



# PolisGnosis Project

Enabling the Computational Analysis of City Performance

---

**Mark S. Fox**

Distinguished Professor of Urban Systems Engineering

University of Toronto

[msf@eil.utoronto.ca](mailto:msf@eil.utoronto.ca), [www.eil.utoronto.ca](http://www.eil.utoronto.ca)

12 October 2015

**Papers:** [eil.utoronto.ca](http://eil.utoronto.ca)

**Ontologies:** [ontology.eil.utoronto.ca](http://ontology.eil.utoronto.ca)



# City Indicators

---

- World Bank and Province of Ontario funded a study at the the University of Toronto of performance indicators at 9 cities.  
(P. McCarney, UofToronto)
- Belo Horizonte, Brazil
- Bogota, Colombia
- Cali, Colombia
- King County, Washington State, USA
- Montreal, Canada
- Toronto, Canada
- Vancouver, Canada
- Porto Alegre, Brazil
- Sao Paulo, Brazil

**Total of 1100 indicators across 9 pilot cities.**

**Only 2 comparable**

# ISO 37120 (May 2014)



## Over 100 indicators defined

### City Indicator Themes

- Economy
- Education
- Energy
- Environment
- Finance
- Fire and Emergency Response
- Governance
- Health
- Recreation
- Safety
- Shelter
- Solid waste
- Telecommunication and innovation
- Transportation
- Urban Planning
- Wastewater
- Water and sanitation

### Profile Indicator Themes

- People
- Housing
- Economy
- Government
- Geography and climate

***ISO 37120 – Sustainable  
Development and Resilience of  
Communities – Indicators for  
City Services and Quality of Life,  
May 2014.***



## 6.4 Primary 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 **preschool** teachers and staff **shall not be included**.
- The number of teachers shall be counted in fifth time increments, for example, a teacher working one day per week should be counted as 0.2 teachers, and a teacher working three days per week should be counted as 0.6 teachers."
- **Grades, courses, Catholic school, ...**

# World Council on City Data (2015)

<http://www.dataforcities.org/>



## Global Cities Registry™ for ISO 37120

The **WCCD Global Cities Registry™** is the internationally recognized list of cities that are certified against ISO 37120 in accordance with the WCCD certification system.

The data for all cities listed has been independently verified and deemed to be in conformity with ISO 37120 according to the WCCD. All cities in the Global Cities Registry™ have provided third-party verified data to the WCCD Open Data Portal, allowing for city-to-city comparisons, cutting-edge visualizations and customized trend analyses.

Once certified, cities will be added to the WCCD Global Cities Registry™ for a period of one year. Cities must apply for certification and registration on an annual basis.

Identification Number	City	Country	Reporting Year	Certification Level
2014-P-0001	Amman	Jordan	2014	Platinum
2014-A-0002	Amsterdam	Netherlands	2014	Aspirational
2014-P-0003	Barcelona	Spain	2014	Platinum
2014-A-0004	Bogotá	Colombia	2014	Aspirational
2014-P-0005	Boston	United States of America	2014	Platinum
2014-G-0006	Buenos Aires	Argentina	2014	Gold
2014-P-0007	Dubai	United Arab Emirates	2014	Platinum
2014-P-0008	Guadalajara	Mexico	2014	Platinum
2014-A-0009	Haiphong	Vietnam	2014	Aspirational
2014-A-0010	Helsinki	Finland	2014	Aspirational
2014-A-0011	Johannesburg	South Africa	2014	Aspirational
2014-P-0012	London	United Kingdom	2014	Platinum

# Toronto ISO 37120 for 2013



## Toronto's Results for Global City Indicators Under ISO 37120

ISO Theme/Indicator Under ISO 37120	Indicator Result/Rate	Absolute Value	Year of Data	Comments
5.6 - Number of businesses per 100,000 population (supporting indicator)	3,360 businesses (per 100,000 population)	93,142 businesses (in total)	2013	
5.7 - Annual number of new patents per 100,000 population (supporting indicator)	8.2 patents (per 100,000 population)	228 patents (in total)	2013	
<b>Section 6 – Education</b>				
6.1 - Percentage of female school-aged population enrolled in schools (core indicator)	100% of girls enrolled	169,740 girls enrolled (in total)	2011 census/ NHS	<ul style="list-style-type: none"> <li>• Compulsory education to age 18</li> <li>• Includes students in private schools and home- schooled</li> </ul>
6.2 - Percentage of students completing primary education: survival rate (core indicator)	100% of students complete primary school		2013	Compulsory education to age 18
6.3 - Percentage of students completing secondary education: survival rate (core indicator)	83.2% of students complete high school		2013	<ul style="list-style-type: none"> <li>• For 2008 grade 9 cohort (graduated by fall 2013)</li> <li>• Blended result combining:                             <ul style="list-style-type: none"> <li>➢ Toronto District School Board (TDSB) - 82.9%</li> <li>➢ Toronto Catholic District School Board (TCDSB) – 83.8%</li> </ul> </li> </ul>
6.4 - Primary education student/teacher ratio (core indicator)	14.6 students per teacher	185,685 students and 12,691 teachers in grades 1-8 (in total)	2013	<ul style="list-style-type: none"> <li>• Blended result combining:                             <ul style="list-style-type: none"> <li>➢ TDSB – 15.02</li> <li>➢ TCDSB – 13.65</li> </ul> </li> </ul>

