RESEARCH ARTICLE

Achieving interoperability of smart city data: An analysis of 311 data

Soroosh Nalchigar[1] **and Mark S. Fox**[1,2*]
1. Department of Computer Science, University of Toronto, Canada
2. Department of Mechanical and Industrial Engineering, University of Toronto, Canada

Abstract: A major challenge in making cities smarter is performing comparative analyses across two or more cities, or within a city across two or more departments. The problem is that data models and the underlying semantics of their content differ, making analysis difficult at best and erroneous at worst. This paper explores the hypothesis that a single, interoperable (i.e., shareable) data model/ontology can be designed for one category of city data: openly published 311 call centre data. 311 is a service provided by many North American cities that responds to non-emergency questions and reports made by the public. It has rapidly become the single point of contact for city services, inquiries, etc. We perform a semantic analysis of the content of 311 open datasets from four cities. The result of the analysis is that existing 311 datasets combine multiple semantic dimensions in their data making it impossible to perform comparative analysis. We then construct a 311 Reference Ontology that separates the semantic dimensions, and show how 311 data from multiple cities can be mapped onto the 311 Reference Ontology. We also demonstrate how the ontology can be used to support analysis.
Keywords: 311, ontology, semantic web, city data, customer service

*Correspondence to: Mark S. Fox, Department of Computer Science and Department of Mechanical and Industrial Engineering, University of Toronto; Email:soroosh@cs.toronto.edu, msf@eil.utoronto.ca

Received: November 17, 2016; **Accepted:** February 25, 2017; **Published Online:** August 20, 2017
Citation: Nalchigar S and Fox M S, 2017, Achieving Interoperability of Smart City Data: An Analysis of 311 Data. *Journal of Smart Cities*, vol.3(1): 1-13. http://doi.org/10.26789/JSC.2017.01.001.

1. Introduction

Cities are primarily service providers. Traditionally focused on providing infrastructure services, such as transportation, public safety, potable water, sewage processing, garbage collection and electricity, they have transformed how their services interact with stakeholders (e.g., citizens and corporations to be web-based. Over the last decade these interactions have evolved from providing stakeholders with service specific touch points, to providing two general points of interaction: emergency services accessed by 911, and non-emergency services accessed by 311. 311 is the name and the telephone number of city departments that receive and process non-emergency municipal information and service requests. The main goal of 311 systems is to enhance accessibility of city services, increase cities effectiveness in responding to public inquiries, and ultimately improve city life. This evolution has greatly reduced the complexity of accessing city information and services from the stakeholders' perspectives[1].

We are now in the midst of the next evolution in city services. Motivated, if not mandated, by the open government movement, municipal information and service delivery data are now being openly published by cities to satisfy two goals: 1) making information and services more open to stakeholder scrutiny, and 2) enabling access to, and visualization and analysis of service information, by third parties. Cities such as Ottawa, Toronto, Vancouver New York, San Francisco, and Chicago all have major efforts underway to make city data, such as 911 and 311 calls, publicly available.

The goal of publishing service data is a great idea, but there are issues that limit its usefulness. Consider existing 311 data. Each city has a unique data model for publishing 311 data. Not only do they differ in the number of relations, but also in their attributes. For example, the attribute "Responsible Agency in San Franciscos dataset roughly corresponds to combination of Toronto's "Division" and "Section-unit". But is Toronto's "Service Request Name" equivalent to "Request Type" in San Francisco's dataset? Are the values equivalent? Assuming that Toronto's "Service Request Name" is equivalent to San Francisco's "Request Type", is San Francisco's "Sign Repair" equivalent to Toronto's "Sign Maintenance" or "Missing/ amaged Signs" or both? Toronto's 311 uses 371 different names for describing the service request types, while New York, San Francisco, and Chicago are using 120 and 25, and 12 different names for representing service requests, respectively.

Why do these differences in syntax and semantics pose a problem? If we are to create tools to enable stakeholders to access, visualize and analyse city services, then it would be easier and cheaper if cities shared the same data model; adopting the same data model enables the creation of generic tools that can be shared across cities. Secondly, if we are to merge, analyse and/or compare data from multiple cities, the results would be questionable without a shared data model. Without identifying and mapping equivalent attributes and values across 311 datasets, it is not possible to integrate, merge, and analyze the data[2].

In this paper, we explore the hypothesis that a single data model, i.e., ontology, can be defined that catures the semantics of the data found in 311 datasets openly published by cities in North America. The result of this exploration is the 311 Reference Ontology (311RO) that provides a unified and extensible terminology with definitions. (The OWL version of this ontology and its documentation can be found at the following addresses, respectively: http://ontology.eil.utoronto.ca/open311.owl and http://ontology.eil.utoroto.ca/open311.html). Existing 311 datasets can be maped onto the ontology thus enabling interoperability, and the creation of city data analysis tools that can be applied across cities.

The rest of this paper is organized as follows. Section 2 provides background on ontologies and their application to 311 data. Section 3 provides an overview of 311 datasets from four cities and describes their data schemas. Section 4 identifies the concepts found in existing 311 datasets and defines an ontology that spans them. Section 5 presents the evaluation of the ontology.

2. Background

Open311 (http://www.open311.org/) standardizes the API used to communicate with 311 departments. They take a minimalist approach to defining concepts and attributes that are common across cities. For example, their GeoReport API "allows developers to build applications to both view and report issues" (http://wiki.open311.org/GeoReport_v2/), such as potholes, graffiti or broken street lights. GeoReport assumes that city services differ from city to city and does not conceptualize what these services are. Instead they provide a "Get Service" API where cities return their specific services, leaving it to the cities to define them. Secondly, it provides a "Get Service Definition" API where the city returns a service specific set of attributes, their datatype and possible valuesa generic Service attribute and value vocabulary or ontology is not provided. Conversely, GeoReport's "Post Service Request" API, which allows for the communication of service requests, introduces attributes that are generic to service requests across all cities, such as: latitude, longitude, address, email, first name, last name, phone, and description. Finally, "Get Service Request", which provides the status of a service request, introduces generic attributes for: start date, end date, status (open, closed), agency responsible and various times of service. These are a first step towards introducing a 311 vocabulary, but only for the attributes of a service request.

With the wide availability of 311 datasets analysis of this data has begun. Appendix A provides a list of cities with 311 datasets found on Namara.io. Many cities perform frequency analyses to determine the types and volumes of requests. Attempts have been made to predict call volume based on 311 dataset attributes and possibly other attributes from other datasets. Zha and Veloso[3] attempt to predict call volume based on 311 call records and weather data. But their results are no better than the common sense "weather model", that is what happened yesterday is the best predictor of today. On the other hand, OBrien[4] demonstrates that useful results can be achieved by analyzing the myriad of attribute values found in a single city's 311 dataset.

