards for transit data will adapt to changes in transit services. These standards adapt themselves, or insist that they can be adapted, to fit the context of their implementation.

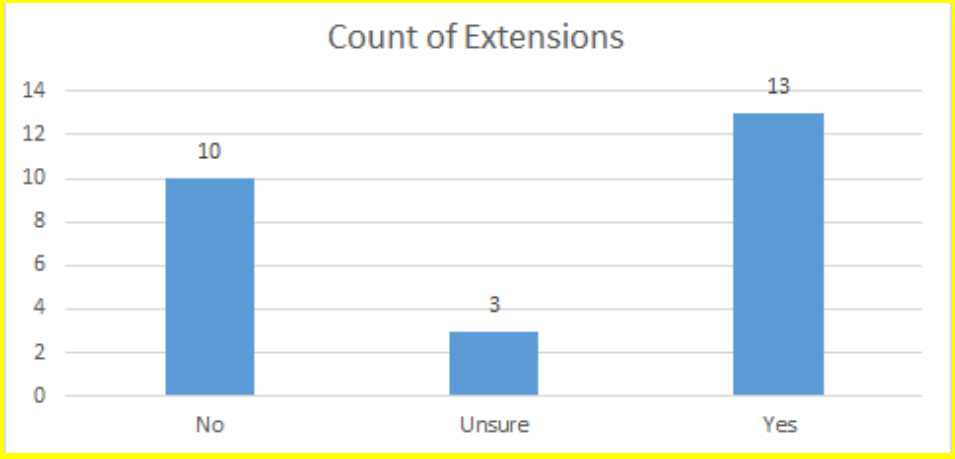

***Figure 5:* Count of Extensions (Spreadsheet 1)**

Machine and human readable

All standards in the list are both machine and human readable. All observations emphasize that standard ease automation of the data; however, not all standards refer to formal libraries to specify meaning of the data. Spain's vocabulary for budget data, referred to as "Vocabulario para la representación de datos sobre presupuestos de entidades locales" in Spreadsheet 1, emphasizes clear semantic components and a data architecture. These standards use clear language that is relatively easy to understand.

Specifies timely data

19 of the 26 standards specify timely data. Up to date data depends on the domain of data. Real-Time transit data standards, such as TCIP and GTFS-RealTime require formats that can handle real time data. Within the domain of expenditure data, standards vary in how they specify frequency of updates to the dataset. For instance, some expenditure standards requires that data be updated quarterly while others specify it should be updated yearly.

Metadata

Some standards designate a CSV or JSON file descriptor to act as a source of information about the data publishers, version of standard, data being published, etc. Some standards suggest that metadata should be incorporated into the standardized data but do not require it. Standards that suggest but don't require specifying metadata for the dataset are categorized as unsure.

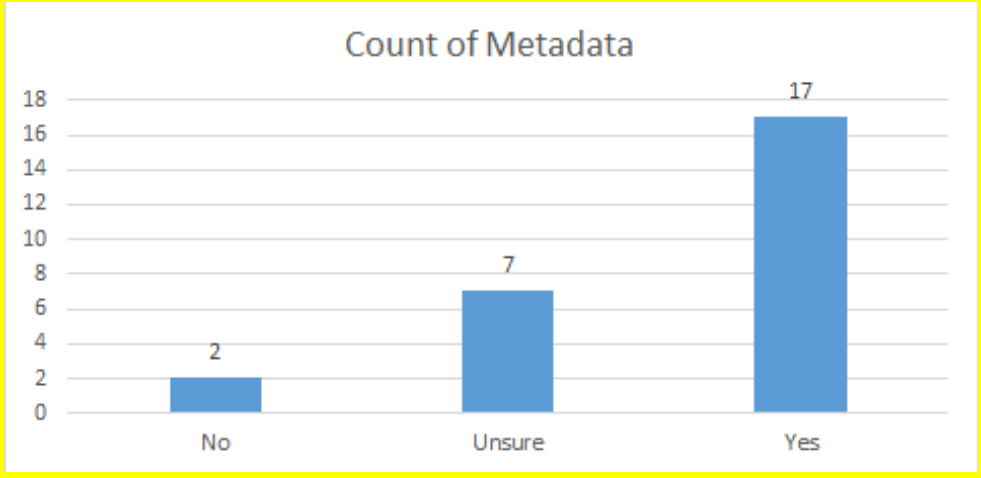

Figure 6: Count of Metadata (Spreadsheet 1)

