A Foundation Ontology for Global City Indicators

Mark S. Fox
Mechanical and Industrial Engineering
University of Toronto
msf@eil.utoronto.ca

ABSTRACT

This paper defines a generic, reusable ontology that forms the foundation for representing the definitions of ISO37120 Global City Indicators and their instantiation on the Semantic Web by cities. It makes two contributions. The first contribution is to the field of city indicators. The ontology makes it possible to:

- create precise definitions of indicators thereby reducing the ambiguity of their interpretation and making them computationally accessible;
- represent the "data behind the data" or supporting data, enabling drilling down;
- determine the consistency of metrics. E.g., is the supporting data of the same scale, refer to the same place, measured in the same way, covering the same time period, etc.; and
- achieve interoperability, namely the ability to access, understand, merge and use measurement data available from datasets spread across the Semantic Web by providing a standard representation using OWL 2.

The second contribution is the successful integration of seven basic ontologies and their extension with trans-foundation axioms. The ontology spans analytical models (e.g., ratios), statistical models (e.g., population measurements), geo-spatial models (e.g., city boundaries), temporal models (e.g., time periods) and description logic models (e.g., definitions of students, teachers, etc.). It also integrates meta-knowledge such as provenance, validity and trust.

Keywords: City Indicator, Ontology, Semantic Web, ISO 37120.

1. Introduction

Cities are moving towards policy-making based on data[1]. Yet it has been recognized by urban researchers, city professionals and political leaders that city level data is both incomplete and inconsistent. In 2007, it was recognized that "there are thousands of different sets of city (or urban) indicators and hundreds of agencies compiling and reviewing them. Most cities already have some degree of performance measurement in place. However, these indicators are

[1] "Data driven decision making is one of the reasons New York City is the safest big city in America," said Mayor Bloomberg. "Just as data helps us reduce crime, prevent fire fatalities and keep incarceration levels low, we believe understanding data can help us work with judges and criminal justice agencies to further improve the effectiveness and efficiency of our criminal justice system." Press Release, New York City, PR-012-13, 7 January 2013.

usually not standardized, consistent or comparable (over time or across cities), nor do they have sufficient endorsement to be used as ongoing benchmarks." (Hoornweg et al., 2007).

In response to this challenge, in 2010 the Global City Indicators Facility (GCIF)[2] was created to work with cities throughout the world in identifying a common set of indicators and establishing standardized definitions and methodologies that can be consistently applied globally (Global City Indicators Facility, 2010a; McCarney, 2011, 2013). The outcome of this effort is the international standard ISO 37120 "Sustainable development of communities — Indicators for city services and quality of life"[3].

The primary problem with indicator development is that definitions are people oriented; they are provided in natural language, i.e., English, and not in a more formal, possibly computer readable language. As the old joke goes, ask two lawyers a question and you will get three opinions. The same generally holds true for most attempts at defining terminology, such as indicators, using an imprecise language like English. The reader of the definition imposes their own interpretation based on their understanding of the language and the environment in which they live (i.e., how their own city may define some terms).

Consider the definition of a Student/teacher ratio as provided in Hoornweg et al. (2007, p. 45):

> "Student/teacher ratio"[4].

This has been expanded by the World Bank (2008, p. 18) to:

> "Student/teacher ratio", where the numerator is "Number of Students", and the denominator is "Number of Teachers".

One problem is whether "student" refers to full time students, or part time students. Are they regular students or special needs students? Do they include kindergarten students or not? It is also difficult to compare an indicator for a single city across time if the definition of student changes. For example, today the educational system includes students with special needs, but 30 years ago they may not have been enrolled. Without a more precise definition of terms, it makes it difficult to compare an indicator across cities where each city interprets what a student is differently, or against itself when definitions change.

Obviously, the definition and documentation of indicators can be expanded, as has been done in ISO 37120, in recognition of the limited value of prior indicator definitions. Following is the definition of student teacher ratio provided by the standard:

> "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

[2] http://www.cityindicators.org/
[3] First edition published 2014-05-15.
[4] Yes, just three words ☺.

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 definition of student teacher ratio clearly addresses some of the issues raised above. Never the less, there will always be a disconnect between the actual value of a city's indicator and the data sources and processes used to measure it; while the indicator's value is recorded in a machine-readable form (e.g., database or semantic web), the sources and measurement processes are buried in datasets and documents that are inaccessible or only human readable. In the end, all we are left with is a record of indicator values without an understanding of what they actually measure and how they were measured. We have to rely on the good will of the people who reported to the data to adhere to the definitions.

Our goal is to formalize the definition of city indicators using Ontologies (Gruber, 1993; Grüninger & Fox, 1995) as implemented in the Semantic Web (Berners-Lee et al., 2001). By doing so we will:

- enable the representation of precise definitions thereby reducing the ambiguity of interpretation and making them computationally accessible,
- take indicators out of the realm of humans and into the realm of computers where the world of Big Data, open source software, mobile apps, etc., can be applied to analyze and interpret the data,
- achieve interoperability, namely the ability to access, understand, merge and use indicator data available from datasets spread across the semantic web, and
- automate the detection of data inconsistency, and the root causes of variations.

In Section 2 we provide background on city indicators. We then present in section 3 the core ontologies for representing place, measurement, populations, and statistics. Section 4 focuses on meta information such as provenance, validity and trust. Section 5 provides an example of a student teacher ratio implemented using the ontology. Section 6 focuses on axioms for determining the consistency of indicators represented using the ontology. Section 7 then evaluates the ontology[5].

2. City Indicators

In this section we review earlier work related to ontologies and City Indicators.

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

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

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 indicator data.

In light of previous efforts to define city indicators, Hoornweg et al. (2007), identified the following aspects a good "indicator must possess to be accurate, timely and relevant for policy purposes:

- **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."

The *raison d'être* for the Global City Indicator Facility and the creation of ISO 37120 is to define city indicators that satisfy these aspects.

The Organization for Economic Co-operation and Development (OECD: www.oecd.org) "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 (Hausenblas et al., 2009; Capadisli et al., 2013).

As part of IBM's Smart Cities initiative, they have developed an Ontology representing various types of city knowledge, including city organization and services, flow of events and messages, and key performance indicators (Uceda-Sosa et al., 2011). OWL definitions of the classes and properties are provided. Axiomatization is limited and so its use of foundational ontologies.

Ghahremanloo et al. (2012) examined the ontologies employed by the cities of Melbourne and Vancouver to determine the classes and properties that represent key indicator concepts.

They identified two models: generic and specific, where in both cases the ontology provides some meta-level concepts, but the definitions of the indicators are represented as strings.

3. Core Ontologies for Global City Indicators

In this section we select and extend the core ontologies for representing Global City Indicators (GCI) definitions (classes) and their instances. We illustrate the construction of the GCI ontology using a single city indicator: Student/Teacher Ratio (STR). A number of issues arise in representing its definition. These issues will be addressed as we build the GCI ontology "one brick at a time" using foundational ontologies.

3.1. Placename Ontology

- What is the city being measured?
- What area does it cover?
- What places does it contain?

The STR is computed over a geographic area. In the case of GCIs, it would be a city. Hence, a requirement of the GCI 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 neighbouring cities. But 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[6] 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[7] 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 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[8] 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. Geonames also provides a web interface that allows anyone to search for and/or add new placenames to its knowledge base.

---

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

[7] 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.

[8] 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.

At the core of the Geonames ontology is the `geo:Feature`. A `geo:Feature` includes the following properties:

- `name`: text name of the feature, e.g., “Toronto”.
- `featureClass` – Class of feature such as Administrative (e.g., state, parish), Hydrographic (e.g., stream, lake), and Area (e.g., Parks).
- `featureCode` – Code for the feature within the class.
- `population` – Population of the feature.
- `postalCode` – One or more postal codes in which the feature resides.
- `wgs84_pos:lat` – Latitude of the feature.
- `wgs84_pos:long` – Longitude of the feature.
- `nearbyFeatures` – Features spatially located nearby.
- `wikipediaArticle` – One or more articles in Wikipedia about the feature.

A key component is the `geo:featureCode` which adapts and extends the feature codes developed by the United States National Geospatial-Intelligence Agency (NGA).

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.

3.2. Measurement Ontology

- What is the quantity being measured?
- What is its units?
- What is its scale? Mega, kilo?
- How is the indicator derived?

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? Though ISO37120 provides a written definition of the indicator, it is for human consumption. The problem is how do we define an indicator in a way that computers can understand?

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 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 STR, the purpose of a measurement ontology is to provide the underlying semantics of the number, such as what is being measured and the unit of measurement. The importance of grounding an indicator in measurement ontology is to assure that the numbers 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 and Pease, 2001) and CYC (Matuszek et al., 2006) provide classes for representing quantities, but the OM ontology[9] (Rijgersberg et al., 2011) provides a more rigorous ontology based on measurement theory. QUDT[10] is an alternative to OM. We chose OM over QUDT for the reasons expressed in Rijersberg et al. (2013).

In the following, we review some of the basics of the OM ontology:

Subset of the OM Measurement On tology		
Class	Definition	Examples
Quantity	Refers to what is being measured. It links the phenomenon (e.g., an object) being measured to the value of the measurement (Measure).	Length, diameter.
Unit_of_measure	"A unit of measure is a definite magnitude of a quantity, defined and adopted by convention and/or by law. It is used as a standard for measurement of the same quantity, where any other value of the quantity can be expressed as a simple multiple of the unit of measure."	"For example, length is a quantity; the meter is a unit of length that represents a definite predetermined length. When we say 10 meter (or 10 m), we actually mean 10 times the definite predetermined length called 'meter'."
Measure	"Combines a number to a unit of measure on an interval or ratio scale."	"3 metres", "10 kilograms"

Figure 1 represents the outcome we wish to achieve in using the OM ontology to represent an indicator. The STR is a subclass of `Quantity` that has a value that is a subclass of Measure whose units are a `Population_ratio_unit` which is a subclass of Unit of Measure. The actual value measured is a property of the instance of `Student_teacher_ratio_measure.`

[9] 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.
[10] http://www.qudt.org.

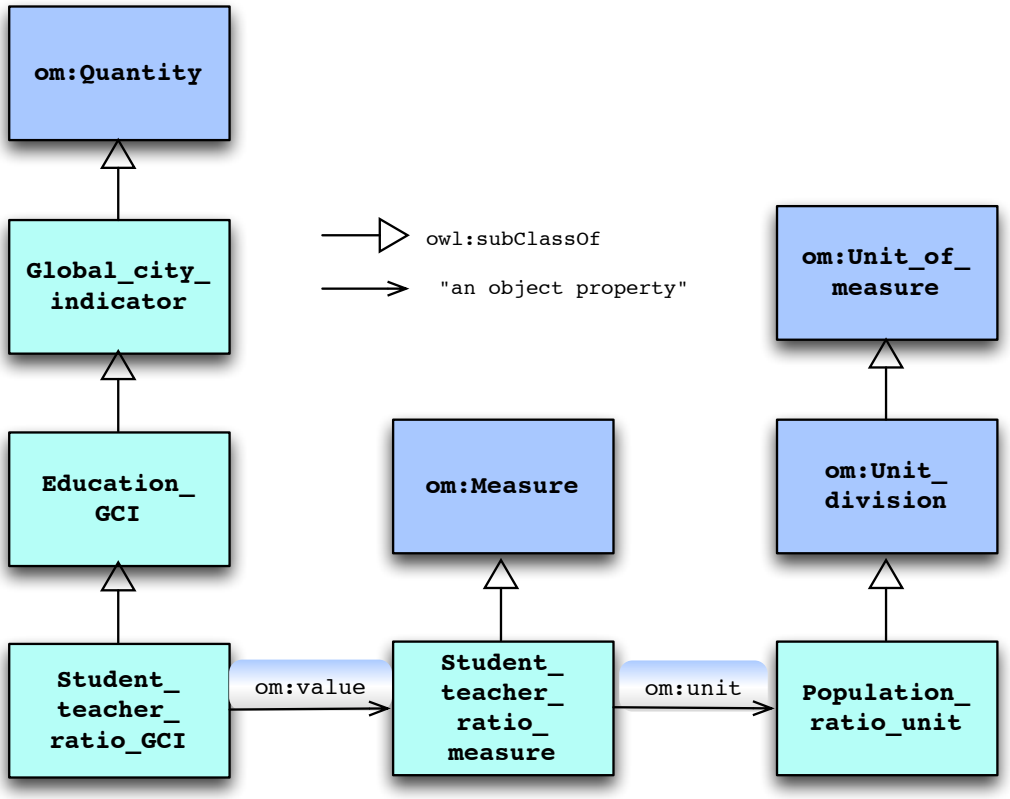

Figure 1

3.3. Measuring Populations

OM provides a good starting point, but to realize the structure in Figure 1, there are several building blocks that need to be put in place. The STR is the ratio of Student to Teacher, which is the ratio of the number of students to the number of teachers (within a city). 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.