If cities wish to compare their 311 experiences, and if we wish to extend predictive analyses to incorporate data from multiple cities, a gap has to be filled, namely the definition of a standard vocabulary of service types, their attributes and their values commonly found in 311 datasets. In other words, what are the services provided by a city, along with their attributes and values.

The definition of city services has been a focus of efforts in creating municipal reference models. The Municipal Reference Model (MRM)[5] documents Canadian efforts over the last 20 years to standardize municipal documentation and information systems around Programs, Services, Processes and Resources. While programs, services, processes and resources are the standard terms, specific services, resources, etc. are left to municipalities to define. More recently standards bodies have focused on standards for city knowledge. PAS 182[6] is a recent British Standard for the representation of high level city concepts. Like the MRM, its standard terms exist at a higher level of abstraction, covering, Service, Event, Resource, Organization, Plan, Agent, Agreement, etc. Specific services are not included.

IBM's SCRIBE is a semantic model for cities[7]. It provides a taxonomy of city services including Education, Justice and Correction, Public Safety. Though not focused on 311, the event class provides an approach for capturing 311 requests and their status, hence overlapping with the Open 311 API attributes.

In order to test our hypothesis, it is necessary to represent the meaning of existing 311 attributes and values. An Ontology is an "explicit representation of shared understanding"[8].It "consists of a representational vocabulary with precise definitions of the meanings of the terms of this vocabulary plus a set of formal axioms that constrain interpretation and well-formed use of these terms"[9] What distinguishes simple vocabularies from ontologies is the latter adds definitions of the terms and constraints on their interpretation using a computational language. Key to the creation of an Ontology is grounding the definitions of terms in lower level, more concrete terms[10]. Languages such as Description Logic and First Order Logic are used to define the terms of the ontology. By adopting an ontological approach, we can provide a clear and precise set of concepts and properties, along with their definitions, that span 311 data.

3. Analysis of Published 311 Data

We analysed the 311 data sets of four cities in order to determine their semantic content. The cities chosen were: Toronto, New York, San Francisco, and Chicago. In selecting these cities, we considered factors such as availability of 311 data as well as existence of enough instances of service requests to ascertain the variety of their attribute values. This section describes the datasets of each of the cities.

3.1. Toronto

Toronto's open 311 dataset (http://www.toronto.ca/311) includes 6 attributes. (After the completion of this research, Toronto 311 revised all of their 311 datasets available on their open data site to include only three attributes: Date, Location, and Service Request Type.) Service Request Name is the unique title of an individual service request. Problem Code is a unique identifier of the service request name. Creation Date indicates the date and time that the corresponding request instance was submitted to 311. The attributes Division and Section-Unit represent the responsible City division and the section or unit within the division. Finally, Internet Self-Serve shows if the service request is reported via the web. Table 1 shows a service request record in this dataset.

Table.1 Toronto Service Request Record

Item	Result
Creation date	02/01/2010 8:55:59 AM
Service Request Name	Missing/Damaged Signs
Division	Transportation Services
Section - Unit	TMC - Signs & Markings
Problem Code	SAM - 01
Internet Self Serve	Yes

3.2. San Francisco

The San Francisco 311 dataset (http://data.sfgov.org) includes 15 attributes. In this dataset, the three attributes Category, Request Type and Request Details split and extend the information found in Torontos Service Request Name, Whereas the single attribute Responsible Agency aggregates Torontos Division and Section-Unit attributes and more. Furthermore, this data set has additional attributes than Toronto, such as Status, Address, and Point (latitude and longitude coordinates). Table 2 shows an example of a service request record in this dataset.

Table 2. San Francisco Service Request Record

Item	Result
CaseID	2441080
Opened	06-03-2013
Closed	06-02-2013
Status	Closed
Work Status	N/A
Responsible	DPT SignShop Survey
Agency	Tech_Sean Philpott Queue
Address	60 ONONDAGA AVE, SAN FRANCISCO, CA, 94112
Category	Sign Repair
Request Type	Sign - Defaced
Request Details	Street_Cleaning - Defaced; support - OK
Source	Twitter
Supervisor District	11
Neighborhood	Outer Mission
Updated	06-03-2013
Point	(37.722021793, -122.438834272)

3.3. New York

New York's open 311 dataset (http://nycopendata.sorcata.com) includes 53 attributes. Unique Key, Created Date, Closed Date, and Agency in this dataset are equivalent to CaseID, Opened, Closed, and Responsible Agency attributes form San Francisco's dataset, respectively. Other attributes such as Complaint Type, Latitude, and Longitude have obvious equivalens, but with a different name than in the San Francisco dataset. Some of the attributes that are appearing only in this dataset are Due Date, Facility Type, Location Type, Cross Street. It should be noted that this dataset, at the time of this research, has some attributes (e.g., Garage Lot Name, School Number) whose values are missing for the majority of the records in the dataset. This could be due to the fact that those attributes are only relevant for a small subset of problem types. Table 3 provides an example of a record in this dataset.

Table 3. Subset of the New York Service Request Record

Item	Result
Unique Key	32840262
Created Date	03/05/2016 13:41
Closed Date	03/05/2016 14:04
Agency	NYPD
Agency Name	New York City Police Department
Complaint Type	Derelict Vehicle
Descriptor	With License Plate
Location	Type Street/Sidewalk
Incident Zip	11209
Incident Address	411 100 STREET
Cross Street 1	4 AVENUE
Cross Street 2	FT HAMILTON PARKWAY
Status	Closed
Resolution	Action
Updated Date	03/05/2016 2:04:03 PM
Borough	BROOKLYN
X Coordinate (State Plane)	975020
Y Coordinate (State Plane)	162481
Latitude	40.61264
longtitude	74.0332
Location	(40.61264440251783, 74.0332439930462)

3.4. Chicago

Chicagos open 311 dataset (http://data.cityofchicago.org) is provided in separate files, where each file contains requests of a specific type (e.g., tree debris, garbage carts, etc.) and has a different set of attributes. While there are 15 common attributes in these files (e.g., Creation Date, Completion Date, Status, Service Request Number), there exist some attributes that are unique and belong to only one specific file. For example the attribute "Licence Plate" appears appears only in the file that keeps requests of type "Abandoned Vehicle Complaint". The 15 attributes that are common in Chicago data files, all have an equivalent attribute in either or both San Francisco and New York datasets. However, they usually have different names. For example the Completion Date here is equivalent to the Closed attribute in San Francisco. If we assume that New York's "Complaint Type" is equivalent to Chicago's "Type of Service Request", it is

Table 4. Chicago Service Request Record

Item	Result
Creation Date	04/15/2014
Status	Completed
Completion Date	05/22/2014
Service Request Number	14-00542162
Type of Service	Request Abandoned Vehicle Complaint
Licence Plate	IOWA 650-ZCZ
Vehicle Make/Model	Mazda
Vehicle Color	Black
Current Activity	FVI-Outcome
Most Recent Action	Create Work Order
How Many Days has the Vehicle Been Reported as Parked?	180
Street address	468 W MELROSE ST
ZIP Code	60657
X Coordinate (State Plane)	1172371.641025
Y Coordinate (State Plane)	1921912.242725
Ward	44
Police District	19
Community Area	6
Latitude	41.9412817441674
longtitude	- 87.6420057129598
Location	(41.94128174416743, - 87.64200571295987)

not clear whether New York's "Derelict Vehicle" equivalent to Chicago's "Abandoned Vehicle Complaint" (See Tables 3 and 4). Table 4 provides an example of a service request record in Chicagos dataset.

