Expert Systems With Applications 186 (2021) 115734

Contents lists available at ScienceDirect

Expert Systems With Applications

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

An ontology engineering approach to measuring city education system performance

Mark S. Fox [1]

Enterprise Integration Laboratory, University of Toronto, 5 King's College Road, Toronto, ON M5S 3G8, Canada

ARTICLE INFO

Keywords:
Education Ontology
City Education System Indicators
Education Data
Architectures for education technology system
Semantic Web, ISO 37,120

ABSTRACT

This paper addresses the problem of how to represent education system measurement definitions and the data used to derive them. ISO 37120 is standard for measuring city performance. It defines 100 indicators of which seven focus on Education. As cities adopt the standard and begin publishing their indicator values, citizens, city bureaucrats and politicians will be able to see how well their education systems perform relative to other cities. Hence, education systems will be subject to greater scrutiny and will have to provide evidence-based explanations for their performance. In order to enable the *automated* analysis of city performance, four problems must be solved: 1) how are indicator definitions represented? 2) how is the data used to derive an indicator value represented? 3) how is educational system specific knowledge represented, and 4) how is city specific educational system knowledge represented? In this paper we present an Education Ontology designed to represent ISO 37120 education indicators and the information delineated above. The ontology is designed to enable an intelligent agent to perform consistency and root cause analysis of city indicators. It can also be used by cities to publish their education indicators and data, using Semantic Web standards, on their Open Data portals.

1. Introduction

ISO 37120 (ISO 37120, 2014) is a global standard for measuring the performance of cities. It is comprised of 100 city indicators, divided amongst 17 Themes including Education, Energy, Health, Safety, Finance, and Shelter. A standard set of indicators makes it possible for cities to comparatively analyse their performance in each of the themes (i.e., longitudinal analysis: analyzing the changing performance of a city over time, and transversal analysis: analyzing the differences in performance of two or more cities), and use the results of the analysis as a basis for improvement. As cities across the world adopt the standard and begin publishing their indicator values (see dataforcities.org) and the data used to derive them on their open data web sites, citizens, city bureaucrats and politicians will be able to see how well their education systems perform relative to other cities, and determine the root causes of performance. Hence, education systems will be subject to greater scrutiny and will have to provide evidence-based explanations for their performance[2].

Consider an indicator that measures the ratio of students to teachers at the primary grade level. On the surface the definition is simple, being the ratio of the number of students to the number of teachers. The ISO 37120 version of this indicator, educational indicator 6.4: "Primary Education Student/Teacher Ratio", contains more detail:

> "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,

E-mail address: msf@eil.utoronto.ca.
[1] ORCID: 0000-0001-7444-6310.
[2] de Freitas and Dixon-Román (2017) and Gulson and Webb (2017) eloquently express their concerns regarding the impact of datafication on education policy.

https://doi.org/10.1016/j.eswa.2021.115734
Received 20 August 2020; Received in revised form 14 July 2021; Accepted 4 August 2021
Available online 21 August 2021
0957-4174/

and a teacher working three days per week should be counted as 0.6 teachers."

A major issue facing any indicator-based analysis is whether cities interpret an indicator in the same way. The assumption that cities will adhere to a standard is a strong one, as cities often interpret definitions differently (Hoornweg et al., 2007; Slack, 2017). Indeed, before any meaningful analysis can be performed, three questions with respect to consistency need to be answered: Is a city's interpretation of an indicator (Wang & Fox, 2017):

- *Definitionally consistent*, Is a city's interpretation of the indicator consistent with its definition? E.g., is the definition of student and teacher populations reported by a city consistent with the indicator's definition?
- *Intra-indicator consistent*, Is the data used to derive an indicator's value, internally consistent? E.g., are the student and teacher populations in the indicator from the same time and location?
- *Inter-indicator consistent*, When comparing two cities, do they interpret the indicator definition in the same way? E.g., are the cities' definitions of student and teachers consistent a with each other?

If we are to construct an intelligent agent to analyse a city's indicators or compare two cities' indicators, it needs to understand the following:

1. *Indicator Definition:* the definition of each indicator. Hence we need to translate the ISO definitions from English into a machine understandable representation – this requires an Ontology;
2. *Indicator Theme Knowledge:* a certain amount of "common sense" education knowledge in order to understand the definition. For example, if dealing with primary school student teacher ratio, it needs to understand the concept of school: primary vs secondary, public vs private, student, teacher, grade, etc. – this too requires an Ontology;
3. *City Specific Knowledge:* how the common sense education knowledge uniquely maps onto a city. For example, what are the primary grades for a specific city? What categories of students are allowed to attend primary school? What are public schools (e.g., do they include US charter schools)? This too requires an Ontology; and
4. *City Data:* a city's specific indicator value and (more importantly) the data used to derive it. For example, information about each school in the city, what grades they teach, students, teachers, etc. This information may be available in PDF files or spreadsheets but needs to be translated into a machine understandable representation – this too requires an Ontology.

This paper presents an education ontology that satisfies the aforementioned requirements. It can be used to represent: the definition of education indicators, common education knowledge, city specific education knowledge, and the data used by a city to derive its indicator values. We apply the ontology engineering methodology defined in Grüninger and Fox (1995) to develop and evaluate the ontology, The ontology conforms to Semantic Web standards, enabling the publishing of both knowledge and data on a city's open data web site.

In the remainder of this paper we first review how existing vocabularies and ontologies representing indicators and education related concepts. We then review the ISO 37120 education theme indicators, and for each we define a set of competency questions the ontology must be able to answer in order to represent each indicator's definition. The next section introduces the Education Ontology, followed by a demonstration of how the ISO/IEC 37120 education indicators are represented using it. Next, the ontology is evaluated by demonstrating its use in answering the competency questions, and by showing how the The Common Education Data Standards (CEDS) (CEDS, 2017) maps on the education ontology. Finally we describe how the ontology is used by an intelligent agent to perform definition, longitudinal and transversal analysis.

2. Background

The following review is divided into two parts. The first focuses on ontologies for the representation of indicator definitions, and the second focuses on standards and ontologies for representing Education information.

2.1. Representing indicator definitions

Earlier work in the development of ontologies for city indicators focused on how to represent the meta data associated with a published indicator value (Fox, 2013; Fox, 2015). 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. The Global City Indicator Foundation Ontology[3], developed as part of the PolisGnosis project (Fox, 2017), integrates and extends existing ontologies depicted in Fig. 1:

The ontologies included are:

- Time (Hobbs & Pan, 2006).
- Measurement (Rijgersberg et al., 2011).
- Statistics (Pattuelli, 2003).
- Validity (Fox & Huang, 2005).
- Trust (Huang & Fox, 2006).
- Placenames (www.geonames.org).

At the core of the Foundation ontology is the OM measurement ontology (Rijgersberg et al., 2011). The purpose of a measurement ontology is to provide the underlying semantics of a number, such as what is being measured and the unit of measurement. The importance of grounding an indicator in a measurement ontology is to assure that the numbers are comparable, not that they are measuring the same thing, but the actual measures are of the same type, e.g., the counts of the student and teacher populations, that comprise the ratio of student and teacher population sizes, are of the same scale (i.e., thousands vs millions).

Fig. 2 depicts the basic classes of the OM ontology used to represent an indicator. There are three main classes in OM: a 'Quantity' that denotes what is being measured, e.g., diameter of a ball; a 'Unit of Measure' that denotes how the quantity is measured, e.g., centimeters; and a 'Measure' that denotes the value of the measurement which is linked to the both 'Quantity' and 'Unit of Measure'. For example, Student Teacher Ratio is a subclass of 'Quantity' that has a value that is a subclass of 'Measure' whose units are a 'population ratio unit' which is an instance of 'Unit Division' which is a subclass of 'Unit of Measure'. The actual value measured is a property of the 'Measure' subclass 'Student teacher ratio measure'. In the figure it is also shown Toronto's 2013 Student-Teacher ratio indicator Quantity and Measure with a value of 14.6.

The Student Teacher ratio indicator is based on a measure of the number of students and teachers (that satisfy the indicators' definition of each) within a city's population. One can view both as a statistical measurement in the sense that there is a population that we want to perform a measurement of, the measurement being a count of the number of members that satisfy a description of a Student and a Teacher, respectively. While the indicators require a count of members of the population, other measures may require statistics such as mean, standard deviation, etc. We have included in our core the GovStat general statistics ontology (Pattuelli, 2003). The core class is the 'Population' to

[3] The GCI Foundation Ontology has been published as standard ISO/IEC 21972:2020 "Information technology — Upper level ontology for smart city indicators"

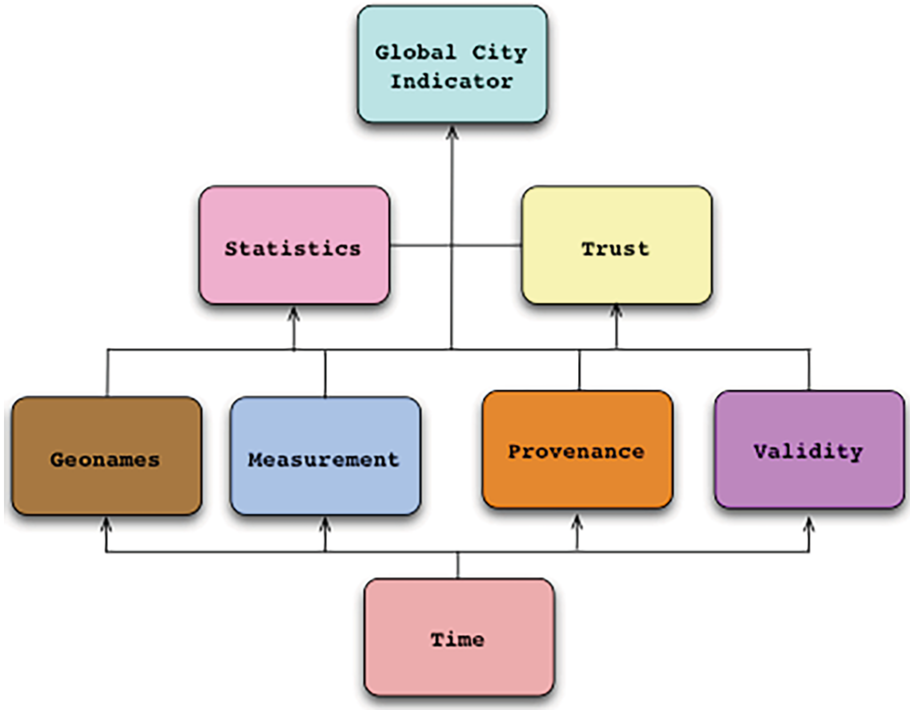

Fig. 1. GCI Foundation Ontology Components.

be measured. 'Population' is linked to a parameter (e.g., mean, standard deviation) by the is_described_by property, and the parameter is a subclass of 'Parameter'. In order to define the what portion of a city we are determining the size of, we extended the GovStat Population class (Fig. 3) with a property located_in, that uniquely identifies (using Geonames IRIs) the area (e.g., city) that the Population is drawn from, the property for_time_interval that identifies the time interval the Population is for, and the property defined_by, that identifies the class that all members of the Population are subsumed by (Fox, 2018).

An example of a Population statistic used in city indicators is the size or cardinality of a Population. Fig. 4 depicts the definition population_cardinality_unit (symbol: pc) which is a singular unit of measure within the OM unit of measure class. It also defines multiples of pc, such as kilopc.

All ISO 37120 indicators are ratios. Fig. 5 depicts an ontology pattern for indicators. The pattern defines an indicator, that is the ratio of two populations, as having a unit of measure defined to be a 'Population Ratio Unit' that specifies that the indicator is the ratio of the sizes (cardinalities) of two populations. One population size is the numerator and the other the denominator. A 'Population Size' is defined as the cardinality of a 'Population', and 'Population' is defined by a 'City' that the population is located in, and by a description of a 'Person' within the 'City'. For example, the 'Person' could be 'Female Student'. Hence the 'Population Size' could be the number of 'Female Student's in a particular 'City'. This structure is used in the indicator definitions in Section 5.

2.2. Education-Related ontologies

In order to represent the the Education indicators defined in Section 3, we need additional concepts, properties and axioms that span:

- the types of educational institutions at the primary and secondary level, including whether they are public or private, and certified,
- the grade structure of schools including cohorts,
- the definition of students and what educational programs they have taken, and
- the definition of teachers, including where and what they have taught.

We reviewed a number of ontologies to determine their relevance. Our review was hindered by their lack of competency questions; in most cases the ontologies were published as RDF or OWL with little documentation. Hence a detailed examination of their axioms was required. In the following we identify some of the vocabularies and ontologies that we reviewed. In Section 4, we identify the concepts that are relevant to the Education Ontology.

Schema.org is an initiative primarily led by the major search engine vendors. Its goal is to enhance search results by providing a vocabulary of concepts and properties that web page creators can embed in their web pages using RDFa. Many of the classes defined in the OWL version of schema.org only have subclassof property specified.

SUMO (Niles & Pease, 2001) is an upper level ontology[4]. It attempts to provide an overarching taxonomy of knowledge. In other words, its taxonomy of concepts is meant to span most of what we may want to represent.

OpenCYC (Matuszek et al., 2006) is a large ontology that is both very broad and very deep. It has been under development for over 20 years. The ontology is very rich in the areas of intelligence/defence.

An ontology has been created for describing the national curricula across the UK (Mohamed et al., 2013). The purpose of the ontology is to:

- "provide a model of the national curricula across the UK",
- "organise learning resources, e.g. video clips and revision content", and
- "allow users to discover content via the national curricula".

It is focused on course content as opposed to the organization and resourcing of its educational system.

Scribe (Uceda-Rosa et al., 2011) is an ontology designed specifically

[4] The SUMO ontology can be found at http://ontologyportal.org/sumo.owl. We will use the prefix "sumo" where needed.

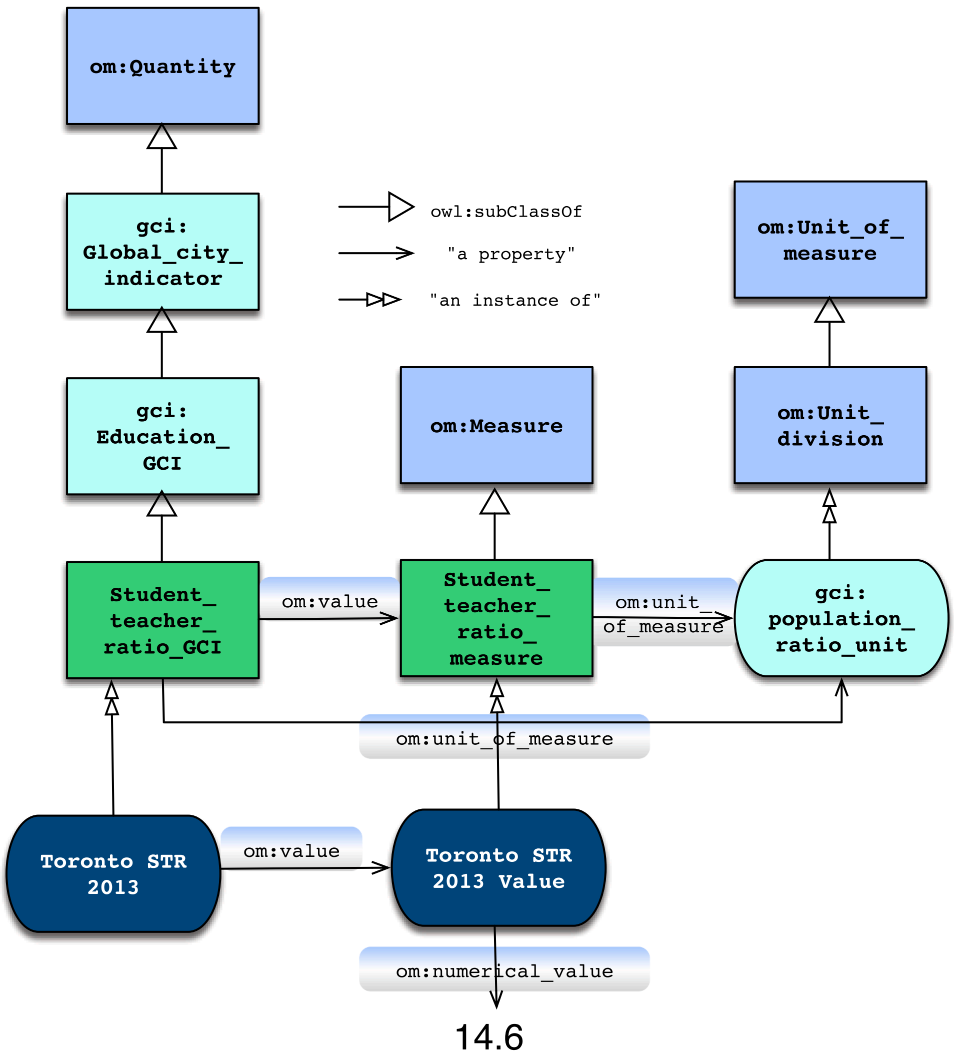

Fig. 2. Measurement Ontology.

