Computers, Environment and Urban Systems 54 (2015) 266–279

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

journal homepage: www.elsevier.com/locate/ceus

The role of ontologies in publishing and analyzing city indicators

Mark S. Fox

Department of Mechanical and Industrial Engineering, University of Toronto, 5 King's College Road, Toronto, ON M5S 3G8, Canada

ARTICLE INFO

Article history:
Received 25 June 2014
Received in revised form 23 September 2015
Accepted 23 September 2015
Available online xxxx

Keywords:
City indicator
Urban metrics
Ontology
Semantic Web
ISO 37120

ABSTRACT

This paper addresses the problem of how city indicators and their supporting data are to be published on the Semantic Web so that automated analysis can be performed. With the publishing of ISO 37120, cities have a standard set of indicators that can be used to compare their performance. The problem is that no standards exist for publishing indicators on the Semantic Web. In this paper we introduce the Global City Indicator Ontology (GCIO). The GCIO addresses five issues: 1) how is meta-data associated with a single indicator value represented? 2) how are indicator definitions represented? 3) how is the data used to derive an indicator value represented? 4) how is indicator theme specific knowledge represented, and 5) how is city specific knowledge represented? The GCIO has been implemented and validated using the City of Toronto ISO 37120 indicators reported for 2013. Research continues in developing ontologies specifically for each of the indicator themes such as: Education, Shelter, Health, Transportation and Innovation.

1. Introduction

Open City Data is part of the broader Open Government movement where the belief is that making data publicly available will lead to more effective public oversight, where waste and inefficiencies can be detected, crowd-based solutions suggested, and citizens and corporations harnessed to implement solutions. Open City Data is published in many forms, with spreadsheets (e.g., xls, csv) and XML being the most pervasive. However the content, including the data model and the values, is idiosyncratic to the city and even to departments within a city. There does not exist global standards for the representation and publishing of Open City Data. Concurrently, there has been a keen interest in making cities "smarter," inevitably leading to comparisons among cities based upon data whose models and content are idiosyncratic, if available at all. Thereby leading to questionable comparisons.

ISO 37120, "Sustainable development of communities — Indicators for city services and quality of life", defines 100 indicators across 17 themes, including Education, Health, Shelter, Safety, and Transportation. There are two important contributions of this standard. First, it has selected a set of indicators from thousands that exist to measure city performance. Second, it provides more precise definitions (in English) of these indicators than previously available. The hope is that by adopting this standard, cities will be able to compare their performance based on metrics that are consistently interpreted and applied. Cities, such as Toronto (2014), have begun to publish the values for all 100 indicators on their web sites.

The development of ISO 37120 traces its roots back to an analysis of existing city indicators that identified the following aspects a good "indicator must possess to be accurate, timely and relevant for policy purposes" (Hoornweg, Nunez, Freire, Palugyai, & Villaveces, 2007, p. 13):

- Objective: clear, well defined, precise and unambiguous, simple to understand.
- Relevant: directly related to the objectives.
- Measurable and replicable: easily quantifiable, systematically observable.
- Auditable: valid, subject to third-party verification, quality controlled data (legitimacy across users).
- Statistically representative at the city level.
- Comparable/Standardized longitudinally (over time) and transversally (across cities).
- Flexible: can accommodate continuous improvements to what is measured and how. Have a formal mechanism for all cities and interested parties to comment on.
- Potentially Predictive: extrapolation over time and to other cities that share common environments.
- Effective: tool in decision making as well as in the planning for and management of the local system.
- Economical: easy to obtain/inexpensive to collect. Use of existing data.
- Interrelated: indicators should be constructed in an interconnected fashion (social, environmental and economics).
- Consistent and sustainable over time: frequently presented and independent of external capacity and funding support.

E-mail address: msf@eil.utoronto.ca.

http://dx.doi.org/10.1016/j.compenvurbsys.2015.09.009
0198-9715/

We believe that "objective" and "relevant" aspects have been reasonably satisfied by ISO 37120, but aspects such as "auditable", "statistically representative", "comparable", "effective" and "consistent and sustainable over time" have not. The reason for this is simple, only the values of the indicators are reported and not the supporting data from which they were derived. We can only know that a city's indicator values may differ over time or in comparison to other cities, but we cannot know why. Consider the indicator "Primary Student Teacher Ratio" from the education theme of ISO 37120. Summarizing the definition, the teaching program must be "primary" and not include "pre-school", the schools must be "public", and the "teachers" must not be "administrators". In order to determine whether the definition is being satisfied and the root cause of differences in indicator values, each city must openly publish its supporting data in order to allow one to drill down to the level of detail necessary, perhaps to the individual school level.

What impedes a city from publishing an indicator's supporting data? Besides internal decisions not to publish, there does not exist the processes and information systems to routinely capture, store and publish many of the data. Often, the supporting data is spread across multiple departments and external organizations with few information systems linking them. Even if they were able to capture and publish the supporting data, what format would they use? If it is a spreadsheet or PDF file then it may be limited to review by people. Or should it be a computer format like JSON or RDF so that it is possible for web applications to read it? And if a computer format is used, what would the data model be? What standards exist, i.e., vocabularies or ontologies, for representing the data?

The ultimate goal of this research, called the PolisGnosis Project (Fox, 2015), is to automate the performance of longitudinal analysis (i.e., how and why a city's indicators change over time) and transversal analysis (i.e., how and why cities differ from each other at the same time), in order to discover the root causes of differences. However before we can focus on analysis, we have to solve the indicator representation problem. The representation problem can be divided into five parts:

1. How do we represent the metadata associated with a published indicator value? For example, its units, scale, when it was created, who created it, what process was used to create it, the degree of certainty in the value, and the degree to which we trust the organization that created it and/or the process they used?
2. How do we represent the (ISO 37120) definition of an indicator? In order for the analysis of indicators to be automated, the PolisGnosis system must be able to read and understand the definition of each indicator, which may change over time.
3. How do we represent the data used to derive an indicator value? An indicator is the apex of a tree of supporting data that is aggregated across place, time, organizations, etc. How is this represented?
4. How do we represent ISO 37120 theme specific knowledge? Each theme, such as Education, Health, Shelter, etc., has a core set of knowledge, that has to be represented in both the definition of an indicator and in publishing an instance of an indicator and its supporting data.
5. How do we represent a city's theme specific knowledge? Each city may define concepts such as "primary school", "grades", "teachers", etc. differently. Differences in indicator values may be due to differences in the interpretation of these terms between cities.

This paper describes the Global City Indicator Ontology (GCIO). The GCIO provides a set of classes and properties that underlie the representation of all indicators defined in ISO 37120. The GCIO is defined using Description Logic and implemented in OWL (Hitzler et al., 2012) using Protegé (Noy et al., 2003). The ontology is available on the web - see Appendix A for details. (For the reader who is unfamiliar with ontologies, linked data and the Semantic Web, Fox (2013a) provides an introduction to these concepts in the context of cities.)

In the following we review previous approaches to the representation of indicators. We then review the definition of the ISO 37120 6.4 Primary Student Teacher Ratio indicator (PSTR), which will be used as our example throughout the paper. Next we present the components of the GCIO that address how we represent an indicator value's meta information. We then present the remainder of the GCIO that focuses on representing the definition of an indicator, and instances of indicators and their supporting data. We then briefly explore the representation of theme and city specific knowledge needed to represent PSTR. We end with an evaluation of our approach, and conclusion.

2. Background

The rapid growth of Asian cities led the Asian Development Bank to launch a city indicator project in 1999. The objectives of the project were to "to establish a policy-oriented urban indicators database for research, policy formulation, monitoring of the development impact of interventions in the urban sector, comparison of performance between cities, and improving the efficiency of urban service delivery." (Westfall and de Villa, 2001 p. x). The result of the project provides the motivation and detailed definition of indicators. It also anticipates an important role for the World Wide Web in the representation and interconnection of indicators and their supporting data.

The Organization for Economic Co-operation and Development[1] "provides a forum in which governments can work together to share experiences and seek solutions to common problems." At the core of their work is a large number of indicators spanning topics such as health, education, environment and trade. The indicators are documented in detail, in English, and the results are published as spreadsheets. Definitions of the indicators using Semantic Web ontologies are not available. On the other hand, some OECD datasets have been the object of research in how to automatically transform statistical databases into linked data (Capadisli, Auer, & Ngonga Ngomo, 2013; Hausenblas, Halb, Raimond, Feigenbaum, & Ayers, 2009).

More recently, the Global Cities Institute at the University of Toronto[2] has led the development of a set of indicators for city services and quality of life, which has been published as ISO 37120 in 2014. The standard provides more precise definitions of 100 indicators covering 17 themes such as education and transportation. Following is the definition of 6.4 Primary Student Teacher Ratio that we will use throughout the paper:

> "The student/teacher ratio shall be expressed as the number of enrolled primary school students (numerator) divided by the number of full-time equivalent primary school classroom teachers (denominator). The result shall be expressed as the number of students per teacher. Private educational facilities shall not be included in the student/teacher ratio. One part-time student enrolment shall be counted as one full-time enrolment; in other words a student who attends school for half a day should be counted as a full-time enrolment. If a city reports full-time equivalent (FTE) enrolment (where two half day students equal one full student enrolment), this shall be noted. The number of classroom teachers and other instructional staff (e.g. teachers' aides, guidance counselors) shall not include administrators or other non-teaching staff. Kindergarten or preschool teachers and staff shall not be included. The number of teachers shall be counted in fifth time increments, for example, a teacher working one day per week should be counted as 0.2 teachers, and a teacher working three days per week should be counted as 0.6 teachers."

The field of Ontology Engineering has started to focus on the representation of city indicators. As part of IBM's Smart Cities initiative, a comprehensive Ontology (SCRIBE) for representing various types of

[1] http://www.oecd.org.

[2] http://www.globalcitiesinstitute.org.

city knowledge, including city organization and services, flow of events and messages, and key performance indicators has been developed (Uceda-Sosa, Srivastava, & Schloss, 2012). OWL definitions of the classes and properties are provided. However the ability to represent the definitions of indicators and precise semantics of the supporting data used to derive an indicator is missing. Axiomatization is limited and so its use of foundational ontologies.