Figure 2 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).

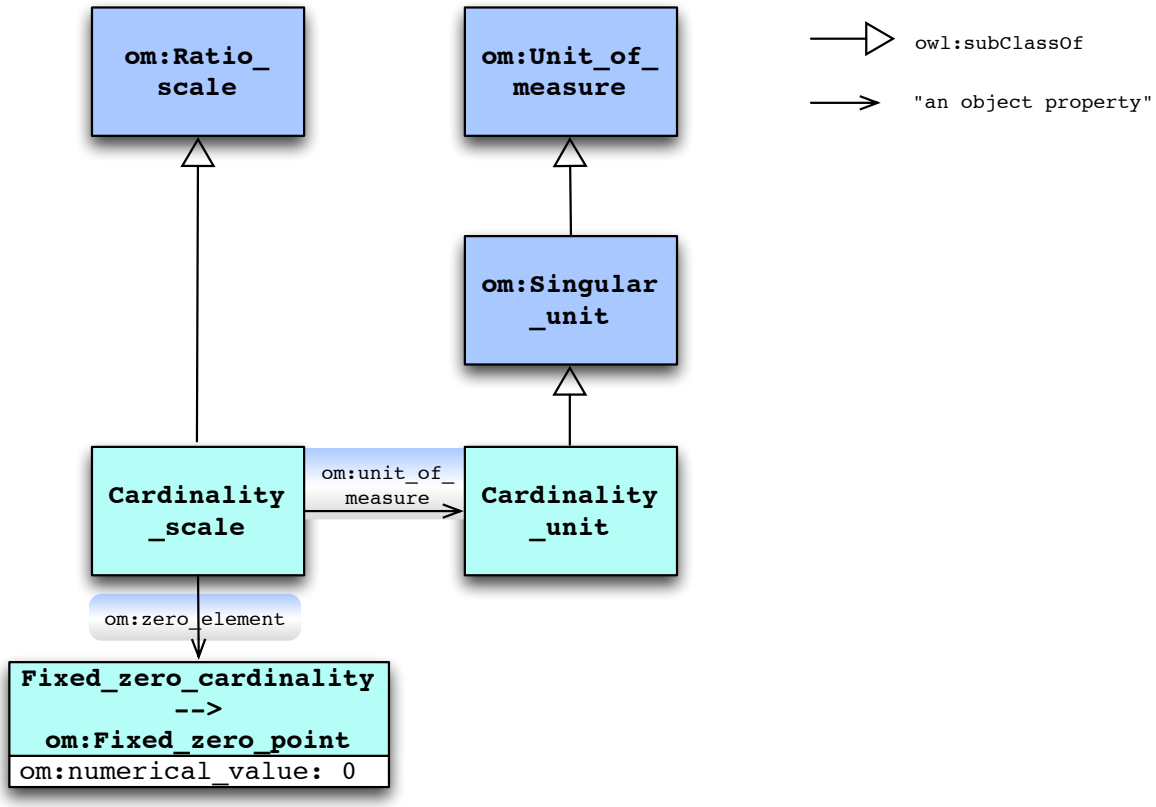

Figure 2

In Figure 3, 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 kilopc represents 1100 pc.

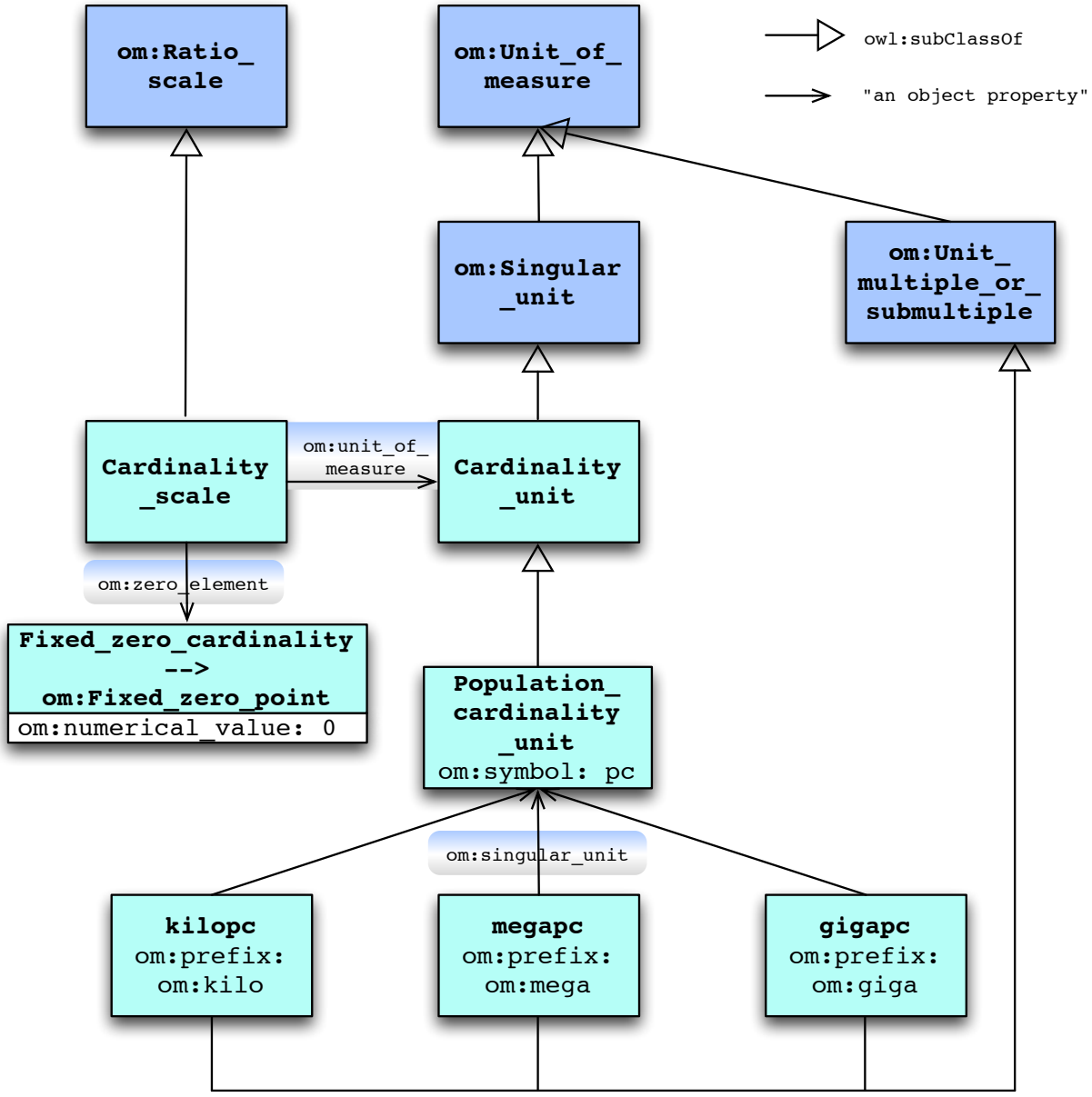

Figure 3

With the above defined, we can now introduce the unit of measure for measuring a population ratio such as STR. `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.

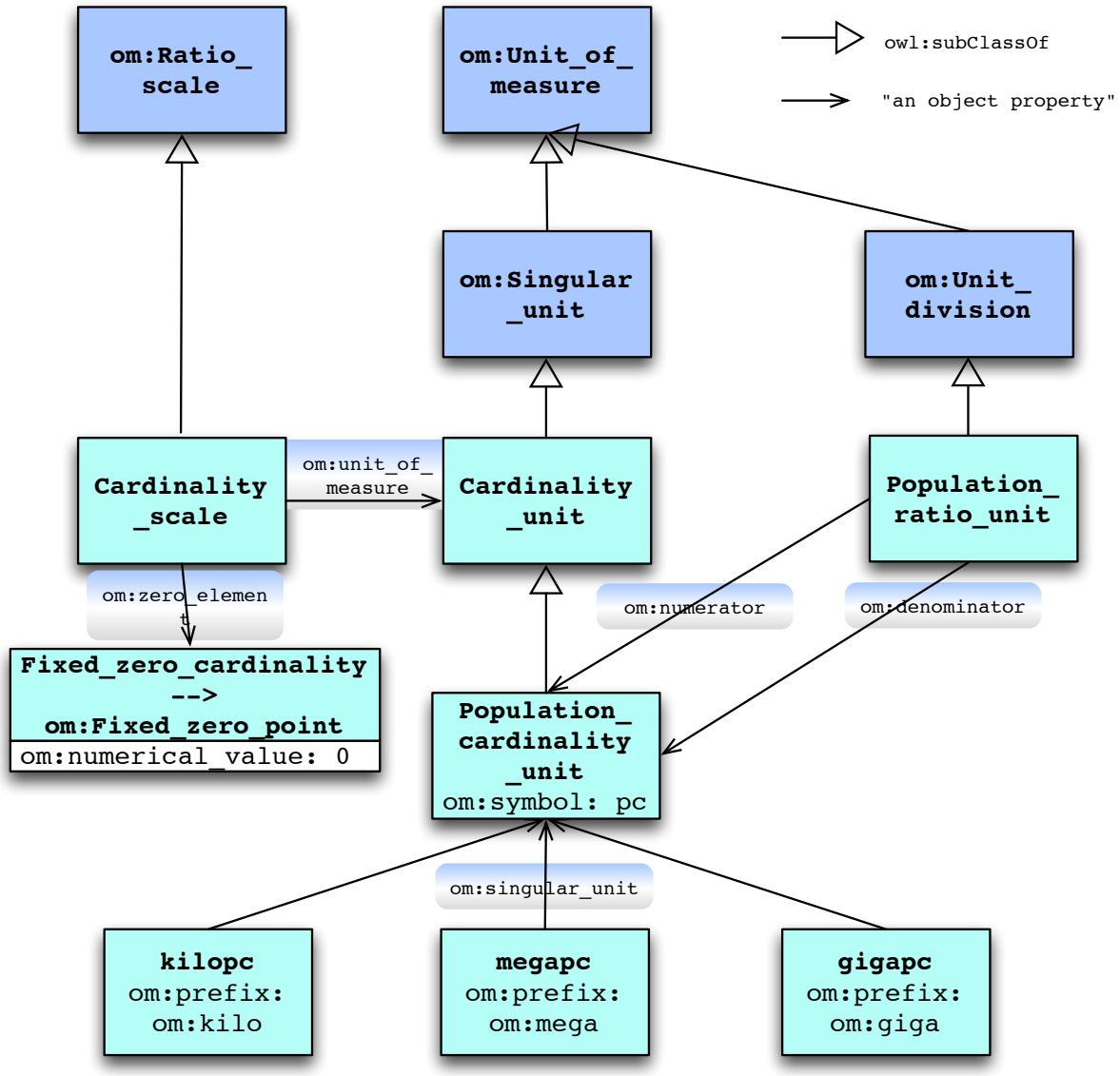

Figure 4

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 STR is derived. In Figure 5 we introduce `gci:Population_size` as a subclass of `om:Quantity`. Its `om:unit_of_measure` is the `gci:Population_cardinality_unit`.