3.5. Observations

There is little commonality across cities in the structure and content of their open 311 data sets. They vary in the number of attributes, the naming of the attributes and the naming of values. Secondly, the intentionality of a 311 record can be vague. For example, is New York's "Derelict Vehicle" meant to convey it is to be removed or to be investigated? The same holds for Chicago's "Abandoned Vehicle Complaint". Finally, some cities are less open than others. For example, Toronto's open 311 data does not contain information on where the problem occurred nor its status.

4. 311 Semantic Analysis

In order to develop a 311 reference ontology, we have to understand both the explicit content and implicit intent of 311 records. Our analysis of the semantics of 311 datasets is divided into two parts. The first part focuses on the semantics of the request type or category. This is the critical component of confirming our hypothesis. If we can clearly identify and represent the semantic components of requests, then we will be able to create a unifying ontology. The second part focuses on "other" attributes for which the semantics is clear, such as address or date.

4.1. Request Type Analysis

The main issue we address in this section is that each city has its own vocabulary for describing a service request. It is often the case that the values associated with a request name, type, category, or however it is referred to by a city, combine two or more concepts. In the following we identify and separate the concepts embedded in service requests found in 311 datasets, and use these as the basis for defining our ontology. Consider the following examples taken from New York(Table 5):

Table 5. Examples from New York

Complaint Type	Descriptor	Location Type
Illegal Parking	Blocked Hydrant	Street/Sidewalk
Noise-Commercial	Loud Music/Party	Club/Restaurant
Vacant Lot	Request to clean	Lot

The first record represents a service request that is about a something illegally parked which implies a vehicle. The location of the vehicle is spread acrross two attributes: by a hydrant on a street/sidewalk. It is implicitly requesting an action from the responsible 311 agency to remove the vehicle. The second record shows a complaint about noise. The location is a club or bar or restaurant. The action that needs to be taken is again implicit. The third record shows a request about a vacant lot (complaint type and location are the same in this case). It represents a request for doing cleaning (action). These examples show that in its current form, the dataset is merging different kinds of information under the same attribute name. For example, the attribute "Descriptor" captures information about subject (e.g., hydrant) and action (e.g., cleaning) and also type of location (e.g., commercial) of the corresponding service request. Next, consider a set of examples from San Francisco(Figure 6):

Table 6. Examples from San Francisco

Category	Request Type	Request Details
Sidewalk Cleaning	Mattress	
Sign Repair Sign-Defaced	Stop-Defaced	support - OK
Graffiti Private Property	Not_Offensive	Building_ other
Property	Graffiti on Private Property	Markings

The first record is about a mattress that is located on the sidewalk. It requests a cleaning action. The second record is about a stop sign. In this case, the location information is not given which can be a street, highway or sidewalk etc. The request is calling for an action which in this case is to repair. The third record is about a graffiti that is located on a private building. Note that the nature of the graffiti (not offensive) is also captured. From these examples, one can see that there is no consistency is the content of the attributes. For example, the attribute "Category" includes information about the action that needs to be taken (e.g., repair) and also it can represent information on the location (e.g., private property), in addition to the subject (e.g., graffiti).Now, consider following examples from Toronto Table 7:

The first record shows a request whose subject is a recycle bin. The location of the request is a residential building. Also, the record is a kind of reporting that something is missing. The action that needs to be taken by 311 is implicit

Table 7. Examples from Toronto

Service Request Name
Residential: Recycle Bin: Missing
Traffic Signal Maintenance
Litter / Sidewalk & Blvd / Pick Up Request

in this case. The second record is about a traffic signal. It shows a request for a maintenance type of action. The third record represent a request that is about litter. The subject (in this case litter) is located on a sidewalk and/or boulevard. Moveover, it is requesting an action, in this case to pick-up. One can observe that, similar to examples from New York and San Francisco, Torontos dataset combines information on different concepts under a single attribute.

Last, for Chicago, service requests are recorded in different files, where each file and its schema correspond to a specific service request type. Examples of types include Tree Trims, Alley Lights Out, and Pot Holes Reported. Again, similar to the issue with previous three cities, information on several concepts related to the requests are all combined and represented in the request type. A service request about a damaged street signs is recorded as “Sign Maintenance” in Toronto dataset, while it appears as Sign Repair and “Street Sign-Damaged” in San Francisco and New York datasets respectively.

1.Our semantic analysis reveals four recurring conepts that comprise a complaint:

2.The subject of the request, e.g., street sign, garbage bin.
3.The type of the location in which the subject exist, e.g., sidewalk, residential building.

4.The type of action that is needed to be taken by the 311 agency, e.g., maintain, clean.

5.The type of message that is being delivered from crowd, e.g., report, complaint.

4.2. Standard Attributes

Our review also identified a set of attributes that are common across the cities. Some examples of those attributes include: open date, close date, status, neihborhood, borough, intersection, and zip code. We incorporate them in the 311 Reference Ontology.

5. 311 Reference Ontology

Based on our analysis in the previous section, we define the 311 Reference Ontology. The core of the ontology is the ServiceRequest, which is composed of a principal property, has311Type, and a set of secondary properties. In the following (Figure 1) we first define the range of has311Type, followed by the secondary properties.

5.1. 311Type

The principal property of ServiceRequest is has 311 Type whose range is 311 Type. 311 Type deconstructs the intent of a 311 request into the following properties:

has311Subject: whose range is the class Subject whose subclasses include but are not limited to RoadSymbol,

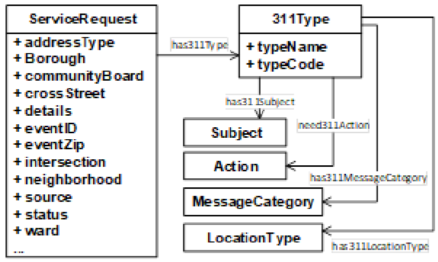

Figure 1. the range of has311Type

GarbageContainer, Plants, Animals, etc. RoadSymbol is further decomposed into Signal, Sign, and Light classes. The following (Figure 2) depicts a portion of the Subject taxonomy:

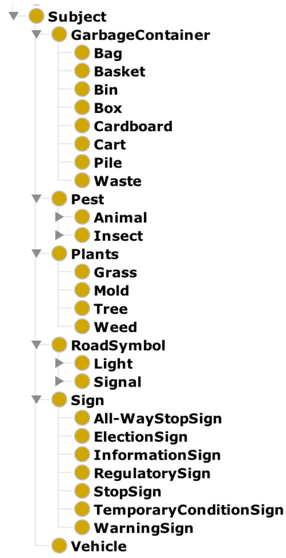

Figure 2. has311Subject

need311Action: whose range is the class Action and represents the action that the 311 agency needs to undertake in response to the corresponding Service Request instance. The class Action has subclasses such as Replace, Repair, Remove, Reinstall, Install, Inspect, etc. The following (Figure 3) depicts the current hierarchy:

has311MessageCategory: whose range is the class MessageCategory which includes four subclasses of Complain, Compliment, Report, and Request. Each of these classes has been decomposed into subclasses. For example the class Request has three subclasses of RequestForInfrormation, RequestForInvestigation, and RequestForAction. The following (Figure 4) depicts the current hierarchy:

hasLocationType: whose range is the class LocationType and represents the type of location of the service request. It has subclasses such as Residential, Commercial, Public, and Private. The following (Figure 5) depicts the current LocationType taxonomy:

We believe that the 311 service requests from the cities we reviewed can be deconstructed into these four properties, resulting in a precise representation of the focus and intent of the request.

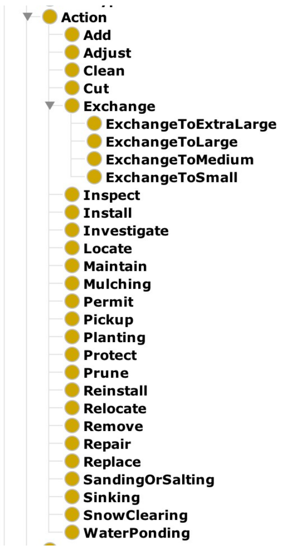

Figure 3. need311Action

Figure 4. has311MessageCategory

Figure 5. hasLocationType

5.2. Secondary ServiceRequest Properties

In addition to has311Type, we include the "standard" data properties that most cities publish. Note that some of the data properties, such as ward and borough can be converted to object properties if a city were to publish IRIs for them:

- **addressType:** type of the address of the service request (e.g., Blockface).
- **borough:** borough of the service request (e.g., Manhattan).
- **communityBoard:** the community board of the service request (e.g., 04 Manhattan).
- **crossStreet:** the two cross streets nearest to the location of event.
- **details:** further information about the request.
- **eventID:** the unique ID for each instance of the service request.
- **eventZip:** the zip code of the service request.
- **intersection:** the intersection streets close to the location of service request.
- **neighborhood:** the neighborhood of the ser-vice request.
- **source:** represents how the service request was made (e.g., voice in).
- **status:** represents the status of the request.
- **ward:** shows the ward number of the service request. Along with these data properties, the class ServiceRequest has following secondary object properties:
- **isHandledBy:** whose range is the class Agency, represent the 311 agency that handles the ser-vice request.
- **hasSPS:** whose ranges is the class SPSPoint, identifying the exact location of the service requests.
- **isSubmittedTo:** whose range is the class org: ivision, showing the 311 responsible division to which the service request submitted.
- **hasOpenDate:** whose range is the class time: DateTimeInterval and represents the submis-sion date and time of the service request.
- **hasDueDate:** similar to previous property, it represents when is the due date and time of the submitted requests.
- **hasCloseDate:** similar to previous property, it captures the closing date and time of the service request.

5.3. Reuse of Ontologies

For many of the secondary object properties, we atempted to reuse other ontologies where possible. The following describes these ontologies.

Organization Ontology. Organization ontology, defined by Fox *et al*[11], focuses on organization structure, roles, authority and empowerment. It was developed as part of the Toronto Virtual Enterprise Project[12]. It is available at: http://ontology.eil.utoronto.ca/organization.owl. (Concepts in the organization ontology appear with the prefix "org".) One of the core classes in this ontology is Organization, defined as a set of constraints on the activities performed by agents. This class contains the following data and object properties:

- **hasName**: a text showing the name of the organization.
- **hasLegalName:** represents the legal official name of the organization
- **hasGoal:** whose range is the class Goal and defines the goals of the organization.
- **consistsOf:** whose range is Division and repesents the subdivisions of the organization.

GeoNames Ontology. The service requests submitted to 311 are associated with a geographic area, which could be a borough, park, cemetery, building, etc. Therefore, a requirement for the 311RO is the ability to identify the geographic area to which the service request is related. The GeoName geographical database includes over 10 million placenames. Beyond names of places in various languages, this database integrates geographical data such as latitude, longitude, elevation, population and postal codes from various sources. All the placenames are instantiations of the GeoNames Ontology that incorporates other ontologies including Schema.org. It is available at http://www.geonames.org/ontology/ontology_v3.1.rdf. (Concepts in this ontology appear with the prefix "gn".) The most fundamental class in GeoNames Ontology is the class gn: feature which includes the following properties:

- **name:** text, representing the main international name of a feature (e.g., New York).
- **altenativeName:** a number of alternative names for the feature.
- **countryCode:** a two letters country code in the ISO 3166 list.
- **population:** population of the feature.
- **wikiPediaArticle:** a Wikipedia article of which subject is the resource.

International Contacts Ontology. Service requests include the address for which the request is made. The address text usually includes number, street name, as well as the postal/zip code. The International Contact (iContact) Ontology provides basic classes and properties for the representation of international street addresses, phone numbers and emails. It is available at http://ontology.eil.utoronto.ca/icontact.owl. (Concepts in this ontology appear with the prefix "ic".) One of the important classes in this ontology is ic:Address that includes following properties:

- **hasStreet:** text, showing the name of the street.
- **hasUnitNumber:** a non-negative integer representing the unit number where the request is located.
- **hasPostalCode:** text, representing the postal code of the location.
- **hasStreetDirection:** shows the direction of the street (e.g., north, east).
- **hasStreetType:** whose range is the class ic: StreetType and shows the type of the street (e.g., avenue, road, boulevard).

Time Ontology. Service requests are associated with temporal information such as the submission date, the closing date, etc. We use the Owl-Time Ontology for representing temporal properties of service requests. It is available at http://www.w3.org/2006/time. (Concepts in this ontology appear with the prefix "time".) Owl-Time provides a standard set of classes and relations for representing relations among instants and intervals, as well as durations and date/time. One of the main classes in this ontology is DateTimeInterval that is connected to the class DateTimeDescription through the object property hasDatetimeDescription. The class DateTimeDescription includes various data properties such as second, minute, hour, day, month, year, etc.

Transportation Ontologies. In order to provide sufficeint expressivity, we needed to define sub-classes of the Transportation Routes. Several urban ontologies, e.g., Towntology ontology[13] and CityGML[14], contain transportation-related classes. Although these hae been created with a specific task in mind, they identify some of the subclasses of transportation routes.