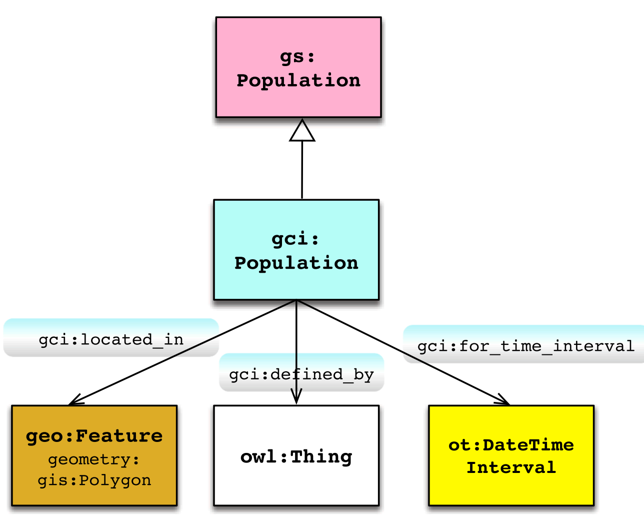

Fig. 3. Population Pattern.

to represent city information. From an education perspective, it refers to Educational Service (a service) and School District (a local government area), but not to schools, grades, teachers nor students.

While some of these ontologies provide some basic classes, e.g., school, teacher, and student, they neither provide the full set of class and properties nor the axioms necessary to model educational indicators. It is clear that there is a need for a well engineered educational ontology with competency questions necessary to support them. A more detailed analysis of their relevant concepts is provided in section 4.

2.3. Education data standards

There exist data standards for the representation and communication of education system data. They are not classified as ontologies because they lack formal specifications of class/object definitions. Never the less, they provide the data that describes students, teachers, schools and school programs. This data would be mapped onto the Education ontology as described in the 4th requirement of section 1.

What is probably the richest education data standard available is the United States' The Common Education Data Standards (CEDS) (CEDS, 2017). CEDS "is a joint effort by the Council of Chief State School Officers (CCSSO) and the State Higher Education Executive Officers (SHEEO) in partnership with the United States Department of Education …"[5] CEDS spans many aspects of an education system, including students, staff, and schools. It provides detailed information on both teachers and their employment, students and the courses they take and

[5] http://www.ccsso.org/Resources/Programs/CIO_Network_.html

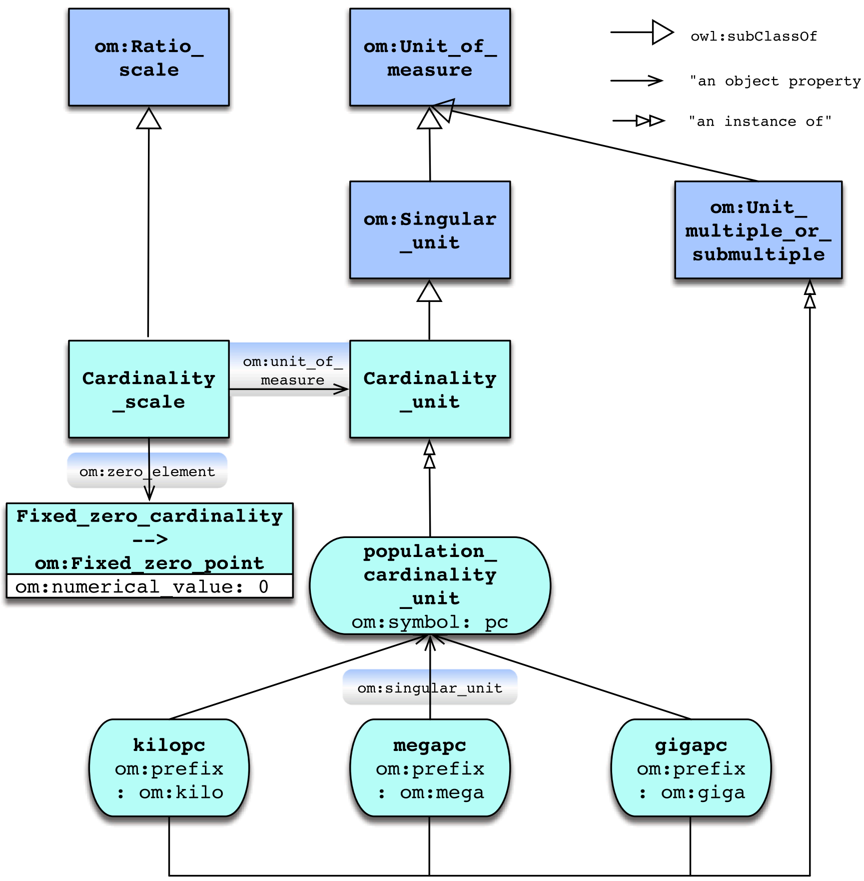

Fig. 4. population_cardinality_unit pattern.

their performance.

A second US data standard has been developed by the Ed-Fi Alliance. It "is a community of educators, technologists, and leaders committed to ensuring that every teacher, school district, and state agency can see, secure, and use their data …"[6] Ed-Fi provides both a standard for publishing data and a set of tools for analyzing the data including a dashboard for displaying a large number of educational indicators spanning students, schools and school districts. Ed-Fi's data standard is aligned with CEDS.

Assuming that CEDS data is publically available for city education systems, it is a rich source of city education data that can be mapped onto the Education Ontology. Never the less, it does not provide all the data that the ontology represents, such as meta data (i.e., provenance, measurement, validity, trust).

3. Use cases and their competency requirements

The uses cases that drive the development of the Education Ontology are the indicators defined in the Education Theme of ISO 37120. In this section we reprint these Education indicators. Using the ontology engineering methodology defined in Grüninger and Fox (1995), we define for each use case (i.e., indicator) a set of competency questions that our education ontology must be able to answer. Competency questions act as requirements that the ontology must satisfy. In other words, they capture what information is needed to represent the indicator definition, indicator theme knowledge, city specific knowledge, and the data used to derive an indicator value. Competency questions fall into the following categories:

- *Factual (F)*: Questions that ask what the value of some property is.
- *Consistency - Definitional (CD)*: Determine whether the instantiation of an indicator by a city is consistent with the ISO 37120 definition.
- *Consistency - Internal (CI)*: Determine whether different parts of the instantiation are consistent with each other.
- *Deduced (D)*: A value or relationship that can be deduced from the instantiation.

3.1. Percentage of female school-aged population enrolled in schools (ISO 37120 6.1)

The first ISO 37120 education indicator focuses on measuring female education:

"The percentage of female school-aged population enrolled in schools shall be calculated as the number of female school-aged population enrolled at primary and secondary levels in public and private schools (numerator) divided by the total number of female school-aged population (denominator). The result shall then be multiplied by 100 and expressed as a percentage.

The definitions of primary and secondary school detailed in Clauses 3.5 and 3.6 shall apply.

The proportion of enrolment in public and private schools should be reported, and cities shall note if private school data are included. In many cities, private schools are a significant component of education in the city. Private schools shall be recognized as providing real, bona fide education; many ministries or departments of education have a program that recognizes such schools. Enrolment in religious schools and home schools should be included if they are recognized.

One part-time enrolment of a half-day or more shall be counted as a full-time enrolment.

[6] https://www.ed-fi.org/what-is-ed-fi/

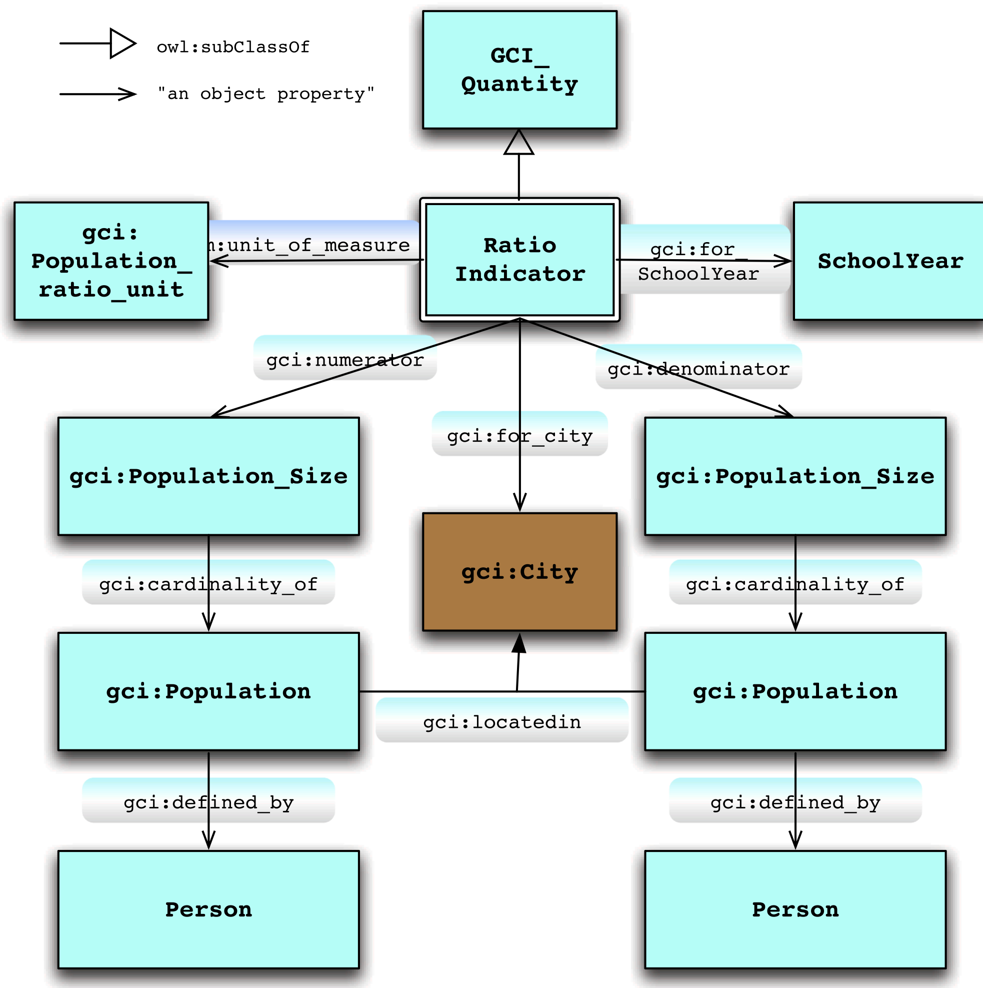

Fig. 5. Ontology Pattern for Population Ratio Indicators.

If the geographies of school districts and the city are different, best judgment should be used to related enrolment data to the city boundaries."

Competency Questions

1. (F) What city is the indicator for?
2. (CD) Are the students residents of the city?
3. (D) What is the age range for school age women?
4. (F) Is a school a private or public institution?
5. (F) Does a school teach Primary or Secondary courses?
6. (F) Is a school a home school? Religious school?
7. (D) Is the private school certified by the government?
8. (F) What grades comprise primary and secondary school?
9. (F) How many hours of school do you have to attend to be full time?
10. (D) What school did person X attend in year Y?
11. (D) What proportion of the students are in private schools for school year Y?

3.2. Percentage of students completing primary Education: Survival rate (ISO 37120 6.2)

Following is the ISO 37120 definition of Percentage of Students Completing Primary Education. We assume the definition of student and primary school as provided in section 2.1.

"The percentage of students completing primary education or survival rate shall be calculated as the total number of students belonging to a school-cohort who complete the final grade of primary education (numerator) divided by the total number of students belonging to a school-cohort, i.e. those originally enrolled in the first grade of primary education (denominator). The result shall then be multiplied by 100 and expressed as a percentage. The survival rate of primary education shall be expressed as the percentage of a cohort of students enrolled in the first grade of primary education who reached the final grade of primary education.

Survival rates for the private education sector should be reported, if known. The user of this International Standard should note if private school data are included."

> "Example: If the city reporting year is 2012 and primary education last five years, report the percentage of students entered primary education in 2006 and reached the final grade of primary education in 2011."

Competency Questions

We extend the competency questions in section 2.1 to include the following:

1. (F) What grades are included in primary school?
2. (D) What students in final primary year X are cohorts Y
3. (D) If a student was in their first grade of primary school in year Y, what would be their final year in primary school?
4. (F) How many students started first grade of primary school in year Y?

5. (D) How many students whose first grade in primary school was year Y, were in the final grade of primary school?
6. (D) What percentage of students who survived were in private school?

3.3. Percentage of students completing secondary Education: Survival rate (ISO 37120:6.3)

Following is the ISO 37120 definition of Percentage of Students Completing Secondary Education. We assume the definitions of student and secondary school are as defined in section 2.1.

"The percentage of students completing secondary education or survival rate shall be calculated as the total number of students belonging to a school-cohort who complete the final grade of secondary education (numerator) divided by the total number of students belonging to a school-cohort, i.e. those originally enrolled in the first grade of secondary education (denominator). The result shall then be multiplied by 100 and expressed as a percentage. The survival rate of secondary education shall be expressed as the percentage of a cohort of students enrolled in the first grade of secondary education who reached the final grade of secondary education."

"Example: If the city reporting year is 2012 and secondary education lasts seven years, report the percentage of students that entered secondary education in 2004 and reached the final grade of secondary education in 2011."

Competency Questions

The competency questions for this indicator are the same as in section 2.2 with the exception of substituting secondary for primary.

3.4. Primary education Student/Teacher ratio (ISO 37120 6.4)

Following is the ISO 37120 definition of Student Teacher Ratio:

"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 pre-school 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."

Competency Questions

1. (F) What city is the numerator/denominator for?
2. (CD) Are the numerator, denominator and indicator for the same city?
3. (F) Is the teacher administrative staff or teaching staff?
4. (F) Is the student part time or full time?
5. (F) Did the teacher work at a public school or private school in Year Y?
6. (F) Did the student attend a public school or private school in Year Y?
7. (D) What grades did teacher X teach in year Y?
8. (D) What schools did student X attend in year Y?
9. (D) What public schools are included in the indicator?
10. (D) Are there any private school included in the indicator?