# Problem

---





# Data Safari



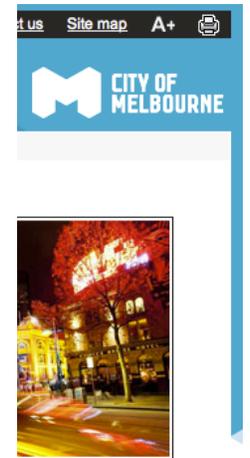
City of Toronto / A

## Open Data

### Data catalogue

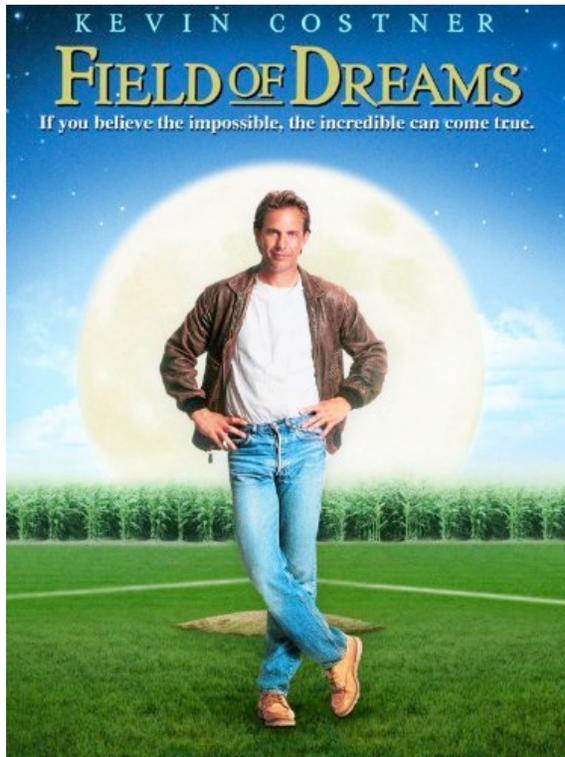
- Business
- City government
- Community services
- Culture and tourism
- Development and in
- Environment
- Finance
- Garbage and recycl
- Health
- Locations and mapp
- Parks and recreator
- Permits and licenser
- Public safety
- Transportation
- Water
- Policy
- Licence
- Frequently asked ques
- Glossary of terms
- Apps
- Data eh?

Hi-res JPG file download - Resolution 5000x3750px - www.psdgraphics.com



# Field of Dreams Effect

- If we publish it, they will come!



- What are the problems you want solved?
- Does the published data support its solution?

# But much (not all) of the data is ...





# PolisGnosis Project

---



# Goal 1: What to Publish

---

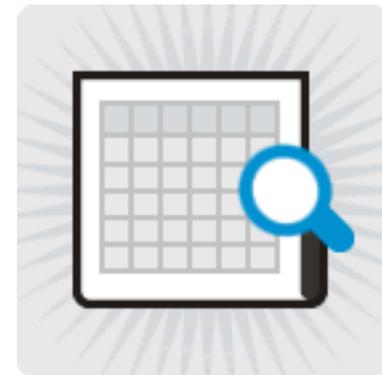
- Cities lack direction on what data sets should be published on their Open Data web.
- Our goal is to identify the content that needs to be published to support the analysis of their indicator performance.

# Capturing the Unwritten Narrative

---

- Over the last 10 years we have seen a growing awareness of the need to independently publish the data upon which research is based.
  - RDA: Research Data Alliance
- While the focus is on providing meta information about a data set as a whole, the problem we face with city indicators is that each datum has a substantial, unwritten narrative that is not captured.

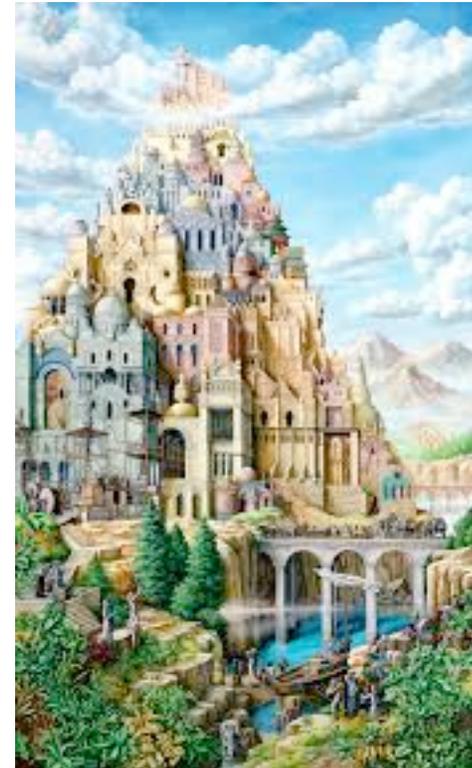
How can you drill down when there is nothing to drill?