Other Ontologies. There are other classes in our ontology, such as Plants, Animal and Insects that are related to other ontologies. Interested readers are referred the OWL file of our ontology for further details (http://ontology.eil.utoronto.ca/311RO.owl).

Figure 6 depicts how the 311RO is related to the Organization ontology. In this figure we specialize the class Organization to the class Agency. This allows the class Agency to inherit the properties of the Organization as defined in Organization ontology, e.g., hasName.

Figure 7 shows how the 311RO integrates the GeoNames and Schema.org ontologies. This figure indicates that the object property hasCity connects the class ServiceRequest to the schema.org class sc:City which inherits the properties of the geonames class gn:Feature.

Figure 8 shows how the 311RO is related to iContact Ontology.

Figure 9 illustrates that the class ServiceRequest form 311RO is connected to Time Ontology through four different object properties, namely hasOpenDate, hasCloseDate, hasUpdateDate, and hasDueDate.

In the 311RO, the class ServiceRequest is connected to the class 311Type via the object property has311Type. The class 311Type is connected to the class LocationType through the object property hasLocationType. One of the subclasses of LocationType is TransportationRoutes, meaning that a service request could be located on a transportation route such as an Expressway. Figure 10 shows how our ontology is connected to the Towntology, CityGML and DBpedia ontologies.

It should be noted that within the 311 Reference Ontology, the class TransportationRoutes has other subclasses that were identified by careful review of 311 city datasets, e.g., Expressway, Boulevard.

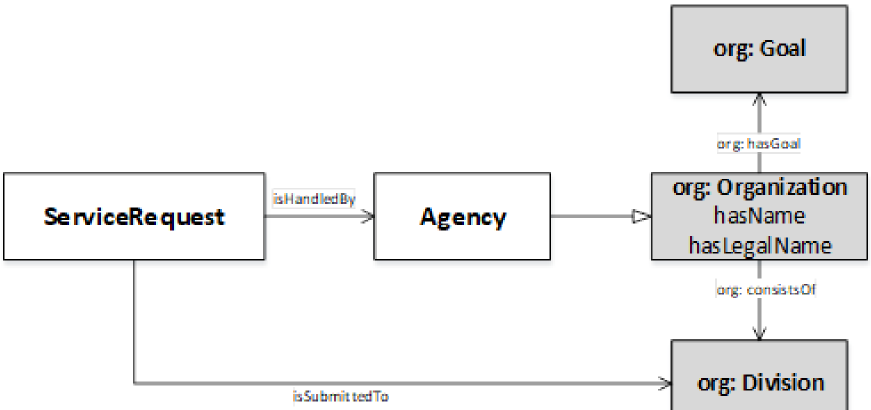

Figure 6. 311 Reference Ontology in relation to Organization Ontology. (In all the figures in this paper, arrows with open arrow head represent the rdfs:subclassof properties. Regular arrows symbolize the object property of the given label.)

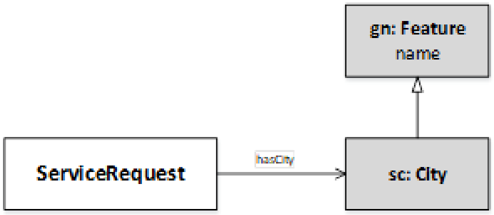

Figure 7. 311 Reference Ontology in relationto GeoNames and Schema.org ontologies.

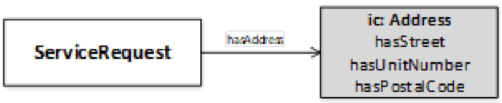

Figure 8. 311 Reference Ontology inrelation to iContact Ontolog.

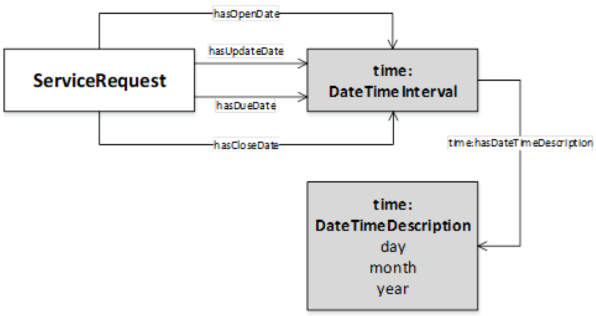

Figure 9. 311 Reference Ontology in relation to Time Ontology)

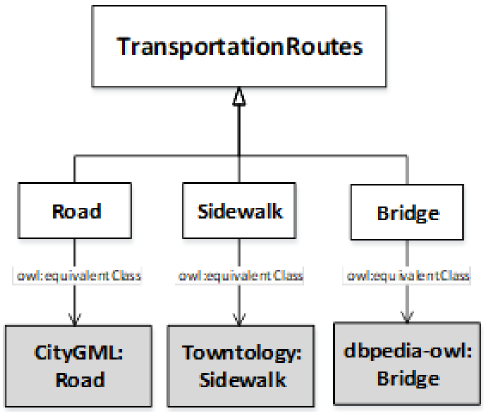

Figure 10. Equivalent Classes of 311 Reference Ontology in Other Ontologies

5.4. Description Logic Definition of Classes

311RO contains 187 classes, 83 object properties, 44 data properties and 37 individuals. The following fomalizes the two main classes, ServiceRequest and 311Type, using Description Logic (DL). To represent the "exactly one" cardinality in these formulations, we contract the≥ 1 and≤ 1 constructors to =1, due to space limitations.:

$$
\begin{array}{l}
\mathrm{ServiceRequest} \equiv \\
\quad =1\ \mathrm{has311Type.311Type} \sqcap \\
\quad =1\ \mathrm{isHandledBy.Agency} \sqcap \\
\quad =1\ \mathrm{isSubmittedTo.Division} \sqcap \\
\quad =1\ \mathrm{hasAddress.Address} \sqcap \\
\quad =1\ \mathrm{hasCity.City} \sqcap \\
\quad =1\ \mathrm{hasOpenDate.DateTimeInterval} \sqcap \\
\quad \leq 1\ \mathrm{hasCloseDate.DateTimeInterval} \sqcap \\
\quad \leq 1\ \mathrm{hasDueDate.DateTimeInterval} \sqcap \\
\quad \geq 0\ \mathrm{hasUpdateDate.DateTimeInterval} \sqcap \\
\quad =1\ \mathrm{eventID.string} \sqcap \\
\quad =1\ \mathrm{source.string} \sqcap \\
\quad =1\ \mathrm{status.string} \sqcap \\
\quad \leq 1\ \mathrm{addressType.string} \sqcap \\
\quad 1\ \mathrm{borough.string} \sqcap
\end{array}
$$

```
≤1 communityBoard. string ⊓
≤8 crossStreet.string ⊓
≤1 deatils.string ⊓
≤1 intersection.string ⊓
≤1 hasSPS. SPSPoint ⊓
≤1 neighborhood.string ⊓
≤1 ward.string
311Type ≡
=1 has311Subject.Subject ⊓
=1 need311Action.Action ⊓
=1 has311MessageCategory.
    MessageCategory ⊓
=1 has311LocationType.
   LocationType ⊓
=1 311typeCode.string ⊓
=1 311typeName.string
```

6. Evaluation

We evaluate the 311 Reference Ontology in two parts. The first part evaluates whether the ontology can represent the data that is needed to answer competency questions. The second part evaluates the ontology by illustrating how data from each city is represented according to the ontology.