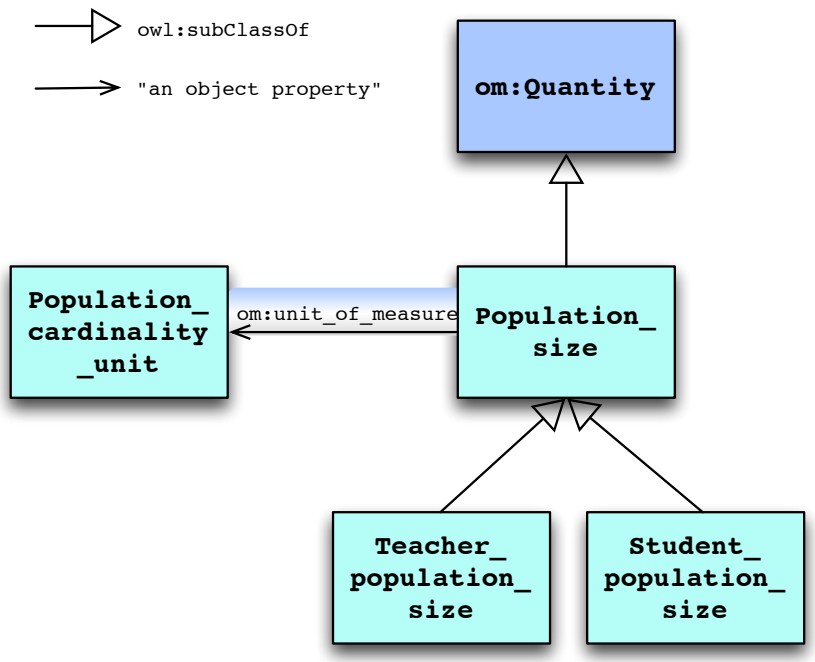

Figure 5

We now have the requisite infrastructure to define GCIs (Figure 6). First we define the class of GCIs, `gci:Global_city_indicator`, as a subclass of `om:Quantity`. All GCIs 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.

The actual value for a city's STR will be an instance of the measure `Student_teacher_ratio_measure`, which is linked to `Student_teacher_ratio_GCI and is` a subclass of `gci: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_unit's`.
- **`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.
- **`gci:teacher_def & gci:student_def`,** whose range are a subclass of `Teacher` and `Student` respectively. These define the properties of the teachers and students that we are measuring. For example, all full time students in grades 1 through 12.

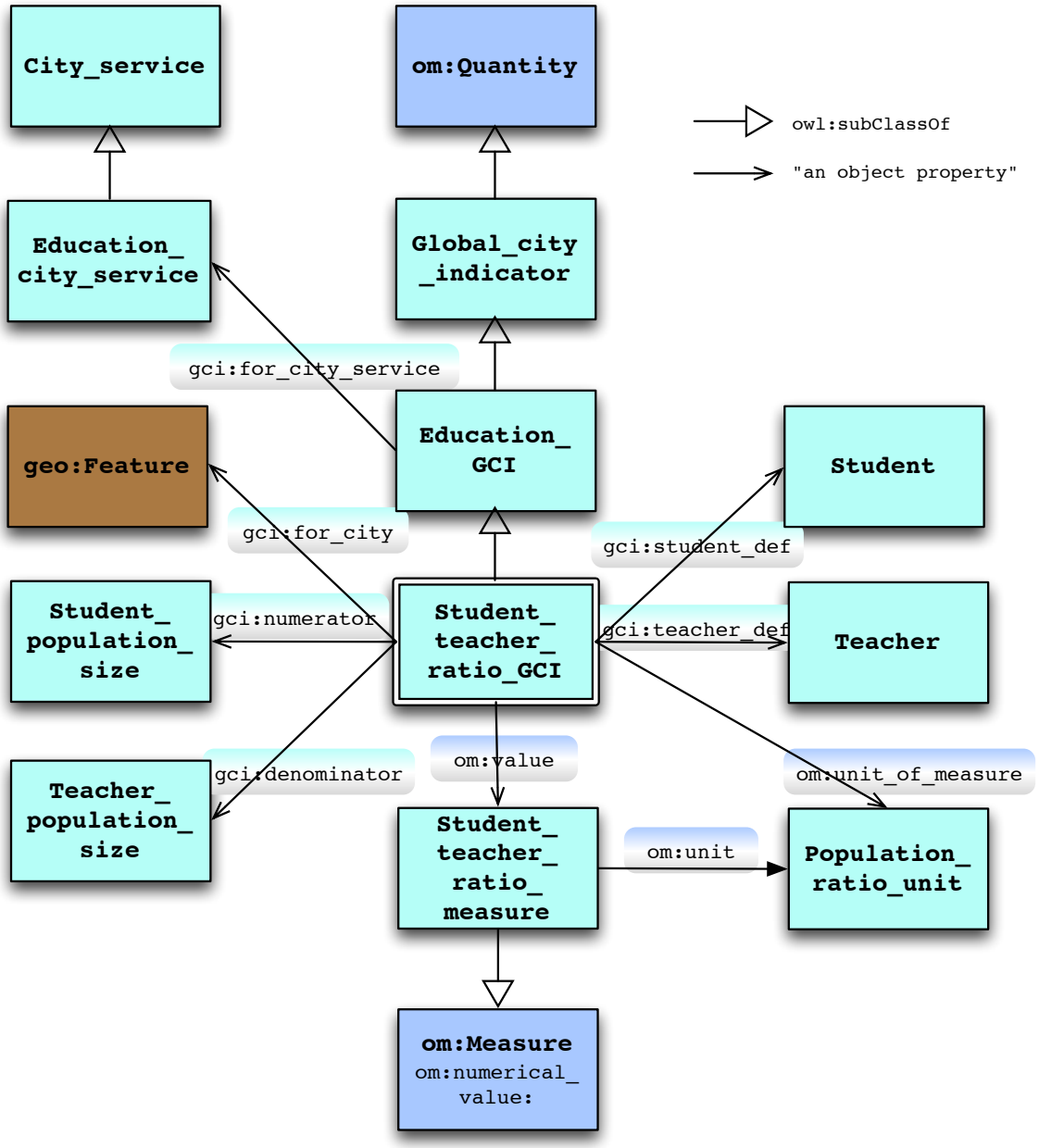

Figure 6

At this point you may have noticed that neither the `gci:Student_population_size` nor `gci:Teacher_population_size` have been linked to the students nor teachers within a city. We do so in the next section where we introduce the statistics ontology.

3.4. Statistics Ontology

- What defines the members of a population?
- Over what area is the population being drawn from?
- What is its unit of measure?
- How is the population size estimated?

The STR 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 STR requires a count of the population, other measures would require statistical measures of mean, deviation, etc. of other characteristics of the population.

Anticipating the larger requirements of the Global City Ontology, we have adopted the GovStat[11] general statistics ontology (Pattuelli, 2009). Figure 7 depicts the main classes and properties of the GovStat ontology. The core class is the `gs:Population` to be measured. (A definition of the population is not provided and will be part of our extension to GovStat in Figure 8.) 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`. Finally, the variable for which the parameter is being measured is defined by the class `gs:Observation` which `gs:Statistic` links to via the property `gs:is_composed_of`, and the actual variable which is a subclass of `gs:Variable` is linked to `gs:Observation` via the property `gs:is_a_characteristic_of`.

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

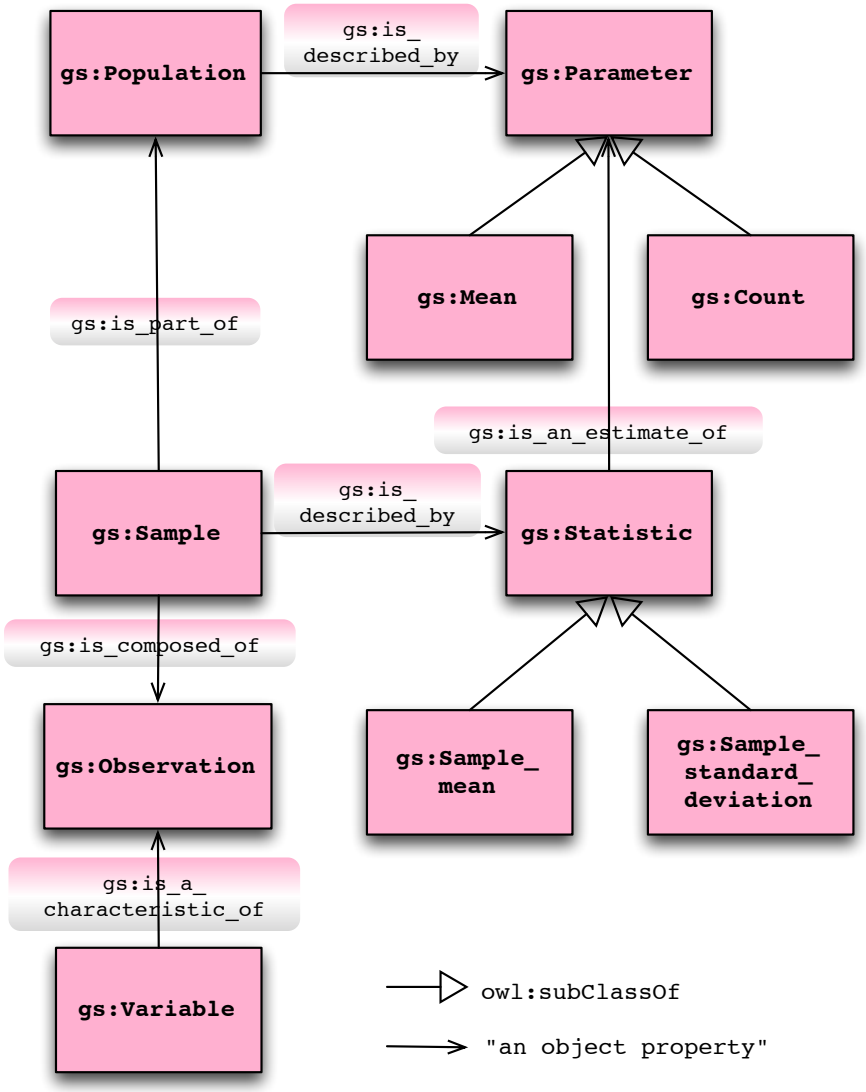

Figure 7

What is missing at this point is a definition of the population that we are measuring or from which a sample is to be taken. For the STR 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.

As depicted in Figure 8, 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 class that all members of the Population are subsumed by.