# Goal 2: How to Publish

---

- Cities lack direction on the Format and Vocabulary for the indicator-related data they publish.
- Our goal is to provide standard ontologies for the publishing of indicator-related city data on the Semantic Web.



# Goal 3: Automated Analysis

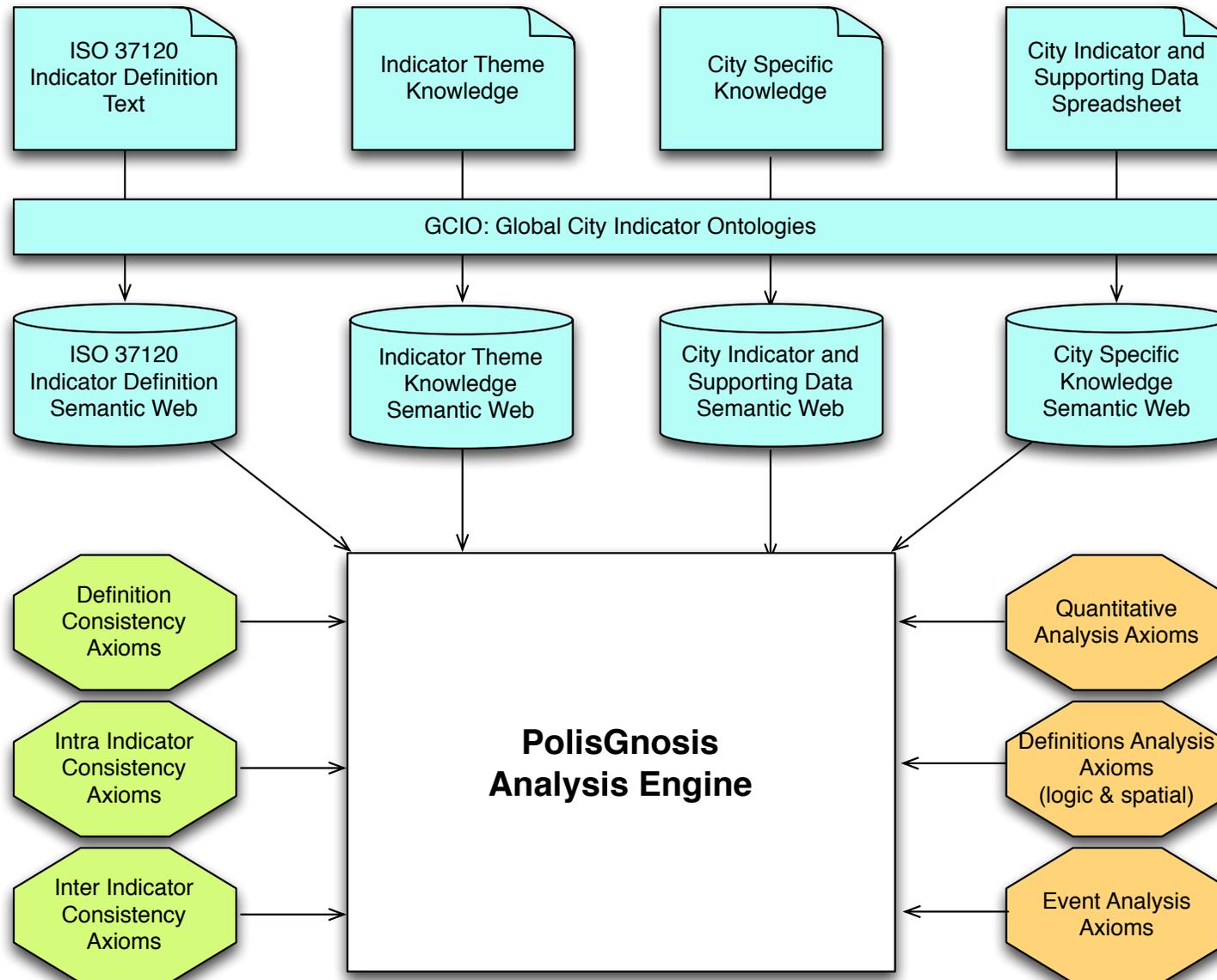
---

- Though the intent of Open Government and Open Data is to enable public oversight, the vast amount of information and the complexity of analysis limits what citizens can do (and they can't be bothered).
- Our Goal is to provide tools to automate the analysis of city performance
  - Perform longitudinal and transversal analyses, and
  - Determine the root causes of differences,using data from across the semantic web.