6.1. Usage Scenarios

In this section we use the ontology engineering methodology defined by Gruninger and Fox[15]. We begin by defining two usage scenarios that identify the user and how they would use information stored using the ontology. Competency Questions are derived from the scenarios. Competency questions serve both as requirements for the ontology and as a means for evaluating the ontology.

Customer inquiries. The contact center of the city 311 department receives numerous calls from customers who have inquiries about their previously reported service requests. Usually, the customers call to check the status of their request and to get updates, based on the reference number of their service request. To answer those inquiries, the contact center needs to access the stored data of service requests. To this end, the city 311 needs to keep records of the date and time in which the request was submitted as well as its latest status (open, closed, etc.).

Performance management. Every day, the 311 call center receives thousands of service requests from the crowd, through various channels such as email, smart phone apps, and phone calls. The mayors office understands that in the current rapidly changing business environment, deriving insights from raw data and making data-driven decisions is important. Towards this end, the 311 department has developed a standard reporting system that addresses the information needs of the mayors office. Among others, the mayor's office wants to know what the busiest agencies are, i.e., which agencies are receiving highest number of service request. This information would help them to assign more employees to busy agencies, balance the workload, and hence reduce the time it takes to address requests. Also, each service request is about a different subject, e.g., garbage bins, traffic signals, etc. The mayor's office wants to know what the most reported service topics are. These will help them in aggregating messages arising from the crowd and use it to gain insights about the city problems. Beside these reports, the mayor's office is interested in comparisons and cross-city analyses. They like to know how other cities are different from them with respect to environmental pollution and crime. In particular, they like to know which cities are having more reports about dead animals as well as reports about law contravention. In order to generate these reports, the city needs to be able to compare their 311 data with the 311 data of other cities.

The usage scenarios identify two types of request knowledge that needs to be represented: knowledge of the status of requests, and knowledge of the content/type of requests. These are addressed in the next section.

6.2. Competency Questions

Based on the above scenarios, we have identified three categories of competency questions. The first category focuses on the simple retrieval of attribute values:

- **QC-1:** What is the submission date of a given service request with the unique code "XYZ"?
- **QC-2:** What is the status of a given service request with the unique code "XYZ"?
- **QC-3:**What is the category of a given service request with the unique code "XYZ"? The second category of competency questions focuses on the aggregation of information within a single city:
- **QC-4:** What are the top five busiest 311 agencies in terms of total number of received service requests?
- **QC-5:** How many service requests about "Subject1" are reported since the beginning of the year?
- **QC-6:** What street of the city has the most number of service requests?
- **QC-7:** What are the most reported service subjects? The third category of competency questions focuses on cross-city comparisons (i.e., transversal analysis):
- **QC-8:** Which city is receiving the most number of service requests from citizens?
- **QC-9:** Which cities have received more than 1000 reports categorized as illegal issues?
- **QC-10:** What are the top three cities with most number of reports of the subject "dead animals"?

6.3 Answering the Competency Questions

In this section we translate the competency questions into SPARQL questions using the 311 Reference Ontology. All queries assume that the namespace prefix 311RO refers to the IRI http://ontology.eil.utoronto.ca/ 311RO.owl.

***QC-1:** What is the submission date of a given service request with the unique code "XYZ"?*

In order to answer the first competency question, we need to retrieve the date in which the given service request was submitted to the city 311. The following query finds the answer:

```
SELECT ?day ?month ?year
WHERE\{
  ?ServiceRequest 311RO:eventID "XYZ".
  ?ServiceRequest 311RO:hasOpenDate
  ?DTInterval.
  ?DTInterval time:
    hasDateTimeDescription
  ?DTD.
  ?DTD time:day ?day.
  ?DTD time:month ?month.
  ?DTD time:year ?year
\}
```

In our ontology, the ServiceRequest class is connected to the class DateTimeInterval (imported from Time Ontology) via the object property hasOpenDate. In Time Ontology, the DateTimeInterval class is connected to the class DateTimeDescription through the object property hasDateTimeDescription. The data that is required to answer the first competency question are represented as data properties of the class DateTimeDescription.

QC-2: *What is the status of a given service request with the unique code "XYZ"?*

The following query answers the question:

```
SELECT ?Status
WHERE\{
  ?ServiceRequest 311RO:eventID "XYZ".
  ?ServiceRequest 311RO:status ?Status
\}
```

In the ontology, the class ServiceRequest has the data property of status whose value is a string. The answer to the second competency question could be obtained from this data property.

QC-3: *What is the category of a given service request with the unique code "XYZ"?*

The following query obtains the answer to the third competency question:

```
SELECT ?Category
WHERE\{
  ?ServiceRequest 311RO:eventID "XYZ".
  ?ServiceRequest 311RO:has311Type
  ?311Type.
  ?311Type 311RO:typeName ?Category
\}
```

The class ServiceRequest is connected to the class 311Type via the object property has311Type. The category of each service request is recorded as a data property of the class 311Type.

QC-4: *What are the top five busiest 311 agencies in terms of total number of received service requests?*

The answer to this competency question is obtained by following SPARQL query:

```
SELECT ?Name (COUNT (?ServiceReuqest)
     AS ? Total)
WHERE\{
  ?ServiceRequest 311RO:isHandeledBy
  ?Agency.
  ?Agency 311RO:hasName ?Name
\}

GROUP BY ?Name
ORDER BY DESC ?Total
LIMIT 5
```

In our ontology, the object property isHandledBy connects the class ServiceRequest to the class Agency. The class Agency has the data property of hasName which is a unique string representing the name of agency that handles the service request. In order to find the answer to the forth competency question, this query counts total number of service requests that are submitted to the city agencies. Then, by ordering and finding the top 5 instances of the class Agency, the answer to the fourth competency question is found.