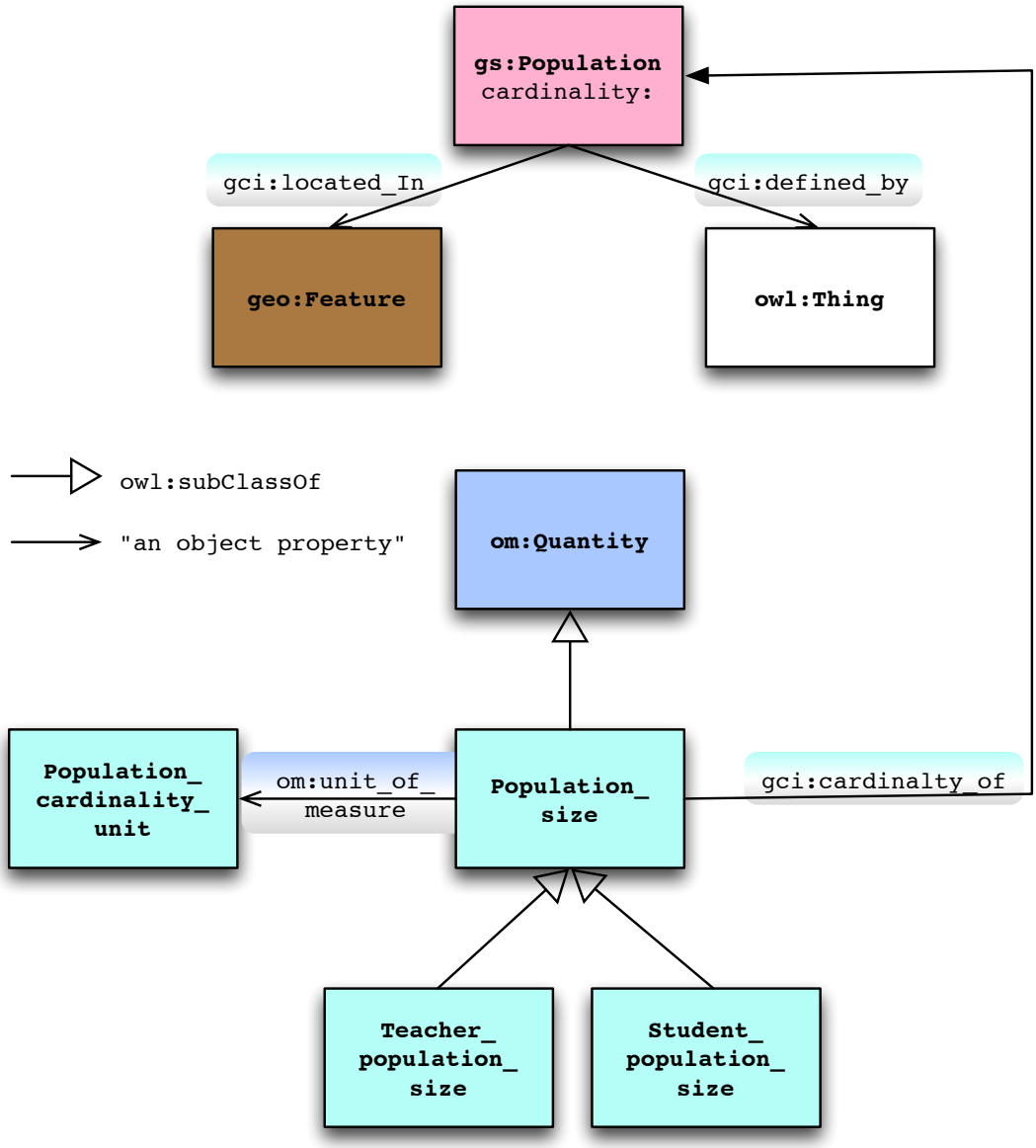

Figure 8

In order to complete definition of `gci:Population_size` pictured above, we need a further constraint. The property of (`gs:is_property_of`) the `gci:Population` must be a `gs:Count` parameter.

3.5. Summary

In this section we provided the foundation ontologies necessary for representing the definitions of GCIs. Our goal is to provide a precise representation of the semantic intent of an indicator's definition as provided in ISO 37120. By providing the definition, we have made it possible to automate tasks such as checking that the data provided by a city is consistent with the ISO definitions. We have also made it possible to automate longitudinal and transversal analysis.

4. Indicator Meta Information

Where the previous section focused on representing indicator definitions as found in ISO 37120, this section focuses on the representation of meta-information associated with the actual data provided by cities.

4.1. Provenance Ontology

- Who created the actual value of the GCI?
- When was it created?
- What process was used to create it?
- What datasets is it based on?
- Has this GCI been revised?

An important aspect of an indicator is its provenance, namely where did it come from and how was it derived. Over the last decade, concerns around information validity, provenance and trust have grown. With the web now containing trillions of documents authored by millions of people, the need to know whether the content is valid, where the content came from and whether to trust its creator, has taken on an increasing importance.

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[12] (Belhajjame et al., 2012), which is based on the work of Hartig & Zhao (2010) and Moreau et al. (2010). We have chosen to use PROV as it satisfies the competency questions above and it is an emerging standard. In the following we outline the basic concepts of the PROV ontology and indicate how it is incorporated into the GCI ontology.

At the heart of the PROV ontology are three classes:

- **`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:

- **`pr:wasGeneratedBy`**: It links an `pr:Entity` (domain) to a `pr:Activity` (range), identifying the activity that generated the entity.

[12] 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.

- **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:wasAssociatedWith**: It links an pr:Activity (domain) to a pr:Agent (range), identifying the agents that play a role in the activity.
- **pr:wasAttributedTo**: It links an pr:Entity (domain) to an pr:Agent (range), identifying the agents that had a role in creating the entity.
- **pr:wasRevisionOf**: Links two pr:Entity's where domain entity is a revision of the range entity.
- **pr:wasDerivedFrom**: Links two pr:Entity's where domain entity was derived from the range entity (without indicating the method of derivation).

Finally, the PROV ontology provides a time property that specifies the time an entity was created.

- **pr:generatedAtTime**: It links a pr:Entity (domain) to a pr:time (range), identifying the time the entity was generated.

Figure 9 depicts the integration of the PROV ontology into the GCI ontology. First, om:Measure is a owl:subClassOf pr:Entity. Consequently, every indicator's Measure we create will be treated as a pr:Entity and inherit its properties, including pr:generatedAtTime which provides us with the time that the indicator was created, and pr:wasRevisionOf which allows us to track revisions to the value of the indicator. It also allows us to link the GCIs to a pr:Activity via a pr:wasGeneratedBy to show what activity generated the GCI measure, and to a pr:Agent via a pr:wasAttributedTo to show who the source of the GCI measure was. Finally, the gci:numerator and gci:denominator are made to be owl:subPropertyOf pr:wasDerivedFrom to show what entities were used to derive the GCI (not shown).

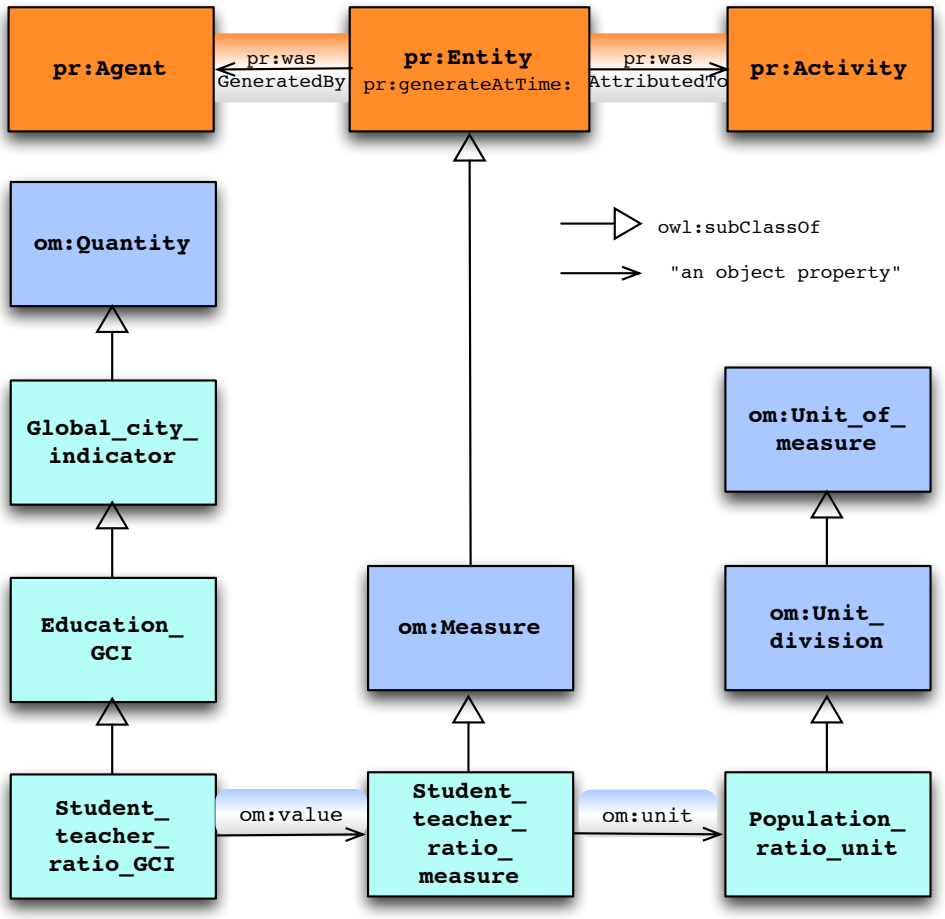

Figure 9

4.2. Time Ontology

- When was this GCI constructed?
- Over what period of time should the GCI be considered valid?
- What the time of the measure before, during or after an event?
- Was the teacher population sizing done during the same time that the student population sizing was done?

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 its 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 amongst them.

Many time ontologies have been developed. We have chosen OWL-Time[13] for its simplicity and ability to represent time as a point or interval. OWL-Time is based on the work of Allen & Ferguson (1997) and described in Hobbs & Pan (2006).

The root class for OWL-Time is the `ot:TemporalEntity`. It has two subclasses:

- **`ot:Instant`**: It represents a time point.
- **`ot:Interval`**: It represents a period of time with a beginning and an end. An `ot:ProperInterval` is an ot:Interval where the start time is less than the end time.

An **`ot:Interval`**'s starting point, ending point and duration are denoted by the following properties:

- **`ot:hasBeginning`**: links a `ot:TemporalEntity` (domain) to an `ot:Instant` (range) where the latter denotes the beginning of the `ot:TemporalEntity`.
- **`ot:hasEnd:`** links a `ot:TemporalEntity` (domain) to an `ot:Instant` (range) where the latter denotes the end of the `ot:TemporalEntity`.
- **`ot:hasDurationDescription`**: links a `ot:TemporalEntity` (domain) to an `ot:Interval` (range) where the latter denotes the duration of the `ot:DurationDescription`.

Two other classes of note are:

- **`ot:DateTimeDescription`**: A specification of a date plus time using a year, month, day, hour, etc. set of properties.
- **`ot:DurationDescription`**: is a class whose instance can combine multiple descriptions such as 2 days and 2 hours to specify a duration.

Finally, there is a set of properties that relate `ot:ProperInterval`'s, including `ot:inside`, `ot:intervalOverlaps`, `ot:intervalAfter`, `ot:intervalContains`, etc.

Figure 10 depicts the integration of the time ontology to the GCI ontology. The integration of the time ontology occurs with the `pr:Entity`. We modify `pr:generatedAtTime` from being a data property to an object property whose range is an `ot:TemporalEntity`. With this change the time that a `pr:Entity` is generated can be either a point or an interval. Secondly, it can take advantage of the relational reasoning (i.e., is the generation time of a `pr:Entity` before, during, or after some other `pr:Entity` or event) supported by the ontology.

[13] 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.

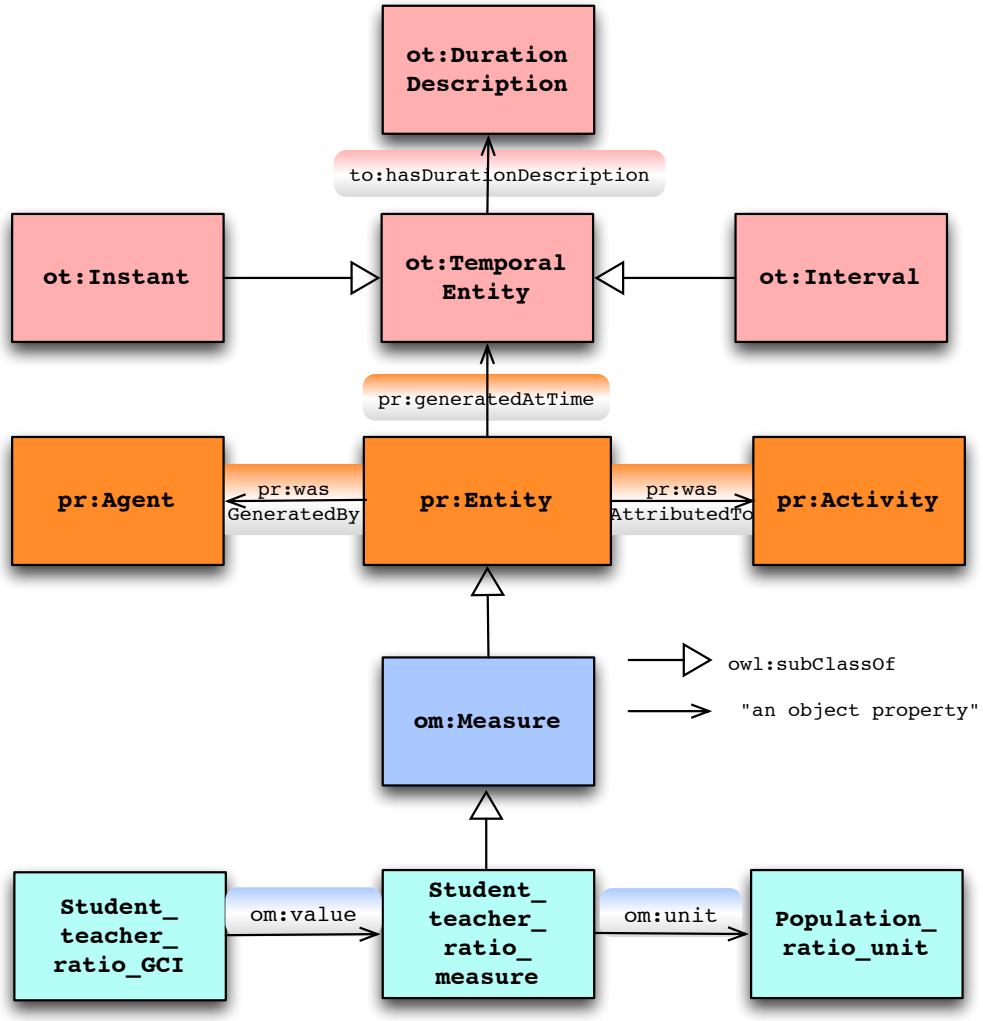

Figure 10

4.3. Validity Ontology

- Is the GCI believed to be an accurate measure by its creator?
- Over what time is it believed to be accurate?

