

Received November 16, 2016; accepted December 11, 2016, date of publication December 30, 2016, date of current version March 2, 2017.

Digital Object Identifier 10.1109/ACCESS.2016.2645658

# Open Data-Set of Seven Canadian Cities

HAIWEI DONG<sup>1</sup>, (Senior Member, IEEE), GOBINDBIR SINGH<sup>2</sup>,  
AARTI ATTRI<sup>3</sup>, AND ABDULMOTALEB EL SADDIK<sup>1</sup>, (Fellow, IEEE)

<sup>1</sup>Multimedia Computing Research Laboratory, School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, ON K1N 6N5, Canada

<sup>2</sup>Department of Electronic Business Technologies, University of Ottawa, Ottawa, ON K1N 6N5, Canada

<sup>3</sup>Department of Electrical and Computer Engineering, Carleton University, Ottawa, ON K1S 5B6, Canada

Corresponding author: H. DONG (hdong@uottawa.ca)

**ABSTRACT** Open data has attracted huge attention for the construction of smart city in terms of delivering useful city information to citizens and interacting with citizens from the city council perspective. In this paper, we present an overview of the current status and issues of open data opened by different seven Canadian cities. We start by presenting the characters of open data, followed by data format conclusion and detailed dataset explanation for each Canadian city (e.g., Calgary, Halifax, Surrey, Waterloo, Ottawa, Vancouver, and Toronto), including the different data catalogues and their detailed characteristics. Next, we discuss the state-of-the-art of the tools and applications developed over each city's open data. Here, we not only illustrate the most successful examples, but particularly consider the potential issues due to the characters of the city datasets. This paper is not only beneficial for a government, which can compare its open data status with that of the Canadian cities but also quite useful for users or companies interested in tool development over open city data.

**INDEX TERMS** Big data, characters of open-data, smart city, city application tools.

## I. INTRODUCTION

Nowadays, all cities in the world are great producers of data, and this data is given a new term which is known as Big Data [1]. Efficient management of this huge data is very essential to constitute powerful tool to have structured form of data and giving back to the public with the data having more utility. This real-world data is a key to the implementation and validation of cities' social, economic and educational structure as it contains useful reviews from public about various activities within the cities. City administrators need integrated the factual data to make better decisions and policies to make cities smarter and sustainable. Thus the availability and accuracy are the major parameters that may affect the reliability of the resulting estimates. Thus there are various steps under which a city open data constantly goes through before it is converted into useful information. The process involves the collection and then storage of data for further processing [2]. The data is closely analysed and segregated in different categories and formats to convert it into a meaningful information. Furthermore the data visualization is important concept to make the information available for citizens in a better structured form. The main purpose to analyse Open Data is to extract the meaningful information from this open government data that can contribute to the betterment of public [3]. There are many efforts made by

researchers and by different organizations to study this huge data [4].

The significance of open data to bring in innovations and developments is given a lot of stress to bring sustainability in a city. Thus many entrepreneurs [5] of IT companies [6] mention the importance of open data in a sense that it helps to figure out present issues in a city and give notion to work towards the betterment of citizens. This is the basis of selecting seven cities from Canada which has a flourishing IT market so that it can be helpful for professionals looking for ideas and specific sectors to work on that ideas. A substantial work has been done by researchers to emphasize on developing e-services from government open data [7], [8]. An effort has been made where a framework has been developed [9] to explore the status of present Open Government Data (OGD) by using the content analysis of web portals of government open data from 35 countries and also actually working briefly on open data portals from four countries Morocco, UAE, Kenya and Ghana listing only number of data-sets and the formats used.

However there is still a void for detailed analysis at city level in Canadian cities, which is one of the driving force for this paper. A city level discussion for open data utilisation in five smart cities namely Barcelona, Chicago, Manchester, Amsterdam, and Helsinki has been done [10] in order to

highlight the significance of open data and its resulting innovations in these cities. In different parts of the world researchers are working on the initiatives taken by governments for OGD and how these initiatives has valued the country in its economic growth as well as democratic empowerment [11]. A considerable effort has been made in [12] where an event management system has been developed to generate events in different categories, after which the huge amount of data generated from events is analysed and is utilised in a manner to extract results from it. Although this platform is presented for big data in general but this can be used as a good platform to study open data from different categories listed by the cities and develop tools that might result in an innovation. It has defined the data collection from different sources such as social media, hard-sensors and public itself. Also this includes analysis of this data in structured form to create events that include various categories as municipal, traffic, medical etc. and thus gives useful information to the public. Therefore it is a public-oriented platform which takes information from public and gives informative knowledge of the city to the public.

However the domain of applications is not limited to city organization or locations only, it can be different based on the category of data provided by the city. Application development has been done on open city data in the field of environmental science [13] where visualizations has been used for Geographical Information System (GIS) in order to show the effect of clusters of buildings in coastal areas on non-mitigation of higher temperatures even if sea breezes flows into inland areas due to obstruction offered by these buildings. The motivation of this paper is to study open-data-sets available in different Canadian cities (Calgary, Halifax, Ottawa, Toronto, Surrey, Waterloo and Vancouver) and perform analysis on this data which can be used by IT professionals to build a platform to use this information.

The paper is organized in the manner that firstly the section II briefs about the characteristics of open data based on which the quality of of open data provided by a city can be judged. Further, section III describes the current status of the data and emphasizes the various characteristics, formats and diverse catalogues describing the open (government) data. Section IV addresses various tools developed for open data in different cities. Further, section V discusses some challenges for open data and briefly touches the proposed model to deal with heterogeneous data.

## II. CHARACTERISTICS OF OPEN DATA

This section will explain a few characteristics of open-data. The idea has been taken form the general characteristics of data discussed in the literature [14] which clearly mention that government if government or any organization wants to makes its open for public then the deep study of the data characteristics is very essential. It helps the organization to look into the resources and then decide what to make open and in which format it should be flooded to the public. Those characteristics are as explained as follows:

### A. COMPLETENESS (VOLUME)

The completeness of open data refers to the amount of open data released by the city. It does not refers to the data in different domains but volume of data in a particular domain. Most of the cities do not release the full-fledged data available for a particular category but instead release a sample of the data which can help the user to work around with the available attributes and sample data with the attribute values. Ideally the public data should be available completely to the public because this data does not have privacy or security issues. The sense of completeness can only be achieved when electronic copies of bulk files for public data are provided to the user.

### B. AVAILABILITY

This characteristic defines the availability of the city open data either through the city open data website or by other means. Every city provides an open license to all the viewers to access the data. Usually there are no restrictions on open data because the sense of open data is lost when the authorities put restrictions on the released data. It may also happen that sensitive categories like criminal records might not be made available without a license or if it is made available, the user must accept that the city is not responsible for any inferences made from the available data.

### C. USABILITY

The data provided should be such that the data can be easily used by the user. Thus the cities are producing data in digital formats (CSV) over internet. Similarly a particular data demonstrating number values is much more usable when represented in tables instead of plain text. So lesser the processing required on data, more usable the data is.

### D. NON-PROPRIETARY

The control of a particular entity or organisation over the data proves the proprietary rights over the data. However the open data must have the feature that no one should have proprietary rights over the data. Therefore, non-proprietary is a characteristic that open data needs to have.

### E. NON-DISCRIMINATORY

The data released by the responsible entity should not be biased towards a group, community, religion or region. The availability of data should not depend whether the user belongs to the same city or country, a particular religion, race or community. Thus city data should be available to anyone who want it without any prior registration.

### F. VARIETY

The categories of open data must not be limited to selected sectors. It should have the variety in terms of the categories for which the open data needs to be collected. The data-sets of the open data should not focus on the same subject. For example, if all the data-sets of the city focus on transportation of the

city and ignore the categories like criminal records, buildings, municipal operations, elections, electricity and water, then the data-sets of the city lacks variety.

### G. TIMELY PROCESSED AND UPDATED

This characteristics is defined in a sense that the data is made available as quickly as necessary to reserve the value of data. Thus the released data for the public must be up to date. The access to these updated data-sets should also be given by providing APIs (Application Program Interface) which can be used directly in applications to fetch the latest data. The usefulness of the data decreases as the data gets outdated.