3.5. Percentage of male school-aged population enrolled in schools (ISO 37120 6.5)

Following is the ISO 37120 definition of percentage of male school-aged population enrolled in schools:

"The percentage of male school-aged population enrolled at primary and secondary levels in public and private schools (numerator) divided by the total number of male school-aged population (denominator). The result shall then be multiplied by 100 and expressed as a percentage.

The definitions of primary and secondary school detailed in (indicators) 6.2 and 6.3 shall apply.

Enrolment in public and private schools should be reported, and cities shall note if private school data are included. In many cities, private schools are a significant component of education in the city. Private schools shall be recognized as providing real, bona fide education; many ministries or departments of education have a program that recognizes such schools. Enrolment in religious schools and home schools should be included if they are recognized.

One part-time enrolment of a half-day or more shall be counted as a full-time enrolment.

If the geographies of school districts and the city are different, best judgment should be used to relate enrolment data to the city boundaries."

Competency Questions

The competency questions for this indicator are the same as in Section 3.1 with the exception of substituting male for female.

3.6. Percentage of school-aged population enrolled in schools (ISO 37120 6.6)

Following is the ISO 37120 definition of percentage of school-aged population enrolled in schools:

"The percentage of school-aged population enrolled in schools shall be calculated as the number of school- aged population enrolled in primary and secondary levels in public and private schools (numerator) divided by the total number of the school-aged population (denominator). The result shall then be multiplied by 100 and expressed as a percentage.

Enrolment in public and private schools should be reported, and cities shall note in the comment section if private school data are included. In many cities, private schools are a significant component of education in the city. Private schools shall be recognized as providing real, bona fide education; many ministries or departments of education have a program that recognizes such schools. Enrolment in religious schools and home schools should be included if they are recognized.

Part-time enrolment of a half-day or more shall be counted as a full-time enrolment.

If the geographies of school districts and the city are different, best judgement should be used to relate enrolment data to the city boundaries."

Competency questions

The competency questions for this indicator are the same as in Section 3.1 with the exception of substituting all people for female.

3.7. Number of higher education degrees per 100 000 population (ISO 37120 6.7)

Following is the ISO 37120 definition of number of higher education degrees per 100,000 population:

"The number of higher education (tertiary education) degrees per 100 000 population shall be calculated as the number of people holding higher education degrees (numerator) divided by one 100 000th of the city's total population. The result shall be expressed as the number of higher degrees per 100 000 population."

Competency Questions

1. (F) What are tertiary degrees?
2. (F) What degrees does a person have?
3. (D) How many people have a tertiary degree X?
4. (D) How many females/males have tertiary degree X?

4. Education ontology

In order to represent the definitions of the ISO 37120 education indicators and answer their competency questions, we need to define educational concepts. This section defines the Education Ontology that can found at http://ontology.eil.utoronto.ca/GCI/Education/GCI-Education.owl.

4.1. Student and teacher

Core to the Education indicators is the description of students and teachers. The following competency questions, selected from section 3, focus on student and teacher knowledge.

1. (D) What is the age range for school age women?
2. (CD) Are the students residents of the city?
3. (F) Is the teacher administrative staff or teaching staff?
4. (F) Did the teacher work at a public school or private school in Year Y?
5. (F) Did the student attend a public school or private school in Year Y?
6. (D) What grades did teacher X teach in year Y?
7. (D) What schools did student X attend in year Y?
8. (F) Was the student part time or full time in year Y?

In reviewing existing ontologies, student and teacher definitions are often limited to taxonomic relations. In SUMO, classes do not exist for student nor teacher. OpenCYC (Fig. 6) defines a 'teacher' to be a subclass of 'academic' and 'person'. It has a subclass 'schoolteacher' that is further specialized as 'government schoolteacher' that is 'affiliating with

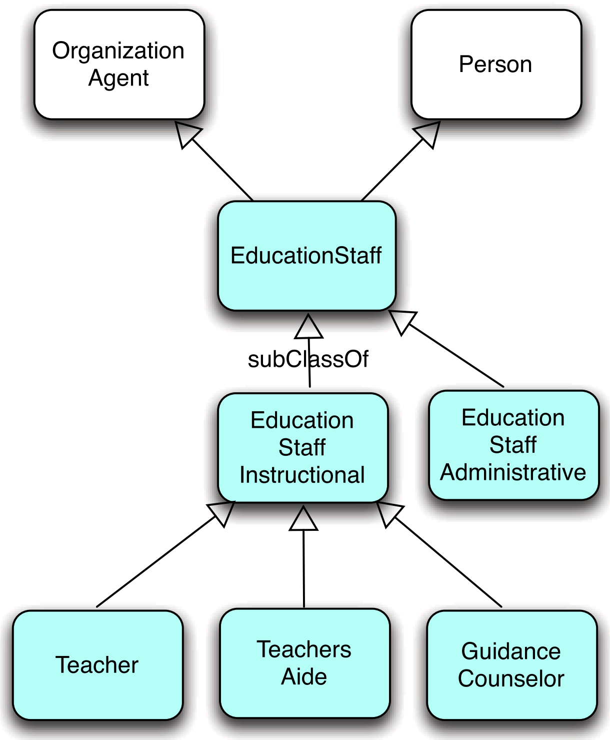

Fig. 7. Education Staff Taxonomy.

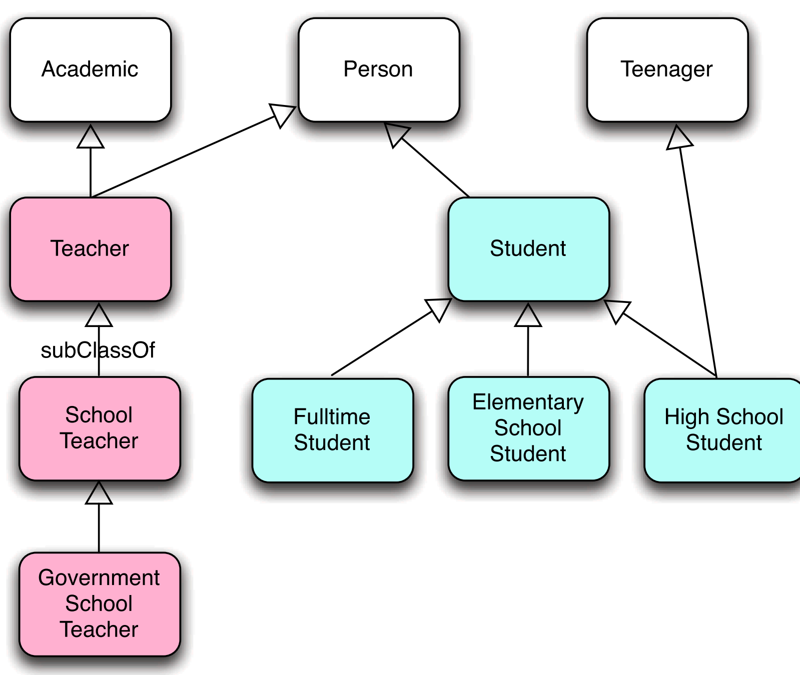

Fig. 6. OpenCYC Student and Teacher Taxonomy.

regional government'. A 'student' is a subclass of 'person' and has many specializations including 'elementary school student', 'full time student', and 'high school student'. These classes do not contain additional axioms other than a 'high school student' being a 'teenager'. The majority of the competency questions cannot be answered using these ontologies.

In the Education Ontology, 'Teacher' is part of an 'Education Staff' taxonomy (Fig. 7). The top level class of all education employees is 'Education Staff' which is a subClassOf 'Person' and 'Organization Agent'. It has two subclasses: 'Education Staff Administrative' and 'Education Staff Instructional'. 'Teacher' is subClassOf 'Education Staff Instructional'.

In the following table[7], we define an 'Education Staff' member as 'Organization Agent' (as defined in the Organization ontology) and a 'Person' (as defined in Schema.org) that has at least one 'Placement'. A 'Teacher' is a subclass of 'Instructional Education Staff' that has a 'Placement' in which they teach a minimum of one 'Course'.

Class	Property	Value Restriction
EducationStaff	rdfs:subClassOf	org:OrganizationAgent
	rdfs:subClassOf	sc:Person
	has_Placement	min 1 Placement
EducationStaffAdministrative	rdfs:subClassOf	EducationStaff
EducationStaffInstructional	rdfs:subClassOf	EducationStaff
Teacher	rdfs:subClassOf	EducationStaffInstructional
	has_Placement	min 1 (Placement and teaches min 1 Course)
	owl: equivalentClass	cyc:Teacher

A 'Placement' provides the details of where an 'Education Staff' member worked, the 'School Year', how many days a week they worked, and 'Course's they taught, if any. An 'Education Staff' member may have many 'Placements', one for each year that they worked at a 'School', or more than one per year if they worked at multiple locations during the same year.

Class	Property	Value Restriction
Placement	educational_Staff_At	EducationFacility
	days_Worked	exactly 1 positiveInteger
	for_SchoolYear	exactly 1 SchoolYear
	teaches	min 1 Course
	Min_Days_Worked	value 1

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

Class	Property	Value Restriction
Student	rdfs: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

Attendance during a school year at a school corresponds to a separate 'Enrollment'. An 'Enrollment' is composed of the 'Program' (defined in the next section) 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., 'Grade') and some comment.

4.2. Educational facility and grade

In this section we define our ontology for educational facilities and grades based on the following competency questions introduced throughout Section 3.

1. (F) Is a school a private or public institution?
2. (F) Does a school teach Primary or Secondary courses?
3. (D) Is the private school certified by the government?
4. (F) What grades comprise primary and secondary school?
5. (D) What students in final primary year X are cohorts?
6. (D) If a student was in their first grade of primary school in year Y, what would be their final year in primary school?
7. (F) How many students started first grade of primary school in year Y?
8. (D) How many students whose first grade in primary school was year Y, were in the final grade of primary school?
9. (D) What percentage of students who survived were in private school?
10. (F) Did the teacher work at a public school or private school in Year Y?
11. (F) Did the student attend a public school or private school in Year Y?
12. (D) What grades did teacher X teach in year Y?
13. (D) What schools did student X attend in year Y?
14. (D) What public schools are included in the indicator?
15. (D) Are there any private schools included in the indicator?

Our competency questions require the distinction between private and public, secondary and primary schools. They also require identifying the education programs they provide, the grades that make up each level, and whether students are enrolled in them.

Schema.org's only relevant class is 'School', which is a subclass of 'EducationalOrganization', and inherits the following properties (http://www.schema.org/School) from 'Organization', which do not address the needs of the competency questions.

- Address
- aggregateRating
- brand
- contactPoint
- department
- duns
- email
- employee
- employees
- event
- events
- faxNumber
- founder
- founders

SUMO has a class 'EducationalOrganization' (no axioms provided) that is a subclass of 'Organization' whose axioms are:

- members of the same 'Organization' share the same purpose, and
- that a member of an 'Organization' is an instance of 'Agent'.

SUMO defines 'School' as a subclass of 'EducationalOrganization'.

[7] The table defines an OWL 2 (Hitzler et al., 2012) class using the Manchester Syntax (Horridge & Patel-Schneider, 2013).

Within 'School' it has 'HighSchool', 'PrivateSchool', 'PublicSchool' and 'SecondarySchool' as subclasses. They have the following axioms associated with them:

- 'PrivateSchool' is disjoint from a 'GovernmentOrganization'.
- 'PublicSchool' is a subclass of 'GovernmentOrganization'.

OpenCYC has 'school' defined as a subclass of 'educational organization', that is a subclass of 'institution'. It has a specialization 'K-12 institution' which in turn has specializations of 'elementary school', 'middle school' and 'high school'. Finally, a 'K-12 institution that is a publically funded thing' is equivalent to the intersection of a 'K-12 institution' and a 'publicly funding thing'. Note the similarity of Schema.org and SUMO to OpenCYC; Schema.org acknowledges portions of their taxonomy are based on openCYC.

In the Education Ontology, we have imported the Organization ontology (Fox et al., 1998), which provides the concepts of Organization, Goal, Activity and Member. Organization is specialized into the following sub classes: "Non Government Organization", "For Profit Organization" and "Government Organization", the latter being used to define publically funded schools.

The basic taxonomy of 'School's is depicted in Fig. 8.

School has the following properties:

1. delivers_Program that identifies the type of 'School Program' that is taught,
2. org:has_Ownership that distinguishes among, public, private, government and charity ownership, and
3. has_SchoolType that distinguishes among religious, secular, home, French Immersion, etc.

Class	Property	Value Restriction
School	rdfs:subClassOf	EducationFacility
	delivers_Program	some SchoolProgram
	org:has_Ownership	exactly 1 Ownership
	has_SchoolType	min 1 SchoolType
	org:hasName	only xsd:string
	org:consistsOf	only org:Division
	org:hasLegalName	exactly 1 xsd:string
	org:hasGoal	only org:Goal

The following defines the 'Public Primary School' class that teaches 'Grade Level Primary'. This defines the grades being taught at the primary level for the corresponding city. The choice of "some" is to allow a school to teach other things than primary grades, e.g., swimming lessons.

Class	Property	Value Restriction
PublicPrimarySchool	rdfs:subClassOf	PublicSchool
	delivers_Program	some GradeLevelPrimary
	has_SchoolType	value secularSchoolType
	has_Ownership	value government_ownership
PrivateSchool	rdfs:subClassOf	School
	has_Ownership	value privately_owned
	has_Certification	some Certification
Certification	certified_By	some GovernmentOrganization
	certification_Date	exactly 1 dateTime

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. 'GradeLevel' has a 'starting_Grade' and 'ending_Grade' that define the first and last grades of the level. Each city defines its own version of 'GradeLevelPrimary' 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 'GradeLevel' 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
SchoolProgram	owl:subclassOf	Program
	has_Course	min 1 Course
	has_SP_Type	all SP_Type
GradeLevel	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

'Grade' 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 'PrimaryGrade' and 'SecondaryGrade'.

We link a 'City' to a 'Grade Level' by the following:

Class	Property	Value Restriction
City	rdfs:subClassOf	geo:Feature
	has_Primary_Grade_Level	exactly 1 GradeLevelPrimary
	has_Secondary_Grade_Level	exactly 1 GradeLevelSecondary

We also define 'School Age Person' by associating them with a 'City' and 'School Year'. Determining whether someone is of school age is defined by a constraint that uses this information along with their birthdate which is a property inherited from 'Person'.

Class	Property	Value Restriction
School_Age_Person	rdfs:subClassOf	Person
	gci:for_City	exactly 1 City
	for_SchoolYear	exactly 1 SchoolYear

The starting grade for all schools in the Province of Ontario at the primary level is 'GradeOne' and the ending grade is 'GradeSix', hence we define:

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

We introduce the concept of a 'Cohort', i.e., the students who started primary or secondary school together and entered the final year of each together. For example, for any given year, e.g., 2014, the cohort is defined to be the subset of students who entered the final year of the grade level in 2014, who also were in the starting grade of the grade level together. If primary school covers grades one through six, then the starting year for the 2014 cohort is 2008.

Class	Property	Value Restriction
Cohort	rdfs:subClassOf	EducationThing
	starting_SchoolYear	exactly 1 SchoolYear
	ending_SchoolYear	exactly 1 SchoolYear
	for_GradeLevel	Exactly 1 GradeLevel

In order to guarantee that the school year of the Education Program matches the school year of the Indicator, we will need to add a consistency axiom in the next section.