An ongoing issue with the web is whether information/data found on a page is correct (true) or incorrect (false). Whether the creator of the information deliberately makes false statements, or unknowingly copies false information from another site, there is no way to discern what is correct from incorrect. The same holds with city indicators. Data and analyses 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 information is true or not, especially if the indicator is based on a sampling of a population, but one can assign a degree of validity to the information. In addition, in the case where data 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, `gci:Student_teacher_ratio` is derived from `gci:Student_population_size` and `gci:Teacher_population_size`, if `gci:Student_population_size` is valid only within an interval of time such as the year for which it is gathered, then outside of that interval, both

gci:Student_population_size's and its dependent gci:Student_teacher_ratio's validity are unknown.

Fox & Huang (2005) provide an ontology, called the Knowledge Provenance Ontology[14] (KP), for representing the validity (certainty) of a proposition. It assigns to a "proposition" 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.

Huang & Fox (2004a; 2004b) define axioms for propagating validity within a dependency network. In the simple case, if a GCI is assigned a validity of 1 (i.e., it is true) but it also has an effective time interval specified for it, then the GCI is valid during that time interval and unknown otherwise. If any GCIs' validity are unknown during a time interval then any GCIs' validity that depend on it are also unknown during the same time interval.

At the core of KP is the kp:KP_prop class which identifies a proposition to which a validity, effective time interval and dependencies can be assigned. We add to the definition of om:Measure that it is a owl:SubClassOf kp:KP_prop (Figure 11). Hence any gci:Global_city_indicator measure is also a proposition to which we can assign a validity, effective time interval and dependencies.

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

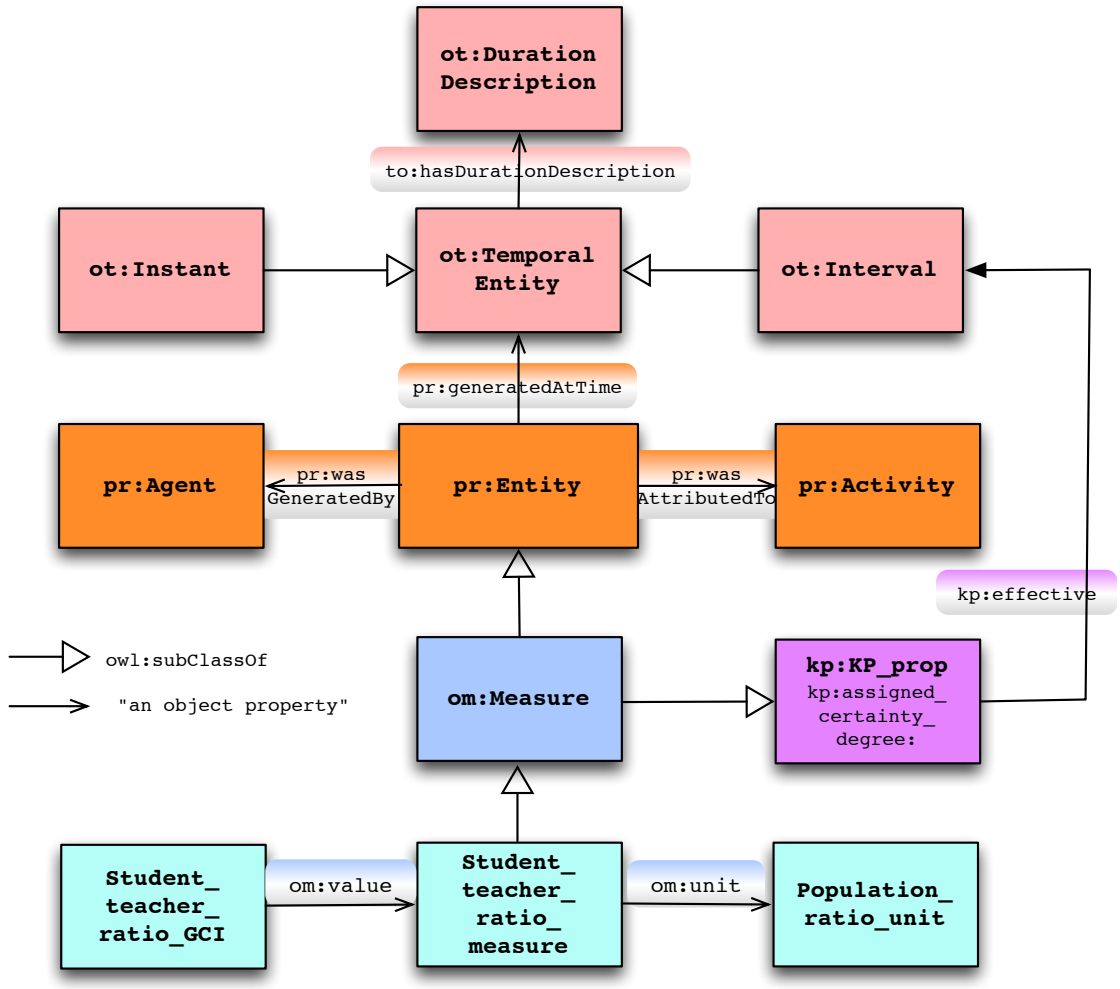

Figure 11

The following properties are associated with a `kp:KP_prop` and are inherited by all `gci:Global_city_indicator measures`'s:

- **`kp:assigned_certainty_degree`**: This is a data property that maps a `kp:KP_prop` (domain) onto a number [0,1] (range) or unknown. It is the degree of certainty that the proposition is valid (true) from the perspective of the creator of the `gci:Global_city_indicator` measure.
- **`kp:effective`**: This is a data property that maps a `kp:KP_prop` (domain) onto a time interval (range). It is the time during which the `kp:assigned_certainy_degree` is valid for the `gci:Global_city_indicator` measure.
- **`kp:is_dependent_on`**: This is a object property that maps a `kp:KP_prop` (domain) onto another `kp:KP_prop` (range). It states that the `kp:assigned_certainty_degree` for the `gci:Global_city_indicator` measure is dependent upon one or more `kp:KP_prop`'s.

Given that `kp:is_dependent_on` is a generalization of `pr:wasDerivedFrom`, we add to the KP ontology that `pr:wasDerivedFrom` is an `owl:subPropertyOf` `kp:is_depedendent_on (not shown)`.

4.4. Dynamic Placenames

- Has the city's boundary changed during the time between two measures of an indicator?

Consider the unique placename for the City of Toronto. If we wish to do a longitudinal analysis of an indicator for Toronto, we run into a problem. The geographic definition of Toronto changed in 1998 after its amalgamation with five adjacent municipalities. Yet in the Geonames ontology there is a single Toronto; there is no representation for how placenames evolve over time. Kauppinen and Hyvönen (2007) have addressed this problem. They propose an ontology based on Spatial Temporal Regions. A placename has associated with it a spatial region, defined by a polygon, and a time interval over which the placename and the region do not change.

In the Global City Ontology we will refer to placenames whose spatial regions can change over time as Dynamic Placenames. Rather than adopt Kauppinen and Hyvönen's terminology directly, we adapt their ideas by reusing the provenance, time and validity ontologies to represent how place names change over time and the cause of their change.

Figure 12 depicts a simplified example of how to represent a dynamic placename for the city of Toronto. First, the placename for each version of the City of Toronto will have to be unique. In the example we append the time period for each version to the name, though just having a unique number is sufficient. We link the [1998-] version of Toronto to the [1967-1998] version via a `pr:wasRevisionOf` property from the provenance ontology to show that the former is a revision of the latter. Secondly, for each placename we link it using the validity property `kp:effective` to a time interval over which the placename is valid.

From a longitudinal analysis perspective, when we compare indicators for a single city over time, we will know the extent to which these comparisons are valid as the city's composition may have changed over time.

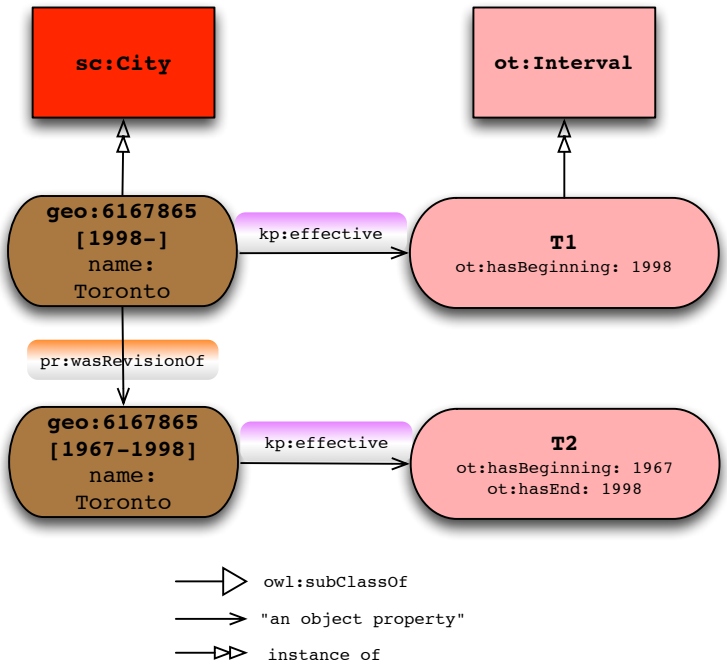

Figure 12

4.5. Trust Ontology

- Do you trust the creator of the GCI?
- Do you trust the process used to create the GCI?
- How does your trust affect the validity of the GCI?

The final piece of the GCI ontology foundation is the representation of trust. The problem we wish to address is how to represent the degree of trust we have in the creator of indicator values and the data from which they are derived. Huang & Fox (2006) define trust as follows:

> *"Trust is the psychological state comprising (1) expectancy: the trustor expects a specific behavior of the trustee such as providing valid information or effectively performing cooperative actions; (2) belief: the trustor believes that expectancy is true, based on evidence of the trustee's competence and goodwill; (3) willingness to be vulnerable: the trustor is willing to be vulnerable to that belief in a specific context where the information is used or the actions are applied."*

This representation of trust differs from degree of validity as trust refers not to the degree of certainty in the data but our trust in the agent/organization that produced the data. 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 data or indicator will be reduced by this lack of trust.

Huang & Fox (2006) and Huang (2008) provide an ontology of trust[15]. The ontology views trust as occurring between two agents, where $agent_1$ has or has not trust in $agent_2$. Trust arises out of direct experience or the experience of others whom you may trust. Trust is also 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. Finally, they identify two types of trust: 1) *trust in belief*, where $agent_1$ believes what $agent_2$ believes, and 2) *trust in performance*, where $agent_1$ believes that $agent_2$ will perform an activity properly.