### H. SUMMARY

Based on the data analysis performed on data-sets for each city, it would be useful to discuss whether the cities has attained majority of the open data characteristics discussed above or not. All the cities except Halifax has maintained variety in their data-sets in different domains. However, in terms of completeness Halifax, Surrey, Vancouver and Waterloo have worked on building open data with considerable number of entries in majority of the data-sets whereas Calgary, Toronto and Ottawa are still struggling in developing in terms of numbers. All the cities have made the data available online on their web portals irrespective of the user's race, gender, religion or a region. None of the cities has any proprietary rights on the open data available on its website. Every city has listed formats of its data which are machine processable and are usable in their original form for the user. Each city is making best of its efforts to process the data timely or frequently so as to provide users with the latest data. However, the update frequency actually depends upon the kind of data-set. For example, "transportation" data is more frequently updated as compared to "elected officials" data-sets as "transportation" data is more vulnerable to changes. For all the cities the laws of usage of open data are governed by respective provinces of the cities. The detailed description and analysis of each city's open data has been done in section IV where these characteristics are discussed with respect to the city's open data.

### III. OPEN DATA BY DIFFERENT CANADIAN CITIES

This section discusses and analyses the open data collected by seven Canadian cities namely Calgary [15], Ottawa [16], Surrey [17], Toronto [18], Vancouver [19], Waterloo [20] and Halifax [21]. The data is collected from their respective websites, on which it is available open to public in the form of different visualization tools and also to make any changes as per their needs. This data is categorized in various catalogues and formats. Several open data catalogues are defined in various categories which are explained in section IV. These catalogues are available in different formats as discussed further in this section. A few of the catalogues from different cities are available in common formats.

### A. DIFFERENT FORMATS OF DATA

There are various formats in which data is represented in different cities. Some of them are machine-readable which are difficult for users to understand but there are some formats which can be easily understood by users. All formats in which data-sets from seven Canadian cities are defined are as follows:

#### 1) CSV (COMMA SEPARATED FILES)

This formats is very useful as it is compatible to define large data-sets. Usually the data-sets like censuses data results, election results, number of parks/beaches, traffic volume etc. within the cities are available in CSV formats.

#### 2) DWG (FROM DRAWING)

It is a binary file format meta data related with various open data-sets. It also includes geometric data information such as maps to define locations and corresponding photos. Data sets such as road networks, transportation bikeways, parking etc. are defined in this format.

#### 3) JSON (JAVA SCRIPT OBJECT NOTATION)

Java Script Object Notation- JSON is one of the format which is easy to read for any programming language. This file format has been used by Ottawa, Surrey, Vancouver, Toronto and Waterloo for listing various data-sets such as Traffic Data (Ottawa), Park Lights (Surrey), Bicycle Stations (Toronto), Business License (Vancouver) etc.

#### 4) KML (KEYHOLE MARKUP LANGUAGE)

Keyhole Markup Language is an XML notation for representation of geographic /geospatial information related with any data-sets for example traffic cameras location, transport route information, traffic count, intersection location, parking location etc.

#### 5) SHP (SHAPEFILE)

It is one of the most widely used format for geospatial data representation specifically for geographic information system (GIS) software. Thus all the data-sets of various cities which involved with GIS software to define them more precisely, are essential to write in this format. Thus the data-sets related with transportation or any service location (health, recreation, parking) etc. are represented in this format.

#### 6) XLS (MICROSOFT EXCEL SPREADSHEET)

Many of the data-sets are available in XLS formats which can be directly used with correct description given by each columns. It is one of the format which is easily understood by users. Data sets related with elections, city boundaries, events etc. are defined in this format.

#### 7) XML (EXTENSIBLE MARK-UP LANGUAGE)

It is the most commonly used file format for data exchange as it keeps the exact structure of the data. It also gives

opportunity to developers to divide and modify different parts of the file. Various data-sets of different cities like city transit routes and schedule, traffic cameras, job opportunities etc. are represented in this format.

## B. OPEN DATA-SETS OF CALGARY

According to the city of Calgary, open data is defined as the information gathered by government on citizen's behalf also including their personal information, which is relevant to use for any purpose including commercial. This information can be found in different data-sets on the website. Users are granted a non-exclusive licence to change the data or modify it according to their own needs but should follow the terms and conditions defined as "open data catalogue terms of use" on the city's website [22]. Transparency of this data is the key factor to promote accountability and provide useful information to citizen about their government and their personal needs. The main purpose of this open data is for the public to explore the whole information regarding city on a single web link. This data is categorized in different data-sets which are available in alphabetical order [15] or in different categories. This data is extracted by different department such as transport, hydro, public security, city welfare societies etc. The main purpose to cover all these sectors of a city is to give an opportunity to reuse this diversified data in innovative way by citizens and other organizations.

There are total 12 categories which are related with social, educational and business areas of citizens and are also based on government activities in these areas. Various data-sets are defined under these categories and data is given in different formats as explained below:

- *Administrative Boundaries*: It includes data related with the city, community, election wards and natural sites boundaries (parks, rivers). The data is defined mainly in DWG, SHP, KML and XML formats.
- *Census Information*: This category defines census data for residential units and total population count in those areas along with the area names. This data is given by community district and by wards from the year 1999-2015. The data is given in CSV, DWG, SHP, KML and XML format.
- *City Facilities*: This category includes data for public provision, thus the data-sets under this category are park amenity equipment, parks monuments, playground equipment, sport equipment, sport surfaces etc. This data is also defined in DWG, SHP, KML, XML formats.
- *City of Calgary 2013 election information*: It defines the city's election information in three different data-sets as 2013 election results by station (XLS and XML), 2013 election results companion guide (XML and DOCX), 2013 election summary results (XML and FILE).
- *City of Calgary Human Resources*: This category defines data related with the career opportunities within the city which helps users to explore different career options as per their needs and skills. This data is available in XML, URL formats.

- *City of Calgary News Feed*: It describes data-sets related with various newsroom of the city in XML, URL formats. This category has the least downloaded data-sets number as only 33 in the month of June 2016 as per data available in July 2016.
- *City Services*: It emphasis on various emergency facilities provided by different government bodies within the city such as 311 customer satisfaction, fire emergency, fire station locations and services etc.
- *Environmental*: It includes various data-sets mainly as habitat, hydrology, natural areas, parks water delivery, waste recycling facility, water features, water single family consumption, land use information. The data is available in DWG, SHP, KML, XML formats.
- *Geospatial Reference*: This category contains only one data-set i.e. high precision network which defines a control network within the city for its development, surveying and mapping. This is monitored by "field surveying service division". It is available in XML and PDF format.
- *311 Requests*: This category has data-sets such as 311 call center activity, march 2015 public service requests by community, road service requests. This data is available in CSV, KML, SHP and XML formats.
- *Subdivision and Development Appeal Board Rulings*: This category contains Subdivision and Development Appeal Board Rulings for 2014-2016. The data is available in XML format. It gives a complete view of all the hearings. Each hearing has a unique file number and also a appeal and decision number. The file includes a summary of hearings, their reason, place and date etc.
- *Transportation*: This is the last category which has maximum data-sets describing all information related with city transportation such as road network, traffic volume, traffic cameras, truck routes etc. Data is defined in CSV, XML, SHP, KML, DWG under this category. This category has maximum download number as 2573 in the month of June 2016 as defined by the data available in July 2016.

Furthermore, RSS feeds for these data-sets are available online and links to Google maps and Bing maps are available in case of geo-spatial data-sets. The website also have a few sections to involve the citizens to have their reviews on the available data-sets. There is a "discussion forum" and a "citizen dashboard" for public to write their ideas about the data-sets to improve existing categories.

There is one major drawback of the data available and that is incompleteness of data i.e. data-sets include only the sampled data. For instance the data set "traffic cameras" under "transportation" category has only 77 total entries in CSV file. This file includes the information about the location of the traffic cameras along with the reference image but from researcher point of view or even for commercial use, this data is not enough to explore it with proper results. This is a prime challenge for the concerned authorities.

### C. OPEN DATA-SETS OF HALIFAX

Halifax (Halifax Regional Municipality-HRM) is a new name in the field of open data. It defines the data as “something for everyone”, which means that the data is available for public including technical and non-technical users. The data is managed by ESRI’s ArcGIS (Aeronautical Reconnaissance Coverage Geographic Information System) Open Data platform and is available in different data-sets which can be accessed on [23] but these data-sets are quite scattered as they are not segregated either alphabetically or clubbed in categories for easy lookup. Only 33 data-sets are defined as open-data catalogues [21] which are named as zoning boundary, transit areas rates, trails, street centrelines, soild waste collection areas, residents association area rates, recreation area rates, local improvements area rates, HRP parks, community boundaries, fire protection area rates, crime, civic addresses, by-law areas, bus-routes, building symbols, building permits, building outlines, BID area rates, HRM park recreation features, contours 5m, bus stops, parking meters, tax designation, transportation area rates, spot heights, pre-amalgamation boundaries, polling districts, polling division and community council boundaries. All these data-sets are avaibale in CSV, KML and Shapefile formats.