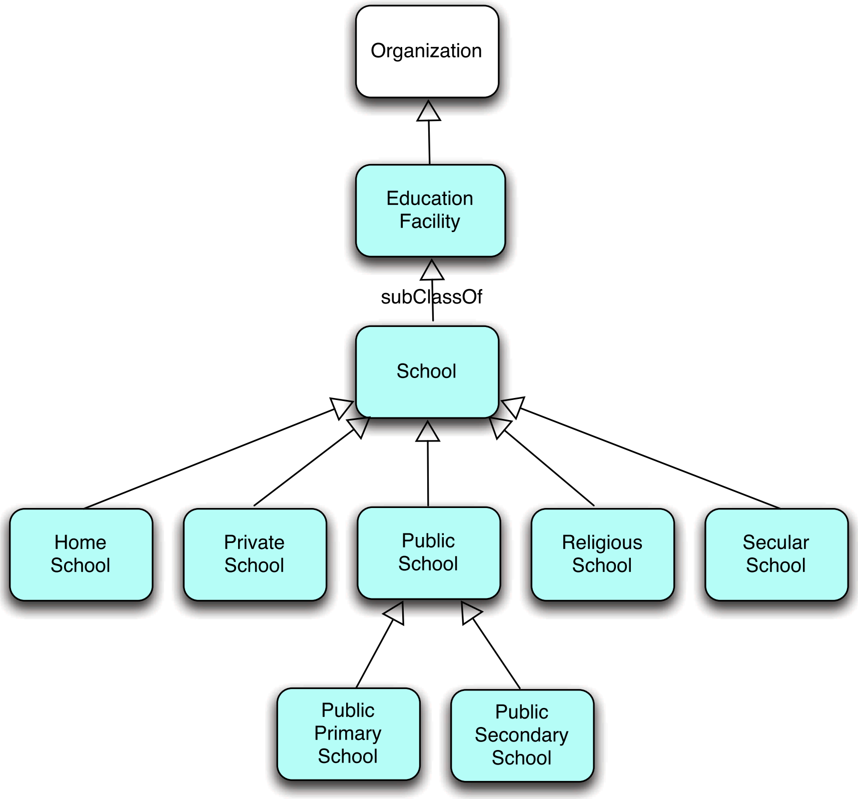

Fig. 8. Education Facility Taxonomy.

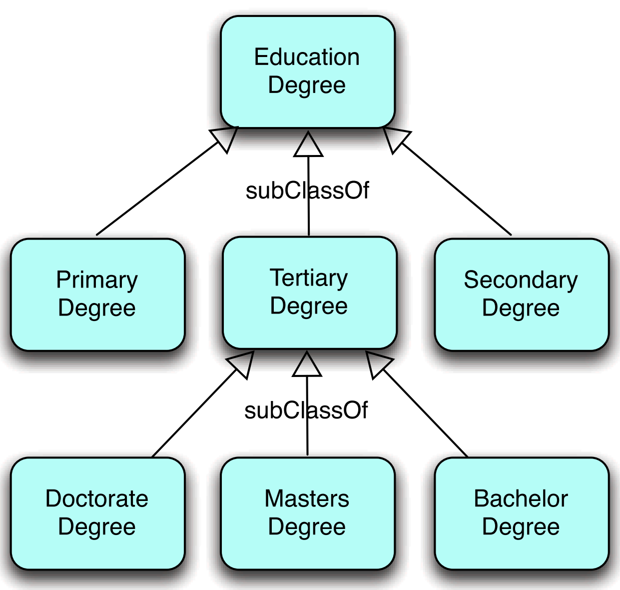

Fig. 9. Educational Degrees.

4.3. Higher education degrees

For the 7th indicator, we need to extend the Education ontology to allow a city to identify what educational degrees count as Tertiary.

1. (F) What are tertiary degrees?
2. (F) What degrees does a person have?
3. (D) How many people have a tertiary degree X?
4. (D) How many females/males have tertiary degree X?

The set of admissible degrees may differ from city to city, but is assumed to be post-secondary. We introduce the concept of Education Degree as follows: Fig. 9

The various types of tertiary degrees can be refined, such as arts, science, engineering, etc. We extend the definition of a Person to include the property: has_EducationDegree, and the 'EducationDegree' has the properties:

Class	Property	Value Restriction
EducationDegree	rdfs:subClassOf	EducationThing
	awarded_Year	exactly 1 Year
	awarding_EducationFacility	exactly 1 EducationFacility
	enrolled_Program	only 1 EducationProgram
	degree_Name	exactly 1 string

The enrolled_Program property allows for the specification of the courses taken in each year of the program.

4.4. Consistency axioms

The following are additional axioms that cannot be formulated in OWL, but in our system are implemented in Prolog:

1. The ending grade of a grade level must be after the starting grade of the same grade level.
2. The starting grade of secondary school is the next grade after the final grade of primary school.
3. All Students who attend a primary school must attend primary grades.
4. All Teachers in a primary school must teach at least one primary course.
5. For students to be a cohort, if they are counted in the final year population then they must be a subset of the students in the first year cohort population.
6. A Grade that is a member of Primary Grade must be contained within the Primary Grade Level (same for Secondary Grade).
7. A student's age must be within the age range of the grade level they are associated with.
8. The difference in years in the start and ending year of a cohort is equal to the difference in years in the starting and ending grade of a grade level.
9. The value of the Educational Program for_School_Year has to be the same as the value for an indicator's for_School_Year.

5. ISO 37120 education indicators ontology

With the Education Ontology defined, we now have the classes and properties necessary to represent the definitions of the ISO 37120 Education indicators. In this section we define the seven ISO 37120 Educational indicators. The OWL 2 definitions can be found in http://ontology.eil.utoronto.ca/GCI/ISO37120/Education.owl.

5.1. Percentage of female school-aged population enrolled in schools (ISO 37120 6.1)

The following diagram shows a partial definition of ISO 37120 6.1. Some of the subClassOf links have been omitted but can be found in the OWL definition file.

Fig. 10 uses the GCI Foundation ontology ratio indicator (Fig. 5) to provide the structure for indicator 6.1. It is an Education Global City

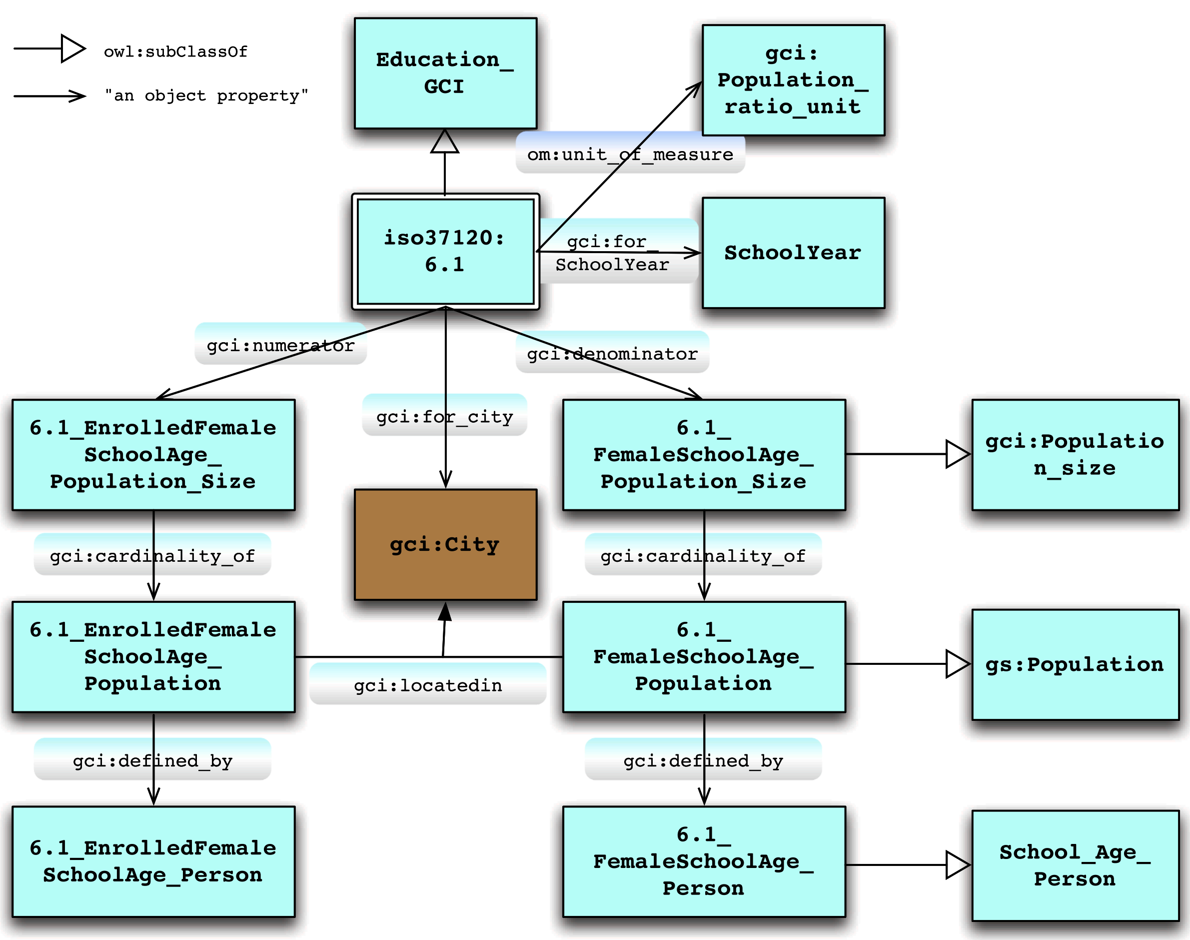

Fig. 10. 6.1 Definition.

Indicator. It is a ratio ('Popultation_ratio_unit') that has a numerator of the size of the population of enrolled school age women. The denominator is the size of the population of all school age women.

What is unique to this indicator is the definition of the people making up the populations (linked using defined_by), namely 'Enrolled Female School Age' and 'Female School Age'. The following defines 'Enrolled Female School Age':

Class	Property	Value Restriction
6.1_EnrolledFemaleSchoolAge_Person	rdfs:subClassOf	Female_Person
	rdfs:subClassOf	School_Age_Person
	has_Enrollment	6.1_Enrollment

There are two issues we have to address in this definition:

1. We have to make sure that for the year the metric is being reported that the student is of school age in that year, and
2. They are enrolled on a full or part time basis, in a public or private school, in a primary or secondary grade.

In the previous section, as part of the 'GradeLevel' class, we introduced a starting and ending age. This allows us to determine the age range for both primary and secondary school. To determine whether a 'Person' is in the range, we have to compute their age using their birthdate defined in the 'Student' class and the year for the metric defined by the 'for SchoolYear' property of the '6.1′ class. This calculation is performed by an axiom.

The '6.1 Enrollment' class defines the properties of an enrolled 'Student'. Namely, it is for 'School Year' that is the same as 6.1, they attend some 'School', the grade is primary or secondary, they are full or part time and they are enrolled.

Class	Property	Value Restriction
6.1_Enrollment	rdfs:subClassOf	Enrollment
	for_SchoolYear	exactly 1 (schoolYear_For 6.1)
	attends	exactly 1 School
	enrolled_Program	exactly 1 (GradeLevelPrimary or GradeLevelSecondary)
	enrolled_Grade	exactly 1 (PrimaryGrade or SecondaryGrade)
	enrolled_Status	exactly 1 (Full_Time or Part_Time)
	enrolled_Courses	some Course

Some axioms (constraints), such as the time interval associated with the Population 6.1_EnrolledFemaleSchoolAge_Population in the numerator of the indicator has to be the same as the interval for the indicator itself, 6.1 can be expressed directly, e.g., using the Manchester syntax:

gci:for_time_interval exactly 1 (time:Interval and
(inverse (gcie:for_SchoolYear) min 1 (iso37120:6.1 and

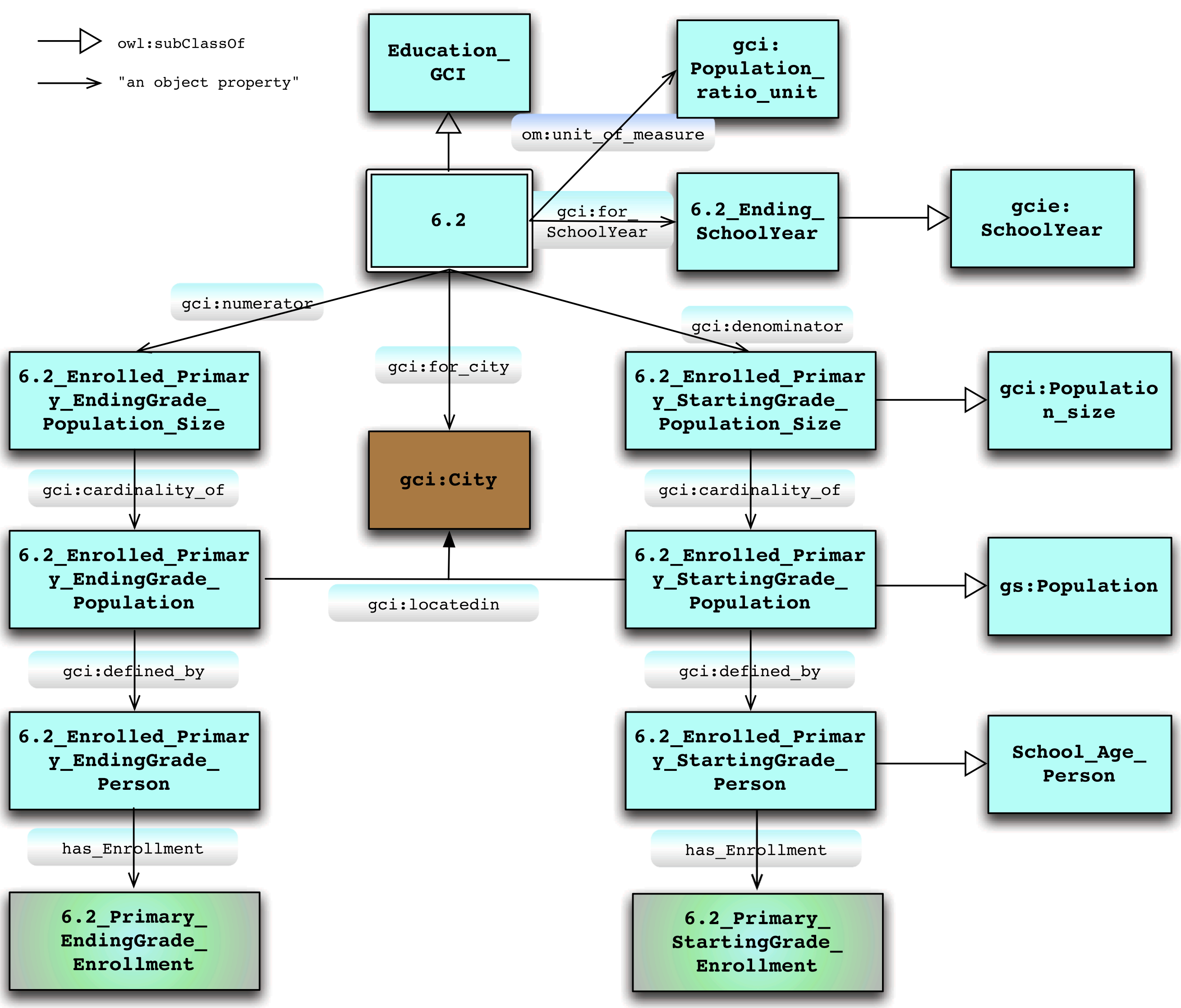

Fig. 11. Education Indicator 6.2.