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

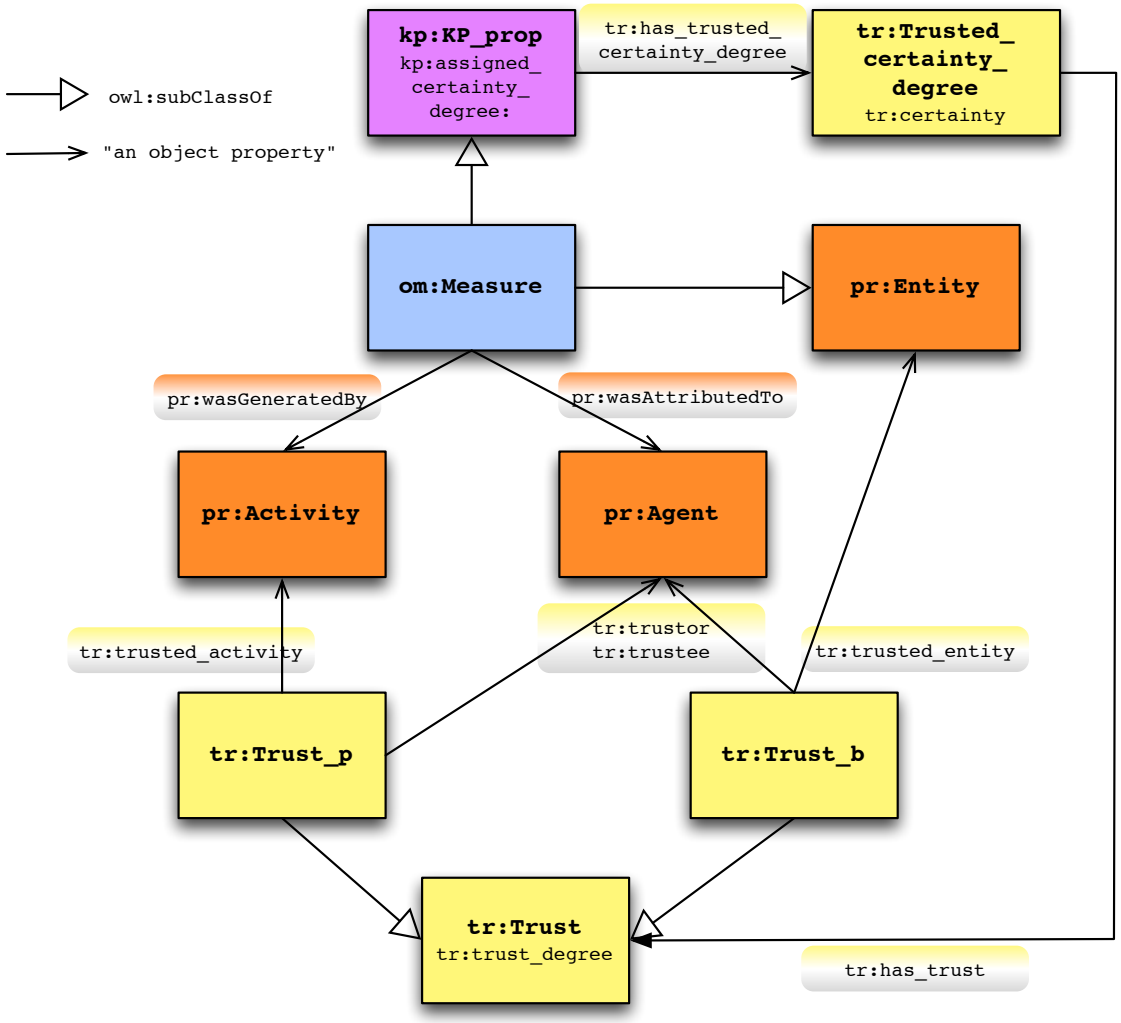

Figure 13

The Trust Ontology also addresses how the validity of an indicator or data changes by taking the original degree of validity, asserted by the creator (agent$_2$), and modifying it by the degree of trust the "user" (agent$_1$) has in the creator. This resultant validity is dependent on agent$_1$ and agent$_2$.

In Figure 13 we integrate the Trust Ontology by defining the class `tr:Trust` which has a data property `tr:trust_degree` which is the degree to which the `tr:trustor` trusts the `tr:trustee`. It is specialized into two classes:

- **`tr:Trust_p`**, which is trust in performance. It has an object property `tr:trusted_Activity` that links it to a `pr:Activity`, and
- **`tr:Trust_b`**, which is trust in belief. It has an object property `tr:trusted_Entity` that links it to a `pr:Entity`, which all `gci:Global_city_indicator measures` are a subclass of.

We then extend the KP Ontology by adding an object property, `tr:Trusted_certainty_degree`, that links a `kp:KP_prop` to a new class that represents the certainty degree computed by combining the GCI's `tr:assigned_certainty_degree` provided by its creator (who is the `tr:trustee`) and the trust that the user (who is the `tr:trustor`) has in the creator. The latter is represented

by an object property `tr:has_Trust` that links the `tr:Trusted_certainty_degree` to the `tr:Trust` that represents the user's belief in the creator.

5. Example

In this section we show how a specific instance of the STR used throughout this paper is represented as instances of the ontology. The instances are represented as bifurcated rectangles where the top part identifies the name of the instance followed by the class it is an instance of (<instance name> -->> <class name>). The bottom part contains data properties and object properties for which, for brevity, we do not depict using a link.

Let's assume that we want to create a STR for the city of Toronto, what we have to do is (Figure 14):

- Create an instance (`ex:TO_str`) of quantity `gci:Student_teacher_ration_GCI`. This will be starting point of the value for the city of Toronto's student teacher ratio indicator. Set the `gci:for_city` object property to the URI of the Toronto placename.
- Create an instance of measure `gci:Student_teacher_ratio_measure` (`ex:TO_str_m`), fill the `om:numerical_value` property with the actual ratio (40 in this example). Link `ex:TO_str` to `ex:TO_str_m` using the `om:value` object property.

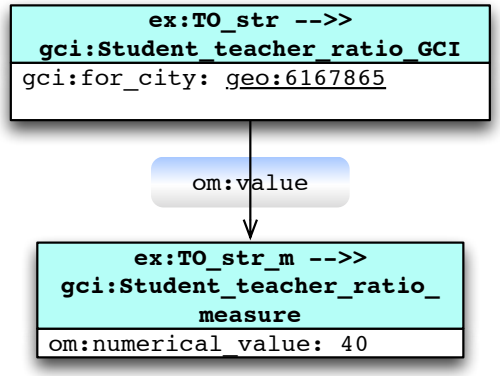

Figure 14

With this representation, based on the classes these are instances of, we know that `ex:TO_str` represents a Student-Teacher ratio for the Education city service for the city of Toronto. The unit of measure is the ratio of two Population sizes, where the populations are defined by the Student and Teacher classes, respectively, which we have yet to define.

Next we will add two types of provenance (Figure 15):

- The date/time this ratio was created (23 January 2013 at 10am), and
- The Agent who created it (Joe Smith),

by instantiating their respective classes.

We then add the validity of the indicator. Since `gci:Student_teacher_ratio_GCI` is a subclass of `kp:KP_prop`, it inherits the `kp:assigned_certainty_degree` data property which we set to 1, i.e., the creator believes the value of the indicator is true.

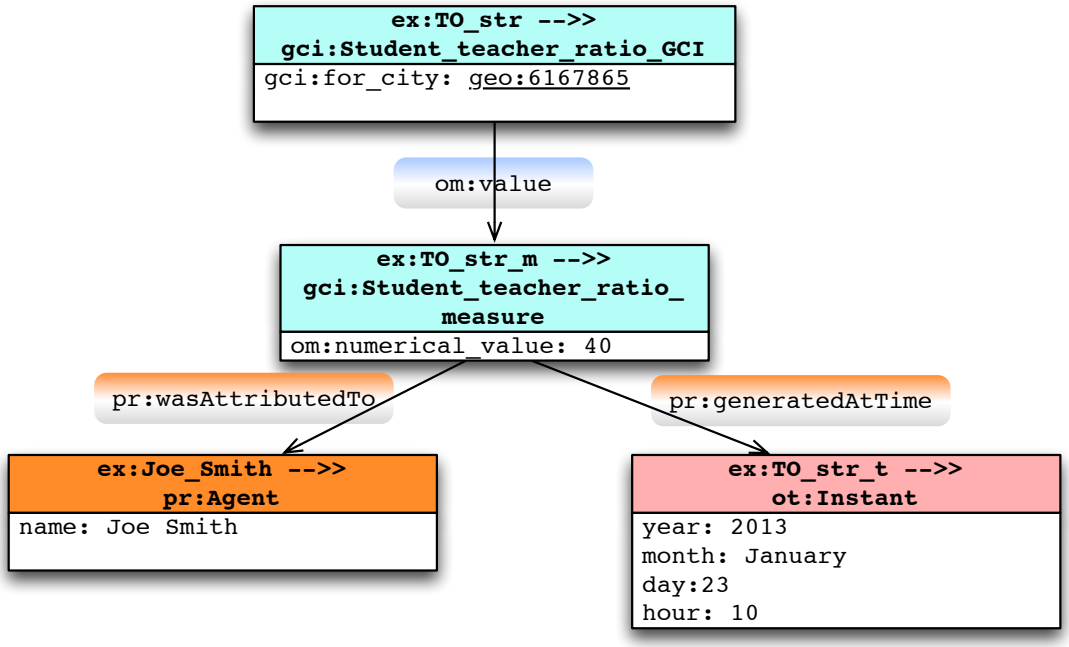

Figure 15

Based on the above, we know that the source of the indicator was Joe Smith, which he created on January 23rd, 2013 at 10am.

Lastly (Figure 16), we add the degree we trust Joe Smith by doing the following:

- Adding the trustor, Mark Fox, and the degree to which he trusts Joe Smith by instantiating the trust in belief class (`tr:Trust_b`).
- Adding the trusted certainty degree to the Student Teacher ratio.

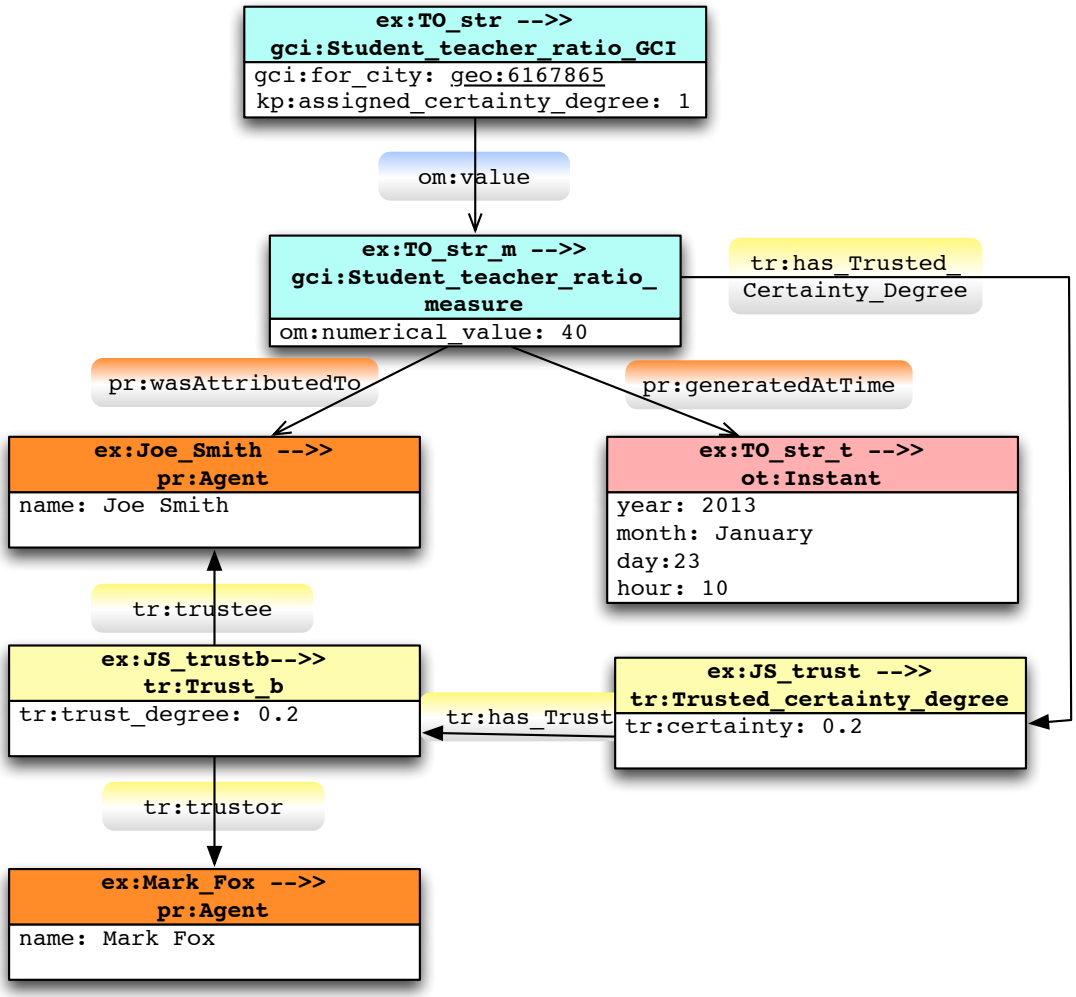

Figure 16

With this additional information, we know that although Joe Smith believes the STR to be correct, Mark Fox assigns a trusted certainty degree of 0.2 because he does not trust what Joe Smith believes.