# PolisGnosis Vision





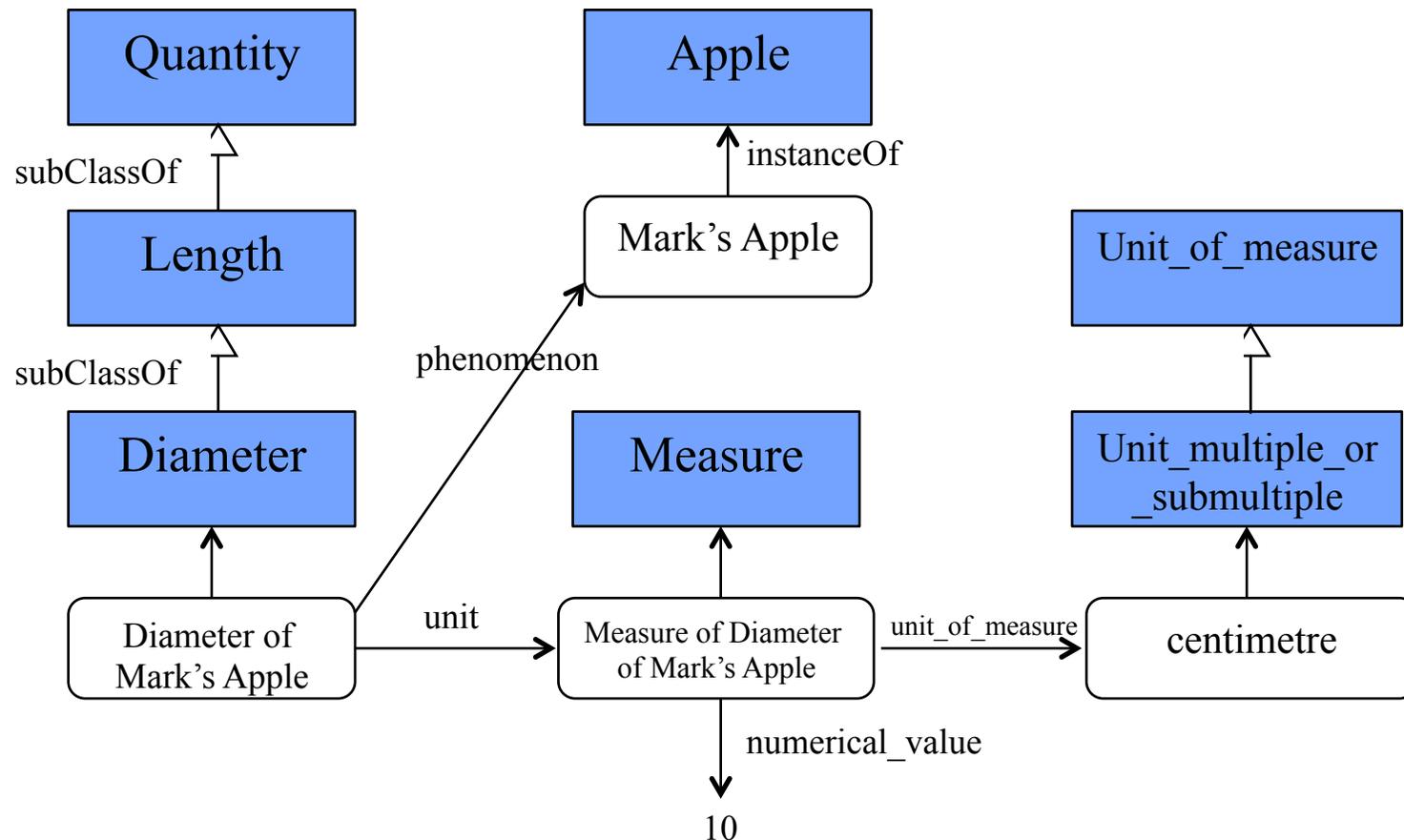
# Indicator Meta-Data

---

14.6

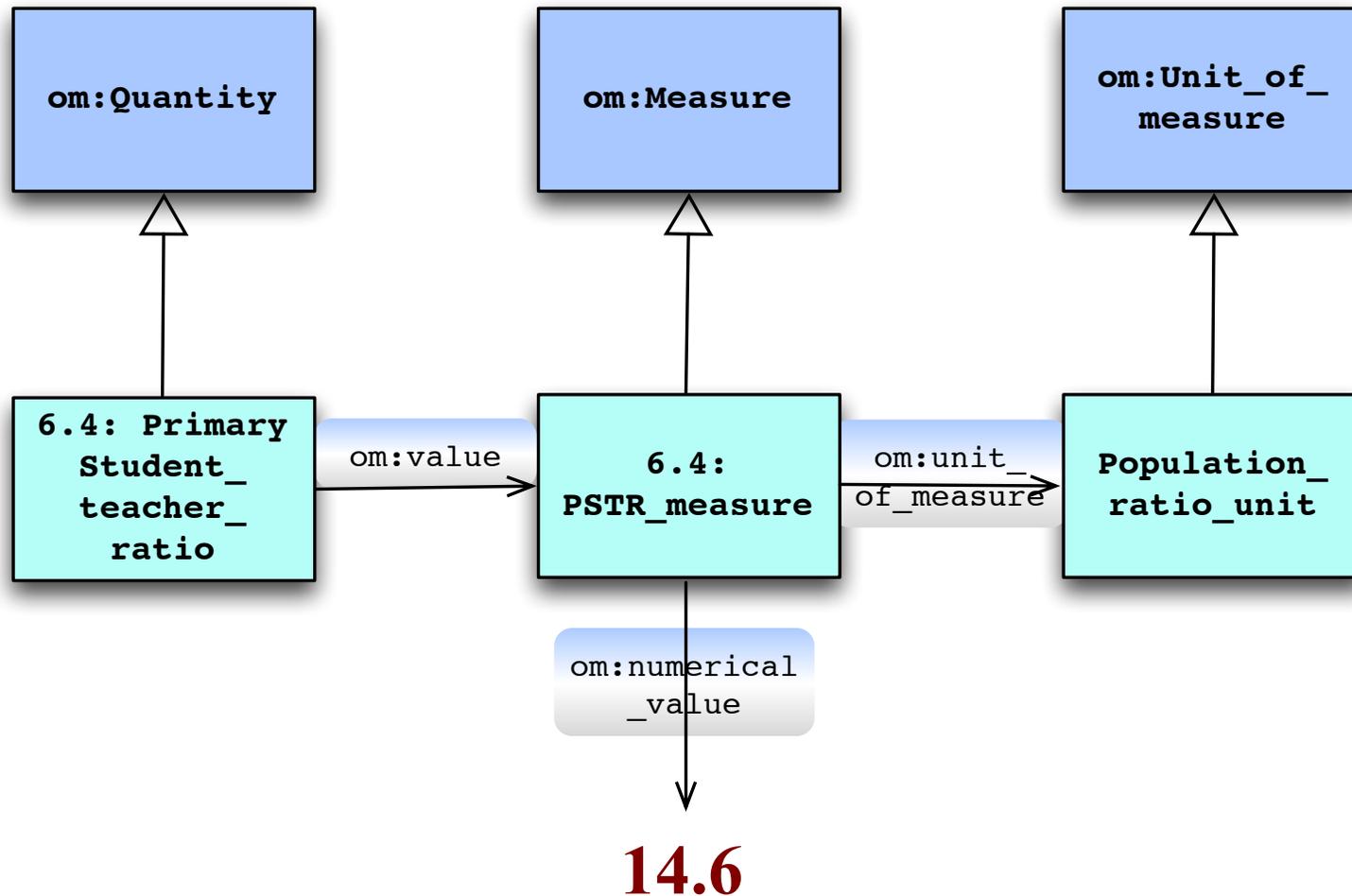
# Quantity, Measure, Unit of Measure

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





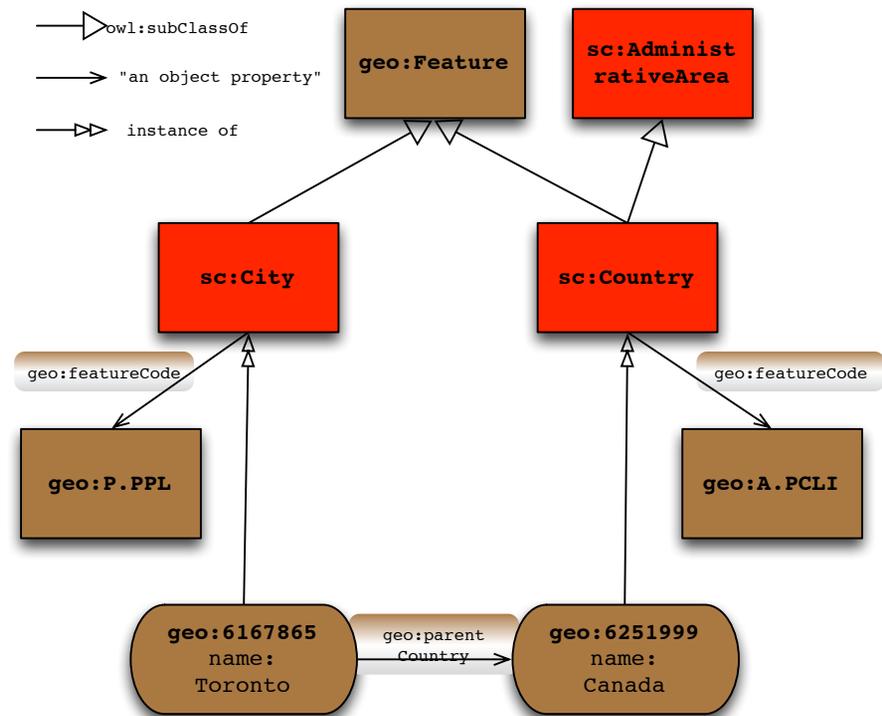
# 6.4 Primary Student Teacher Ratio





# Identifying Place

- What is the city being measured?
- What area does it cover?
- What places does it contain?
- Geonames provides a database of over 10M “placenames”. Builds on the classes defined in:
  - [schema.org](http://schema.org): city, country
  - [Linkedgeodata.org](http://www.geonames.org/ontology/ontology_v3.1.rdf#): hospitals, valleys, prisons, rivers, ...