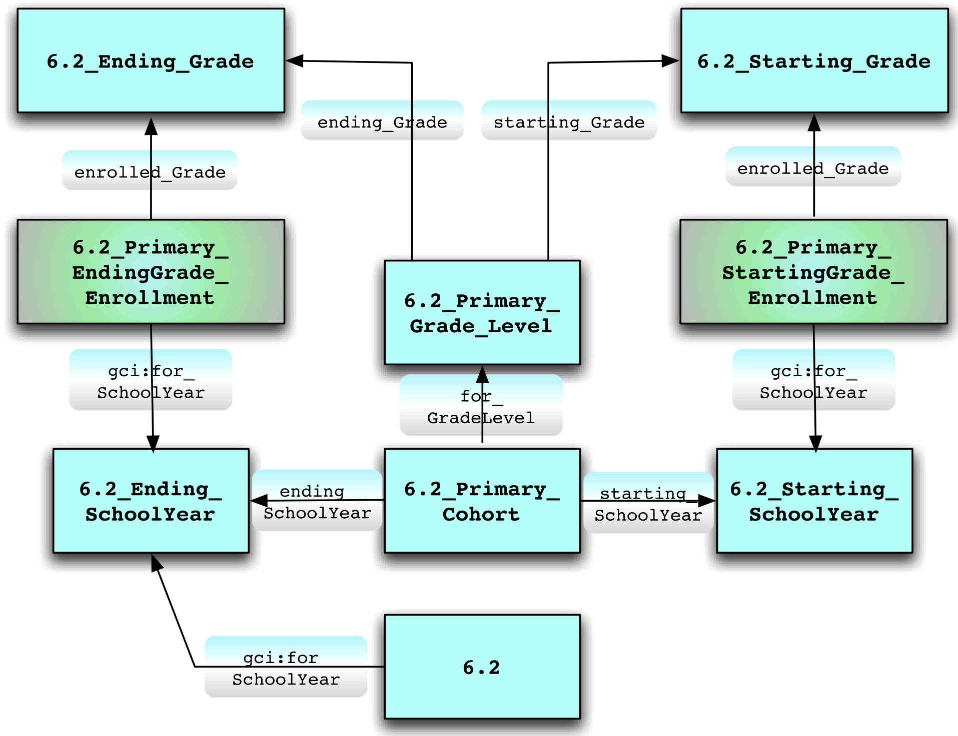

Fig. 12. 6.2 Cohort Definition.

(om:numerator exactly 1 (isoe:6.1_EnrolledFemaleSchoolAge_Population_Size and
(gci:cardinality_of Self)))))

Other axioms (constraints) need to be specified separately as either rules or SHACL constraints:

1. For the specified year, the age of the 'Student' is within the range defined by the grade levels.
2. The school year of the '6.1_SchoolProgram' is the same as the school year for the '6.1' indicator.
3. The 'School' teaches the enrolled 'Program'.
4. The 'Grade' attended in an 'Enrollment' is consistent with the 'Grade's taught at the corresponding School.
5. The Person counted in each Population resides in the Population's city.

The definition of the denominator can be found in the OWL file.

5.2. *Percentage of students completing primary education (ISO 37120 6.2)*

This indicator relies on the definition of 'Cohort'. The basic structure of the ratio is the same as in '6.1', but the definition of the 'Enrolled Primary Ending Grade Person' that defines the population we are taking the size of, is where it differs. In particular, it is constrained by the 'Program' they attend being the ending year of their grade level. Similarly for the starting grade. Fig. 11

The definitions of the Enrolled Primary Starting and Ending Grade Programs are depicted in Fig. 12. The definitions of these 'Program's are where both the 'Grade Level' and 'Cohort' classes come into play.

An 'Enrollment' defines both the 'Grade Enrolled' and the 'Year' of enrollment. To satisfy the definition of '6.2', the Starting and Ending Grades, and the Starting and Ending years have to be consistent with the 'Cohort' specification that includes the 'Grade Level'. In order for this to work properly, we have to define the following axioms:

1. The ending school year of the 'Primary Ending Grade Enrollment' is the same as the 'School Year' of '6.2' and the ending school year of the '6.2 Primary Cohort'.
2. The starting school year of the 'Primary Starting Grade Enrollment' is the same as the starting school year of the '6.2 Primary Cohort'.
3. The ending grade of the '6.2 Primary Ending Grade Enrollment' has to be the same as the ending grade of the'6.2 Primary Grade Level'.
4. The starting grade of the '6.2 Primary Starting Grade Enrollment' has to be the same as the starting grade of the '6.2 Primary Grade Level'.

5.3. *Percentage of students completing secondary education (ISO 37120 6.3)*

6.3′s definition is similar to 6.2 except for the substitution of Secondary for Primary. The OWL 2 definitions can be found in: http://ontology.eil.utoronto.ca/GCI/ISO37120/Education.owl.

5.4. *Primary education Student/Teacher ratio (ISO 37120 6.4)*

6.4 has the same structure as 6.1 but varies in the definition of Student and Teacher. Fig. 13

The numerator is the cardinality of the 'Student Population'. The denominator is the cardinality of the 'Teacher Population'. '6.4 Student' is defined to be a subClassOf 'Student'. The restriction that they attend a

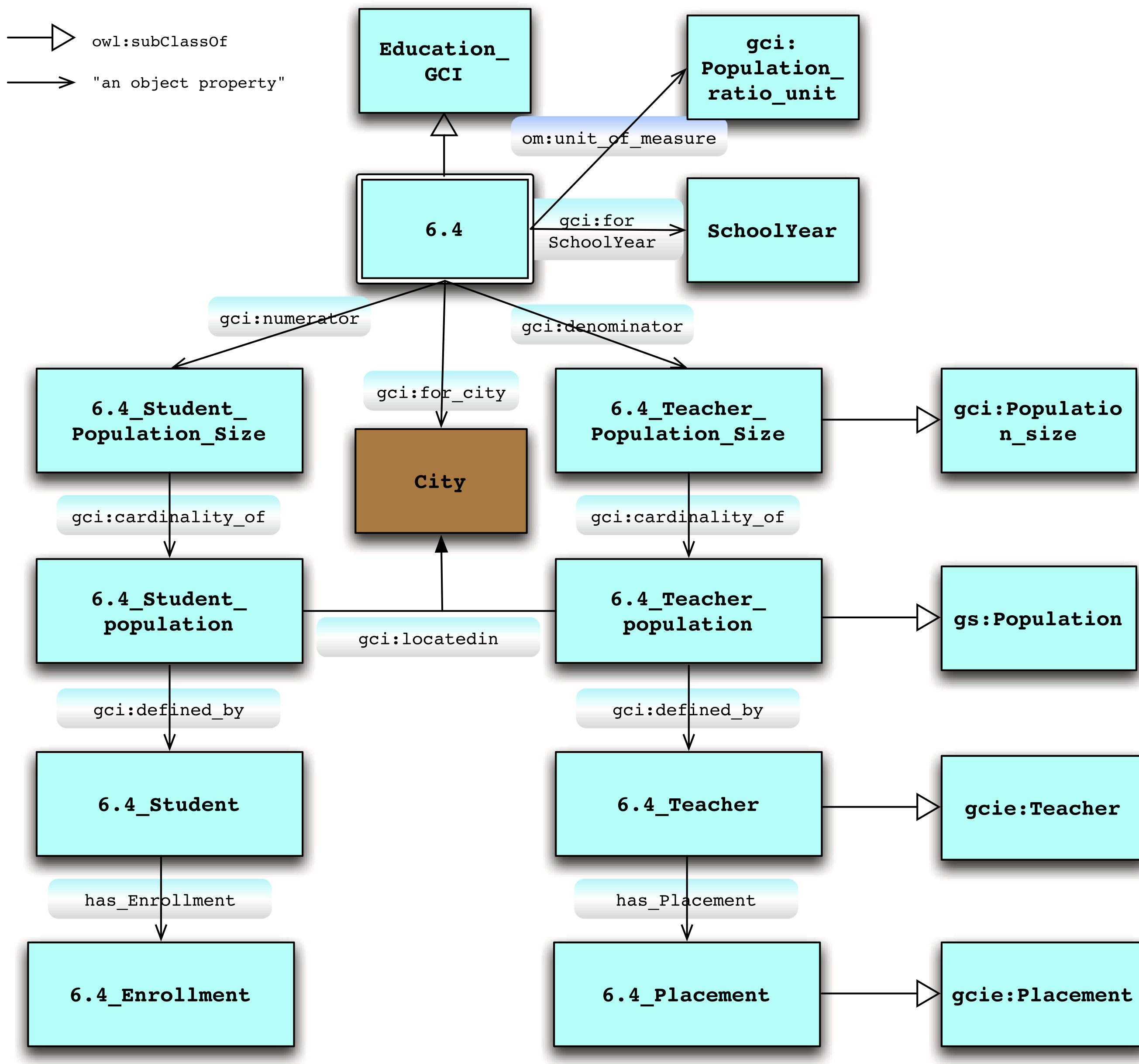

Fig. 13. Education Indicator 6.4.

'Public Primary School' for the designated 'School Year' is defined by their enrollment:

Class	Property	Value Restriction
6.4_Student	rdfs:subClassOf	Student
	has_Enrollment	6.4_Enrollment
6.4_Enrollment	rdfs:subClassOf	Enrollment
	for_SchoolYear	exactly 1 (schoolYear_For 6.4)
	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

6.4_Teacher is defined as follows as having at least one Placement in a Public Primary School.

Class	Property	Value Restriction
6.4_Teacher	rdfs:subClassOf	EducationalStaffInstructional
	has_Placement	6.4_Placement
6.4_Placement	rdfs:subClassOf	Placement
	days_Worked	exactly 1 positiveInteger
	min_Days_Worked	value 1
	org:memberOf	exactly 1 PublicPrimarySchool

Axioms

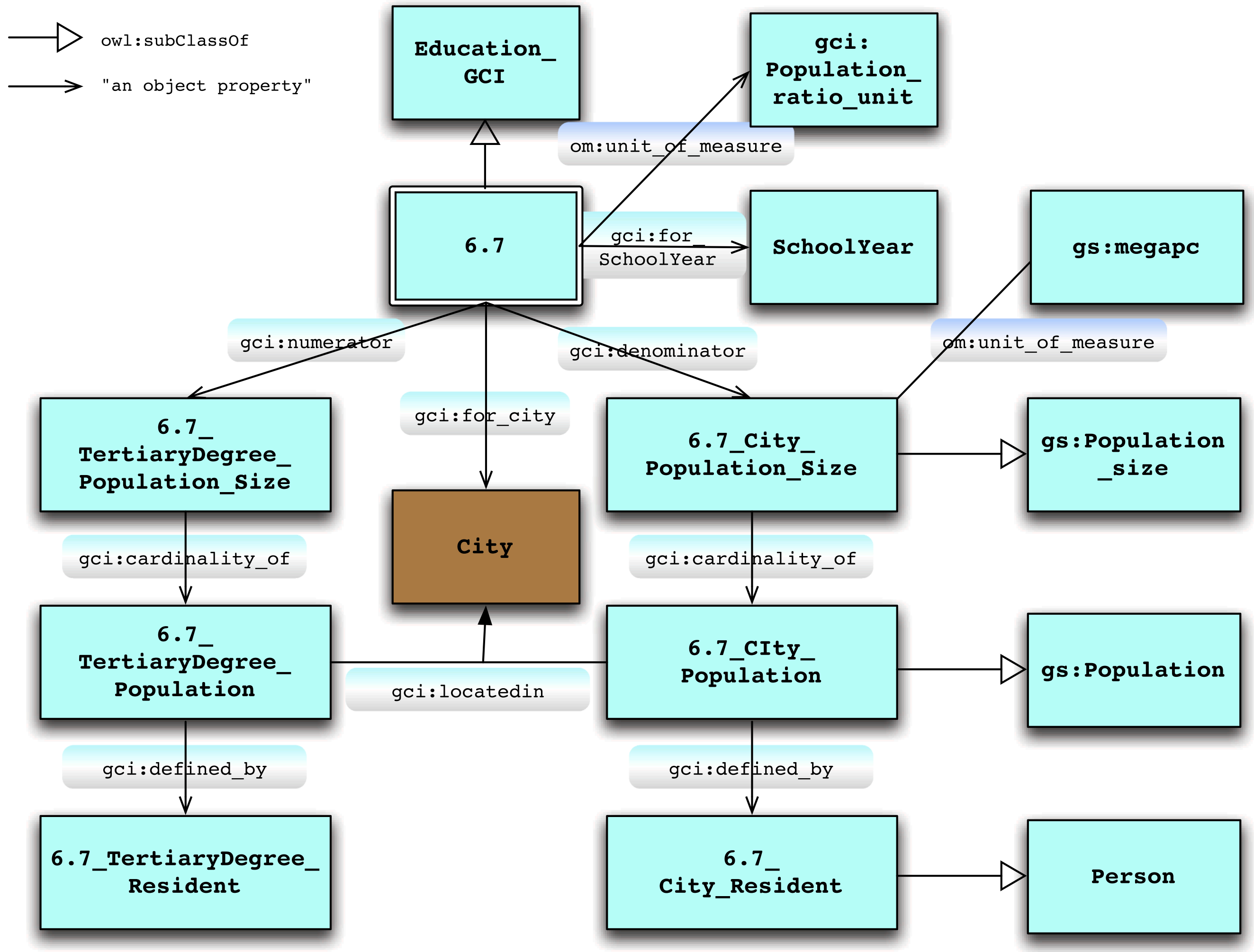

Fig. 14. Education Indicator 6.7.

1. Each teacher has to satisfy the minimum days worked requirement.
2. A teacher is counted as one fifth for each day worked.

5.5. *Percentage of male school-aged population enrolled in schools (ISO 37120 6.5)*

This is defined in the same way as 6.1, except for substituting Male for Female. The OWL 2 definitions can be found in: http://ontology.eil.utoronto.ca/GCI/ISO37120/Education.owl.

5.6. *Percentage of school-aged population enrolled in schools (ISO 37120 6.6)*

This is defined in the same way as 6.1, except for removing the Female restriction. The OWL 2 definitions can be found in: http://ontology.eil.utoronto.ca/GCI/ISO37120/Education.owl.

5.7. *Number of higher education degrees per 100 000 population (ISO 37120 6.7)*

The structure of this indicator is similar to '6.1'. There are two significant differences. First the '6.7' City Population Size has its unit of measure constrained to hectokilopc (100,000) in order to assure that when we take the ratio of number of people with tertiary degrees in the city to total population of the city, it is to 100,000 of population. Second, the definition of the '6.7 Tertiary Degree Resident' is constrained to having a tertiary degree as defined by the city. Fig. 14

The key difference with '6.7' is the definition of the numerator which depends upon the a resident of the city having a tertiary degree:

Class	Property	Value Restriction
6.7_TertiaryDegree_Resident	rdfs:subClassOf has_EducationalDegree	Person some TertiaryDegree

The following axioms are defined to satisfy the definition:

1. Resident tertiary degrees are restricted to those defined by the city.
2. The date of the degree awarded has to be on or before the year for the indicator.
3. The city that the resident resides in is the same as the city for the indicator.