QC-5: *How many service requests about "Subject1" are reported since the beginning of the year?*

Regarding the fifth competency question, we need to retrieve and count service requests of the given subject that are reported in the current year. To do that, following SPARQL query can be used:

```
SELECT (COUNT(?ServiceRequest)
     AS ?Total)
WHERE\{
   ?ServiceRequest 311RO:has311Type
   ?311Type.
   ?311Type 311RO:has311Subject
   ?Subject.
   ?Subject a 311RO:Subject1.
   ?ServiceRequest 311RO:hasOpenDate
   ?DateTimeInterval.
  ?DateTimeInterval time:
     hasDateTimeDescription
  ?DTD.
  ?DTD time:year ?Year.
  FILTER (?Year = "2015")
\}
```

In our ontology, each instance of the class ServiceRequest is associated with its Type through the object property has311Type. Moreover the class Service Request is associated with the class DateTiemInterval from Time Ontology, to keep the time information in which a request is reported. Within Time Ontology, the class DateTimeInterval is connected to the class DateTimeDescription through the object property hasDateTimeDescription. The first step in answering this competency question is to retrieve the set of all service requests whose Type instance has the subject that is the given in the competency questions. Having this set, the next step is to exclude those instances which are not submitted in the current year and count total number of service requests that are remained.

QC-6: *What street of the city has the most number of service requests?*

The following query can be used to find the an-swer to this competency question:

```
SELECT  ?Street  (COUNT(?ServiceRequest)
      AS ? Total)
WHERE\{
    ?ServiceReuqest 311RO:hasAddress
    ? Address.
          ?Address 311RO:hasStreet ?Street
\}
GROUP BY ?Street
ORDER BY DESC(?Total)
 LIMIT 1
```

The object property hasAddress connects the class ServiceRequest to the class Address. The class Address represents the street in terms of the data property hasStreet. This query counts total number of service requests for each street and reports the one which has the highest number of requests.

QC-7: *What are the most reported service subjects?*

The following query can be used:

```
SELECT  ?Subject  (COUNT(?Subject)
    AS ?Total)

WHERE\{
   ?ServiceRequest 311RO:has311Type ?311Type.
   ?311Type 311RO:has311Subject
   ?Subject Instance.
   ?SubjectInstance rdf:type ?Subject.
\}
GROUP BY ?Subject
ORDER BY (?Total)
```

The object property has311Subject represents what a given service request is about (e.g., garbage container). This query can be used by mayor's office for gaining insights about the city problems and for understanding the major topics in the requests submit-ted by crowd.

QC-8: *Which city is receiving the most number of service requests from its citizen?*

This competency questions compares cities with regard to the total number of service requests received by their 311 agencies. The following query can be used:

```
SELECT  ?City  (COUNT (?ServiceRequest)
      AS ? Total)
WHERE\{
    ?ServiceRequest 311RO:hasCity?City.
\}
GROUP BY ?City
ORDER BY DESC(?Total)
 LIMIT 1
```

The class ServiceRequest is connected to the class City via the object property hasCity. This query counts total number of ServiceRequest instances for each city, sorts the results and report the first city.

QC-9: *Which cities have received more than 1000 reports categorized as illegal issues?*

The following query computes the answer to this competency question:

```
SELECT  ?City  (COUNT (?ServiceRequest)
      AS ? Total)
WHERE\{
    ?ServiceRequest 311RO:hasCity ?City.
    ?ServiceRequest 311RO:has311Type
   ?311Type.
   ?311Type 311RO:has311MessageCategory
   ?Category.
   ?Category rdf:type 311RO:IllegalIssue
\}
GROUP BY ?City
HAVING (COUNT(?ServiceRequest) > 1000)
```

In our ontology, the class Type is connected to the class MessageCategory via the object property has311MessageCategory. The class MessageCategory represents the type of message that is being delivered by the service request. It has various subclasses such as Complain, Compliment, Report, etc. The class Complain has IllegalIssue as a subclass. To find the answer for this competency question, for each city, all the instances of ServiceRequest whose category is Complain is first retrieved so that those that are illegal issue are then retrieved and counted.. The last step is to exclude those cities which have less than 1000 service requests of the specified category.

QC-10: *What are the top three cities with most number of reports of the subject "dead animals"?*

In this competency question, cities should be compared with regarding to the number of submitted service requests about dead animals. The answer to this question results from following query:

```
SELECT  ?City  (COUNT (?ServiceRequest)
      AS ? Total)
WHERE\{
    ?ServiceRequest 311RO:hasCity ?City.
    ?ServiceRequest 311RO:has311Type
    ? 311Type.
    ?311Type 311RO:has311Subject ? Subject.
    ?Subject a 311RO:DeadAnimal
\}
GROUP BY ?City
ORDER BY DESC(?Total)
LIMIT 3
```

In the ontology, the class Type is connected to the class Subject via the object property has311Subject. The class Subject has various subclasses one of which is Pests. The class Pests has two subclasses, namely Animal and Insects. The DeadAnimal class is a subclass of the class Animal. In this question we look for ServiceRequest instances that are connected to the DeadAnimal class via the object property has311Subject.

6.4 Mapping Datasets to the 311 Reference Ontology

In this section, we illustrate the mapping of existing datasets into the 311 reference ontology. As previously explained in Section 2, Tables 1 and 2 present examples of a service requests in the Toronto and San Francisco datasets. Figures 11 and 12 show how these examples are represented in the 311 Reference Ontology. (It should be mentioned

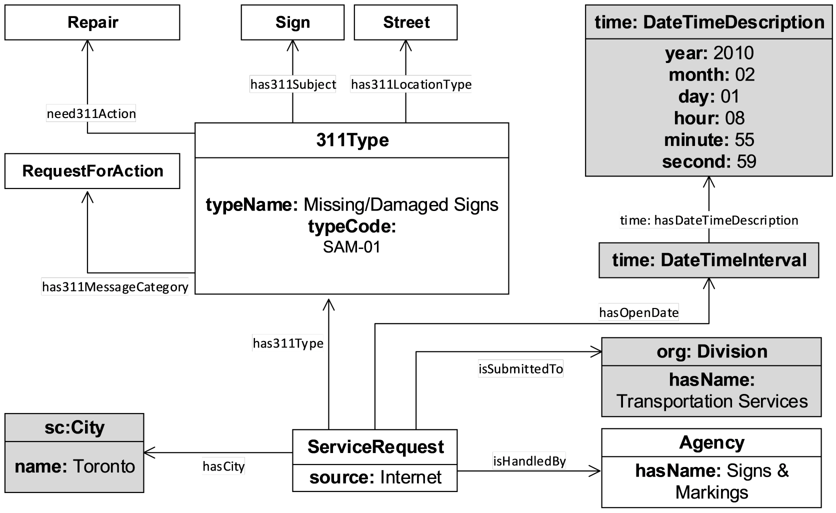

Figure 11. Mapping Toronto's Dataset to the 311 Reference Ontology. (Arrows represent object properties between instances of classes. Gray-colored boxes are classes from related ontologies identified by prefixes. White boxes are classes in 311 Reference Ontology