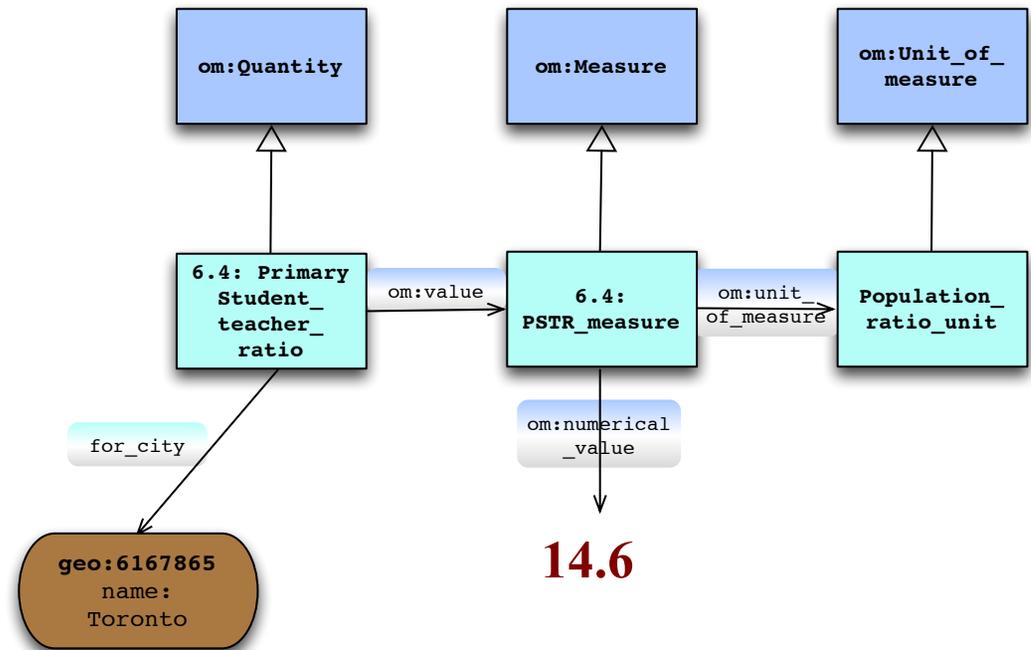
**sc:** <http://schema.org/> ,

**geo:** [http://www.geonames.org/ontology/ontology\\_v3.1.rdf#](http://www.geonames.org/ontology/ontology_v3.1.rdf#)



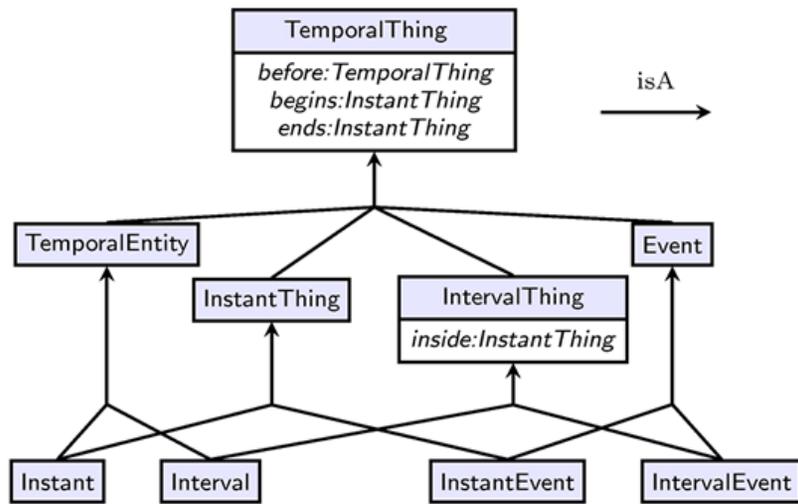
# Placenames

- To uniquely identify a city, we use the IRI (unique International Resource Identifier) provided by Geonames.
- The IRI is defined in terms of the classes provided in schema.org, linkedgeodata, etc.



# Time

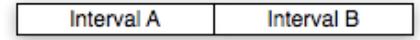
- How do we represent time?