Ghahremanloo, Thom, and Magee (2012) examined the indicators employed by the cities of Melbourne and Vancouver in order to identify a taxonomy of indicators, leaving the definitions of the indicators as strings. The focus of the work is primarily on methodology. The capability to represent the definitions of indicators and precise semantics of the supporting data used to derive an indicator is not addressed.

In summary, city indicators represent a new area for the application of ontologies. None of the existing efforts provide an ontology for representing indicator definitions nor the meta-data and supporting data from which they are derived.

3. PolisGnosis Project

The PolisGnosis Project addresses the gaps identified in the previous section. The goal of the PolisGnosis Project is to automate the longitudinal and transversal analysis of city indicators based on the ISO 37120 standard. Its approach is to develop the PolisGnosis Engine (Fig. 1) that takes as input:

- All of the information and knowledge with respect to an indicator,
- A set of consistency axioms, and
- A set of diagnosis axioms,

and applies the axioms to determine why indicators change.

The blue boxes focus on the representation of indicators and their supporting data. The goal being to transform existing data formats into a standard ontology that can be published on the Semantic Web. The green boxes focus on consistency analysis of published indicator data to verify that it is consistent with the definitions of the indicators, and that indicators for a city over time or comparing two cities are consistent. The orange boxes focus on the theories to diagnose the root causes of longitudinal and transversal differences.

The general method we used to create the GCIO is the methodology defined in Grüninger and Fox (1995). The process begins by defining a set of usage scenarios. The scenarios in this case are based on the diagnosis of city indicators. Based on the scenarios, we identify a set of competency questions that the ontology must answer. These are the requirements for what is to be represented and the deductions to be performed. Next, we search for existing ontologies that contain classes and properties that satisfy portions of the competency questions. The relevant classes and properties are selected for inclusion in the GCIO. Where necessary, the selected classes and properties are extended, and new classes and properties are created to satisfy the competency questions that are not covered. Next we specify the semantics of the terminology by constructing a set of axioms that define and/or constrain the interpretation of the classes and properties. The axioms are important as they precisely define the indicators and their supporting data, and can determine whether the data that underlies the indicators are consistent (e.g., the time periods during which the student and teacher populations are the same). More detail can be found in Fox (2013b).

4. Representing indicator values

Over the last decade, funding agencies have increasingly required the publishing of datasets that were used or generated by the funded project. These datasets are to survive the project, be publically available and contain additional (meta) information that describes who, how and why the dataset was created. Concurrently, there has emerged "data journals" that publish "data papers" that describe publically available datasets (Candela, Castelli, Manghi, & Tani, 2015). We mention this because it represents a growing need in the scientific community to not

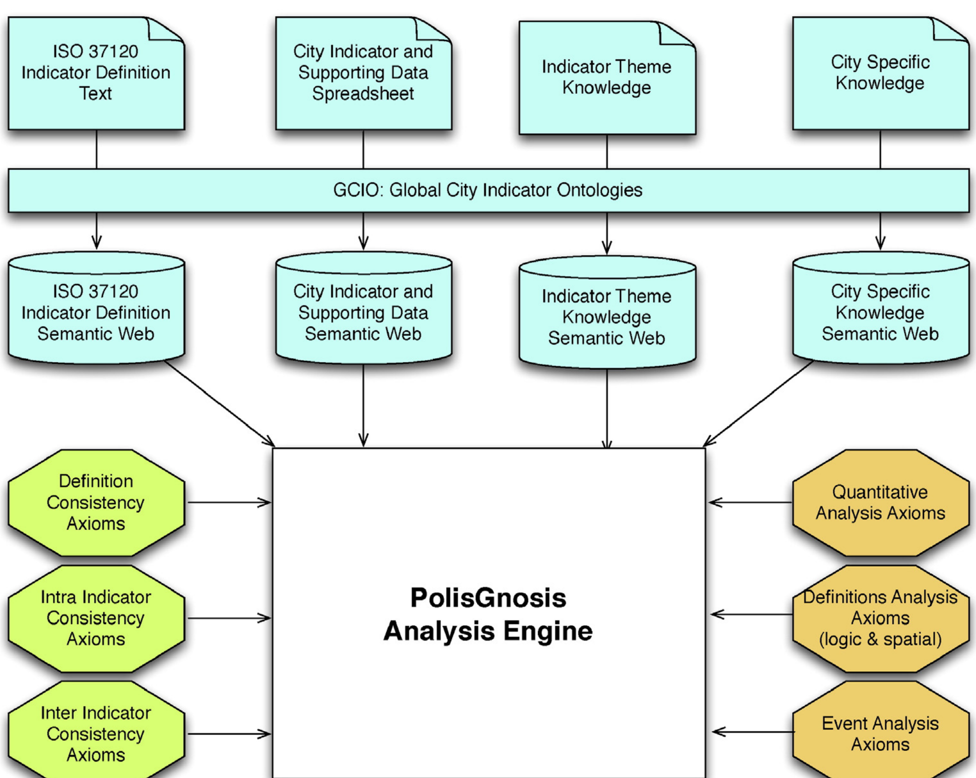

Fig. 1. PolisGnosis Architecture. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

only understand the data upon which scientific results have been based but to enable the reuse of that same data by other scientists.

A set of indicators published by a city is also a dataset for which meta information should be provided. However what distinguishes an indicator dataset from many other scientific datasets is that each datum in the indicator dataset has its own unique meta information covering who, how and why the datum was created. However the sources and measurement processes for an indicator are buried in datasets and documents that are mostly inaccessible. In the end, we are left with indicator values that we cannot verify; we have to rely on the good will of the people who reported to the data to adhere to the definitions. This missing meta-information amounts to an "unwritten narrative" which can be found in the minds of the people who created the indicator values and is verbally passed on to others.

Why is this "unwritten narrative" important? If we want to compare the PSTR of two cities, the simplest consistency test is to see they have the same units. Without this measurement meta information, the test cannot be performed. Without provenance meta information, we cannot verify that the process used by one city is consistent with another; a difference in process can introduce significant variation.

The goal of this section is to describe the portion of the GCIO that allows for the representation of this "unwritten narrative". This portion of the GCIO spans:

- Placenames: unique identifiers for cities,
- Measurement: quantities and units of indicators,
- Provenance: how an indicator was derived and by whom,
- Time: when an indicator is valid, or when it was produced,
- Validity: the degree to which an indicator is believed to be correct, and
- Trust: the degree to which the individual or organization is trusted to produce an indicator correctly.

We do not provide the description logic definitions of the ontology (see Appendix A for links to the definitions) but summarize the ontology's classes and properties using a graph representation.

4.1. Placenames

A requirement of the ontology is the ability to identify the geographic area over which the indicator has been calculated. That is, to associate a "placename" with a geographic area. Such placenames could conceivably be applied to areas larger than a city, such as a region, state or country, or smaller than a city, such as a neighborhood or postal code. For example, a reference to Toronto should cover the city of Toronto but a reference to the Greater Toronto Area should cover the larger area encompassing neighboring cities. However it must be clear which each refers to. A second requirement is that when two indicators are supposed to be computed over the same geographic area, they are in fact the same area. This means that an area has to have a unique identifier.

There are a number of ontologies that represent geographic and place information. Schema.org[3] provides classes of placenames such as 'sc:City', 'sc:Country', and 'sc:State'. It also provides classes for 'sc:GeoCoordinates' (i.e., elevation, latitude, and longitude) and 'sc:GeoShape' denoted by a polygon or circle. The Linkedgeodata.org ontology[4] extends what can have a placename by providing classes for 'gd:neighborhood', 'gd:building', 'gd:bridge', 'gd:hospital', 'gd:airport', 'gd:prison', etc. The GeoNames project (www.geonames.org) provides over ten million placenames spanning the world. It provides an International Resource Identifier (IRI) for every placename so that they can be uniquely referred to. The GeoNames' placenames are instantiations of the Geonames Ontology[5] that integrates a number of ontologies, including Schema.org and Linkedgeodata.org, to provide a broad set of classes that span almost every conceivable type of place. For example, the unique IRI for the city of Toronto is: http://www.geonames.org/6167865. It is asserted to have a 'geo:parentCountry' of 'geo:6251999' which is the unique IRI for Canada.

4.2. Measurement

A city indicator is a measure of some property of a city. At the core of an indicator lies a number. The question is what does that number represent? Measurement ontologies provide the basic concepts that underlie numbers. They divide measurement into a 'Quantity' such as length (the what) and a 'unit of measure' such as meters (the how). A 'unit of measure' has a scale classified as nominal, ordinal, interval or ratio, and whether the number is the composition of dimensions such as velocity being composed of speed and direction, and whether it has a starting point such as absolute zero on the Kelvin scale.

In the case of the PSTR, the purpose of a measurement ontology is to provide the underlying semantics of the number. The importance of grounding an indicator in a measurement ontology is to assure that the indicator values are comparable, not that they are measuring the same thing (which is dealt with later), but the actual measures are of the same type, e.g., ratio of student and teacher population counts, or that the counts of the student and teacher populations are of the same scale (i.e., thousands vs millions).

Upper level ontologies such as SUMO (Niles & Pease, 2001) and CYC (Matuszek et al., 2006) provide classes for representing quantities, but the OM ontology[6] (Rijgersberg, Wigham, & Top, 2011) provides a more rigorous ontology based on measurement theory. In addition to covering ratio scales, it covers nominal, ordinal and interval. QUDT[7] is an alternative to OM. We chose OM over QUDT for the reasons expressed in Rijgersberg, van Assem, and Top (2013). The core classes of OM include:

- 'om:Quantity': What is being measured, such as a length or diameter.
- 'om:Unit_of_measure': It is the units and type of the quantity, such as a meter or yard.
- 'om:Measure': It combines the number with both the unit of measure and what is being measured (i.e., Quantity).

Fig. 2 depicts 6.4 Primary Student Teacher Ratio as a subclass of 'om:Quantity' with a value that is a subclass of 'om:Measure' and a unit of measure that is a subclass of 'om:Unit_of_measure'. Every indicator value would be represented as instances of these three classes.