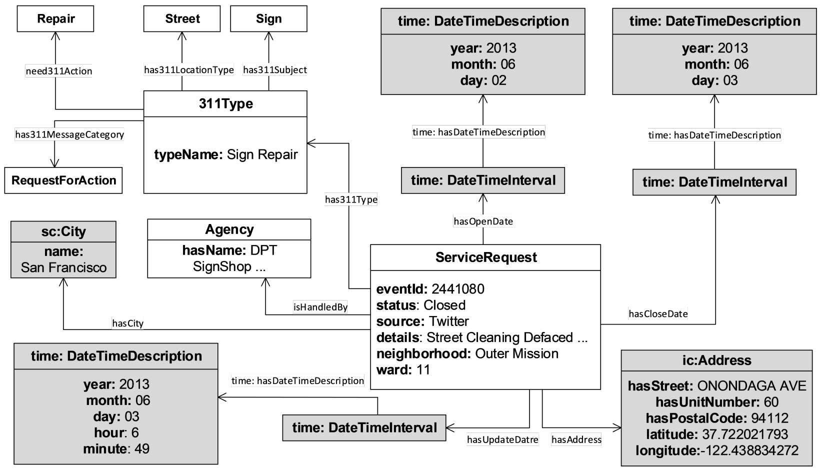

Figure 12. Mapping San Francisco's Dataset to 311 Reference Ontology.

that we also have mapped datasets of the cities New York and Chicago to our ontology. However, the examples are not presented here). Moreover, a mapping of the Toronto dataset has been performed using software we have developed. The input is a tab delimited raw dataset and the output is in RDF/XML format. The result can be found at: http://ontology.eil.utoronto.ca/311/311-Toronto-2010.rdf.

Figure 11 depicts the Toronto example (Table 1) of Missing/Damaged Signs which is allocated to Transportation Services.

Figure 12 depicts the San Francisco example (Table 1) of Sign - Defaced, which is allocated to DPT Sign Shop.

7 Conclusion

This paper confirms the hypothesis that an ontology can be designed to support interoperability among 311 datasets. In order to confirm this hypothesis, two problems had to be addressed. First, we had to perform a semantic analysis to deconstruct the service requests based on the cities studied. The deconstruction identified four concepts intermingled across the datasets:

1. the message category, e.g., complaint, compliment, request;
2. the subject, e.g., graffiti, vehicle, pest;
3. the action to be performed, e.g., remove, repair; and
4. the type of location of the request.

The second problem is the variety of values the cities use for each of these properties. As part of the 311Reference Ontology, we included a wide range of message type, subjects, actions and location types. These values were based on the cities we studied. In addition, we included other ontologies, such as time and organization, to be used as ranges of object properties.

Returning to the New York's derelict vehicle and Chicago's abandoned vehicle, the correct interpretation is that the message category is "request", the subject is "vehicle", and the action is "remove". Whether the vehicle is abandoned or derelict may be immaterial, or can be captured by having "derelict vehicle" and "abandoned vehicle" as subclasses of "vehicle".

To evaluate the ontology, we illustrated that it satisfies the competency questions, and we showed how a city's 311 data is mapped onto it, thereby making it possible to perform aggregate and comparative analyses of multicity 311 data.

The process of creating this ontology clearly illustrates the need for common vocabularies and ontologies for 311 and other city data. The challenge now is to persuade cities to adopt this standard. One path for achieving adoption is for municipal software providers to incorporate the 311 Reference Ontology into their software. This is beginning to happen. LocaliData, a company located in Madrid (http://www.localidata.com/), has incorporated the 311 Reference Ontology into their Open Data Portal for the publishing of city data.

References

1 Nam T and Pardo T A, 2012, Transforming City Government: A Case Study of Philly311. *Proceedings of ICEGOV2012*, Albany NY.
https://doi.org/10.1145/2463728.2463787

2 Fox M S, 2013, City Data: Big, Open and Linked. *Municipal Interfaces*, November: 19-25.

3 Zha Y and Veloso M, 2014, Profiling and Prediction of Non-Emergency Calls in New York City. *Proceedings of the Workshop on Semantic Cities: Beyond Open Data to Models, Standards and Reasoning*, AAAI.

4 O.Brien D T, 2015, Custodians and Custodianship in Urban Neighborhoods: A Methodology Using Reports of Public Issues Receive a City's 311 Hotline. *Environment and Behavior*, vol. 47(3): 304-327.
https://doi.org/10.1177/0013916513499585

5 Wiseman R, 2015, Canadian Governments Reference Models. *Service Systems Science*, Springer: 109-128.
https://doi.org/10.1007/978-4-431-54267-4_7

6 BIS, 2014, Smart City Concept Model, PAS 182: 2014, British Standards Institute.

7 Uceda-Sosa R, Srivastava B and Schloss R, 2011, Building a Highly Consumable Semantic Model for Smarter Cities. *Pro-*

8 Gruber T, 1993, Towards Principles for the Design of Ontologies used for Knowledge Sharing. *Proceedings of the International Workshop on Formal Ontology*, Padova, Italy.

9 Campbell A E and Schapiro S C, 1995, Ontologic Mediation: An Overview.*Proceedings of the IJCAI Work-shop on Basic Ontological Issues in Knowledge Sharing*, AAAI Press, Menlo Park CA, USA.

10 Jakulin A and Mladenic D, 2005, Ontology Grounding. *Proceedings of the Conference on Data Mining and Data Warehouses*, Ljubljana, Slovenia.

11 . Fox M S, Barbuceanu M, and Gruninger M, 1996, An organisation ontology for enterprise modeling: prelimi-nary concepts for linking structure and behavior. *Computers in Industry*, vol. 29(1): 123-134.
https://doi.org/10.1016/0166-3615(95)00079-8

12 Fox M S, 1992, The TOVE project towards a commonsense model of the enterprise. *Industrial and En-gineering Applications of Artificial Intelligence and Expert Systems*. Springer Berlin Heidelberg: 25-34.
ceedings of the Workshop on AI for an Intelligent Planet, Association for Computing Machinery.
https://doi.org/10.1145/2018316.2018319

13 Keita A, Laurini R, Roussey C and Zimmerman M, 2004, Towards an ontology for urban planning: The towntology project. *CD-ROM Proceedings of the 24th UDMS Symposium*, Chioggia, vol.12(1) .

14 Grger G, Kolbe T and Czerwinski A, 2007, Candidate OpenGIS/r CityGML Implementation Specification (City Geography Markup Language).*Open Geospatial Consortium Inc, OGC*.

15 Gruninger M and Fox M S, 1995, Methodology for the design and evaluation of ontologies.*In Proceedings of the Workshop on Basic Ontological Issues in Knowledge Sharing*, International Joint Conference on Artificial Intelligence.

16 Flowers M, 2013, Beyond Open Data: The Data-Driven City, in *Beyond Transparency: Open Data and the Future of Civic Innovation*, Chapter 15: 185-198.

17 PrudHommeaux E and Seaborne A, 2008, SPARQL query language for RDF, *W3C recommendation*, 15.