6. Evaluation

In this section we verify the Education Ontology in three ways:

1. Is the Education Ontology along with the ontologies it relies on and imports logically consistent.

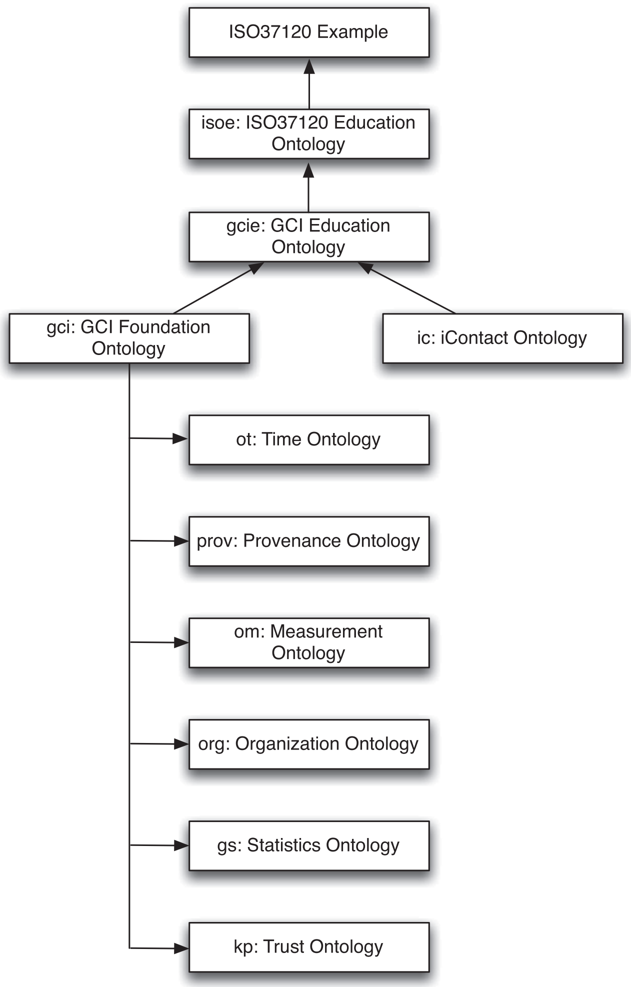

Fig. 15. Education Ontology import hierarchy.

2. Can competency questions be represented, as SPARQL queries, using the Education Ontology.
3. Can the relevant data found in Education data standards be mapped onto the Education Ontology.

6.1. *Consistency*

The consistency of our Education ontology is dependent upon the ontologies it imports. The following diagram depicts the ontology import hierarchy. Fig. 15

The following identifies the URI for each of the imported ontologies:

- gcie: http://ontology.eil.utoronto.ca/GCI/Education/GCI-Education.owl#
  The Education ontology defined in section 4
- gci: http://ontology.eil.utoronto.ca/GCI/Foundation/Foundation-v2.owl#
  The Foundation ontology
- gs: http://ontology.eil.utoronto.ca/govstat.owl#
  Statistics ontology
- ic: http://ontology.eil.utoronto.ca/icontact.owl#
  International contact ontology
- isoe: http://ontology.eil.utoronto.ca/ISO37120/Education.owl#
  The ISO 37,120 education indicators definitions defined in section 5.
- kp: http://ontology.eil.utoronto.ca/trust.owl#
  Trust and validity ontology
- ot: http://www.w3.org/2006/time#
  Time ontology
- om: http://www.wurvoc.org/vocabularies/om-1.8/
  Measurement ontology
- org: http://ontology.eil.utoronto.ca/organization.owl#
  Organization ontology
- prov: http://www.w3.org/ns/prov#
  Provenance ontology

Using Protégé's Hermit reasoner, we can test an ontologies consistency. In this case, the ontologies in Fig. 15 were found to be consistent.

6.2. *Competency*

Our second approach to evaluating the Education Ontology is to see whether the competency questions can be represented as SPARQL queries using the ontology. This is part of the ontology evaluation process defined in Grüninger and Fox (1995). The ability to define the competency questions in SPARQL demonstrates the ontology is sufficient to meet the needs of representing city data of relevance to city education indicator-based analysis.

The definitions of the education indicators determine the kinds of information a city would need to publish in order to support the answering of competency questions, and perform definitional, longitudinal and transversal analysis. As part of this research, a search was conducted for education data to use as a test case. In spite of the proliferation of city open data websites, almost all education data remains inaccessible (Fox & Pettit, 2015; Hugh & Fox, 2018). For example, aggregate education data can be found in open city data sites, e.g., New York,[8] but data at the individual student and teacher level, is either not available, or only for approved research purposes due to privacy restrictions. The World Council for City Data,[9] a non-profit that maintains a repository of indicator data for cities that are certified as ISO 37120 compliant, restricts access to the data reported by member cities. When their data was available, it only contained the indicator values, and none of the supporting data.

The data we used for verifying competence is based on the Toronto education system. The City of Toronto was one of the few cities to openly publish their ISO 37120 indicators[10]. We contacted the City to see if we could have access to the data used to derive their Education indicators. We were informed that the Toronto District School Board was the source of the data. We contacted the School Board, but after 4 months of discussion were informed that a freedom of information request would have to be submitted. We did not pursue the matter further. Hence, much of the data is based on our knowledge of the education system versus data available in open city datasets.

In the following example, we use the City of Toronto to illustrate the competency questions. For ease of understanding we will show the instances in table form. Prefixes are defined as follows:

- iso: http://ontology.eil.utoronto.ca/ISO37120.owl#
  URIs for each ISO 37120 indicator
- isoe: http://ontology.eil.utoronto.ca/GCI/ISO37120/Education.owl#
  The ISO 37120 education indicators definitions defined in section 5.
- gcie: http://ontology.eil.utoronto.ca/GCI/Education/GCI-Education.owl#
  The Education ontology defined in section 4
- gn: http://sws.geonames.org/
- sc: http://schema.org/

This first table defines the Toronto-specific education data that provide city specific information about Toronto, and its schools, grades, etc. It defines the Ontario Primary School level education program, including starting and ending grades, and starting and ending ages.

Instance	Property	Value
gn:6251999	rdfs:label	Canada
	rdfs:type	gn:Feature
	rdfs:type	sc:Country
gn:6093943	rdfs:label	“Ontario”
	rdfs:type	gn:Feature
	rdfs:type	sc:Province
gn:6167865	rdfs:label	“Toronto”
	rdfs:type	gn:Feature
	rdfs:type	sc:City
ontarioPrimaryProgram	rdfs:type	gcie:GradeLevelPrimaryCanada
	gcie:has_Certification	opp_certification
	gcie:has_Fulltime_Hours	35
	gcie:has_Fulltime_Period	om:week
	gn:parentCountry	gn:6251999
	gcie:starting_Grade	ontarioGradeOne
	gcie:ending_Grade	ontarioGradeSix
	gcie:starting_Age	6
	gcie:ending_Age	13
opp_certification	rdfs:type	ProgramCertification
	gcie:certified_By	omet
	gcie:certification_Date	1951–01-01
ontarioGradeOne	rdfs:type	gcie:GradeOne
	gn:locatedIn	gn:6093943 (Ontario)
ontarioGradeSix	rdfs:type	gcie:GradeSix
	gn:locatedIn	gn:6093943 (Ontario)
cedar_grove	rdfs:type	gcie:PublicPrimarySchool
	gcie:delivers_Program	ontarioPrimaryProgram
	gcie:has_Certification	cg_certification
omet	rdfs:type	GovernmentOrganization
	rdfs:label	“Ontario Ministry of Education and Training”
cg_certification	rdfs:type	SchoolCertification
	gcie:certified_By	omet
	gcie:certification_Date	1951–01-01

[8] https://datasetsearch.research.google.com/search?query=education%20%20New%20York&docid=L2cvMTFocmR3NjRtNg%3D%3D

[9] https://www.dataforcities.org/

[10] https://www.toronto.ca/wp-content/uploads/2017/11/989c-Final-Summary-of-Torontos-WCCD-ISO-37120-Results-6-AODA-.pdf

The following table defines the instances that instantiate the 6.1 indicator, its numerator and denominator, and their corresponding Populations, and enrolled female definition.

Instance	Property	Value
6.1_ex	rdfs:type	iso:6.1
(instance of 6.1)	gci:numerator	6.1_EF_size
	gci:denominator	6.1_F_size
	gci:for_City	gn:6167865
	om:value	6.1_ex_value
6.1_ex_value	rdfs:type	om:Measure
(the value of 6.1)	om:numerical_value	30
	om:unit_of_measure	gci:Population_ratio_unit
6.1_EF_size	rdfs:type	isoe:6.1_EnrolledFemaleSchoolAge_Population_Size
(numerator of 6.1)	gci:cardinality_of	6.1_EF_pop
	om:value	6.1_EF_size_value
6.1_EF_size_value	rdfs:type	om:Measure
(value of the numerator of 6.1)	om:numerical_value	1000
	om:unit_of_measure	gci:Population_cardinality_unit
6.1_F_size	rdfs:type	isoe:6.1_FemaleSchoolAge_Population_Size
(denominator of 6.1)	gci:cardinality_of	6.1_F_Pop
	om:value	6.1_F_size_value
6.1_F_size_value	rdfs:type	om:Measure
(value of the denominator of 6.1)	om:numerical_value	30,000
	om:unit_of_measure	om:Population_cardinality_unit
6.1_EF_pop	rdfs:type	isoe:6.1_EnrolledFemaleSchoolAge_Population
(Numerator population)	gci:locatedin	gn:6167865
	gci:defined_by	6.1_EF_person
6.1_F_pop	rdfs:type	isoe:6.1_FemaleSchoolAge_Population
(Denominator population)	gci:locatedin	gn:6167865
	gci:defined_by	6.1_F_person
6.1_EF_person	rdfs:subClassOf	isoe:6.1_EnrolledFemaleSchoolAge_Person
	gcie:has_Enrollment	6.1_EF_enrollment
6.1_F_person	rdfs:subClassOf	isoe:6.1_FemaleSchoolAge_Person
6.1_EF_enrollment	rdfs:subClassOf	isoe:6.1_Enrollment
	gcie:for_SchoolYear	6.1_SchoolYear
	gcie:attends	cedar_grove
	gcie:enrolled_Courses	oc1, oc2, oc3, oc4, oc5, oc6, oc7
	gcie:enrolled_Grade	og1, og2, og3, og4, og5, og6
	gcie:enrolled_Program	ontarioPrimaryProgram
	gcie:enrolled_Status	fulltime
6.1_SchoolYear	rdfs:type	SchoolYear
	starting_Year	2014
jane_smith	rdfs:type	FemaleStudent
	org:memberOf	6.1_F_pop
	org:memberOf	6.1_EF_pop
	has_Primary_Residence	js_home
js_home	rdfs:type	ic:HomeAddress
	ic:hasCity	gn:6167865

The following shows how each consistency question for education indicator "6.1 Percentage of female school-aged population enrolled in schools" is implemented in SPARQL.

1. (F) What city is the indicator for?

SELECT ?city WHERE {6.1-ex gci:for_City ?city}

2. (CD) Are the students residents of the city?

Identifies each student that is a member of the Enrolled Female Population and verifies that their primary residence is the same city as the indicator's.

```
SELECT ?student WHERE
  { 6.1_ex gcie:for_City ?city .
  ?student org:memberOf 6.1_EF_pop .
  ?student gcie:has_Primary_Residence ?PR .
  ?PR ic:hasCity ?city }
```

3. (D) What is the age range for school age women?

```
SELECT ?start ?end WHERE
  { 6.1_ex gci:for_City ?city .
  ?city gcie:has_Primary_Grade_Level ?pgl .
  ?pgl gcie:starting_age ?start .
  ?city gcie:has_Secondary_Grade_Level, ?sgl .
  ?sgl gcie:ending_age ?end }
```

4. (F) Is a school a private or public institution?

SELECT ?status WHERE { cedar_grove org:has_Ownership ?status }

5. (F) Does a school teach Primary or Secondary courses?

SELECT ?ctype WHERE { cedar_grove gcie:has_SPType ?ctype }

6. (F) Is a school a home school? Religious school?

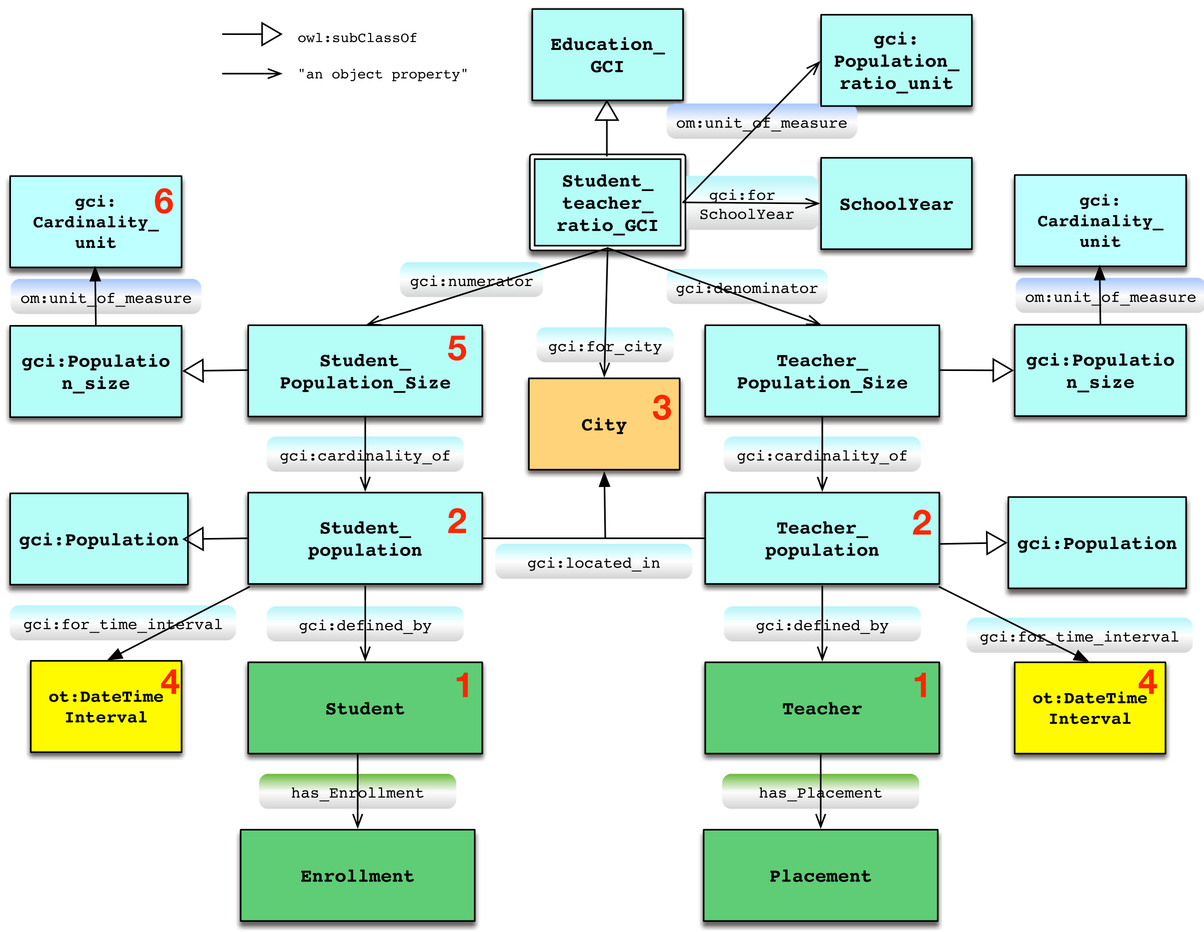

Fig. 16. Student-Teacher Ratio Pattern (simplified).

SELECT ?type WHERE { cedar_grove gcie:has_SchoolType ?type }

7. (D) Is the private school certified by the government?

SELECT ?govorg WHERE
{ cedar_grove has_Ownership privately_owned .
cedar_grove has_Certification ?cert .
?cert certified_By ?govorg .
?govorg rdfs:subclassOf GovernmentOrganization }

8. (F) What grades comprise primary (and secondary) school?