4.3. Provenance

An important aspect of an indicator is its provenance, namely where did it come from and how was it derived. Much of the research into provenance has grown out of workflow management where the focus has been the evolution of a document as it proceeds through a sequence of edits, perhaps by different people and/or systems. Tracking the various versions created, who did what and when has been the primary concern. This research has culminated in the proposed Semantic Web ontology: PROV[8] (Belhajjame et al., 2012), which is based on the work

[3] The Schema.org ontology is available at: http://schema.org/. We will use the prefix "sc:" to identify classes and properties from the ontology.

[4] The Linkedgeodata.org Ontology is available at: http://www.linkedgeodata.org/ontology/. We will use the prefix "gd:" to identify classes and properties from the ontology.

[5] The Geonames Ontology is available at: http://www.geonames.org/ontology/ontology_v3.1.rdf#. We will use the prefix "geo:" to identify classes and properties from the ontology.

[6] The OM ontology can be found at: http://www.wurvoc.org/vocabularies/om-1.8/. We will use the prefix "om:" to identify classes and properties from the ontology. Definitions and examples are taken directly from the ontology where quoted.

[7] http://www.qudt.org.

[8] The PROV Ontology can be found at: http://www.w3.org/ns/prov#. We will use the prefix "pr:" to identify classes and properties from the ontology.

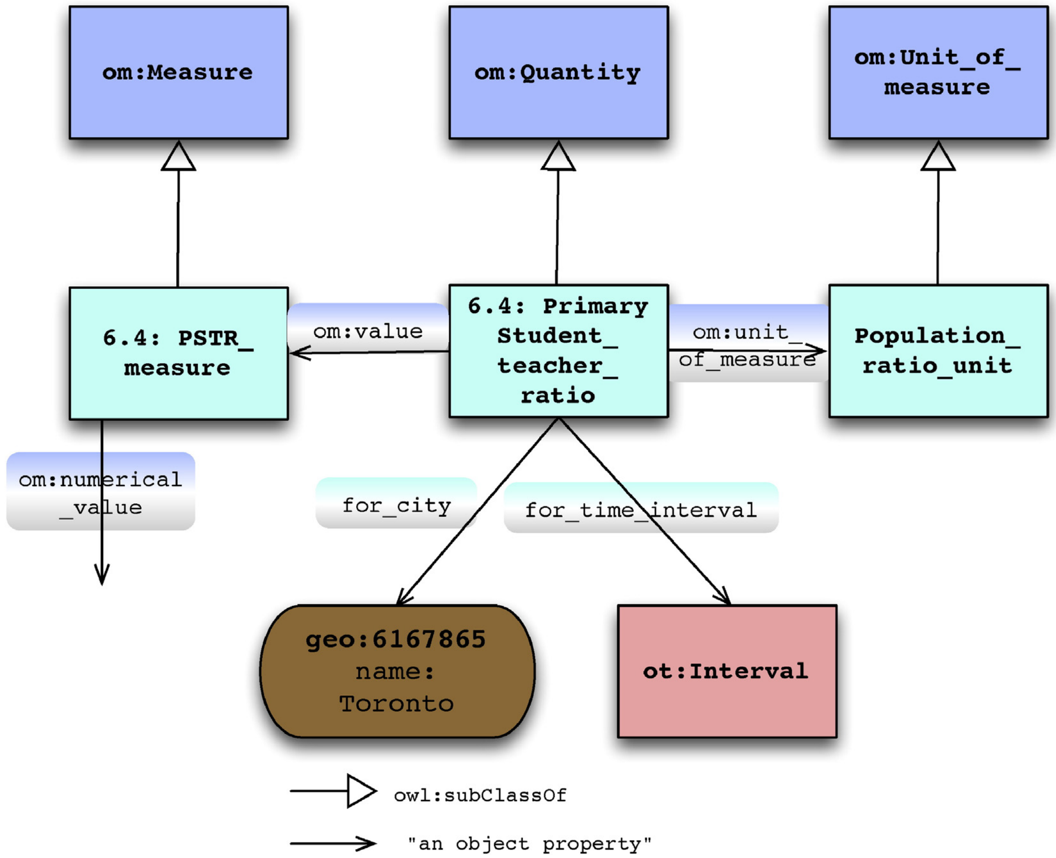

Fig. 2. Measurement ontology. Depicts the 6.4 Primary Student Teacher Ratio class as being linked to the geoname of a city by the 'for_city' property. Note that we have not explicitly defined any spatial ontologies as they are referred to within the geonames ontology.

of Hartig and Zhao (2010) and Moreau et al. (2010). We have chosen to use PROV as it is an emerging standard.

At the heart of the PROV ontology are three classes (Fig. 3):

- 'pr:Entity': represents any artifact for which we want to specify its provenance. In our case it would be an indicator or the data from which the indicator was directly or indirectly derived.
- 'pr:Activity': the action (or sequence of actions) that creates or transforms an entity. In our case it may be a computation performed over some data set such as census data.
- 'pr:Agent': the person, organization, or system that performs or plays some role in the activity that transforms an entity. In our case it may be a software application that mines a data set or a person who reviews a data set.

Along with these classes are defined a set of properties that define the causal relationship among entities and activities, including:

- 'pr:wasGeneratedBy': It links an 'pr:Entity' (domain) to a 'pr:Activity' (range), identifying the activity that generated the entity.
- 'pr:used': It links an 'pr:Activity' (domain) to an 'pr:Entity' (range), identifying the entities used by an activity to produce a new entity.
- 'pr:wasDerivedFrom': Links two 'pr:Entity's' where domain entity was derived from the range entity (without indicating the method of derivation).
- 'pr:generatedAtTime': It links a 'pr:Entity' (domain) to a 'pr:time' (range), identifying the time the entity was generated.

Fig. 3 depicts how provenance information is associated with a 'om:Measure' by making it a subclass of 'pr:Entity'. An indicator value, represented as an instance of a measure, inherits the 'pr:entity' properties defined in the PROV ontology.

4.4. Time

Fundamental to the concept of provenance is the time at which measurements are taken, computed or derived. Questions may arise regarding the temporal relationship among indicators and among their supporting data. Not just at what time something occurred, but whether something occurred before, after or during some external event. For example, was "Total Employment" of New Orleans determined before or after Hurricane Katrina? Or did Katrina take place during the interval that the indicator was determined? To answer these questions, we need a much richer notion of time that supports reasoning about time points, time intervals and the relationships among them. Many time ontologies have been developed. We have chosen OWL-Time[9] for its simplicity and ability to represent time as a point or interval. OWL-Time is based on the work of Allen and Ferguson (1997) and described in Hobbs and Pan (2006). In Fig. 3 we show how 'pr:Entity' is linked to a temporal entity such as a time instance or interval via the 'pr:generatedAtTime' property.

4.5. Validity

The publishing of an indicator carries with it the implication that it is valid. However indicators that are believed to be true at the time they are gathered or computed, may be found over time to be incorrect. Or it may not be clear whether the indicator is true or not, especially if it

[9] The OWL-Time Ontology can be found at: http://www.w3.org/2006/time. We will use the prefix "ot:" to identify classes and properties from the ontology.

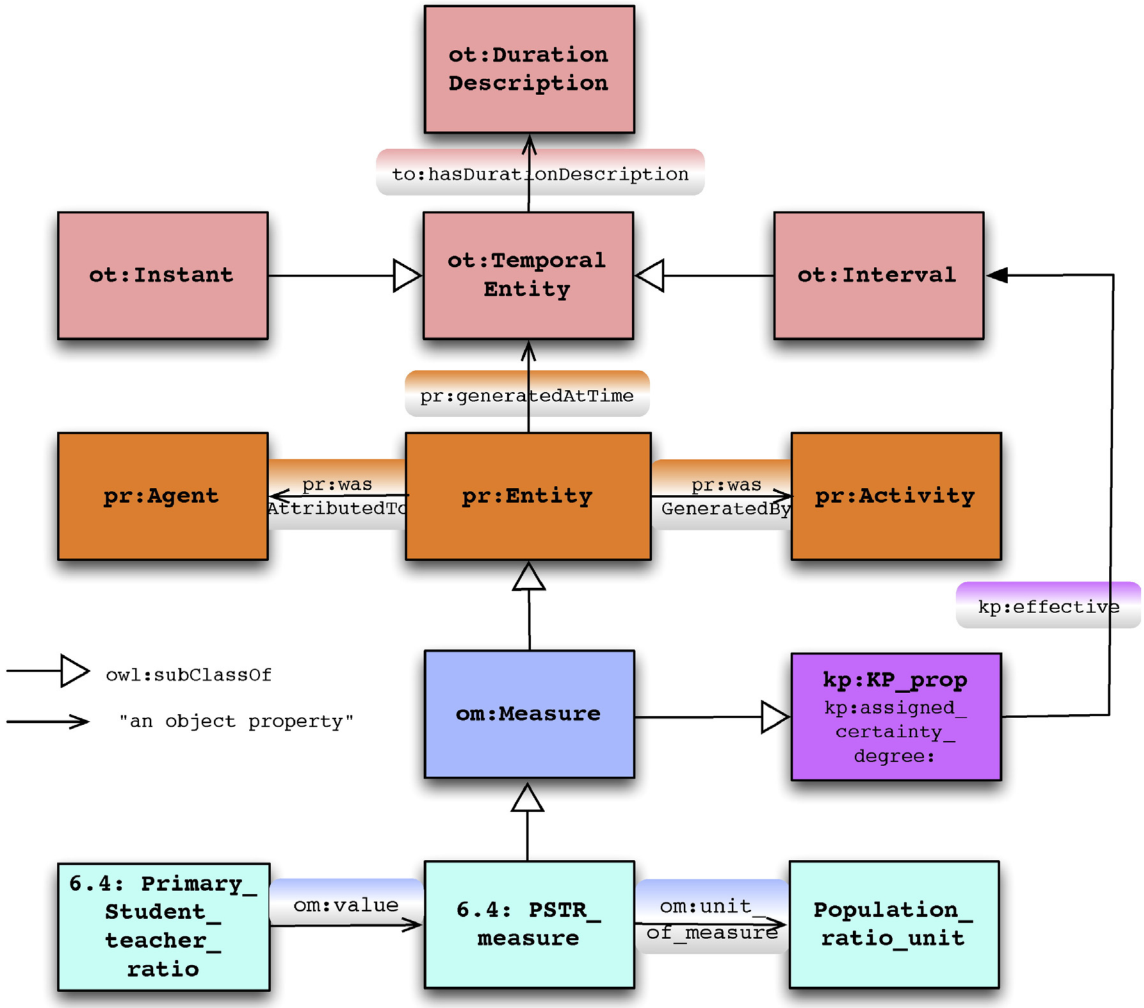

Fig. 3. Measure meta information.

is based on a sampling of a population, but one can assign a degree of validity to the indicator. In addition, where an indicator is derived from other data, and the latter is no longer valid at some point of time, then the former becomes invalid for that same point of time. For example, if the PSTR is derived from a count of students and teachers, and if the student count is valid only within an interval of time such as the year in which it is gathered, then outside of that interval, both the student count and its dependent PSTR's validity are unknown.