Applications built from standard

Majority of standards do not have applications built from them. Open311 Georeport API and GTFS are standards with the most applications built from them. Checkbook NYC acts as a platform that visualizes the data. From this platform users may pull the "raw" expenditure data from the website.

Standard's documentation

Most documentation of these standards are hosted on GitHub or on a webpage. GitHub allows anyone to file issues and pull requests. Some repositories offer examples of implementation of standards and code to help automate the process of implementation.

References

Publishers create web pages to explain why and how to use their standard online. These web pages offer "how-to" information in order to ease the implementation of the standard. In addition, there are blog posts posted by city governments and publishers regarding the initiative and implementation of standards.

OPENNESS OF PROCESS

Standards that use open licenses

Standards compiled in the inventory either do not specify license information or use open licenses. The most common license used by open data standards publishers is a version of the Creative Commons Attribution license.

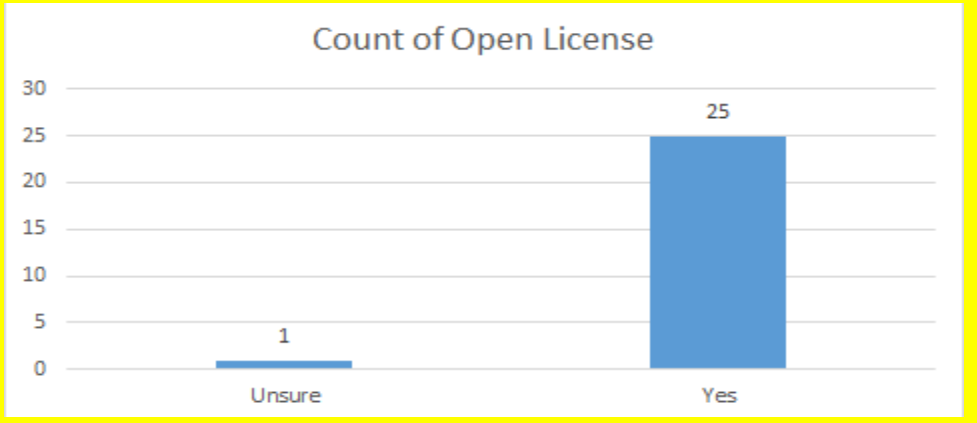

Figure 7: Count of Open License (Spreadsheet 1)

Stakeholder participation

15 of the 26 standards in the inventory utilize multi-stakeholder participation in their conception (*see figure 8*). Governments will work together with private technology companies and non-governmental organizations to create a new standard. Input by governments ensures that the their interests and needs in implementing the standard are accounted for by the standard's publishers.