Furthermore, the city of Halifax has failed to bring much variety into the categories but there are data-sets such as data defined by transportation department that contains enough data for users (citizens, researchers, analysts, developers) to extract useful information as per their needs. Thus the city has definitely provided quality data having considerable number of entries in most of the data-sets. This data is easy to use by citizens, researchers and analysts as almost all data-sets are defined in tabular format. Also different APIs (Application Programming Interface) are available for software developers to easily access the data directly. A lot of detailing about the attributes, data-sets and the creation of data-sets has been provided as metadata.

### D. OPEN DATA-SETS OF SURREY

The city of Surrey in the province British Columbia has defined the open data to be an idea according to which some data must be available for every one free of cost and can be used and republished. The data is available worldwide under the open non-exclusive license provided by the City of Surrey and governed by the laws of province of British Columbia and the applicable laws of Canada which allows the user to copy, modify, reuse, publish and translate the data without giving any warranty over the errors, omissions and completeness of the data. The objective of the city to promote open data is to empower citizens with a good quality data, help small businesses flourish, build a chance to develop its health and education facilities, economic productivity and create more scope for scientific research. Moreover the city of Surrey also expects that a useful research or data evaluation done on this data could come handy for the city in future. The city has taken the help of open data portal platform

CKAN (Comprehensive Knowledge Archive Network) for maintaining its open data.

Surrey has broadly classified its 325 data-sets among 12 categories which are explained as follows:

- *Business and Economy*: It has mainly 11 data-sets which are business licenses, population estimation, employment in arts, culture, individuals with low income, land in food production, availability of employment, businesses by sector, employees by sector, rental market, business improvement areas, restaurants and majority of these data-sets are available in CSV formats.
- *Community Services*: The data-sets under this category are places of interest (available in CSV), licensed child care (in CSV), low cost and free resources (in CSV, KML, FGDB), registration in city programs (in CSV), social housing (CSV), schools (available in DWG, FGDB, KML, JSON, API), garbage recycling collection days (available in DWG, FGDB, KML, JSON, API) and collection route boundaries (available in DWG, FGDB, KML, JSON, API).
- *Environment*: It has various data-sets related with environmental bodies such as water consumption, trees planted, parks, drainage water bodies, drainage flood control, ecosystem sites, community waste etc. and the majority of the data-sets are available in CSV formats thus easy to understand and use by the users.
- *Finance*: It includes three major data-sets namely city spending on public art, city tax base and city funding on beautification projects and all these three data-sets are available in CSV format.
- *Health and Safety*: It includes data-sets for public health based facilities such as availability of doctors, crime and collision incidents and criminal offenses. All of these data-sets are available in CSV formats thus easy to explore by public.
- *Infrastructure*: It is mainly defined to give the idea of whole network connectivity within the city such as water supply, sanitary manholes, sanitary valves, signs etc. Almost all of these data-sets are defined in the formats which are specifically built to locate the geographic location of the entity such as KML, JSON, DWG, FGDB.
- *Land Use and Development*: It is built to define proportion of land utilization for different purposes such as buildings, farming, urban centers etc. This data is need to be defined with exact location thus the formats which are used for these data-sets are DWG, FGDB, KML, JSON, API.
- *Recreation and Culture*: It includes data-sets as heritage sites (CSV), public art (CSV), arts and culture groups (CSV), youth centered events (CSV), events (JSON) and heritage routes (available in CSV, FGDB, KML, JSON, API).
- *Transportation*: It defines various data-sets related with traffic network within the city such as traffic cameras (CSV), traffic signals (CSV, JSON, KML, DWG),

poles (CSV, JSON, KML, DWG), railway crossing (CSV, JSON, KML, DWG), traffic count (CSV), etc.

Thus Surrey has surely tried to achieve variety in its data-sets and has covered many different domains. Further a lot of serious efforts have been committed by the city to build and maintain this much amount of data.

#### E. OPEN DATA-SETS OF WATERLOO

The city of Waterloo has maintained a very good quality of open data covering all the aspects of open data. The data is open to everyone irrespective of the region, age, race or community. Hence, the open data is purely non-discriminatory. The work done by the city of Waterloo to collect and organize the data is clearly remarkable in contrast to other Ontario cities like Toronto or Ottawa. Waterloo has focused on maintaining the variety as well as the volume in the city open data. The city has provided access to 13 categories which further have various data-sets. These categories are explained below:

- *Events*: This category contains data-sets related to different events in the city on different dates along with locations helping the public to plan their activities according to their interests. The data available in these data-sets is provided in spreadsheet format.
- *Base Data*: It mainly has data-sets as buildings, railway contours 2012 city boundary historical, addresses and roads which are available in spreadsheet, KML and SHP file format.
- *Boundaries*: It is having various data-sets such as city boundary, polling 2014, wards 2014, wards, 2010, neighborhood associations, district plans and all are available in spreadsheet, KML, shapefile format.
- *Closures*: This category is available in spreadsheet, KML and SHP file. The spreadsheet consists with only sample data with a very few entries. It gives information regarding closure name (sidewalk, road, intersection), street name, date for closure and some specific information as “sidewalk closed- please use other side of street” to for user’s convenience.
- *Community*: It has four data-sets which are community access bikeshare stations, neighborhood matching fund, neighborhood associations, older adult housing directory. All are available in spreadsheet, KML and SHP format. The first data-set gives complete information regarding bike station such as station name, location, access time and bike counts etc. The second data-sets defines projects name which are focusing on one community such as arts, education, environment, history, public safety, community building or recreation. They have received neighborhood fund matching grant. The third data-set describes the boundaries within neighborhood association within the city. The last data-sets defines information about residential areas for older adults in the city.
- *Elections*: It is having 19 data-sets equipped with data from different wards related to candidates, results etc for different years in the city.

- *Environment*: It is mainly for the users to explore the current status of the data and help in maintaining sustainability of the resources. The data is available in spreadsheet, KML, SHP file.
- *Heritage*: It defines data-sets as walkability network, heritage buildings, city boundary historical, historical streets and all of these data-sets are available in three formats- spreadsheet, KML, SHP file.
- *Parks and Recreation*: It has 9 data-sets namely parks, bicycle parking, bylaw parking infraction, parking lots, sports field and diamonds, trails, playground, recreation points, outdoor rinks. All these data-sets are easily accessible by all users as it is available in simple tabular format too.
- *Points of Interest*: It is available in spreadsheet, SHP, KML formats and it contains the data-sets which describe the points of interest in this city like important dates and places.
- *Records*: Its data is also in the same three formats as define above.
- *Transportation*: It has maximum data-sets as a total of 23 which are related with the transportation activities within the city such as bicycle counts, walkability network, major transportation routes, parking etc.

Thus each of the mentioned categories have subsets falling under that particular category which sums up to 122 data-sets overall.

#### F. OPEN DATA-SETS OF OTTAWA

The capital city of Canada has provided access to its open data with the help of CKAN (Comprehensive Knowledge Archive Network) and it can be downloaded in multiple formats as well as the data can be fetched by using API functions provided on the website which ensures that every time the API is hit, the most recently updated data is fetched. The city of Ottawa provides an open non-exclusive license to the user worldwide to use, distribute and modify the data without granting any proprietary rights to the user. This user license is governed by the laws of the Province of Ontario. However the city does not give any warranty for the completeness or accuracy of the data.