A is before B or  
B is after A



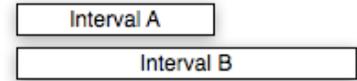
A meets B or  
B is met by A



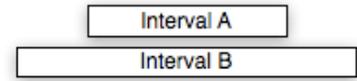
A overlaps with B or  
B is overlapped by A



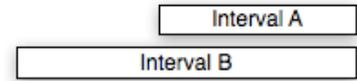
A starts B or  
B is started-by A



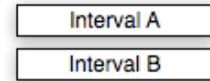
A during B or  
B contains A



A finishes B or  
B is finished-by A



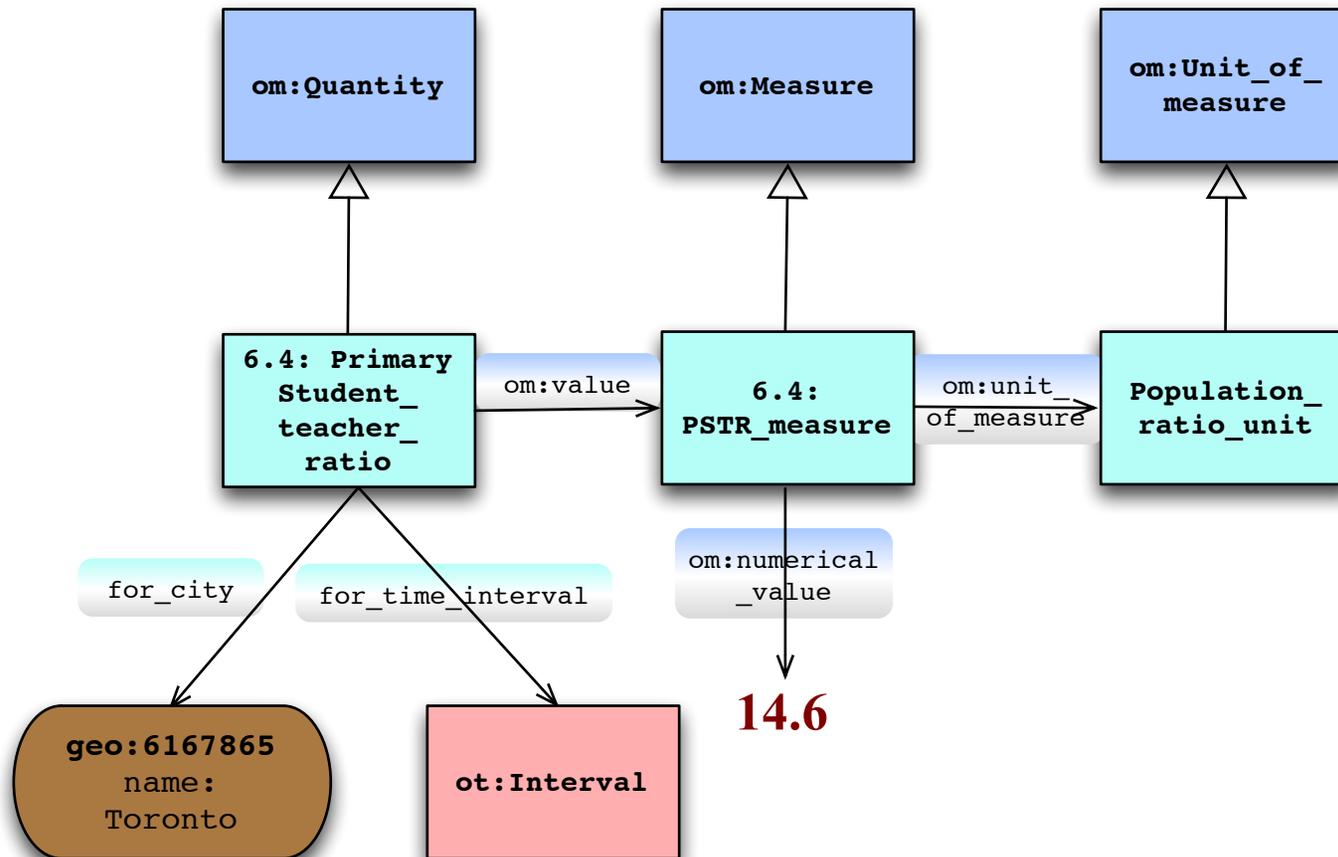
A and B are cotemporal



Hobbs, J.R., and Pan, F., (2006), “Time Ontology in OWL”, <http://www.w3.org/TR/owl-time/>.