Fox and Huang (2005a, 2005b) define the Knowledge Provenance Ontology[10] (KP), for representing the validity (certainty) of a proposition, and axioms for propagating validity within a dependency network Huang and Fox (2004a, 2004b). With roots in the early work of Assumption-Based Truth Maintenance (de Kleer, 1986) and the representation and propagating of certainty or belief (Shortliffe, 1976), each "piece of information" or proposition is viewed as a logical statement that is derived from one or more other statements. All derived statements are linked to their assumptions and if the truth value of their assumptions change, so does the truth value of the derived statement.

In the context of city indicators, KP assigns to an indicator value or supporting data (i.e., Measure) a validity between [0,1] or "unknown." This validity may be dynamic in that it changes over time. An example of the latter is a population count that is representative of a population only at a point of time or for an interval of time. The time interval during which the proposition's validity is known is called the "effective" time interval. Fig. 3 depicts the extension of 'om:Measure' to include validity information by making it a subclass of 'kp:KP_prop'. 'om:Measure' inherits a data property representing the validity of the measure and an object property linking it to a time interval over which it is valid. Building on the PROV ontology, validity of a measure can be derived using the 'pr:Entity' and 'pr:Activity' dependencies.

In theory, this approach to representing the validity of an indicator's value and/or supporting data is very general. However in practice there are problems. How does a city determine their indicator or support data's validity? What does a validity of 0.5 mean? How comparable are cities' validity values?

4.6. Trust

How do we represent the degree of trust we have in the creator of indicator values and the data from which they are derived? This representation of trust differs from validity as trust refers not to the degree of certainty in the indicator but our trust in the agent/organization that produced it. The obvious example is how to represent the trust we have in an organization that has a history of "cooking the numbers." The consequence of not having trust in the producer of data is that the validity one assigns to an indicator will be reduced by this lack of trust.

Some of the earliest formalizations of trust are due to Marsh (1994) and Demolombe (1998) where trust is inter-individual, i.e., trust occurs between two agents, where $agent_1$ has or has not trust in $agent_2$, and

[10] The Knowledge Provenance Ontology can be found at: http://ontology.eil.utoronto.ca/kp.owl. We will use the prefix "kp:" to identify classes and properties from the ontology.

arises out of direct experience or the experience of others whom you may trust. Huang and Fox (2006) and Huang (2008) extend the conceptualization of trust in three directions: 1) trust is context dependent, for example, $agent_1$ may trust $agent_2$ in providing information on topics relevant to their expertise, such as a meteorologist characterizing the climate of a city, but lacks trust in $agent_2$ outside of their field of expertise; 2) certain types of trust can be transitive; and 3) the distinction between: trust in belief, where $agent_1$ believes what $agent_2$ believes, and trust in performance, where $agent_1$ believes that $agent_2$ will perform an activity properly. This third point is important in that we would want to distinguish between trust in the process used to derive an indicator value and trust in the person or organization that asserts the value. It is possible that we can trust the organization but their processes are poor, and vice versa. The Trust ontology[11] defines how the validity of an indicator or data changes by taking the original validity, asserted by the creator ($agent_2$), and modifying it by the degree of trust the "user" ($agent_1$) has in the creator or process. This resultant validity is dependent on $agent_1$ and $agent_2$. The representation of an indicator's trust can be found in Fox (2013b).

Similar to validity, the approach to the representation and computation of trust is very general. However in practice determining the degree of trust a city has in another city's people, or organizations or processes is very subjective. Reducing trust values to ordinals values such as: low, medium, high might be useful.

4.7. Summary

The GCIO provides the basic representation for the "unwritten narrative" of an indicator's value. The representation covers placenames, measurement, provenance, validity and trust. By using the GCIO to represent an indicator value, we can perform basic analyses without having to refer back to the city for additional information that is usually inaccessible.

5. Representing indicator definitions, instances and supporting data

One of Hoornweg et al. (2007) requirements is that indicators be objective and auditable. Automation of the analysis of a city's indicator requires that the analysis system understand the definition of the indicator. There have two ways of accomplishing this. One way is to embed the definition of the indicator into the analysis software code. There are three issues with this approach:

1. Verifiability: the difficulty in verifying that the code faithfully implements the definition of the indicator.
2. Modifiability: when the definition of the indicator changes, the code will have to be modified too, which can be complicated and lead to problems with verifiability.
3. Extensibility: If we want to extend the analysis code to emerging indicators, new code would have to be written.

Consequently, we have pursued a second approach, namely to represent the definition using the GCIO so that an indicator's definition becomes an input to the PolisGnosis system. If a definition changes, then the definition of the indicator in the GCIO is modified, which is much simpler than modifying code. If new indicators are introduced, then the definitions of the new indicators are constructed using our ontology, and are input to PolisGnosis.[12]

In order to represent indicator definitions, we will need a representation that is semantically rich enough to represent the various components of the definition. Let's consider what some of those components are in the Primary Student Teach Ratio indicator:

- The indicator is the ratio of two numbers whose units and scale must be the same (measurement theory: ratio scale).
- The number of students (numerator) and teachers (denominator) are cardinalities of two different sets, each which may be a sample of the complete population (measurement theory, statistics).
- The sets are based on a population defined within a geographic area (geolocation/placename).
- The populations being sampled are determined by a (logical) definition of a student or teacher (description logic).
- A student is defined as a full time student in primary school (description logic).
- Administrative staff are not to be included in the teachers counted (description logic).
- A Primary School has to be a public school that teaches primary grades (description logic).

The indicator "Primary Student Teacher Ratio" is the root of a tree where the constituent definitions branch out below it. The tree is heterogeneous in that its nodes span various types of representations including analytical, statistical, spatial, logical and events.[13]

The GCIO provides a core set of classes and properties for representing the structure of an indicator definition. It can also be used to represent an instance of an indicator and the supporting data used to derive its value. In other words, when a city publishes an indicator value, they can use the portion of GCIO in Section 4 to publish its meta information, and then use the definitional portion of the GCIO to publish the derivation tree comprised of supporting data used to derive the indicator's value.

In the remainder of this section we explore the types of knowledge necessary to represent the definition of Primary Student Teacher Ratio. A more detailed survey and analysis can be found in Fox (2013b).

5.1. Measurement theory revisited

In order to represent a definition, several building blocks need to be put in place. The PSTR is the ratio of Student to Teacher, which is the ratio of the number of students to the number of teachers. Both students and teachers represent sets, i.e., the set of all students within a city (Placename) and the set of teachers within the same city (Placename). We need to represent the cardinality of these sets.

Fig. 4 depicts the new unit of measure classes required to represent the number of students and teachers. We start by defining a unit of measure: 'gci:Cardinality_unit'. Just as the meter is the unit of measure for length, a 'gci:Cardinality_unit' is the unit of measure for the size of a set. The 'gci:Cardinality_unit' is a ratio scale: 'gci:Cardinality_scale', which is a subclass of 'om:Ratio_scale' and is has a zero element (namely zero).

In Fig. 4, we specialize the 'gci:Cardinality_unit' to the class 'gci:Population_cardinality_unit' which is the unit of measure for the cardinality of set defined by a Population (defined in the next section), and associate the symbol "pc" with it. For example, 1100pc represents a population cardinality (or size) of 1100. We can take full advantage of prefix notations available in OM to scale the numbers by defining units of measures: 'gci:kilopc', 'gci:megapc' and 'gci:gigapc' which are multiples of 'gci:Population_cardinality_unit'. 1.1 'gce:kilopc' represents 1100 pc.

[11] The Trust Ontology can be found at: http://ontology.eil.utoronto.ca/trust.owl. We will use the prefix "tr:" to identify classes and properties from the ontology.

[12] It is also possible that there is more than one definition for an indicator: the definition provided in ISO 37120, and the definition used by a city to generate a report. Even though the ISO definitions are much more precise than previously available, there still exists ambiguities in the definition, for example, what are primary grades? Is a school private even if it receives partial funding from the government? We will address this issue in the next section.

[13] Events play an important role in the analysis of indicator data as it may be an event that has a significant impact on an indicator in a particular year. For example, air quality in a given year may be affected by a volcano eruption. But we do not include them in our discussion as they do not form part of an indicator's definition.