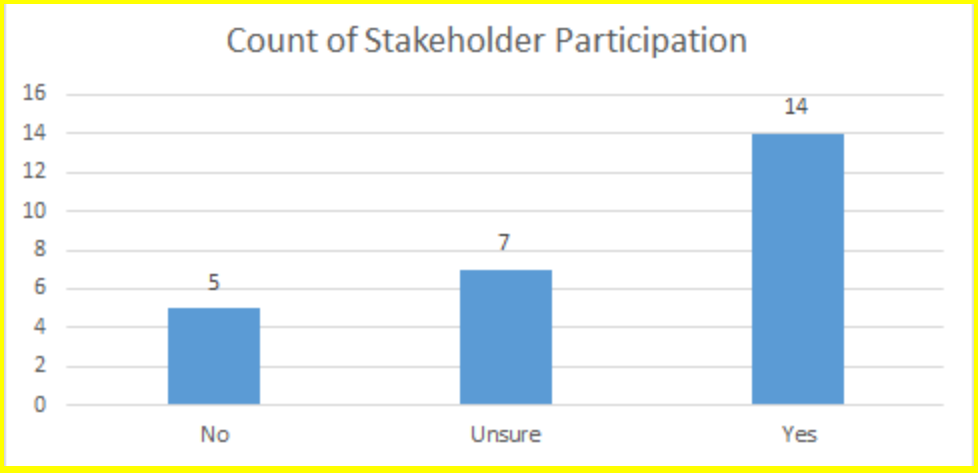

***Figure 8*:** Count of multi-stakeholder participation (Spreadsheet 1)

Consensus-based governance

15 of the 26 standards use a consensus-based approach for their maintenance (*see figure 9*). 9 of 26 standards do not utilize a consensus-based approach for their maintenance. (*see figure 9*). Standards that do not use a consensus-based approach usually rely on a closed technical committee to deal with the standard's future development.

A consensus-based approach entails that anyone may contribute feedback about the standard to its developers. Receiving feedback by anyone requires some platform for communication. Many of the standards' publishers use GitHub or Google Groups to receive feedback about the standard. Publishers that actually use these platforms encourage feedback from the public in their documentation. However, it is less clear that feedback is seriously considered and enacts real change on the part of developers. Therefore, future research may be interested in observing if consensus-based governance really includes a two way exchange between the general public and developers.

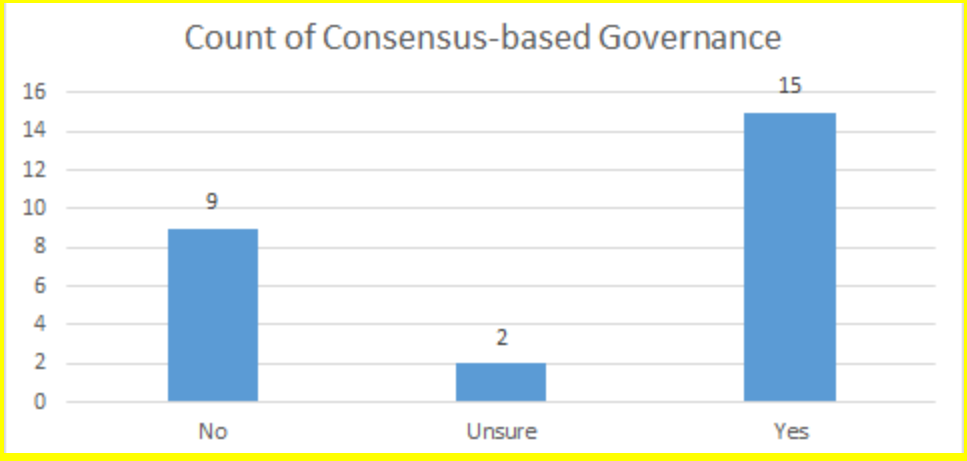

***Figure 9*: Count of Consensus-based Governance (Spreadsheet 1)**

Spreadsheet 2

Results of the second spreadsheet indicate that most municipally published data does not apply open data standards. According to results of Spreadsheet 2, GTFS and Open311: GeoReport API are the most common standards applied to civic datasets for the surveyed 10 domains. In addition, among these 10 domains, CSV and XLS format is a commonly used to publish municipal data. However, many of these files are not structured to be linked in a relational database.

Across cities, schema and semantics of the data vary within domains. Some cities use proprietary file formats to publish their data, while other cities publish in open formats. For instance, the city of Toronto only publishes their zoning data in ESRI shapefile. In contrast, the cities of Edmonton, Vancouver, and Surrey published their zoning data in KMZ file formatting. KMZ may be a more open way of publishing, since it does not require an expensive software to access the data. The city of Edmonton uses Socrata Open Data API (SODA) to publish their open data. The SODA ensures that the data utilizes unique identifiers and supports various response formats for the data (CSV, GeoJSON, JSON, RDF-XML, XML).