In order to answer this question properly, we would have to loop through the grades from the starting to the ending grade. We do not show that looping here. Instead we show the starting and ending grades.
SELECT ?sgrade ?egrade WHERE
{ ?gradelevel for_City toronto .
?gradelevel starting_Grade ?sgrade .
?gradelevel ending_Grade ?egrade. }

9. (F) How many hours of school do you have to attend to be full time?

The following will print out the hours and period for every program associated with the school cedargrove.
SELECT ?hours ?period WHERE
{ cedargrove delivers_Program ?program .
?program has_Fulltime_Hours .
?program has_Fulltime_Period ?period }

10. (D) What school did person X attend in year Y?

We answer this for a specific person, johnsmith, for the primary grade level for school year 2010.
SELECT ?school WHERE
{ johnsmith has_Enrollment ?enrol .
?enrol enrolled_Program ?gradelevel .
?gradelevel rdfs:subclassOf GradeLevelPrimary .
?enrol for_SchoolYear ?sy .
?sy starting_Year 2010 .
?enrol attends ?school }

11. (D) What proportion of the students are in private schools for school year x?

The following will return a count of students who enrolled in a primary grade level and taught at a private school for year 2010.
SELECT (COUNT(?studentpriv) AS ?Num) WHERE
{ ?studentpriv has_Enrollment ?enrol.
?enrol enrolled_Program ?gradelevel.
?gradelevel rdfs:subclassOf GradeLevelPrimary.
?enrol for_SchoolYear ?sy .

?sy starting_Year 2010 .
?enrol attends ?school .
?school has_Ownership privately_owned }

6.3. Education data mapping

Our final form of evaluation is to determine the extent to which education system data is a potential source of indicator-related data. We focus solely on the CEDS education data standard as it appears to be the most developed. In addition, it contains many properties beyond that required for the Education Ontology

In the following, we repeat each group of Education ontology classes and describe for each grouping of classes a possible mapping from CEDS properties to the Education ontology properties[11].

Teacher

Class	Property	Value Restriction
EducationStaff	rdfs:subClassOf	org:OrganizationAgent
	rdfs:subClassOf	sc:Person
	has_Placement	min 1 Placement
EducationStaffAdministrative	rdfs:subClassOf	EducationStaff
EducationStaffInstructional	rdfs:subClassOf	EducationStaff
Teacher	rdfs:subClassOf	EducationStaffInstructional
	has_Placement	min 1 (Placement and teaches min 1 Course)
	rdfs:equivalentClass	cyc:Teacher
Placement	educational_Staff_At	EducationFacility
	days_Worked	exactly 1 positiveInteger
	for_SchoolYear	exactly 1 SchoolYear
	teaches	min 1 Course
	min_Days_Worked	value 1

In CEDS, the core entity for staff, students, etc. is *Person*.[12] A *Person* has one or more *Roles* such as *K12StaffAssignment*'s which maps to 'Placement'. It specifies the linkage between a staff member and a course they teach. CEDS does not distinguish between administrative and instructional staff directly, but has a *RefTeachingAssignmentRoleid* which can be used to infer this. 'educational_Staff_At' can be inferred from the *CourseSectionLocation*, 'days_Worked' and 'min_Days_Worked' map to *RoleAttendance*, and 'for_School_Year' would have to be inferred from *RoleStatus* which specifies the start and end dates for a role. 'teaches' is specified in the *K12StaffAssignment* entity as the *Organizationid* (*Course* and *CourseSection* are sub entities of *Organization*).

Student

Class	Property	Value Restriction
Student	rdfs: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
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

'Student' is specified in a similar way as 'Teacher'. Instead of *K12StaffAssignment*, the *Role K12StudentEnrollment* is equivalent to 'Enrollment', linking a *Person* to a *CourseSection*. Basic information about a student, e. g, birthdate and resident, are provided by the *Person* entity. 'Enrollment' grade, courses, status and result are provided by the *K12StudentEnrollment*.

Educational Facility

Class	Property	Value Restriction
School	rdfs:subClassOf	EducationFacility
	delivers_Program	some SchoolProgram
	org:has_Ownership	exactly 1 Ownership
	has_SchoolType	min 1 SchoolType
	org:hasName	only xsd:string
	org:consistsOf	only org:Division
	org:hasLegalName	exactly 1 xsd:string
	org:hasGoal	only org:Goal
PublicPrimarySchool	rdfs:subClassOf	PublicSchool
	delivers_Program	some GradeLevelPrimary
	Has_SchoolType	value secularSchoolType
	has_Ownership	value government_ownership
PrivateSchool	rdfs:subClassOf	School
	has_Ownership	value privately_owned
	has_Certification	some Certification
Certification	certified_By	some GovernmentOrganization
	certification_Date	exactly 1 dateTime

CEDS' *K12School* entity provides the basic information about a school. Ownership is divided into types relevant to the US, e.g., charter school. *K12School RefSchoolLevel* maps to 'SchoolProgram'. 'hasGoal' and 'consistsOf' are not represented in CEDS. CEDS does not distinguish between public and private schools at the entity level, but can be inferred by the types. *AccredationAgencyName* maps to 'Certification'.

Program and grade level

Class	Property	Value Restriction
Program	has_Certification	some Certification
	has_Fulltime_Hours	exactly 1 positiveInteger
	has_Fulltime_Period	exactly 1 TimePeriod
SchoolProgram	rdfs:subclassOf	Program
	has_Course	min 1 Course
	has_SP_Type	all SP_Type
GradeLevel	rdfs: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
Cohort	rdfs:subClassOf	EducationThing
	starting_SchoolYear	exactly 1 SchoolYear
	ending_SchoolYear	exactly 1 SchoolYear
	for_GradeLevel	Exactly 1 GradeLevel

In CEDS there are programs (e.g., migrant programs, special education), but they are not related to GradeLevels as in the Education ontology. It appears to be assumed that grade levels are the same throughout the US, which is inappropriate when representing school systems globally. *K12StudentCohort* maps onto 'Cohort'.

Higher Education Degrees

Class	Property	Value Restriction
EducationDegree	rdfs:subClassOf	EducationThing
	awarded_Year	exactly 1 Year
	awarding_EducationFacility	exactly 1 EducationFacility
	enrolled_Program	only 1 EducationProgram
	degree_Name	exactly 1 string

CEDS refers to Post Secondary programs (*PsProgram*) with a *RefPSProgramLevelid*, Assuming *RefPSProgramLevelid* can refer to higher education degrees, then the remainder of the 'EducationDegree' properties can be inferred from *PsProgram*.

In summary, most of the information required by the ISO 37120 Education theme indicator definitions can be extracted from the CEDS data format. But a search of the web site Namara.io[13] turned up zero

[11] There exist alternative maps due to ambiguities and repetition of certain types of attributes in the CEDS data model specification.
[12] We will denote CEDS objects using the *italic* font.

[13] Namara.io is a web site that crawls open data web sites in cities across North America. It downloads the datasets it finds into its repository. It contains over 100,000.

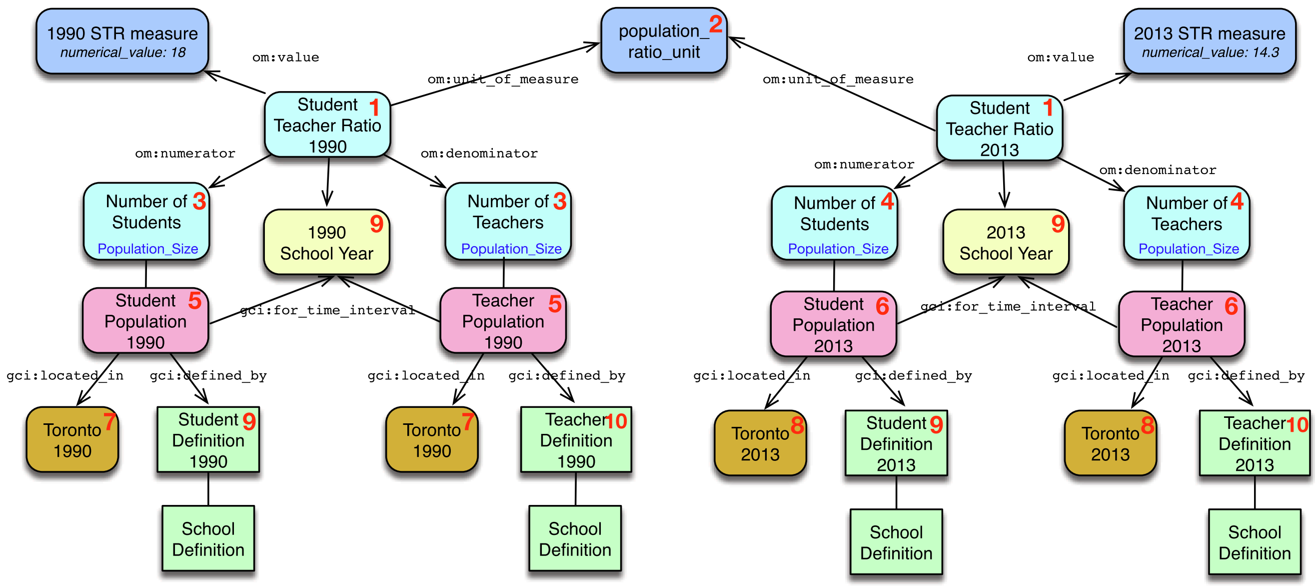

Fig. 17. Longitudinal Analysis of Toronto's Student-Teacher Ratio.

CEDS datasets.

7. Comparative analysis of city performance

In the previous section, multiple methods were employed to evaluate the ontology. These methods verified the consistency of the ontology, whether it is sufficient to answer the competency questions, and whether the data required to instantiate Education indicators can be extracted from a major education data model. In this section we describe how the ontology can be used to support comparative analysis of city performance, in particular:

- *Definition analysis:* analyzing whether a city's interpretation of an indicator is consistent with its published definition,
- *Longitudinal analysis:* analyzing the changing performance of a city over time, and
- *Transversal analysis:* analyzing the differences in performance of two or more cities.

In the remainder of this section we will use the following terms:

- *Standard Indicator* refers to an indicator defined and published by a Standards Development Organization (SDO). For example, ISO 37120 and the UNSDG's indicators. It is assumed that the standard indicator is represented and published using the GCI ontologies.
- *Reported Indicator* refers to the application of a standard indicator by a city and openly published. For example, the City of Toronto reporting the application of the ISO 37120 standard Education indicators to Toronto for 2013. It is assumed that the reported indicator is represented and published using the GCI ontologies. This includes any specializations of classes in the standard indicator, new classes and all instances.

7.1. The challenge of inconsistencies in indicator interpretation

A major goal of defining city indicators, whether for measuring smartness, sustainability or resilience, is to be able to compare cities or to evaluate how a city has changed over time. In both cases the intent is to discover root causes of success/failures and ultimately identify best practices. The challenge with indicator-based analyses, is that cities rarely interpret and apply them consistently (Hoornweg et al., 2007; McCarney, 2013; Slack, 2017). It is these inconsistent interpretations/applications of indicators that lead to flawed comparisons. Hence, before any meaningful "apples to apples" comparison can be performed, either longitudinally or transversally, the reported indicators must satisfy three consistency requirements (Wang & Fox, 2017):

- *Definition consistent*: Is a city's reported indicator consistent with the indicator's definition. E.g., is the definition of student and teacher populations reported by a city consistent with the indicator's definitions? Does a city's definition of a teacher conform to the definition of a teacher in the standard's indicator?
- *Intra-indicator consistent*: Is the data used to derive an indicator's value, internally consistent? E.g., are the student and teacher populations in the indicator from the same time and place? Note that a city's interpretation of an indicator can be definitionally consistent, but not intra-indicator consistent. Secondly, while there exists temporal and spatial relations in the indicator patterns that restrict temporal and spatial values between branches (e.g., numerator and denominator) of an indicator definition, they are usually not enforced in data published by a city.
- *Inter-indicator consistent*: When comparing an indicator longitudinally or transversally, do they interpret the indicator's definition in the same way? E.g., is the city's definitions of students and teachers consistent across time, or between two cities? Note that a city's interpretation of an indicator can be definitionally consistent, but not inter-indicator consistent.

Moreover, it will be shown that the existence of these inconsistencies can be the root cause for differences in reported indicators.

7.2. Consistency pittfalls

A city's published indicator data can be consistent, potentially consistent, or inconsistent. A city's reported indicator is consistent if it satisfies the standard indicator's definition, and is consistent with the data it is being transversally or longitudinally compared to. A city's indicator data may be potentially consistent if there is a possible interpretation of the indicator data that is consistent. In this section, four

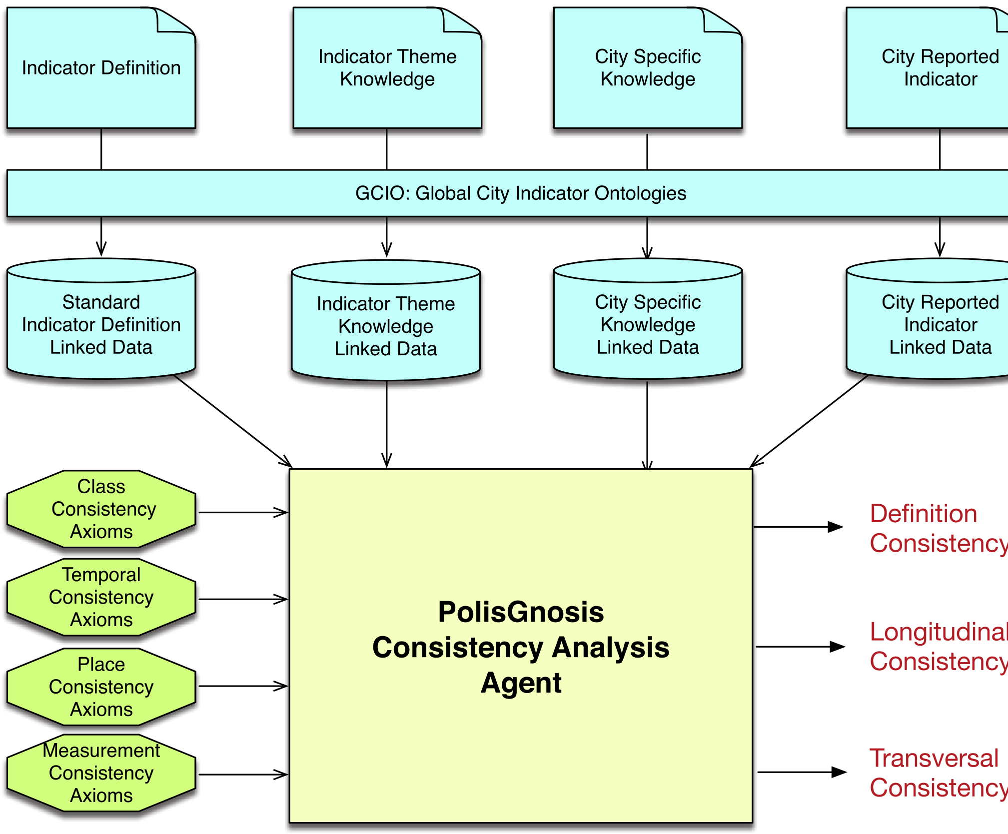

Fig. 18. PolisGnosis Agent Information Architecture.

types of consistency pittfalls[14] are described. Examples of each pitfall are made with reference to Fig. 16, which depicts a simplified version of the student–teacher ratio pattern

7.2.1. Class consistency pitfall

Class consistency determines whether corresponding classes across two indicator patterns are consistent. Two classes correspond if they are reached by the same property path from the root of the indicator pattern. For example, two Population classes correspond if they are reached by the path numerator → cardinality_of in Fig. 16.