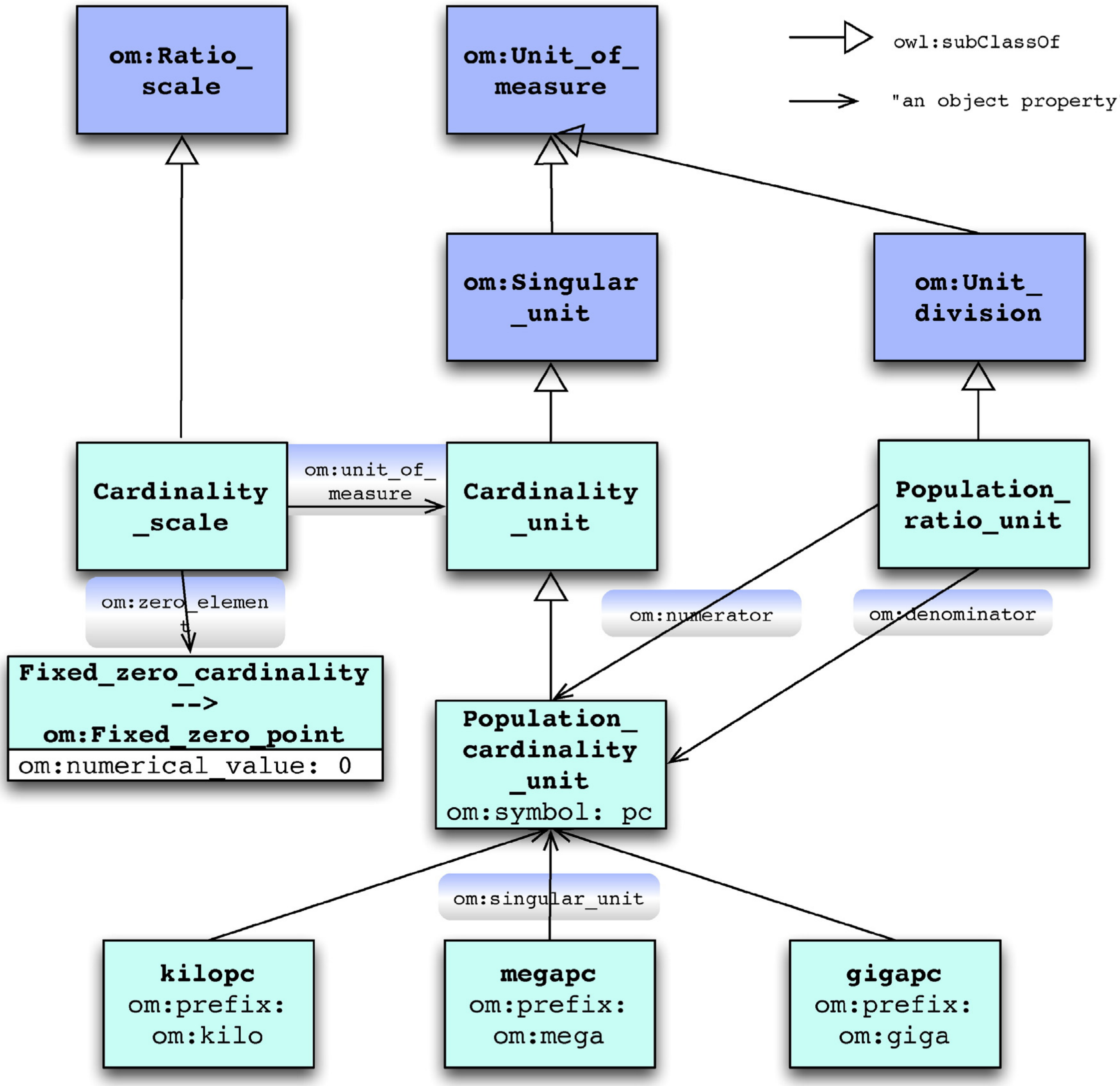

Fig. 4. Population units of measure.

With the above defined, we can now introduce the unit of measure for measuring a population ratio such as PSTR. 'gci:Population_ratio_unit' is defined to be a subclass of 'om:Unit_division'. It has two properties:

- 'om:numerator' whose range is restricted to being a 'gci:Population_cardinality_unit'.
- 'om:denominator' whose range is restricted to being a 'gci:Population_cardinality_unit'.

In other words, a population ratio is derived from two population cardinalities.

The above, provides the unit of measures for populations (pc) and population ratios (pc/pc) (the how). We now have to define what we are measuring which is referred to as a 'Quantity' in the OM ontology. First, we need to define the 'om:Quantity' for the size of the teacher and student populations from which the PSTR is derived. In Fig. 5 we introduce 'gci:Population_size' as a subclass of 'om:Quantity'. Its 'om:unit_of_measure' is the 'gci:Population_cardinality_unit'.

We now have the requisite infrastructure to define an indicator (Fig. 6). First we define the class of 'gci:Global_city_indicator', as a subclass of 'om:Quantity' (not shown). All indicators will be a subclass of 'gci:Global_city_indicator'. 'gci:Education_GCI' is introduced as a subclass of 'gci:Global_city_indicator' with a property that it is a 'gci:for_city_service' 'gci:Education_city_service'. Simply, this denotes that this indicator is for the education city service.

PSTR (shown as 6.4) is defined as a subclass of 'Education_GCI'. It has the following properties:

- 'om:unit_of_measure', whose range is the 'gci:Population_ratio_unit'. This signifies that the quantity is a ratio with a numerator and denominator that are restricted to being 'gci:Population_cardinality_units'.
- 'gci:numerator' & 'gci:denominator', whose ranges are 'gci:Student_population_size' and 'gci:Teacher_population_size' classes respectively, which satisfy the 'gci:Population_ratio_unit' numerator and denominator constraints.
- 'gci:for_city', whose range is a 'geo:Feature' that uniquely identifies the city for which this is an indicator.

When a city publishes their number for indicator 6.4 (PSTR), they create an instance of 6.4 and link it to the object being measured (i.e., City) with the actual measurement being an instance of a Measure. The instance of Measure then contains the measurement's numeric value and a link the unit of measure (Fig. 2).

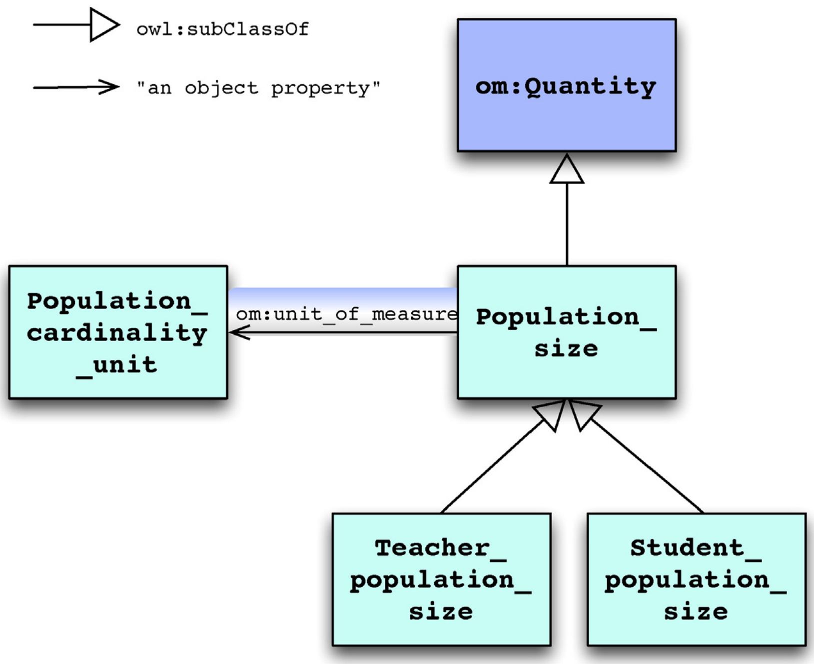

Fig. 5. Population size quantity.

5.2. Populations and statistics

The PSTR indicator is based on a measure of the number of students and teachers within a population designated by a city (placename). One can view both as a statistical measurement in the sense that there is a population that we want to perform a measurement of, namely a city population, and we are counting the number of members that satisfy a description of a Student and a Teacher, respectively. While the PSTR requires a count of the population, other indicators may require statistical measures of mean, deviation, etc. of other characteristics of the population.

Anticipating the broader requirements of the GCIO, we have adopted the GovStat[14] general statistics ontology (Pattuelli, 2003). In GovStat, the core class is the 'gs:Population' to be measured. A 'gs:Population' is linked to a parameter (e.g., mean, standard deviation) by the 'gs:is_described_by' property, and the parameter is a sub class of 'gs:Parameter'. In statistics it is almost always the case that only a portion of the population is measured. This portion is represented by the class 'gs:Sample', and the parameter being measured is represented as a subclass of 'gs:Statistic'.

What is missing is a definition of the population that we are measuring or from which a sample is to be taken. For the PSTR indicator the 'gs:Population' must identify the area in which the population resides, i.e., the city, and what characterizes a member of the population, namely the characteristics of a Student or Teacher. For example, the characteristics of a Teacher could be:

- Fulltime, defined as teaching 30 or more hours per week, and
- Teaches at the primary or secondary level, where primary spans grades 1 thru 8 and secondary spans 9 thru 12.

[14] The GovStat Ontology is not available online, but a version with the GCI extensions can be found at: http://ontology.eil.utoronto.ca/govstat.owl. We will use the prefix "gs:" to identify classes and properties from the ontology.

We have extended the GovStat ontology as follows:

- Added a property to 'gs:Population, gs:located_in', that identifies the area that the Population is drawn from.
- Added a property to 'gs:Population', 'gs:defined_by', that identifies the members of the Population that are to be counted.

This is depicted in Fig. 6 where both student and teacher populations are linked to a city and definition of a student and teacher respectively.

5.3. Summary

The GCIO provides the basic constructs for defining an indicator, including:

- Defining populations by their geoname and the entity.
- Defining the size of a population and its unit of measure.
- Defining the ratio of the size of two populations and its unit of measure.

With the classes and properties defined in Sections 6 and 7, it is possible to precisely define a city indicator, with enough detail to automate analysis. Secondly, the GCIO can be used to publish the derivation tree composed of supporting data that was used to derive the indicator's value.