The city of Ottawa has listed 15 organizations which have helped the city to build and maintain this data. These organizations are- City Clerk and Solicitor (7 Data-sets), Community and Social Services (8 Data-sets), Crime Prevention Ottawa (1 Data-set), Emergency and Protective Services (2 Data-sets), Environmental Services (5 Data-sets), Financial Services (2 Data-sets), Human Resources (3 Data-sets), Infrastructure Services (35 Data-sets), OC Transpo (3 Data-sets), Ottawa Public Library (5 Data-sets), Parks, Recreation and Cultural Services (23 Data-sets), Planning and Growth Management (12 Data-sets), Ottawa Public Health (3 Data-sets), Public Works (7 Data-sets), Service Ottawa (7 Datasets). All these organizations belong to different domains and have provided data according to that particular domain. The data is placed under 9 groups or categories

namely *Business and Economy*, *City Hall*, *Demographics*, *Environment*, *Geography and Maps*, *Health and Safety*, *Living*, *Planning and Development* and *Transportation*. These 9 categories contain 129 data-sets from different sectors ensuring the variety in the open data. Those are explained as follows:

- *Business and Economy*: This is the first category which has defined 2 data-sets namely business improvement areas (available in SHP, GeoJSON formats) and Job Opportunity (available in XML, JSON format).
- *City Hall*: It has information related with the elections, voting places and nominated candidates in SHP, CSV, GeoJSON and CSV formats respectively.
- *Demographics*: This category has mainly two data-sets providing data for 311 monthly service requests submission from 2013-2014 (available in XLS, XLSx formats) and ward data from census for 2006 and 2011 (available in CSV format).
- *Environment*: It mainly emphasise on natural bodies within the city as water, rivers, beach water, water quality, street trees and various formats are used for these data-sets such as XML, DWG, CSV, SHP, GeoJSON for water, XML, DWG, CSV, KMZ, SHP, GeoJSON for rivers and XLS for beach water sampling data, CSV format for water quality, KMZ, SHP, CSV, XML for street trees and XLS for drinking water.
- *Geography and Maps*: It defines various routes with exact locations for various means of transportation such as bus, rails, airports runways. It also has data-sets related with beaches within city, also the sports field, basketball, tennis, volleyball courts, truck routes and pedestrian network in all over the city. The main formats for these data-sets are XML, DWG, KMZ, SHP, GeoJSON.
- *Health and Safety*: It mainly includes data related with health clinics which is mainly in tabular format.
- *Living*: It defines the data for public facilities such as cultural resources, garbage schedule, library, street food vendors, library programs, library hours and locations, museums etc.
- *Planning and Development*: It has main three data-sets as large buildings, drainage and neighborhood names. All these data-sets are available in XML, DWG, KMZ, CSV, SHP, GeoJSON formats.
- *Transportation*: It has the data-sets as O-train stations, tracks, cycling network, OC-transpo schedule, parking lots, truck routes, railway, traffic data etc. The data related with OC-Transpo gives live changes in the schedule of buses to the users.

Thus the city of Ottawa has diversified data however, as stated above the city clearly states that it provides no warranty of the completeness of the data which is pretty much clear by looking at the data. The data entries range maximum up to a few hundreds and with very less detailing. The city of Ottawa has left a lot of room for the workers to make more progress in building a good quality of city data.

The city of Ottawa has maintained the metadata for each set describing the creation date as well as the last update for the data-set. The city has not failed to deliver the details about the dataset and its attributes or fields and also the frequency of updates for each dataset and has also listed the accuracy for each dataset.

### G. OPEN DATA-SETS OF VANCOUVER

The city of Vancouver defines open data as the the data that people want, i.e., the data that is of the most interest to the community. The notion of this project came into existence when in May 2009, Vancouver City Council passed a project named as Open3 which was to make data public for everyone to explore and make changes as per their requirements. The prime root for this draft was to do research to find out which data would be more useful for public such as legal, public services, technical and business related data. After the successful efforts, the city was able to launch its open data website in September 2009 [19] and have been adding new data-sets on the website since then. It is for public to use or make changes as per the open government license defined on the website. The important thing to be noticed here is that this city have not defined any category for its data rather the data is made available to public in data-sets. Those are on the city's website in alphabetical order. There are total 186 data-sets and are related with social, business and government areas within the city.

The first data-set is accessible Parking that is available in DWG, SHP, KML format. It defines all locations which is meant for parking. The next data-set is parks, rivers. The apartment recycling area in which data is defined mainly in SHP, KML formats. The next data-set is bike ways that is available in DWG, SHP, KML. The data-set business license defines all the business related license in CSV, XLS, JSON, XML format. The next data-set is census local area profile 2001, 2006, 2011 including data for public provision. This data is defined in CSV, XLS formats. The next data-set is city boundaries which is available in DWG, SHP, KML. Further alphabetically defined data-sets are community centers (CSV, DWG, SHP, KML, XLS), community gardens and food trees (CSV, XLS), crime (CSV, XLS, JSON, SHP), drinking fountains (DWG, SHP, KML, CSV, XLS, JSON), employee remuneration and expenses (CSV, XLS), elementary school boundaries (DWG, SHP, KML, XML), food vendors (CSV, KML, XLS), garbage collection schedule zones (DWG, SHP, KML), heritage property (CSV, KML, XLS), intersections (DWG, SHP, KML), libraries (CSV, DWG, KML, SHP, XLS), local area boundary (CSV, DWG, XLS, SHP), motorcycle parking (DWG, KML, SHP), municipal election results (CSV, XLS), noise control areas (DWG, KML, SHP), Olympic city site (DWG, KML, SHP), one way streets (DWG, KML, SHP), parks listing (CSV, XLS, XML), parking meter (DWG, KML, SHP), public art series (DWG, KML, SHP), public streets (DWG, KML, SHP), public washrooms (CSV, XLS, KML), railway (KML, SHP), road ahead closures (DWG, KML, SHP), road ahead under

construction (DWG, KML, SHP), sanitary mains (DWG, KML, SHP), sanitary manhole (DWG, KML, SHP), schools (CSV, XLS, DWG, KML, SHP), street lighting pole (DWG, KML, SHP), street trees (CSV, XLS, JSON, XML), traffic count directional (DWG, KML, SHP), traffic signals (DWG, KML, SHP), truck route (SHP), voting places (CSV, XLS, KML, SHP), water control valves (DWG, KML, SHP), weekend play-field status (CSV, XLS, JSON, XML), water transmission mains (DWG, KML, SHP), zoning districts and labels (DWG, KML, SHP). The data is available in different formats and most of the data-sets have the entries in thousands which is good for researchers and also for analysts for evaluation purpose.

#### H. OPEN DATA-SETS OF TORONTO

Toronto, the largest city of Canada is working alongside Montreal, Vancouver, Ottawa and Edmonton on improving the quality standards and maintenance of open city data. This project is being called G4Plus. Toronto has definitely achieved a good rating in developing tools over its city open data but it still has not provided the users with a good volume of data. The city has provided access to 214 data-sets which are defined under ten categories which are explained as follows:

- *Business*: This is the first category which has mainly five data-sets namely bicycle shops (available in SHP file), business improvement areas (SHP file), economic indicators (excel file), Toronto economic bulletin (excel file), Toronto employment survey summary table (excel file).
- *Community Services*: It has seven data-sets which are defined to give information to public for various social activities such as human right office service statistics, licensed child care centers, marriage licensed statistics, Ontario early year centers (Toronto), school locations, social housing, sports and recreation, Toronto public library branch locations. These data-sets are defined in different formats such as CSV, Excel file, SHP, KML etc.
- *Culture and Tourism*: This category is defined with six data-sets which gives useful information to citizens and tourists about city's cultural life. The data-sets which are involved with this category are bicycle stations (XML, JSON), cultural hot-spots (SHP), festival and events (XML), places of interest and Toronto attractions (excel, SHP), places of worship (SHP), sports and recreation (excel).
- *Development and Infrastructure*: It defines the city's infrastructure and also give an idea of construction-standards within the city. Main data-sets under this category are building permits (available in CSV, XML), heritage districts (SHP), intersection file (SHP), urban centers (DWG, FGDB, KML, JSON, API), farming protection development permit area (DWG, FGDB, KML, JSON, API), town center land use plan (FGDB, KML, JSON, API), legal plan boundaries (DWG, FGDB,

KML, JSON, API), agriculture land reserve (DWG, FGDB, KML, JSON, API).

- *Environment*: This category includes data-sets related with various environmental activities like chemical tracking, renewable energy installation and many more. These data-sets are available in SHP, excel, XML formats.
- *Finance*: It defines data-sets which provides city's capital budget (excel file), tax information (excel file), parking ticket price (CSV format), water billing (excel file) etc.
- *Health*: This category is mainly defined to give essential information related with various health services such as ambulance station locations, care centers etc.
- *Parks and Recreation*: This category is mainly design to give a brief knowledge of parks location and recreation activities within cities. The main data-sets are forest and land cover, parks, parks drinking fountains.
- *Public Safety*: It has data-sets for public safety related information such as fire station location, police station locations, Also it has data-sets called road restriction to provide safety instructions to public.
- *Transportation*: This is the last category which has maximum number of data-sets. Those are traffic cameras (CSV), traffic signals (CSV, DWG, FGDB, KML, JSON, API), railway crossing (CSV, FGDB, KML, JSON, API), poles (CSV, DWG, FGDB, KML, JSON, API), mode of travel to work (CSV), traffic counts 2013-2015 (CSV), traffic volume (CSV), walking routes (DWG, FGDB, KML, JSON, API), truck routes (DWG, FGDB, KML, JSON, API).