The example to this point shows how a the single number that represents the student teacher ration is represented using the OM ontology, along with the meta information representing provenance, validity and trust. If the supporting information is available, we can extend this example to represent what data was used to derive the STR.

Figure 17 depicts how the denominator of the STR is represented. `ex:TO_tpopsize` is an instance of `gci:Teacher_population_size` with the value specified by the `om:value` property linking it to a measure `ex:TO_tps_m`, which in turn specifies the teacher population size to be 1 `kilopcs`. The teacher population size is the cardinality of the population `ex:TO_tpop` linked to by `gci:cardinality`. This population is located in Toronto, and its membership is defined by the class of teachers who teach fulltime in primary school (this class definition has been abbreviated).

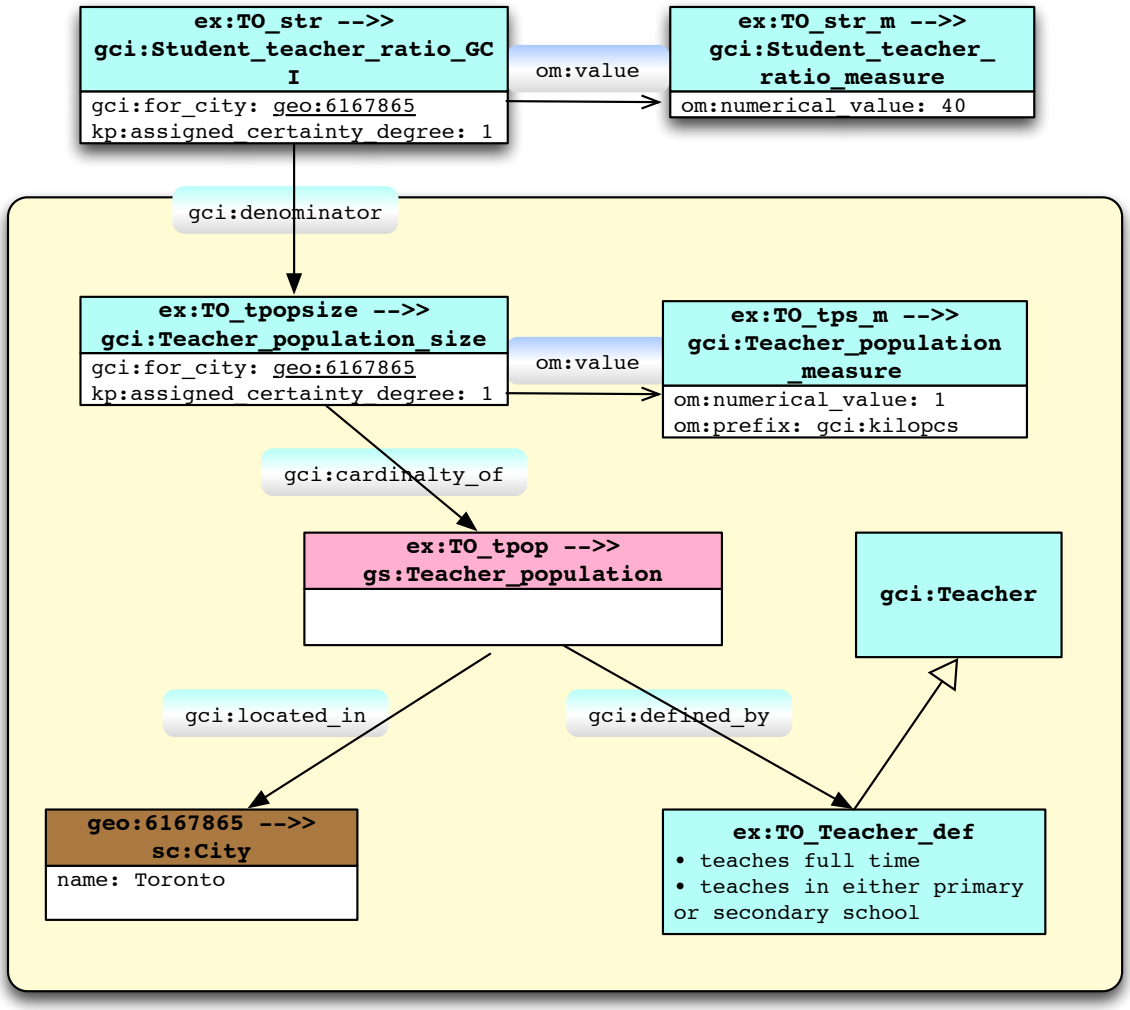

Figure 17

The numerator, i.e., the number of students, is defined similarly. Where they differ is in the definition of population membership, namely the definition of students and teachers. In order to precisely define a student and a teacher, we need an Education Ontology that covers the main points raised in the ISO 37120 definition, such as, primary vs. secondary school, administrative vs. teaching staff, school age students, cohorts, etc. An Education Ontology along with the complete definitions of the ISO37120 educational indicators can be found in Fox (2014).

6. Consistency

Given an indicator definition and indicator data provided by a city in the form of instances of the GCI ontology, it is now possible to automatically detect inconsistencies in the data using a set of consistency rules (axioms). These rules make sure the various parts of an indicator's instance data refer to the same places, have the same units, etc. They are obvious and simple extensions to the ontology, but necessary in that they can detect errors that commonly occur in datasets.

In the following we informally describe each rule which is implemented in prolog.

6.1. Placename Rules

The purpose of these rules is to check that the City associated with the STR is consistent with the cities associated with each of the Teacher and Student populations. We want to assure that they are referring to and measuring the populations for the same geographic area.

> **Rule G1:** The city for the STR being measured is the same as the cities where its numerator and denominator are measured.

The city for the STR is defined by its `gci:city` property. The placename for the city must match the placename specified in the `gci:located_in` property attached to the `gs:Population` that is linked to the `gci:Population_size` class via a `gci:cardinality_of`. See Figure 18.

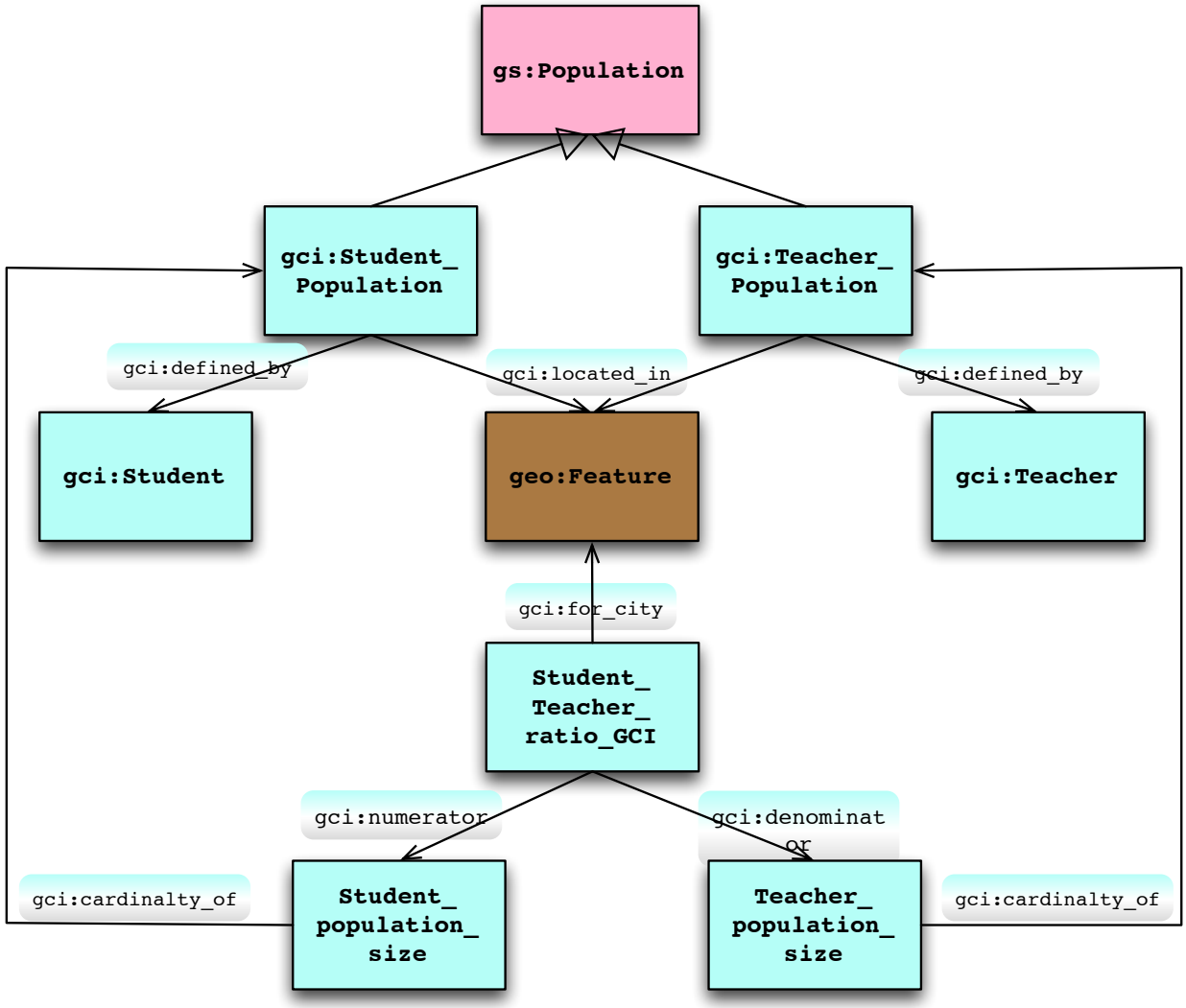

Figure 18

6.2. Measurement Rules

These rules define the consistency of measurements used in a STR.

Rule M1: The numerator and denominator of a `gci:Student_teacher_ratio_GCI` are the correct type.

Rule M1 verifies that the numerator and denominator are of the types specified by the definition of the `gci:Student_teacher_ratio_GCI`, namely that the numerator is a `gci:Student_population_size` and the denominator is a `gci:Teacher_population_size`. The rule is generalized to apply to any GCI that is a ratio.

Rule M2: The numerator and denominator of the `gci:Student_teacher_ratio_GCI` are consistent with the numerator and denominator of its unit of measure.

Rule M2 verifies that the unit of measures of the STR's numerator and denominator, are consistent with the units specified by the `gci:Population_ratio_unit`. As defined, the `gci:Population_ratio_unit`'s numerator and denominator have to be `gci:Population_cardinality_unit`'s. The rule is generalized to apply to any GCI that is a ratio.

Rule M3: If the numerator and denominator of a `gci:Student_teacher_ratio_GCI` are the same type, then they should have the same unit of measure.

Since the numerator and denominator of the STR are population counts, then Rule M3 verifies that the numerator and denominator are of the same units, e.g., you cannot have `gci:Student_population_size` measured in `gci:kilopcs` and `gci:Teacher_population_size` measured in `gci:pcs`. The rule is generalized to apply to any GCI that is a ratio.

Rule M4: The units of the actual measurement are the same as defined by GCI it is a measure of.

In the STR case, the unit of measure of the `gci:Student_teacher_ratio_GCI` and the unit for the `gci:Student_teacher_ratio_measure` must be the same, namely a `gci:Population_ratio_unit`.