6. Defining theme specific knowledge

Significant portions of the indicator definition remain to be represented. Within the education theme, words are used without definition because they are assumed to be understood by the reader. For example, both student and teacher, are used in the education indicators, but are not defined. However they appear in Fig. 6 linked to the 'Student_ Population' and 'Teacher_Population' respectively via a 'defined_by' property. In order for the PolisGnosis system to understand and analyze an education indicator, it must also understand these commonly

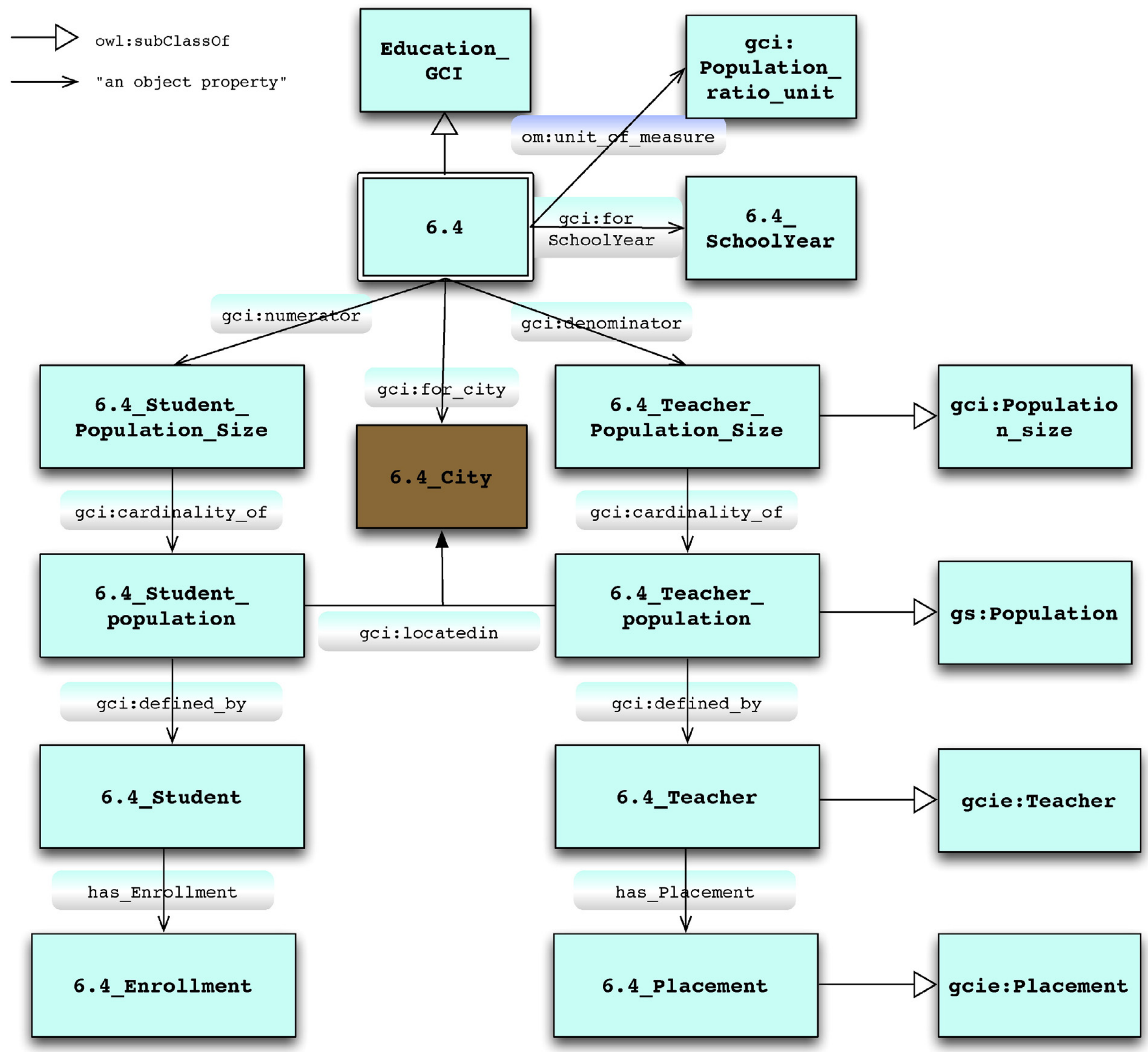

Fig. 6. Primary Student Teacher Ratio Definition.

understood words. The approach we have taken is to construct a separate ontology for each ISO 37120 theme. In the PSTR example, we have created an education ontology that captures all of the concepts that appear in the theme's indicators (Fox, 2014).

In the following we describe a portion of the educational concepts needed to represent the definition of a Student and show how they are used to represent a Student in the PSTR indicator. The definitions provided are more general than what is required by the PSTR as they have to be applied to all educational indicators. More information on this can be found in Fox (2014).

We define a 'Student' to be a subclass of 'Person' and has been enrolled in one or more 'Educational Program's. Each 'Grade' they attend is represented as a separate 'Enrollment' due to the need to represent associated information. For example, a 'Student' may be enrolled in a different 'Grade' at different 'School's, they may be part time in one grade and full time in another, etc.

Class	Property	Value restriction
Student	owl:subClassOf	sc:Person
	has_Enrollment	min 1 Enrollment
	has_Birthdate	exactly 1 xsd:dateTime
	owl:equivalentClass	cyc:Student
	has_primary_residence	exactly 1 ic:HomeAddress

An 'Enrollment' is composed of the 'Program' the student is enrolled in, an 'Educational Facility' they attend, 'School Year', 'Course's they took, 'Grade', and an 'enrolled Status' of full or part time.

Class	Property	Value restriction
Enrollment	attends	exactly 1 EducationFacility
	enrolled_Program	exactly 1 Program
	for_SchoolYear	exactly 1 SchoolYear
	enrolled_Courses	min 1 Enrolled_Course
	enrolled_Grade	exactly 1 Grade
	enrolled_Status	exactly 1 Enrollment_Status
Enrolled_Course	for_Course	exactly 1 Course
	has_Result	exactly 1 xsd:string
	has_Comment	only xsd:string

'Enrolled Course' is defined by identifying the 'Course' that was enrolled in, having a result (i.e., a grade) and some comment.

A 'Program' is anything that requires 'Certification'. It also defines what it means to be 'Fulltime' in terms of the number of hours required over a designated period of time, such as a 'day', 'week', 'month' or 'year'. A 'School Program' defines the 'Course's' that are taught and whether the program is primary, secondary, etc. 'Grade Level' is a subclass of 'School Program'. The 'Grade Level' class allows each city to define the grades that correspond to primary and secondary school. 'Grade Level' has a 'starting_Grade' and 'ending_Grade' that define the first and last grades of the level. Each city defines its own version of 'Grade Level Primary' that is appropriate for their school system. In the case of Toronto, the starting and ending grades are constrained by the definitions provided by the Province of Ontario. A 'Grade Level' also has a starting and ending age to represent the range of ages that can attend this level of school.

Class	Property	Value restriction
Program	has_Certification	some Certification
	has_Fulltime_Hours	exactly 1 positiveInteger
	has_Fulltime_Period	exactly 1 TimePeriod
School program	owl:subclassOf	Program
	has_Course	min 1 Course
	has_SP_Type	all SP_Type
Grade level	owl:subClassOf	SchoolProgram
	starting_Grade	exactly 1 Grade
	ending_Grade	exactly 1 Grade
	gci:for_City	exactly 1 City
	starting_age	exactly 1 positiveInteger
	ending_age	exactly 1 positiveInteger

The 'Grade' class has subclasses covering all possible grades, e.g., Grade One, Grade Two. Each grade is connected to another via the 'next_Grade' property to define the ordering. It also has sub classes 'Primary Grade' and 'Secondary Grade'.

We can now define a student as provided in the PSTR indicator as follows: '6.4_Student' is a class that defines a student according to the definition found in ISO 37120 6.4. '6.4_Student' is a subclass of the more general 'Student'. It represents the indicator's definition of a student by defining an enrollment '6.4_Enrollment'. This is defined as having to 'attends' exactly one public primary school, and is enrolled in at least one primary grade. The definitions of the concepts 'PrimaryGrade', 'PublicPrimarySchool', etc. are defined in Fox (2014).

Class	Property	Value restriction
6.4_Student	subClassOf	Student
	has_Enrollment	6.4_Enrollment
6.4_Enrollment	owl:subClassOf	Enrollment
	for_SchoolYear	exactly 1 6.4_SchoolYear
	attends	exactly 1 PublicPrimarySchool
	enrolled_Grade	some PrimaryGrade
	enrolled_Status	exactly 1 (Full_Time or Part_Time)
	enrolled_Program	exactly 1 GradeLevelPrimary
	enrolled_Courses	some Course

7. Representing city specific knowledge

The fourth problem we wish to solve is how to represent a city's specific knowledge. The definition of an indicator uses language whose interpretation may differ from one city to another. For example, the definition of the grades that make up Primary School in Toronto may differ from the grades that make up Primary School in London. The same would be true of Public School, where Toronto's are government funded and London's are private. In order to successfully diagnose why one's city's indicator value differs from another, we need to understand how the interpretation of an indicator differs over time or between cities. This requires a representation of city specific knowledge.

Consider an analysis of PSTR between two cities. Toronto has a lower PSTR than Calgary. Toronto defines that primary grades are from one through six. Calgary defines primary grades as one through nine. In the latter case, the higher grades, such as seven and eight, may have larger student teacher ratios than in the lower grades due to students being mature enough to learn in larger class settings. This may be the root cause of the latter city having a higher PSTR. However in order for the PolisGnosis analysis engine to determine this, it has to have access to city specific knowledge of how primary grades are defined in the two cities.

The following are examples of city specific education knowledge. They are not included in the GCIO but are defined in separate ontologies that extend the GCIO and the indicator's theme specific knowledge (discussed in Section 6).

Similar to the approach in the previous section, we define a 'GradeLevel' as a sub class of a 'SchoolProgram' which is a sub class of a 'Program'. The 'GradeLevel' concept introduces the properties of starting and ending grades.

Class	Property	Value Restriction
Program	has_Certification	some Certification
	has_Fulltime_Hours	exactly 1 positiveInteger
	has_Fulltime_Period	exactly 1 TimePeriod
School program	owl:subclassOf	Program
	has_Course	min 1 Course
	has_SP_Type	all SP_Type
Grade level	owl:subClassOf	SchoolProgram
	starting_Grade	exactly 1 Grade
	ending_Grade	exactly 1 Grade
	gci:for_City	exactly 1 City
	starting_age	exactly 1 positiveInteger
	ending_age	exactly 1 positiveInteger

In the case of Toronto, grade levels are defined by the province. So we define the Ontario primary grade level as follows:

Class	Property	Value Restriction
GradeLevelPrimaryCanadaOntario	owl:subClassOf	GradeLevelPrimaryCanada
	starting_Grade	exactly 1 GradeOne
	ending_Grade	exactly 1 GradeSix

Similarly, definitions of public versus private schools and other components of the PSTR definition can be defined. See Fox (2014) for more detail on education indicators.