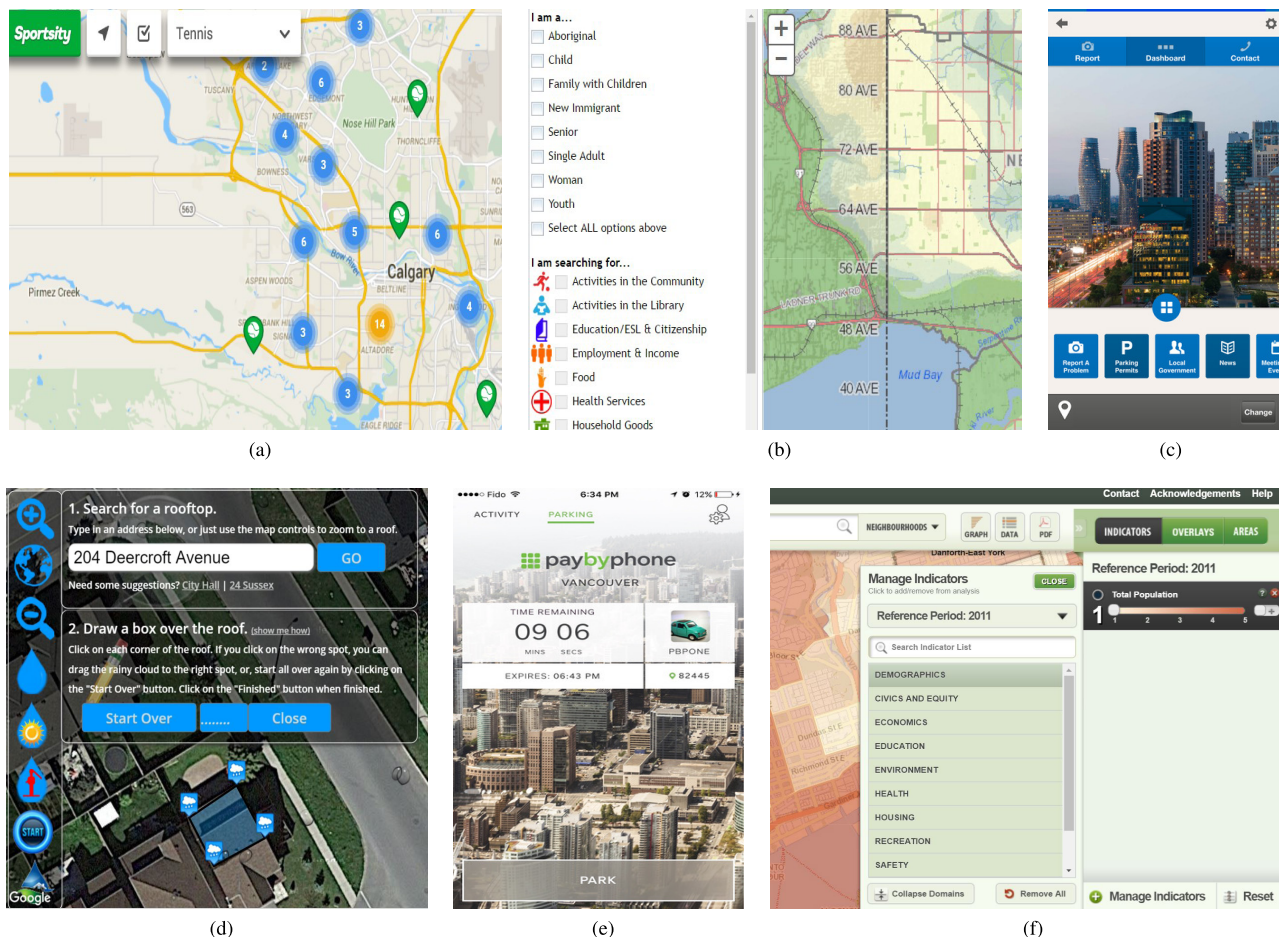
Furthermore, as explained above various data-sets are defined under these categories for users to explore complete information of the city. The users are given access to all data-sets along with a very descriptive metadata for that particular data-set having information about the data itself, data-set publish date, update frequency, data owner and the available formats. The data is updated frequently and API (Application Programming Interface) functions have also been provided by the organizations to get the latest data with a single API hit.

As discussed in the conclusion of Section II, Halifax has left a plenty of room to be filled both on maintaining quality open data. The best practice every city has followed is to provide a descriptive metadata containing details about the dataset, update frequency, creation date, attribute description etc. CKAN (Comprehensive Knowledge Archive Network) is helping most of the cities to maintain its data and the license for all the cities is governed by laws of respective provinces.

#### IV. TOOLS DEVELOPED OVER OPEN DATA OF DIFFERENT CITIES

Data visualization is a very important aspect in this open data study. The raw data has a very limited utility for a user who is not working on processing and analysis of this data. Therefore, there must be a method to create or derive meaningful data from this raw data. This data needs pre-processing





**FIGURE 1. Tools developed on open data of different cities. (a) Sportsity Application: The city of Calgary. (b) Low Cost and Free Resources Application: The city of Surrey. (c) PingStreet Toronto Application: The city of Waterloo. (d) Save the Rain Application: The city of Ottawa. (e) PayByPhone Application: The city of Vancouver. (f) Wellbeing Toronto Application: The city of Toronto.**

and analysis to get inferences from this data. However the inferences made can best be presented in form of visualizations and comparisons. Hence almost every city has worked on building applications aiming at doing the analysis part and then visualizing the analysed data. Not only the cities has built applications but also users has contributed by working on data-sets from different categories and getting useful results. Most of these cities has listed the applications built on its city open data and have opted various visualization tools and methods. Out of all those, one of the most widely used is information graphics (Infographics) which is the combined version of illustration methods and text representation. It is good enough to give a clear idea to any user. The basis for each application and visualization method is converting the raw data into a form such that it is understood by a tool on which the user is working. For example, a user Lauren Archer has used the Garbage and Recycling Data provided by the City of Toronto to produce a Web App named Garbage and Recycling Day Google Calendars having visualisations in form of Google Calendars [24]. The user has managed the raw data by converting it into a data that can be directly used into

Google calendars application and has developed a schedule for garbage and recycling days which can be downloaded by other users. The following sub-sections illustrate the tools developed by the respective cities on city open data.

**A. TOOLS DEVELOPED OVER CALGARY’S OPEN DATA**

There are various data-visualization tools which are formulated with the city’s open data. Sports is the one of highlighted area within the city as the city of Calgary is one of the sports rich city which is proved by the fact that *sport business group*, London shortlisted Calgary as “ultimate sport city 2016” in the month of January, 2016. Thus it has attracted developers’ attentions and they made one applications which is quite famous in the city, called *Sportsity*. This is meant to give the information about various sports fields within the city all in one place. The pictorial view of one of the application *Sportsity* is shown in (Figure 1(a)) which is basically designed to help users to find out any sport court or athletic park in the city and also provides the user with the directions with the help of integrated Google Maps in the application. This application uses City of Calgary’s open dataset “city amenities” to figure

out various sports recreation centres for soccer, basketball, cricket, golf, tennis and many more available throughout the city. It also lists the reviews provided by the visitors for each location and also provides the user with an option to write a review.

Another tool developed with city's open data is *Calgary Traffic Alerts* which is designed to give live traffic alerts to users. It includes information such as notification alerts for traffic incidents as accidents, construction etc., paid and unpaid parking lots, construction closures detours, traffic cameras location. Similarly work has been done on other data-sets as well to develop tools such as *+15 Walkway*, *TransitGo* and *Live Transit*.