In addition, the ways that cities store their metadata varies. These cities mostly embed the dataset's metadata in the html script of the catalogue web page. However, Toronto stores some the the metadata to the datasets in readme text files. Surveying the datasets demonstrate that there is opportunity to improve interoperability and comparability of data at the civic level through implementation of open data standards.

Conclusion

Open data standards make data interoperable and comparable across jurisdictions. This project aims to build two resources to assist in the standardization of data at the city level. The first resource inventories open data standards by data domains and then evaluates the quality of available domain-specific standards. The second resource identifies how data is published on the civic open data catalogues of five Canadian cities and then evaluates the use of open data standards.

Results of the second spreadsheet suggest that most published municipal data does not utilize open data standards. An inventory of available domain-specific data standards may better inform municipal governments about the potential for data standardization. Standardization would enable more usage of data internally and externally. We further argue that standards, which

evolve openly and include multiple stakeholder participation, will be more successful for establishing optimal interoperability than standards that developed behind closed doors.

Open standards are not the only factor in achieving optimal interoperability of high-value data. However, standards remain an important approach for achieving interoperability among information systems. We hope that this project will continue to evolve in a way that is useful to the open standards initiative. Future evolution may include expanding to more municipalities, including more high-value datasets, and updating any emerging standards. Any future development of this initiative will have beneficial consequences for the public good. Palfrey and Gasser (3, 2012) remind us that,

> More often than not, our future success in addressing the big societal challenges of our time, from healthcare to climate change, will depend heavily on our ability to create better interfaces and connections among complex systems and our ability to share information appropriately.

References

Bauer, Florian, and Martin Kaltenböck. "Linked open data: The essentials."*Edition mono/monochrom, Vienna* (2011).

Bennett, Daniel, and Adam Harvey."Publishing Open Government Data." *W3C* (2009).

Bloomberg, M., and Rahul N. Merchant. "Open Data Policy and Technical Standards Manual." *New York City Information Technology & Telecommunications* (2012).

Guidoin, Stéphane and James McKinney. 2012. Open Data, Standards and Socrata. Available at http://www.opennorth.ca/2012/11/22/open-data-standards.html. November 22, 2012

Huijboom, Noor and Tijs Van den Broek. 2011. Open data: an international comparison of strategies. European Journal of ePractice. www.epracticejournal.eu 12, March/April 2011 13pp.http://unpan1.un.org/intradoc/groups/public/documents/UN-DPADM/UNPAN046727.pdf

Nicklin, Andrew. “Enabling Civic Data Standards,” GovEx, January 4, 2016, http://govex.jhu.edu/enabling-civic-data-standards/.

Open Knowledge. *Open Definition 2.0*. *Opendefinition.org*. Retrieved 5 November 2016, from http://opendefinition.org/od/2.0/en/

Public Health Standards Consortium, n.d. "Health Information Technology Standards: Data Standards" *General Information about the Public Health Data Standards Consortium*. Available at http://www.phdsc.org/standards/health-information/D_Standards.asp

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

Open North. *Domain-specific standards project*. https://docs.google.com/a/opennorth.ca/document/d/1ez3UXtBs_YpIZKEmmnDaeMHJ3oX3D0uahAdG8HKU_sw/edit?usp=sharing.

Palfrey, John Gorham, and Urs Gasser. *Interop: The promise and perils of highly interconnected systems*. Basic Books, 2012.

Russell, Andrew L. *Open Standards and the Digital Age*. Cambridge University Press, 2014.

Suszan, Brittany. "The SpotCrime Open Crime Standard (SOCS)." *SpotCrime - The Public's Crime Map* (2014).

Dyson, Lauren, ed. *Beyond transparency: Open data and the future of civic innovation*. Code for America Press, 2013.