A class in a reported indicator is *definition class consistent,* if it is a subclass of the corresponding class in the standard indicator.
A class in a reported indicator is *inter-indicator class consistent* if is equivalent to the corresponding class in the second reported indicator. If one is a subclass of the other, then they are not measuring the same thing.

Assume Fig. 16 represents the standard indicator containing the Teacher class (labeled "1″). The corresponding class in a reported indicator would be definition class consistent if it is equivalent to or a subclass of Teacher. If longitudinal or transversal analysis is being performed, then two reported indicators, separated by time or space respectively, would be evaluated for inter-indicator class consistent. If class consistency fails, then the indicators are not comparable. For example, they have incompatible definitions of Teacher.

7.2.2. Temporal consistency pitfall

Temporal consistency determines whether two parts of a single reported indicator were measured at that same time (intra-indicator temporal consistency), or whether the reported indicators of two different cities where measured at the same time.

The instances of two corresponding classes (that have time intervals) are *temporal consistent* if their time intervals are equal. If the intervals are related by overlap, during or meet, then they are potentially temporal consistent.

For example,Teacher_Population and Student_Population classes in Fig. 16 (labeled "2"), have a for_time_interval property. For intra-indicator temporal consistency, the DateTimeIntervals (labeled "4") associated with instances of both Teacher and Student populations would have to be equal. They are potentially temporally consistent if they overlap, are during or meet, For inter-indicator temporal consistency, the DateTimeInterval in the class in the first reported indicator would have to be equivalent to the DateTimeInterval of the instance of the corresponding class in the second reported indicator. They are potentially temporal consistent if they overlap, meet or are during.

7.2.3. Place consistency pitfall

Place consistency determines whether two parts of a single reported indicator were measured at that same place (intra-indicator place consistency), or whether the reported indicators of the same city at different times were measured at the same place (inter-indicator place consistency).

The instances of two corresponding classes (that have located_in properties) are *place consistent* if the geometry of where they are located are equal. If the geometries are related by overlap then they are potentially place consistent.

For example, Teacher_Population and Student_Population classes in Fig. 16 (labeled "2"), have a located_in property that links to a City (labeled "3") which is a subclass of Feature which has an associated

[14] The term pittfalls, in the context of validating ontologies, originates with Poveda-Villalón et al. (2012).

geometry. For intra-indicator place consistency, the City geometries associated with instances of both Teacher and Student populations, in the reported indicators, would have to be equal. They are potentially place consistent if they overlap, are during or meet, For inter-indicator place consistency, the City's geometry in the first reported indicator would have to be equivalent to the geometry of the instance of the corresponding class in the second reported indicator. They are potentially place consistent if they overlap.

7.2.4. Measurement consistency pitfall

Measurement consistency determines whether the quantities specified in the two indicators are measuring the same thing from unit of measurement perspective. It determines whether the quantities in a reported indicator is consistent with the standard indicator or quantities in two reported indicators are measurement consistent.

> Two Quantities are measurement consistent if their unit of measure are equivalent.

A simple example of measurement consistency is if a distance quantity in one reported indicator is being measured in meters and the other reported indicator in feet. In the student–teacher ratio example, Student_Population_size (labeled "5″) may be measured (labeled "6") by one city reported indicator as a multiple population cardinality unit (i.e., kilopc) versus another city's reported indicator as a non-multiple population cardinality unit (i.e., pc). An indicator has multiple quantities that have to be checked for measurement consistency.

7.3. Case Study: Consistency in indicator longitudinal analysis

To understand how consistency pitfalls can be used in the comparative analysis of city indicators, consider two (simplified) reported student–teacher ratio indicators for the City of Toronto, at two points in time: 1990 and 2013, depicted in Fig. 17 (rounded boxes are instances, squared boxes are classes). The task is to perform a longitudinal analysis to understand why the student–teacher ratio *decreased* between 1990 and 2013. If the reported indicators only provided a Measure for the ratio quantity (18 for 1990 and 14.3 for 2013), then all we could say is that the value of the indicator increased or decreased. If in this case it decreased, we do not know why. If additional information was provided in the reported indicators about the numerator and denominator for 1990 and 2013, all that could be said is that the number of students and teachers increased, decreased or stayed the same from 1990 to 2013. Again, no insight into the root cause. It is only by digging deeper into the indicator definition that possible root causes may emerge.

The first step in comparing the two indicator patterns is to perform measurement consistency on quantities. For example, the Student Teacher Ratio quantities for 1990 and 2013, labeled "1", have their unit_of_measure values, labeled "2", tested for equivalence. In this case they are equivalent, i.e., population_ratio_unit. Other quantities such the number of teachers and students in the 1990 indicator (labeled "3") are evaluated to see that their corresponding quantities in the 2013 indicator (labeled "4") are measurement consistent. Descending through the two indicators, we reach Student and Teacher Populations for 1990 (labeled "5") and 2013 (labeled "6"). There are three types of Population-related consistency checks performed:

1. *Place consistency.* Two versions of place consistency can be performed, intra and inter indicator. In the intra case, the value of the located_in property for 1990 Student and Teacher Populations, i.e., Toronto 1990 (labeled "7"), are checked for equivalent geometry, that is the boundaries of the city were the same when determining members of the populations of students and teachers. In the inter case, the 1990 located_in places are checked for equivalent geometry with their corresponding places in 2013 (labeled "8").
2. *Temporal consistency:* Similar to place consistency, inter- and intra-indicator temporal consistency is performed for the values of the for_time_interval property (labeled "9").
3. *Class consistency*: Both the student and teacher Populations are defined_by a Student and Teacher class respectively. Class consistency is performed on the 1990 and 2013 corresponding Student classes, labeled "9", and the 1990 and 2013 Teacher corresponding classes, labeled "10".

There are two potential root causes for the decrease in student–teacher ratio between the reported indicators. First is due to place consistency. In the beginning of 1998, the City of Toronto was amalgamated with surrounding municipalities. The administrative boundaries of Toronto changed along with its physical boundaries. Hence the city's spatial area changed. A possible result of the amalgamation is that student–teacher ratios of surrounding municipalities were lower than the original city. By combining them the overall ratio was reduced.

A second potential root cause is revealed using class consistency. In 2009 the province of Ontario, which oversees education across the province including Toronto, issued its "Ontario's Equity and Inclusive Education Strategy" (OMoE, 2009), which focused on all types of inclusions, including special needs. Consequently, the definition of student may have changed to include special needs students. This would result in a small increase in the student population, and a larger increase in the teacher population required to support them. Hence, from 1990 to 2013, the teacher population would have grown more quickly.

Intra-indicator consistency is also a persistent problem. Given the difficulty of measuring populations, it is often the case that reported indicators contain intra-indicator place and temporal inconsistencies. For example, not all sections of a city may be included in a population due to lack of data. Or, the time in which one population is measured differs for another. We have seen cities report ratio indicators where the time interval for the numerator is different than the denominator, often by more than one year as data may not be gathered annually. This is more common than one might expect.

7.4. PolisGnosis consistency analysis agent

An "intelligent agent" has been created that implements definition, longitudinal and transversal consistency analysis of city indicators, as described in the previous section (Wang, 2016; Wang & Fox, 2017). Fig. 18 depicts the information architecture for the PolisGnosis consistency analysis agent.

The agent's inputs fall into four categories. Each of these inputs are represented using the GCI Foundation and Theme ontologies, which in the latter case is the Education theme:

1. *Indicator Definition*: The definition of the indicator that is being analysed. For example, ISO 37120 definition of a Student-Teacher ratio.
2. *Indicator Theme Knowledge:* The definition of classes for a theme. For example, definitions of teacher, student, grade, etc. found in the ISO 37120 Education theme indicators definitions.
3. *Indicator City Specific Knowledge*: The definition of classes unique to the city for a theme. For example, the ISO 37120 definition of Secondary school may be specialized to identify the grades that comprise secondary school.
4. *City Indicator:* This is a reported indicator which includes the information an indicator definition requires, but may be partial due to unavailability.

For definition consistency, the agent performs consistency analysis between the standard indicator definition and the reported indicator. For longitudinal analysis, it performs consistency analysis between two reported indicators for the same indicator, for the same city, for different times. For transversal analysis, it performs consistency analysis between two reported indicators for the same indicator, for different cities, and may or may not be for the same time. For longitudinal and transversal analysis it uses the indicator definition and theme knowledge to guide the consistency analysis process.

8. Conclusions

As more cities align with the United Nations Sustainable Development Goals, and adopt standards such as ISO 37120, for measuring their performance across themes including education, the level of awareness and scrutiny of a city's education system's performance both within the city and by other cities will grow. Whether you agree with the specific indicators chosen or not, their influence will grow and educational systems will be required to openly report their data and explain their performance. With this future fast approaching there are a number of questions that need to be answered:

1. What knowledge and data needs to be gathered and published?
2. What format should the data be published in?
3. What vocabulary/ontologies should be used to publish the data?

The ISO 37120 education indicators identify, in a broad sense, the data to be gathered and published. But it was the process of developing the Education Ontology that a clearer understanding of the variety of knowledge/information required emerged. In order to represent the definitions and data of what appear to be rather simple indicators, an ontological infrastructure spanning educational institutions, programs, certification, cohorts, etc. had to be created; it was unexpected that the existing education ontologies, at least the ones we could find, would provide little of what was needed.

As to the format, the future is known: many cities have come to realize that their open data portals have to publish their data using Semantic Web standards. As to what vocabulary/ontologies to use, the education ontology presented in this paper is the first comprehensive ontology for representing the knowledge and data required by the ISO 37120 Education theme indicators.

In summary, this research makes three contributions:

1. Defines an Education Ontology that is broader and deeper than existing education ontologies, but focused on supporting the definition of ISO 37120 education indicators;
2. Defines each ISO 37120 education indicator using the Education Ontology, thereby providing a computationally precise definition; and
3. Publishes the ISO 37120 education indicator definitions using Semantic Web standards, thereby making it possible to reason about the definitions and instances using existing ontology/semantic web/linked data tools.

The capability to both represent and openly publish education indicator definitions, education common sense knowledge, and the data used to derive a city's indicator values, makes it possible to create intelligent agents that can verify that a city's indicators conform to the indicator definitions, and identify root causes of performance.

Never the less, a major hurdle remains, namely the adoption of the Education ontology by cities. This hurdle will be cleared when cities specify in their RFPs for their next of generation enterprise systems that standards, such as above, be incorporated into their systems.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was supported, in part, by the Natural Science and Engineering Research Council of Canada. The PolisGnosis analysis agent is based on joint work with Yetian Wang.

Appendix

The Global City Indicator Foundation ontology can be found in: http://ontology.eil.utoronto.ca/GCI/Foundation/GCI-Foundation-v2.owl.

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

URIs for all of the ISO 37120 indicators can be found in: http://ontology.eil.utoronto.ca/ISO37120.owl.

Definitions of the ISO 37120 education indicators, using the GCI Foundation and Education ontologies can be found in: http://ontology.eil.utoronto.ca/GCI/ISO37120/Education.owl.

References

CEDS. (2017), Common Education Data Standards (CEDS) Version 7 Data Model Guide. https://ceds.ed.gov/pdf/CEDS-7-0-Data-Model-Guide_for_508.pdf.

de Freitas, E., & Dixon-Román, E. (2017). The Computational Turn in Education Research: Critical and Creative Perspective on the Digital Data Deluge. *Research in Education, 98*(1), 3–13.

Fox, M.S. (2013). A Foundation Ontology for Global City Indicators. Working Paper No. 3, Global Cities Institute, University of Toronto. Revised 12 April 2015: http://eil.utoronto.ca/smartcities/papers/GCI-Foundation-Ontology.pdf.

Fox, M.S. (2015). The Role of Ontologies in Publishing and Analyzing City Indicators. Computers, Environment and Urban Systems, Vol. 54, pp. 266-279. http://eil.mie.utoronto.ca/wp-content/uploads/2015/06/fox-ceus15.pdf.

Fox, M. S. (2017). *The PolisGnosis Project: Enabling the Computational Analysis of City Performance*. Institute for Industrial and Systems Engineering.

Fox, M. S. (2018). Semantics of Populations: A City Indicator Perspective. *Journal of Web Semantics, 48*, 48–65.

Fox, M. S., Barbuceanu, M., Gruninger, M., & Lin, J. (1998). An Organisation Ontology for Enterprise Modeling. In M. Prietula, K. Carley, & L. Gasser (Eds.), *Simulating Organizations: Computational Models of Institutions and Groups* (pp. 131–152). Menlo Park CA: AAAI/MIT Press.

Fox, M.S., and Huang, J. (2005). 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 Pettit, C.J. (2015). On the Completeness of Open City Data for Measuring City Indicators. Proceedings of the IEEE Conference on Smart Cities, Guadalajara Mexico. http://eil.mie.utoronto.ca/wp-content/uploads/2015/06/CIDOM-IEEE15.pdf.

Grüninger, M., & Fox, M. S. (1995). *Methodology for the Design and Evaluation of Ontologies*. Montreal, Canada: IJCAI-95.

Gulson, K. N., & Webb, P. T. (2017). Mapping an Emergent Field of 'Computational education policy': Policy Rationalities, Prediction and Data in the Age of Artificial Intelligence". *Research in Education, 98*(1), 14–26.

Hitzler, P., et al. (2012). *OWL 2 Web Ontology Language Primer* ((2nd Edition).).

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., & Villaveces, M. (2007). City indicators: Now to Nanjing. World Bank policy research working paper 4114.

Horridge, M., and Patel-Schneider, P.F. (2013). OWL 2 Web Ontology Language Manchester Syntax (2nd Edition). W3C Working Group Note 11 December 2012, http://www.w3.org/TR/owl2-manchester-syntax/.

Huang, J., & Fox, M. S. (2006). An Ontology of Trust – Formal Semantics and Transitivity. In *Proceedings of the International Conference on Electronic Commerce* (pp. 259–270).

Hugh, S., and Fox, M.S. (2018). Homelessness and open city data. in Open Cities Open Data : collaborative cities in the information era, S. Hawkin, H. Han, C. Pettit (Eds), Palgrave MacMillan Pub. 2018.

ISO 37120 (2014). ISO 37120: Sustainable Development of Communities – Indicators for City Services and Quality of Life. International Organization for Standardization, First Edition, 2014-05-15, ISO 37120:2014(E).

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. (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.

Mohamed, A., Liu, D., Summerfield, J., Rissen, P., Lee, R., and Rose, Z. (2013). Curriculum Ontology, 2013-04-03. http://www.bbc.co.uk/ontologies/curriculum/.

Niles, I., & Pease, A. (2001). *Towards a Standard Upper Ontology*. Ogunquit, Maine. New York: ACM Press.

OMoE. (2009). Ontario's Equity and Inclusive Education Strategy, Ontario Ministry of Education. http://www.edu.gov.on.ca/eng/policyfunding/equity.pdf.

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.

Poveda-Villalón, M., Suárez-Figueroa, M. C., & Gómez-Pérez, A. (2012). *Validating ontologies with OOPS!, EKAW2012*, 8–12.

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*(2), 276–287.

Slack, E. (2017). How Much Local Fiscal Autonomy Do Cities Have? A Comparison of Eight Cities around the World. IMFG Perspectives, No. 19, Institute on Municipal Finance and Governance, University of Toronto.

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

Wang, Y. (2016). *City indicator consistency analysis: An approach to inconsistency detection.* University of Toronto.

Wang, Y., and Fox, M.S. (2017). Consistency Analysis of City Indicator Data. in S. Geertman et al. (eds.), Planning Support Science for Smarter Urban Futures, Lecture Notes in Geoinformation and Cartography, DOI 10.1007/978-3-319-57819-4_20.