#### B. TOOLS DEVELOPED OVER HALIFAX'S OPEN DATA

The city of Halifax worked progressively towards developing tools over its open data. The city has conducted the open data application contest named "Apps4Halifax" with the help of IBM where the users are open to post their ideas as well as submit their developed tools [25]. By the end of this contest, a winner is decided amongst both the categories-Ideas and Apps. In the 2013-14 contest, a total of 275 ideas and 38 apps were submitted online. The major sponsors of this contest were IBM (Main title sponsor), Esri Canada (Challenge sponsor), Global Halifax (Media Sponsor) and Telus (Category sponsor). The Ideas and Apps were submitted under 4 broad categories namely Your City, Go Green, Live It Up and Keep'er Movin'. One of the winners in the idea category was *Garbage When?* which uses the GPS of the mobile device and based on the location it notifies the user about the garbage day and also sends notifications in case of cancellations. The app which won in the Go Green category was *Halification* which is again based on municipal notifications for the city residents. The residents can choose from a number of subscriptions such as crime, power outages, school closures, weather warnings and traffic feeds. The sources of these alerts are the Open data catalogue of Halifax city, Twitter, Environment Canada and Halifax.ca. The data-sets used by this app from the Halifax data catalogue are "civic addresses", "solid waste collection areas" and "crime statistics". A contest ensures more participation and interest from the users as it instils the competitive spirit amongst the participants due to which they thrive for achieving the best. Hence it a progressive step by the city of Halifax for promoting the tool development over its open data.

#### C. TOOLS DEVELOPED OVER SURREY'S OPEN DATA

There are various applications which are built with various open data-sets available on the open data website of the city. Out of those application, one is built on the idea of the *Surrey's poverty reduction plan* i.e. which was made up by Surrey's planning and development department along with the engineering department. This plan was based on the vision to make "low cost and free resources" available to every citizen of Surrey. The application and its name is based on the same concept. The pictorial view of the application is shown

in Figure 1(b), the complete view of one of the application is shown which is to find out the location of low cost services as food, health services etc. The system provides users to apply filter as per their needs and thus shows the result accordingly.

There are many other tools which are developed from the various open data-sets to help people in various ways. The most popular tools are *My Surrey App*, *Surrey Request App*, *Rethink Waste Mobile App*, *Building Inspection Request App*, *COSMOS App*, *Surrey Libraries App*. One of the application that uses variety of data-sets is *My Surrey App* as it gives informative access to users to have a almost complete information about the city. It is the combination of all the other applications that are mentioned above. The main page includes news, point of interest, events, jobs, Surrey request, COSMOS, rethink waste, library, parking, bike routes etc. It has also introduced new services like Contact Surrey, which allows users can write their questions and they can get answers for those in pilot project within the city that is handled by IBM Watson technology. This application has gained a lot of popularity within the city of Surrey.

#### D. TOOLS DEVELOPED OVER WATERLOO'S OPEN DATA

Waterloo uses different mobile applications for users to visualize the data which is collected from different government bodies and from the citizens of the city. As the data collected is very complex and unstructured, so software applications are required to re-use this data in structured manner before giving to users to explore it. One of the famous mobile application which uses city's open data is *Pingstreet* (Figure 1(c)), which is designed for daily interaction of citizens and different government organization, social media etc. It is quite popular within the city as it provides real time access to various activities such as road closures, events, garbage and recycling, overnight parking etc. It is a location based discovery tool and thus all information is delivered directly to user's mobile devices without any cost cost.

*Waterloo park finder* is an another tool that is using open data-set parks that comes under *Parks and Recreation* category to help users to find out the exact location of parks within city. There are three options to do the search easily on the application i.e. search can be made by name (albert green, Alexandra lot etc.) of the park or by type of park (neighborhood park, environmental reserves, special agreement parkland, culturally significant parks etc.) or by facilities (benches, playground, hydro, water service, etc.) included in the park. After this selection the next page will show all the exact location of the park. It can be seen as a small icon, which on selection provides a short description about the park to the user. Another popular tool is *Public Art Waterloo* which is developed to give a complete knowledge about public art and its location within the city. It also gives information about nearby public art locations.

#### E. TOOLS DEVELOPED OVER OTTAWA'S OPEN DATA

There are more than 60 applications which are developed based upon the open data of Ottawa by the users participating

in Ottawa Open Data App Contest sponsored by Microsoft. There are four categories under which these applications were developed [24]. Those are listed as follows:

- On the Move (Sponsored by Telus)
- Having Fun (Sponsored by Nova Networks)
- Your City (Sponsored by CGI)
- Data Analysis and Visualization (Sponsored by Oracle)

All these categories have a few applications to give simple view of information as per their needs. The application that has pleased the users is *Save the Rain*. It uses the open data-sets Drinking Water Summary and Ontario Well Record Data. The significance of this application is to make the users realise that how much rain water could be harvested over their roof tops in a year. The reports generated from this application is amazing the users and the users like it as well. Moreover this tool has been equipped with an attractive user interface as shown in Figure 1(d).

There are many other tools developed over Ottawa's open data-sets in the contest. Another example of such tools is *Ottawa events* which is designed to have a complete knowledge of Ottawa's present and upcoming events. The events are grouped together on the basis of date, month and week. It also shows the address and tickets for each events. The main menu of this application gives two options to users to explore events either by event type (dance, fair/festival, film/new media etc.) or by location (Ottawa urban area, Ottawa rural area etc.). It also gives a short description about the event such as dance party timing, meal menu etc. Thus it is good way to explore social activities within the city. There are many other applications like *Ottawa Garbage collecting schedule*, *Recreation in Ottawa*, *OC Transpo Tracker*, *Ottawa Construction permits*, *Environmental Inspection App*, *Libraries Ottawa*, *Ottawa 311 Service Request* etc. which are using various open data-sets to give a useful information to citizens.

#### F. TOOLS DEVELOPED OVER VANCOUVER'S OPEN DATA

There are main four mobile applications developed by open data of Vancouver namely *VanConnect*, *VanCollect*, *VanGolf*, *PayByPhone*. These are meant to provide required updated information to the users such as road closures, emergency alerts and many more. *PayByPhone* is a very useful tool with maximum downloads by users that has been released by the City of Vancouver that allows the user to pay for the street parking by phone and also extend the parking via call or text. The user is charged only for the specific time he occupies the parking slot which costs less as compared to a specific time interval charge. There is also a provision to set alerts for the expiry of the parking ticket and also lets the user to get the receipts on the phone. The user interface of the application looks as shown in Figure 1(e).

City of Vancouver has developed another application *Vanconnect* that proves its utility to build an interaction bridge between Vancouver city hall and the citizens of the city. It allows the users to submit there requests as per their need or they can submit any complaints regarding garbage, street light, traffic light, abandoned vehicle etc. It gives users

various options to clarify their issues such as type, location, time and description and users can also upload a pictures for the same. It also updates users with news, emergency information etc. One of the biggest advantage of this application is that it connects users with Vancouver city hall 24×7.

#### G. TOOLS DEVELOPED OVER TORONTO'S OPEN DATA

The city of Toronto has made a huge progress in terms of visualisations of its available data. Many different organisations have come forward to utilise the open data and make inferences from it and use various kinds of visualisations for representation. The city has listed a number of mobile applications as well as web applications which have used the city data for visualisations. For example, a user Lauren Archer has used the Garbage and Recycling Data provided by the city to produce a Web Application named *Garbage and Recycling Day Google Calendars* having visualisations in form of Google Calendars [26]. The user has managed the raw data by converting it into a data that can be directly used into Google calendars application and has developed a schedule for garbage and recycling days which can be downloaded by other users. Similarly other users have contributed and worked on various data-sets at different platforms to develop visualisation tools for the given data.

City of Toronto has officially released a web based map visualisation application that is designed to ensure as well as evaluate the city's well being and is thus named as *Wellbeing Toronto (WT)* as shown in Figure 1(f). This visualisation tool is targeted for residents who need to have a good understanding of the area and communities they are living in and working and also for businesses or organizations that require parameter evaluations in order to be informed about their customers. *WT* provides a central platform for discussion of issues at neighbourhood level. It covers multiple domains such as transportation, crime, safety, education, culture etc to select indicators or parameters and provides an option to select the reference period. *WT* allows to select parameters from listed domains within the framework with a flexibility of changing the weight of a particular parameter and provide corresponding data for the users. It does not provide the open data in its raw form but in a business processed form which is more usable for the user. It also gives the user an option to use *WT* as a basis of developing other tools by using indicators in *WT*.

One more tool build on Toronto's open data is *Open Toronto* which is build to provide an easy access to Events and Festivals information within the city of Toronto. It allows users to browse the events and festivals on a particular date or in a particular months as it gives three options to users for browsing the data i.e. all, today and selected date. Thus as per the predefined information the next will show the results accordingly. Users can add information to their personal calendars in their mobile phones. Also they can share the information to their friends and relatives by sending them emails, SMS etc.