The amount of city specific knowledge that needs to be added depends upon the indicator. ISO 37120 has 100 indicators, but there are literally thousands of indicators more if we include other sources such as the OECD. For Education theme indicators, we find that additional information needs to be provided. However Energy theme indicators rely upon data that is more standardized globally. Hence the amount of city specific knowledge that is needed depends upon the indicator theme. While tools like Protege make it possible to specify this knowledge they are inappropriate due to their complexity and the expertise required. Cities will need tools that are designed for the acquisition of indicator knowledge to simplify the process.

8. Indicator ontologies hierarchy

The following diagram (Fig. 7) depicts the organization of ontologies used to define the ISO 37120 indicators. At the highest level, i.e., ISO 37120 ontology level, the ISO 37120 module[15] contains the globally unique identifier (IRI) for each ISO 37120 indicator. For example, the IRI for the Student/Teacher Ratio indicator is: "http://ontology.eil.utoronto.ca/ISO37120.owl#6.5".

For each indicator theme in the ISO 37120 specification, for example Education, there is a separate ontology that provides the definition of each indicator in that theme. For example, ISO37120/Education.owl[16] provides a complete OWL definition for all seven of the indicators in the ISO 37120 specification.

The GCI Ontology level provides the theme specific ontologies required to define each theme's indicators. For example, to define the

[15] http://ontology.eil.utoronto.ca/GCI/37120.owl.

[16] http://ontology.eil.utoronto.ca/GCI/ISO37120/Education.owl.

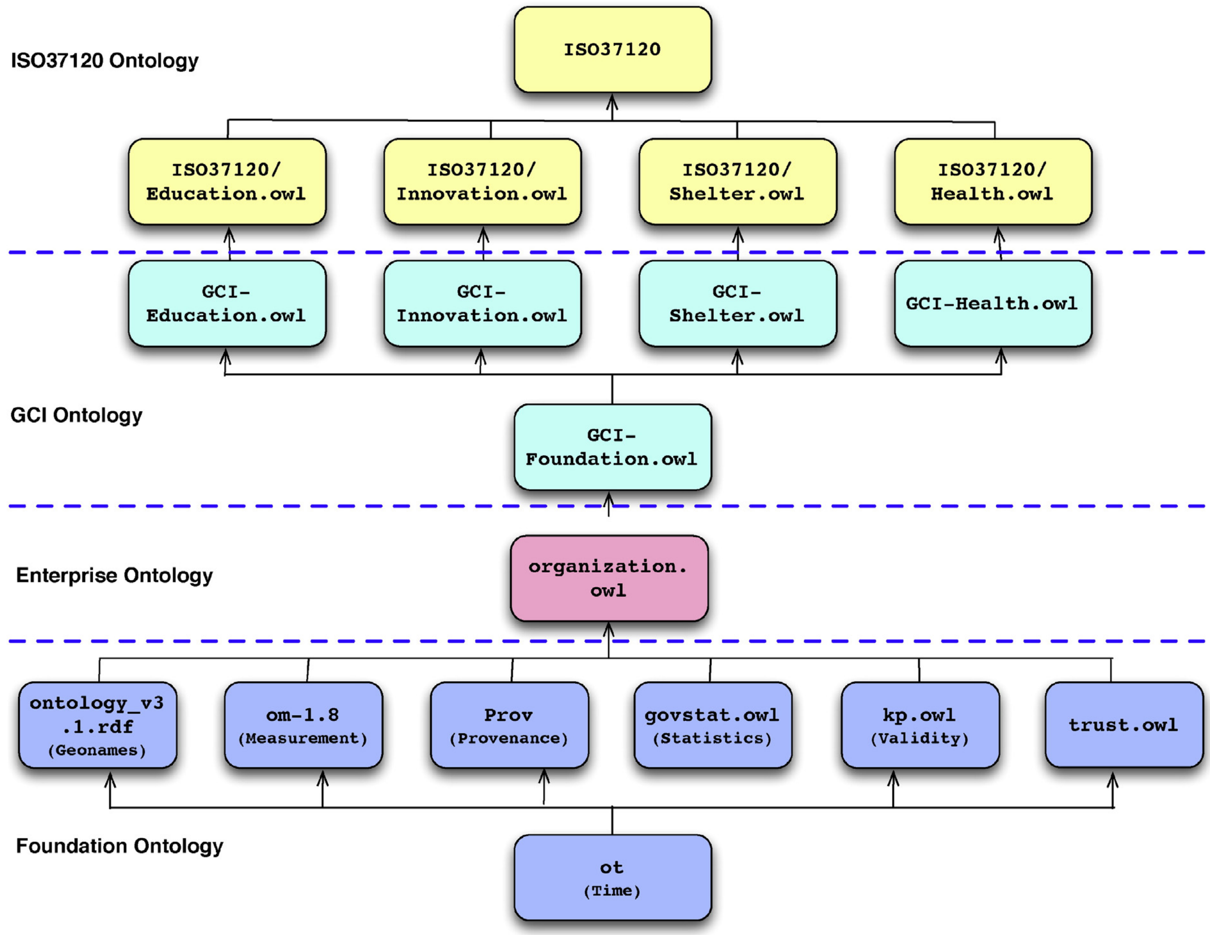

Fig. 7. ISO 37120 ontology modules.

ISO 37120 education indicators, we need an educational ontology covering concepts such as schools, teachers, students, cohorts, etc. GCI-Education.owl[17] provides the classes used by ISO37120/Education.owl. All of the theme specific indicator ontologies rely on the GCIO[18] for more generic concepts such as population counts and ratios, meta-information, etc.

City specific ontologies are not shown as they are provided by the city that publishes indicators.

9. Evaluation

We approach the evaluation of the above ontology from multiple perspectives. The first is the competency of the ontology. In Grüninger and Fox (1995), the requirements for an ontology are defined by a set of competency questions. These questions define how the ontology is to be used by applications. In order for an ontology to be competent with respect to a set of questions, it must be able to correctly deduce answers assuming the model has been instantiated correctly. In Fox (2013b) the competency of the GCIO is demonstrated.

A second approach to evaluating the GCIO is determining how well it fulfills its foundational role in the representation of the ISO 37120 indicators. In Fox (2014) it is shown that the combination of the GCIO and GCI Education ontology is able to faithfully and precisely represent the ISO 37120 education indicators. In Forde and Fox (2015) it is shown that the combination of GCIO and GCI Innovation ontology is able to faithfully and precisely represent the ISO 37120 innovation indicators. It has also been shown indicators. Other indicator themes such as shelter (Wang & Fox, 2015) and health (Falodi & Fox, 2015) have been completed, and environment, transportation and finance are nearing completion and demonstrate the relevance and generality of the GCIO.

A third approach to evaluation is determining the extent that cities are using the ontology to publish their indicator data. The major challenge we face in getting cities to adopt and publish indicator data using any ontology is that most open data and information systems city staff are unaware of the Semantic Web and the role that ontologies play. Consequently, adoption requires a lengthy education process to get to the point that city staff understand its importance.

Never the less, In the fall of 2014, the City of Toronto published its ISO 37120 indicators (Toronto, 2014). The publication is a PDF file whose content does not conform to Semantic Web standards. At the beginning of 2015, the Open Data Group of the City of Toronto agreed to publish the ISO 37120 data based on the GCIO. At the time of writing, the Education indicators have been completed. The published version of these indicators is available at http://ontology.eil.utoronto.ca/ISO37120/Toronto/2013//ISO37120_6_2013_TO.owl.

The City Protocols project,[19] led by the City of Barcelona, has adopted the GCIO for its representation of city indicators, which include the ISO 37120 indicators.

[17] The GCI Education ontology can be found at http://ontology.eil.utoronto.ca/GCI/Education/GCI-Education.owl along with its documentation at http://ontology.eil.utoronto.ca/GCI/Education/GCI-Education.html. We will use the prefix "gcie" where needed.

[18] The GCIO can be found at http://ontology.eil.utoronto.ca/GCI/Foundation/GCI-Foundation.owl along with its documentation at http://ontology.eil.utoronto.ca/GCI/Foundation/GCI-Foundation.html. We will use the prefix "gci" where needed.

[19] http://cityprotocol.org.

The Research Data Alliance's[20] Quality of Urban Life Interest Group[21] has adopted the GCIO for its representation of city indicators.

10. Conclusion

Enhancing the quality, effectiveness and efficiency of the operations and services of a city depends upon the ability to measure them. The development of city metrics faces many challenges. The first challenge is the selection and definition of the metrics. The second challenge is the adoption and use of these metrics by a large number of cities. These first two challenges have been the focus of the Global City Institute for the last five years and has resulted in the creation of ISO 37120 and the adoption of the standard by over 250 cities worldwide. The third challenge is to publish the indicators using a representation that can be linked, merged, mashed, and analyzed based on the principles of the Semantic Web. This work addresses this third challenge. It provides for the first time a set of ontologies for: 1) how meta-data associated with a single indicator value is to be represented; 2) how an indicator's definition is to be represented; 3) how data used to derive an indicator value is to be represented; 4) how indicator theme specific knowledge is to be represented; and 5) how city specific knowledge is to be represented.

As identified in the background section, little work has been done to date on the development of ontologies for city indicators — providing little to compare to. Never the less, this work demonstrates that the decades of research into measurement, provenance, validity and trust ontologies provides a necessary and important foundation for their representation. However this work also highlights that these foundation ontologies are not sufficient. For each ISO 37120 theme, such as Shelter, Education, Transportation, etc., there is the need for theme specific ontologies for representing both the indicator definitions and the city's supporting data. The development of each of these ontologies requires significant effort.

From an ontology development perspective, the GCIO re-uses portions of several ontologies. This should have led to inconsistencies in the merged ontology, but it did not, as the available formalization did not lead to any logical conflict. This lack of conflict is due, in part, to the constituent ontologies having a narrow focus, rarely including classes and properties that overlap with each other, and a lack of available formalization.

There are two directions that our current research is heading. The first is to complete the theme specific ontologies and the definitions of each theme's indicators to span the entire set of ISO37120 Indicators. The second direction is to develop the theories that the PolisGnosis system will use to automate the longitudinal and transversal analysis of city indicators. Work on consistency analysis is nearing completion and will appear in future papers.