Rule M5: The value of the `gci:Student_teacher_ratio_measure` is equal to the value of the `gci:Student_teacher_ratio_GCI` numerator divided by the denominator.

Though rule M5 is obvious, it is still necessary to check that calculations are performed correctly.

6.3. Statistics Rules

The statistics rules assure that the populations being measured are consistent with the indicator in which they are being used.

> **Rule S1:** The definitions of student and teacher as specified by the `gci:Student_teacher_ratio_GCI` are the same as used by its numerator and denominator.

This rule checks to see that the `gci:Student_population` pointed to by the STR's numerator and the `gci:Teacher_population` pointed to by the STR's denominator have `gci:defined_by` ranges that are consistent with the STR's `gci:student_def` and `gci:teacher_def` respectively.

6.4. Provenance Rules

Provenance is used to document the participants and the process used to generate an indicator. It is used for forensic purposes such as determining whether the same methods were used to generate different versions of the same indicator for the same city.

> **Rule P1:** Two versions of the same indicator are inconsistent with each other if different methods were used to generate them.

Consider the situation where the same indicator is measured annually. Each version of the indicator is linked to the prior year's version via `pr:wasRevisionOf` property. In order to assure that the two versions are comparable, we have to assure that the same methodology was used for each. This is done by comparing the `pr:Activity` used to generate numerator for each version, and doing the same for the denominator.

> **Rule P2:** Two versions of the same indicator are inconsistent with each other if the cities are not the same.

A difference of cities can arise because the wrong placenames have been used, or in a dynamic placename situation, the city itself has undergone a change, such as a merger, during the last year.

One could imagine a rule that relates the time the indicator was generated, as recorded by its provenance, with its effective period, but its effective period could either be prior to generation or after, depending on policy.

6.5. Validity Rules

The validity rules assure that the time period for which the STR is specified to be valid, is consistent with the data from which it is derived.

> **Rule V1:** The effective time period for which the STR is valid is contained within the effective time periods of its numerator and denominator.

The numerator and denominator of the STR, namely the `gci:Student_population_size` and `gci:Teacher_population_size`, must have effective time periods that at least overlap, and the STR's effective time period must be contained within that overlap.

> **Rule V2:** The `kp:assigned_certainty_degree` of the STR is less than or equal to the max of the `kp:assigned_certainty_degree` of its numerator and denominator, and greater than or equal to the min of its numerator and denominator.

The STR is a function of the Student and Teacher population sizes. Hence it's trust degree cannot be more/less than the max/min of the individual trusts of the measures it depends on.

6.6. Trust Rules

The trust rules assure that the agents in the trust relations are consistent and they refer to the same Entities.

> **Rule T1:** The trustee in a trust relationship is the same as the `pr:wasAttributedTo pr:Agent` for an indicator.

The STR indicator must be linked to a `tr:Trusted_certainty_degree` which is in turn linked to a `tr:Trust_b` (i.e., trust in the belief of an `pr:Agent`), which in turn points to a `tr:trustee pr:Agent`, that trustee must also be the `pr:Agent` that created the indicator (which is pointed to by the `pr:wasAttributedTo` link of the STR indicator).

> **Rule T2:** The trusted certainty degree of an indicator should have a Trust instance that links to the indicator via the `tr:trusted_entity` property.

> **Rule T3:** The trusted certainty degree of an indicator is less than or equal to the indicator's certainty assigned by its creator.

The point here is that the trusted certainty degree cannot be greater than the certainty assigned by the indicator's creator. It can only be reduced.

7. Evaluation

We approach the evaluation of Global City Indicator Foundation ontology from four perspectives:

1. Is the ontology Competent? In Grüninger & Fox (1995), the requirements of 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.
2. Is the ontology Consistent? An OWL ontology is inconsistent if it contains a class that cannot possibly have any instances.
3. Is the ontology General? Are the classes general enough to represent other indicators and can it be easily extended where necessary?
4. Do we satisfy the aspects of a good indicator as identified in Hoornweg et al. (2007)?

7.1. Competency

The competence of an ontology is defined by a set of questions the ontology must be able to answer. These questions fall into three categories:

1. Questions that require a simple retrieval of the value of a property. For example, the city of a particular indicator.
2. Questions that require the following of one or more links (properties) in the network. For example, measurement consistency rule M1.
3. Questions that require some type of computation. For example, longitudinal or transversal analysis.

Regarding category 1, it is clear from the representation what data can be directly retrieved. Regarding category 2, the section on consistency defines a set of consistency competency questions in the form of rules. These rules have been implemented and tested, and are available for review (see Appendix). Regarding category 3, our future work will explore the types of analysis questions that need to be answered and any further extensions to the ontologies required.

7.2. Consistency

Each of the ontologies used herein were tested for consistency using RacerPro v2.0. Individually and together all of the ontologies was found to be consistent as described in Fox (2013).

7.3. Generality

A major goal of the development of the Global City Indicator ontology for STR is to make it as general as possible so that it can be reused across the remaining indicators. Seven ontology modules were used in the STR example: Placename, Measurement, Statistics, Provenance, Time, Validity and Trust. Some of these modules were externally developed and used without change or extensions, such as Time and Provenance, some were extended significantly, such as Measurement.

An analysis of the complete set of ISO 37120 indicators has shown that the foundation ontologies described in this paper apply across all categories of indicators (Fox, 2013). But additional ontologies will have to be developed, specific to the indicator category, in order to represent the remaining indicators. 17 new modules have been identified:

- Census
- Economy
- Education
- Energy
- Environment
- Geography
- Fire
- Health
- Municipal Finance
- Municipal Governance
- Recreation
- Safety
- Shelter
- Waste
- Telecommunication
- Transportation
- Urban Ecology

7.4. Aspects of a Good Indicator

In Section 2, we discussed the aspects of a good city indicator defined by Hoornweg et al. (2007). In this section we revisit these aspects from the perspective of what and how the Global City Indicator Ontology achieves these aspects.

- **Objective**: *clear, well defined, precise and unambiguous, simple to understand.*

  The ontology provides a clear, precise representation of an indicator that is grounded in more foundational ontologies such as measurement theory, statistics, etc., This reduces, if not removes in most cases, ambiguity in the interpretation of the indicator.

- **Measurable and replicable**: *easily quantifiable, systematically observable.*

This aspect is not addressed by the ontology.

- **Auditable**: *valid, subject to third-party verification, quality controlled data (legitimacy across users).*

  With the inclusion of provenance, validity and trust information in the ontology, the ability to audit the information is greatly enhanced. Add to it the more detailed information on the populations from which the data is drawn from, the quality of the data can be further verified.

- **Statistically representative** *at the city level.*

  While this aspect is not addressed by the ontology, the detailed representation of the place and populations sampled enables the audit function determine whether the information is statistically representative.

- **Comparable/ Standardized** *longitudinally (over time) and transversally (across cities).*

  The incorporation of dynamic placenames, measurement, time, statistics, and provenance makes it possible to perform longitudinal and transversal analysis and to verify that the data being compared is consistent with each other.

- **Flexible:** *can accommodate continuous improvements to what is measured and how. Have a formal mechanism for all cities and interested parties to comment on.*

  The ontology can be easily extended to include other measures as demonstrated by the generality of the underlying modules (i.e., placenames, provenance, measurement, etc.).

- **Interrelated**: *indicators should be constructed in an interconnected fashion (social, environmental and economics).*

  The Semantic Web's network representation is fundamentally an integrated representation, and enables the integration of indicators and the information they are based on.

- **Consistent and sustainable over time**: *frequently presented and independent of external capacity and funding support.*

  An important aspect of publishing indicators and their supporting data on the Semantic Web is the universal access it provides and its availability over time.

8. Conclusions

Industrial Engineering and Management Science both share the view that you cannot manage what you do not measure. Enhancing the quality 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 Indicator Facility 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 represent the indicators and their definitions so that they can be published, linked, merged, mashed, and analyzed based on the principles of the Semantic Web. This work addresses this third challenge. In doing so it makes two unique contributions:

- It provides an ontology for the precise representation for the definitions of city indicators that can be published on the semantic web, and
- It demonstrates the integration and extension of several core ontologies on a significant and important problem.

There are two directions that our current research is heading. The first direction is to complete the Global City Indicator Ontology to span the entire set of ISO37120 Indicators. For example, Fox (2014) defines an Education Ontology and combines it with the Foundation Ontology described herein to define ISO 37120 education indicators. The second direction is to automate the longitudinal and transversal analyses of city data.

9. Acknowledgements

I would like to thank Hajo Rijgersberg for his feedback on the use of the OM Ontology, and Patricia McCarney for her feedback on Global City Indicators.

10. References

Allen, J. F., and Ferguson, G., (1997), "Actions and events in interval temporal logic." In *Spatial and Temporal Reasoning*. O. Stock, ed., Kluwer, Dordrecht, Netherlands, pp. 205-245.

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

Berners-Lee, T., Hendler, J., and Lassila, O., (2001), "The Semantic Web", *Scientific American*, Vol. 284, No. 5, pp. 28-37.

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", Submitted to *Semantic Web*, IOS Press.

Fox, M.S., (2013), "A Foundation Ontology for Global City Indicators", Working Paper, Global Cities Institute, University of Toronto.

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., and Huang, J., (2005a), “Knowledge Provenance in Enterprise Information”, *International Journal of Production Research,* Vol. 43, No. 20., pp. 4471-4492. http://www.eil.utoronto.ca/km/papers/fox-ijpr05.pdf

Fox, M.S., and Huang, J., (2005b), “An Ontology for Static Knowledge Provenance”, In *Knowledge Sharing in the Integrated Enterprise,* IFIP, Vol. 183/2005, pp. 203-213.

Global City Indicators Facility (2010a). “Overview Report of the Global City Indicators Facility”, GCIF, University of Toronto.

Global City Indicators Facility (2010b). “Global City Indicators©: Definitions and Methodologies” September 2010, GCIF, University of Toronto.

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.

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

Grüninger, M., and 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., and 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.” In *The Semantic Web: Research and Applications* (pp. 708-722). Springer Berlin Heidelberg.

Hobbs, J.R., and 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., and 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., and 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., and Fox, M.S., (2004b), "Uncertainty in Knowledge Provenance", *Proceedings of the European Semantic Web Symposium*, Springer Lecture Notes in Computer Science.

Huang, J., and 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

Kauppinen, T., and Hyvönen, E., (2007), "Modeling and Reasoning About Changes in Ontology Time Series", In *Ontologies: A Handbook of Principles, Concepts, and Applications to Information Systems*, Integrated Series in Information Systems, Vol. 14, pp. 319-338, Springer.

Matuszek, C., et al., (2006), "An Introduction to the Syntax and Content of Cyc", In *Proceedings of the 2006 AAAI Spring Symposium on Formalizing and Compiling Background Knowledge and Its Applications to Knowledge Representation and Question Answering*.

McCarney, P. L., (2011) "Cities and climate change: The challenges for governance" Coordinating Lead Author with H. Blanco, J. Carmin and M. Colley, Ch 9 in *Climate Change and Cities: First Assessment Report of the Urban Climate Change Research Network*, Cambridge University Press.

McCarney, P. L., (2013), "Why Cities? Why Metrics? – Tracking Cities in the 21st Century" Global Cities Institute Working Papers, Vol. 1 No. 1, pp. 1-32, University of Toronto.

Moreau et al., (2010), "The Open Provenance Model Core Specification (v1.1)", *Future* Generation Computer Systems. Also see openprovenance.org.

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

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., Wigham, M., and Top, J.L., (2011), "How Semantics can Improve Engineering Processes: A Case of Units of Measure and Quantities", *Advanced Engineering Informatics*, Vol. 25, pp. 276-287.

Uceda-Sosa, R., Srivastava, B., and Schloss, B., (2012), "Building a Highly Consumable Semantic Model for Smarter Cities", *In Proceedings of the workshop on AI for an Intelligent Planet,* ACM.

Westfall, M.S., and d Villa, V.A., (2001), *Urban Indicators for Managing Cities*, Asian Development Bank.

World Bank, (2008), "Global City Indicators Program Report: Preliminary Final Report", April 2008.