**H. SUMMARY**

From the discussion about tools developed by different cities, it can be deduced that researchers, business organisations, city authorities and users as well understand the importance of application development on city’s open data. All cities discussed above have worked upon developing tools on open data either directly or through the users. Toronto leads all the seven city in this department as not only the users have participated in enormous numbers but also city of Toronto has developed extremely useful tools like *Wellbeing Toronto* that has proved its utility in the users, businesses and researchers. It can also be concluded about these tools that most of the tools developed are using Maps as its basis for visualisation and lesser of previously used methods like graphs and charts as Maps look more presentable to display the useful information. This analysis also leaves scope for developers and organizations to set foot in and develop tools based on city open data.

**V. OPEN DATA CHALLENGES**

This section talks about the open challenges for open data. This includes mainly the open data challenges/problems for academia and industries who want to work on it on a large scale.

**A. RESEARCH CHALLENGES**

Based on the discussion on open data-sets and its characteristics so far, some research challenges can be drawn. This sub-section addresses the shortcomings in the study of open data for educational and research prospective. Today the various technical/non-technical organizations and also different educational institutes are working on open data and trying to figure out some dynamic way to deal with such a huge open data. Researchers are trying to find out that the solutions for the following questions:

- How much data is available for the public to access and view?
- Is the released open data useful enough for users with different backgrounds (technical/non-technical professionals etc.)?
- How can applications be developed in such a way that visualization, usage and comparison of open data can be made easy accessible and understandable to everyone?
- What amount of data and in which format should open data in cities be opened to the public by governmental and non-governmental organizations and how can they manage this public data in an efficient way?

There are limitations related to the access of this data as some cities either did not publish the complete data sets or did not make the open to the public. Some of the existing

**TABLE 1. Characteristics of open data and tools/applications used.**

City Name	Open Data Characteristics	Visualization Tools/Applications Used
Calgary	Availability, Non-Proprietary, Non-Discriminatory, Timely Processed and Updated	Sportsity, Calgary Trafic Alerts, +15 Walkway, TransitGo and Live Transit
Halifax	Availability, Non-Proprietary, Non-Discriminatory, Timely Processed and Updated, Variety	Halication
Surrey	Completeness, Availability, Non-Proprietary, Non-Discriminatory, Timely Processed and Updated	My Surrey App, Surrey Request App, Rethink Waste Mobile App, Building Inspection Request App, COSMOS App, Surrey Libraries App, My Surrey App, low cost and free resources
Waterloo	Completeness, Availability, Non-Proprietary, Non-Discriminatory, Timely Processed and Updated, Variety	Pingstreet, Waterloo park nder, Parks and Recreation, Public Art Waterloo
Ottawa	Availability, Non-Proprietary, Non-Discriminatory, Timely Processed and Updated	Ottawa Garbage collecting schedule, Recreation in Ottawa, OC Transpo Tracker, Ottawa Construction permits, Environmental Inspection App, Libraries Ottawa, Ottawa 311 Service Request, Save the Rain
Vancouver	Availability, Non-Proprietary, Non-Discriminatory, Timely Processed and Updated, Variety	VanConnect, VanCollect, VanGolf, Pay-ByPhone
Toronto	Availability, Non-Proprietary, Non-Discriminatory, Timely Processed and Updated	Garbage and Recycling Day Google Calendars, Wellbeing Toronto (WT), Open Toronto

data-sets are only available with a few lines of field entries. To make city data available (opened) completely there is a need to refine and bundle complete set of data in public APIs. The developers should publish the software patches related with the APIs as an open source so that the researchers can contribute to public code projects.

### B. OPEN CHALLENGES FOR INTEGRATION OF OPEN DATA

Integration of this heterogeneous open data is a very powerful approach. It allows to put the open data together across various data-sets in such a way that it can be easily explored by the users. As the section III has explained the various data-sets available in different cities, it is quite understood that every city is working independently in its own way to put data available for users and therefore these data-sets have generally not been designed to integrate together. The first problem to deal with integration process of this open data is a lack of common open source platform to study this data.

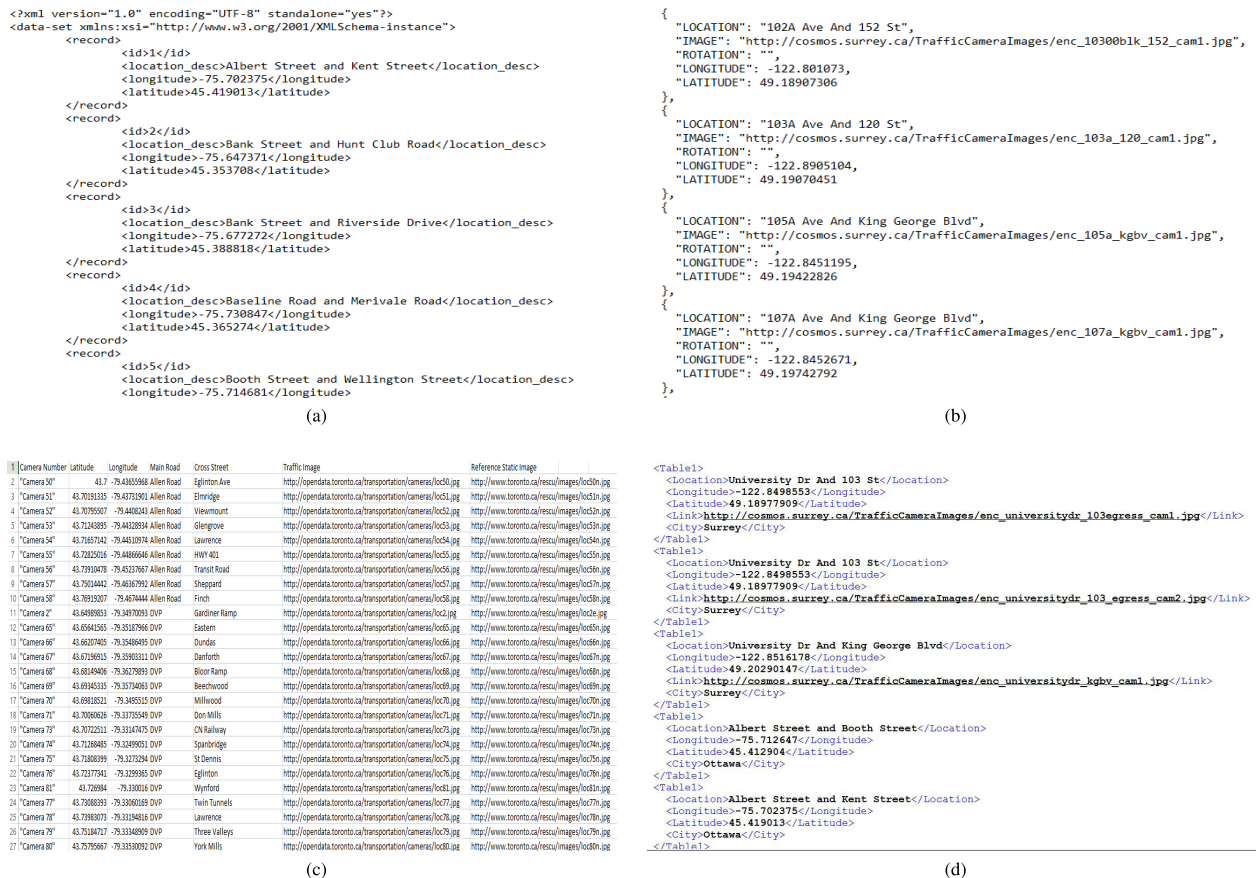
Furthermore, the differences in formats (CSV, XML, DMG, KML etc.) is the another problem in the integration of open data sources. It is easy to download the data from the respective websites of the cities but it difficult and challenging to recognize the common fields between data collected from various cities. Moreover the characteristics of

this open data and tools/applications build on the data-sets are different in all cities as mentioned in table 1. Thus it is difficult for a developer to integrate the data and make a common visualization tool for this huge open data. Also the diversity of characteristics and tools used in these seven cities create usage difficulties for the users.

Also the other kind of problems that are shown up for the integration of open data may include improper data entry, missing data or a lack of common attributes between data from different sources. But on the high level view of integrated open data may see these problems in a way that cannot be seen by individual data-sets and thus lead to data quality improvements without the need for extensive polishing.

### C. PROPOSED MODEL: DATA WRAPPER

We discuss an example of the performed study of open data in seven Canadian cities by extracting common information regarding some data in these cities. The data files are shown in Figure 2 which specifies the three formats of same data-set i.e. *traffic camera*. We developed a common model called Data Wrapper which parses various datasets from different domains and with different structure and formats (CSV, XML, JSON) and fields (latitude, longitude, id, image, location etc.) and produce a unified output in XML format.