One of the takeaways from this research is that indicator supporting data is extremely sparse. This has led to the development of an Open Data Completeness Model (Fox & Pettit, 2015) that measures both the degree to which the data supporting an indicator is opening published, and the format and ontologies that are used to publish it. The most likely path to reducing the sparseness, is to incorporate these ontologies into a city's enterprise software; the acquisition, representation and publishing of indicator data has to be part of the enterprise software systems used to run a city.

Acknowledgments

This research is supported in part by the Natural Sciences and Engineering Research Council of Canada. I would like to thank Hajo Rijgersberg for his feedback on the use of the OM Ontology, and Patricia McCarney for her introduction to Global City Indicators. I would also like to thank the reviewers for their important input.

[20] http://rd-alliance.org/.
[21] https://rd-alliance.org/groups/urban-quality-life-indicators.html.

Appendix A. Ontologies

The Global City Indicator Foundation ontology can be found in: http://ontology.eil.utoronto.ca/GCI/Foundation/GCI-Foundation.owl.

The Global City Indicator Education ontology can be found in: http://ontology.eil.utoronto.ca/GCI/Education/GCI-Education.owl.

The Global City Indicator Shelter ontology can be found in: http://ontology.eil.utoronto.ca/GCI/Shelter/GCI-Shelter.owl.

The Global City Indicator Health ontology can be found in: http://ontology.eil.utoronto.ca/GCI/Health/GCI-Health.owl.

The Global City Indicator Innovation ontology can be found in: http://ontology.eil.utoronto.ca/GCI/Innovation/GCI-Innovation.owl.

URIs for all of the ISO37120 indicators can be found in:
http://ontology.eil.utoronto.ca/ISO37120.owl.

Definitions of the ISO37120 education theme indicators, using the GCI Foundation and Education ontologies can be found in:
http://ontology.eil.utoronto.ca/GCI/ISO37120/Education.owl.

Definitions of the ISO37120 shelter theme indicators, using the GCI Foundation and Shelter ontologies can be found in:
http://ontology.eil.utoronto.ca/GCI/ISO37120/Shelter.owl.

Definitions of the ISO37120 health theme indicators, using the GCI Foundation and Health ontologies can be found in:
http://ontology.eil.utoronto.ca/GCI/ISO37120/Health.owl.

Definitions of the ISO37120 innovation theme indicators, using the GCI Foundation and Innovation ontologies can be found in:
http://ontology.eil.utoronto.ca/GCI/ISO37120/Innovation.owl.

References

Allen, J. F., & Ferguson, G. (1997). Actions and events in interval temporal logic. In O. Stock (Ed.), *Spatial and temporal reasoning* (pp. 205–245). Dordrecht, Netherlands: Kluwer.

Belhajjame, K., Deus, H., Garijo, D., Klyne, G., Missier, P., Soiland-Reyes, S., & Zednik, S. (2012). PROV Model Primer. http://www.w3.org/TR/prov-primer

Candela, L., Castelli, D., Manghi, P., & Tani, A. (2015). Data journals: A survey. *Journal of the Association for Information and Science Technology*.

Capadisli, S., Auer, S., & Ngonga Ngomo, A. -C. (2013). *Linked SDMX data: Path to high fidelity statistical linked data for OECD, BFS, FAO and ECB*. IOS Press (submitted for publication).

Demolombe, R. (1998). To trust information sources: A proposal for a modal logical framework. *Proceedings of the autonomous agents workshop on deception, fraud and trust in agent societies* (pp. 9–19).

Falodi, J., & Fox, M. S. (2015). A health ontology for global city indicators (ISO 37120). *Working paper*. Enterprise Integration Laboratory, University of Toronto (To appear).

Forde, A., & Fox, M. S. (2015). An innovation ontology for global city indicators. *Working paper*. Enterprise Integration Laboratory, University of Toronto (to appear).

Fox, M. S. (2013a). City data: Big, open and linked. *Municipal interfaces, November. A longer version appeared as a working paper, enterprise integration laboratory*. (pp. 19–25). University of Toronto (http://eil.utoronto.ca/smartcities/papers/City-Data-v5.pdf).

Fox, M. S. (2013b). A foundation ontology for global city indicators. *Working paper 3*. Global Cities Institute, University of Toronto (http://eil.utoronto.ca/smartcities/papers/GCI-Foundation-Ontology.pdf).

Fox, M. S. (2014). An education ontology for global city indicators. *Working paper*. Enterprise Integration Laboratory, University of Toronto (http://www.eil.utoronto.ca/smartcities/papers/fox-eil14.pdf).

Fox, M. S. (2015). The PolisGnosis project: Representing and analysing city indicators. *Working paper*. Enterprise Integration Laboratory, University of Toronto (http://eil.utoronto.ca/smartcities/papers/PolisGnosis-11mar2015.pdf).

Fox, M. S., & Huang, J. (2005a). Knowledge provenance in enterprise information. *International Journal of Production Research*, *43*(20), 4471–4492 (http://www.eil.utoronto.ca/km/papers/fox-ijpr05.pdf).

Fox, M. S., & Huang, J. (2005b). An Ontology for Static Knowledge Provenance. *Knowledge sharing in the integrated enterprise, IFIP. 183/2005*. (pp. 203–213).

Fox, M. S., & Pettit, C. (2015). On the completeness of open city data for measuring city indicators. *Proceedings of the IEEE smart cities conference, October, Guadalajarah Mexico, October*.

Ghahremanloo, L., Thom, J. A., & Magee, L. (2012). An ontology derived from heterogeneous sustainability indicator set documents. *Proceedings of the Seventeenth Australasian Document Computing Symposium* (pp. 72–79).

Grüninger, M., & Fox, M. S. (1995). Methodology for the design and evaluation of ontologies. *Proceedings of the workshop on basic ontological issues in knowledge sharing, IJCAI-95, Montreal, Canada*.

Hartig, O., & Zhao, J. (2010). Publishing and consuming provenance metadata on the web of linked data. *Proceedings of the third international provenance and annotation workshop*.

Hausenblas, M., Halb, W., Raimond, Y., Feigenbaum, L., & Ayers, D. (2009). Scovo: Using statistics on the web of data. *The Semantic Web: Research and applications* (pp. 708–722). Berlin Heidelberg: Springer.

Hitzler, P., et al. (2012). *OWL 2 web ontology language primer* (2nd edition) (http://www.w3.org/TR/owl-primer).

Hobbs, J. R., & Pan, F. (2006). Time ontology in OWL. (http:www.w3.org/TR/owl-time/).

Hoornweg, D., Nunez, F., Freire, M., Palugyai, N., Herrera, E. W., & Villaveces, M. (2007). City indicators: Now to Nanjing. *World Bank policy research working paper 4114*.

Huang, J. (2008). *Knowledge provenance: An approach to modeling and maintaining the evolution and validity of knowledge.* (PhD Thesis) Dept. of Mechanical and Industrial Engineering, University of Toronto.

Huang, J., & Fox, M. S. (2006). An ontology of trust — formal semantics and transitivity. *Proceedings of the international conference on electronic commerce* (pp. 259–270) (http://www.eil.utoronto.ca/km/papers/huang-ec06.pdf).

Huang, J., & Fox, M. S. (2004a). Dynamic knowledge provenance. *Proceedings of business agents and semantic web workshop* (pp. 372–387). National Research Council of Canada.

Huang, J., & Fox, M. S. (2004b). Uncertainty in knowledge provenance. *Proceedings of the European Semantic Web Symposium. Springer Lecture Notes in Computer Science*.

de Kleer, J. (1986). An assumption-based TMS. *Artificial Intelligence*, *28*, 127–162.

Marsh, S. P. (1994). *Formalising trust as a computational concept.* (Ph.D. Thesis) University of Stirling.

Matuszek, C., et al. (2006). An introduction to the syntax and content of Cyc. *Proceedings of the 2006 AAAI Spring Symposium on formalizing and compiling background knowledge and its applications to knowledge representation and question answering*.

Moreau, C., et al. (2010). The open provenance model core specification (v1.1). *Future generation computer systems* (Also see http://openprovenance.org).

Niles, I., & Pease, A. (2001). Towards a standard upper ontology. *Proceedings of the 2nd International Conference on Formal Ontology in Information Systems FOIS-2001, Ogunquit, Maine*. New York: ACM Press.

Noy, N., et al. (2003). Protege-2000: An open-source ontology-development and knowledge-acquisition environment. *Proceedings of the AMIA Annual Symposium*.

Pattuelli, M. C. (2003). The GovStat ontology: Technical report. *The GovStat Project, integration design laboratory, school of information and library science*. University of North Carolina at Chapel Hill (http://ils.unc.edu/govstat/papers/govstatontology.doc).

Rijgersberg, H., van Assem, M., & Top, J. (2013). Ontology of units of measure and related concepts. *Semantic Web J.*, *4*(1), 3–13.

Rijgersberg, H., Wigham, M., & Top, J. L. (2011). How semantics can improve engineering processes: A case of units of measure and quantities. *Advanced Engineering Informatics*, *25*, 276–287.

Shortliffe, E. H. (1976). *Computer-based medical consultations: MYCIN.* New York: American Elsevier.

Toronto (2014). *Toronto's 2013 results under ISO 37120 indicators of city service delivery and quality of lifehttp://www1.toronto.ca/CityOfToronto/CityManagersOffice/TorontoProgressPortal/ISO37120/Final-SummaryofToronto%27sWCCD-ISO37120Results-6.pdf*

Uceda-Sosa, R., Srivastava, B., & Schloss, B. (2012). Building a highly consumable semantic model for smarter cities. *Proceedings of the workshop on AI for an Intelligent Planet, ACM*.

Wang, Y., & Fox, M. S. (2015). A Shelter Ontology for Global city Indicators (ISO 37120). *Working paper, enterprise integration laboratory*. University of Toronto (http://eil.mie.utoronto.ca/wp-content/uploads/2015/06/GCI-Shelter-Ontology-v3.pdf).

Westfall, M. S., & de Villa, V. A. (2001). *Urban indicators for managing cities.* Asian Development Bank.