**FIGURE 2. Different Formats of same data-set (Traffic Camera). (a) CSV format for the city of Toronto. (b) XML format for city of Ottawa. (c) JSON format for city of Surrey. (d) the combined output in XML format.**

TABLE 2. Data wrapper fields.

City Name	Data-set	Format	Parameters
Surrey	Traffic Camera	JSON	Location, Image, Rotation, Longitude, Latitude
Ottawa	Traffic Camera	XML	Id, LocationDesc, Longitude, Latitude
Toronto	Traffic Camera	CSV	Camera Number, Latitude, Longitude, Main Road, Cross Street, Traffic Image, Reference, Static Image

Table 2 shows the various data-sets fields used for traffic cameras in a few cities (Toronto, Ottawa and Surrey).

## VI. CONCLUSION

In this paper, the current status of seven Canadian cities have been depicted with respect to open data. It closely represents the open data-sets and open data tools in those cities. One of the biggest advantage of this data is to use it for making applications which keeps its users up-to-date for all activities within cities and also in their neighbourhood. This data helps to improve the lifestyle of citizens as there is feedback column on each city's websites to get reviews of citizens for data and also to know about needs and ideas of citizens to improve the data presentation. Thus it intensifies citizen engagement.

Furthermore, the research on open data of seven cities is quite complex. Users are inundated with this huge data. As different cities have different data-sets and further which are in various formats. Therefore to study this huge data on one single platform is very tricky because of diversified data collected from different cities. This is an open challenge for researchers and for cities' authorities as to bring all data-sets in single format is the first thing that researchers and analysts have to build out and to work on the same data-sets, the same formats are the biggest tasks for cities. Thus cities have made this data to be opened for researchers, analysts, IT companies to work on it and make useful tools for the city's social, cultural betterment and governmental, technological development.

## REFERENCES

- [1] D. Takaishi, H. Nishiyama, N. Kato, and R. Miura, "Toward energy efficient big data gathering in densely distributed sensor networks," *IEEE Trans. Emerg. Topics Comput.*, vol. 2, no. 3, pp. 388–397, Sep. 2014.
- [2] C. Perera, C. H. Liu, and S. Jayawardena, "The emerging Internet of Things marketplace from an industrial perspective: A survey," *IEEE Trans. Emerg. Topics Comput.*, vol. 3, no. 4, pp. 585–598, Dec. 2015.
- [3] C. Millette and P. Hosein, "A consumer focused open data platform," in *Proc. 3rd MEC Int. Conf. Big Data Smart City*, 2016, pp. 1–6.
- [4] X. Hu, T. H. S. Chu, H. C. B. Chan, and V. C. M. Leung, "Vita: A crowdsensing-oriented mobile cyber-physical system," *IEEE Trans. Emerg. Topics Comput.*, vol. 1, no. 1, pp. 148–165, Jun. 2013.
- [5] Y. Tammisto and J. Lindman, "Open data business models," in *Proc. 34th Inf. Syst. Seminar*, Scandinavia, 2011.
- [6] E. Lakomaa and J. Kallberg, "Open data as a foundation for innovation: The enabling effect of free public sector information for entrepreneurs," *IEEE Access*, vol. 1, pp. 558–563, 2013.
- [7] C. Chan, "From open data to open innovation strategies: Creating e-services using open government data," in *Proc. 46th Hawaii Int. Conf. Syst. Sci.*, 2013, pp. 1890–1899.
- [8] S. Qanbari, N. Rekabsaz, and S. Dustdar, "Open government data as a service (GoDaaS): Big data platform for mobile app developers," in *Proc. 3rd Int. Conf. Future Internet Things Cloud*, 2015, pp. 398–403.
- [9] S. Djoko, P. Theresa, and M. Cook, "A framework for benchmarking open government data efforts," in *Proc. 47th Hawaii Int. Conf. Syst. Sci.*, 2015, pp. 1896–1905.
- [10] A. Ojo, E. Curry, and F. A. Zeleti, "A tale of open data innovations in five smart cities," in *Proc. 48th Hawaii Int. Conf. Syst. Sci.*, 2015, pp. 2326–2335.
- [11] T. Vračić, M. Varga, and K. Čurko, "Effects and evaluation of open government data initiative in croatia," in *Proc. 39th Int. Conv. Inf. Commun. Technol.*, 2016, pp. 1521–1526.
- [12] C. Xu, D. Chu, and C. Li, "City event management system based on multiple data source," in *Proc. Int. Conf. Service Sci.*, 2015, pp. 169–173.
- [13] K. Yamamoto, "Visualization of GIS analytic for open big data in environmental science," in *Proc. Int. Conf. Cloud Comput. Big Data*, 2015, pp. 201–208.
- [14] (Oct. 2016). *Open Government Data Principles*. [Online]. Available: <https://public.resource.org>
- [15] (May 2016). *The City of Calgary-Open Data Catalogue-Datasets Alphabetical*. [Online]. Available: <https://data.calgary.ca/>
- [16] (May 2016). *Groups—Open Data Ottawa*. [Online]. Available: <http://data.ottawa.ca/en/group>
- [17] (May 2016). *Welcome—City of Surrey Open Data Catalogue*. [Online]. Available: <http://data.surrey.ca/>
- [18] (May 2016). *Open Data—Accessing City Hall—City of Toronto*. [Online]. Available: <http://www1.toronto.ca/wps/portal/>
- [19] (May 2016). *Data Catalogue: City of Vancouver Open Data Catalogue—Beta Version*. [Online]. Available: <http://data.vancouver.ca/datacatalogue/index.htm>
- [20] (May 2016). *Home—City of Waterloo Open Data*. [Online]. Available: <http://opendata.city-of-waterloo.opendata.arcgis.com>
- [21] (Jun. 2016). *Search—Halifax Open Data Catalogue*. [Online]. Available: <http://catalogue.hrm.opendata.arcgis.com/datasets>
- [22] (Jun. 2016). *The City of Calgary—Open Data Catalogue—Open Data Terms of Use*. [Online]. Available: <https://data.calgary.ca/stories/s/u45n-7awa>
- [23] (Jun. 2016). *Halifax Open Data*. [Online]. Available: <http://www.halifax.ca/opendata/>
- [24] (Jul. 2016). *AppsContest—An Open Data App Contest*. [Online]. Available: <http://www.apps4ottawa.ca/en>
- [25] (Aug. 2016). *Apps4Halifax—Open Data App Contest*. [Online]. Available: <http://apps4halifax.ca/>
- [26] (Jun. 2016). *Toronto Waste Pickup Calendars by Laurenarcher*. [Online]. Available: <http://laurenarcher.github.io/iCaTOWaste/>



**HAIWEI DONG** (M'12–SM'16) received the Dr.Eng. degree in computer science and systems engineering from Kobe University, Japan, in 2010, and the M.Eng. degree in control theory and control engineering from Shanghai Jiao Tong University, China, in 2008. He was a Post-Doctoral Fellow with New York University, a Research Associate with the University of Toronto, a Research Fellow (PD) with the Japan Society for the Promotion of Science, a Science Technology Researcher

with Kobe University, and a Science Promotion Researcher with the Kobe Biotechnology Research and Human Resource Development Center. He is currently with the University of Ottawa. His research interests include robotics, haptics, control, and multimedia.



**GOBINDBIR SINGH** received the B.Tech. degree from the National Institute of Technology, Jalandhar, India. He is currently pursuing the master's degree in electronic business technologies from the University of Ottawa. His interests include social media, big data, and electronic business.



**AARTI ATTRI** received the B.Tech. degree from Punjab Technical University, India. She is currently pursuing the master's degree in electrical and computer engineering with Carleton University, Ottawa. Her interests include computer networking, cloud computing, and wireless communication.



**ABDULMOTALEB EL SADDIK** (M'01–SM'04–F'09) is currently a Distinguished University Professor and a University Research Chair with the School of Electrical Engineering and Computer Science, University of Ottawa. His research focus is on multimodal interactions with sensory information in smart cities. He has authored and co-authored four books and over 550 publications and chaired over 40 conferences and workshop. He has received research grants and contracts totaling over \$18 M. He has supervised over 120 researchers and received several international awards, among others, are the ACM Distinguished Scientist, the Fellow of the Engineering Institute of Canada, and the Fellow of the Canadian Academy of Engineers, and the IEEE I&M Technical Achievement Award and the IEEE Canada Computer Medal.

...