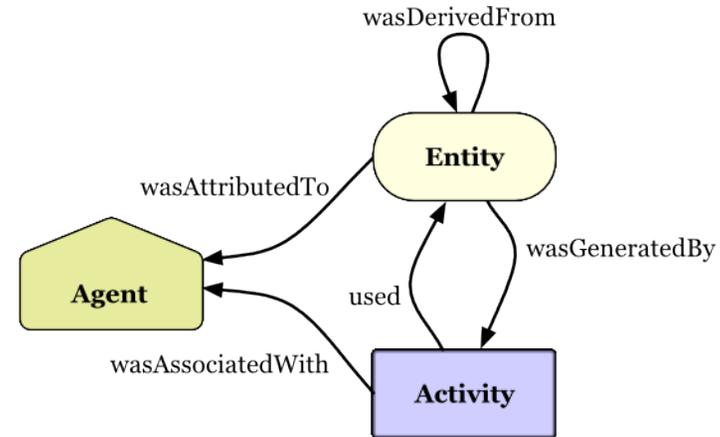


# Specifying the Year of the Indicator

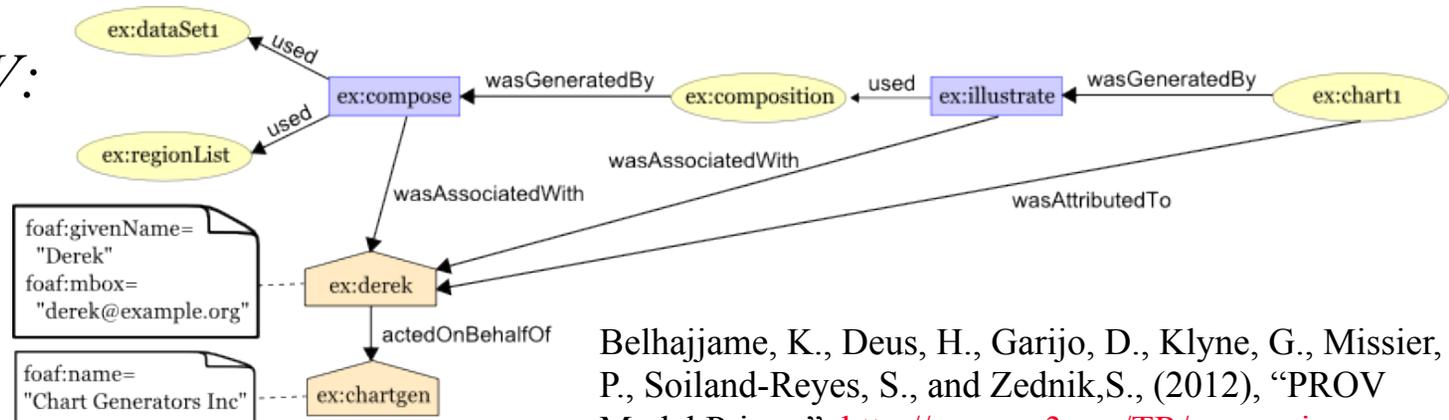


# Specifying Provenance

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



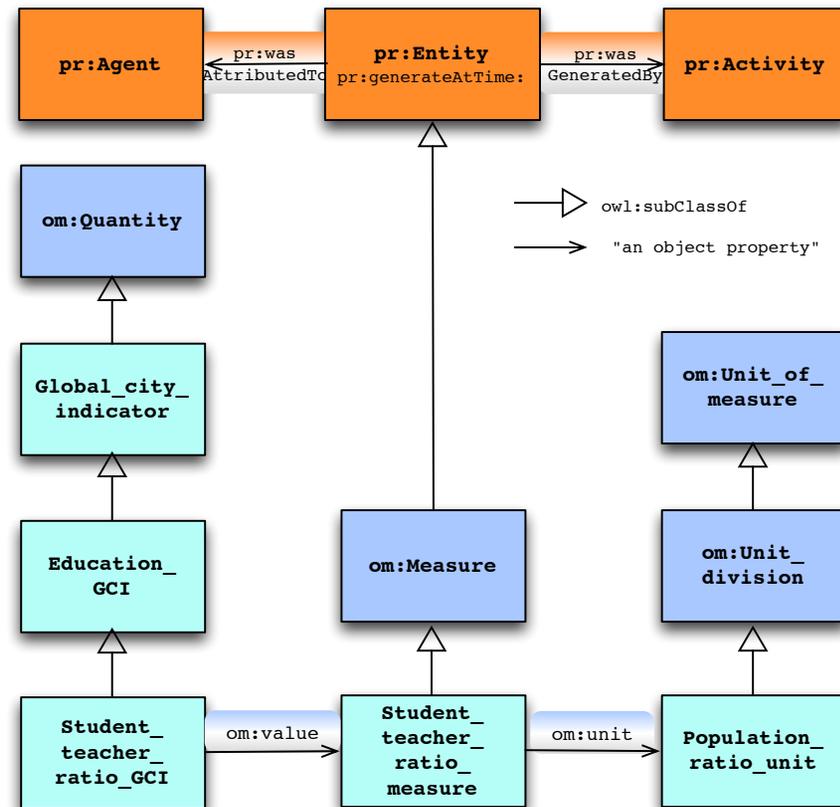
- *PROV*:



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

# Provenance

- By making a Measure a subclass of a provenance Entity, it inherits all of the attributes of it.
- Can then link indicators and source numbers to activities, agents, time, etc.

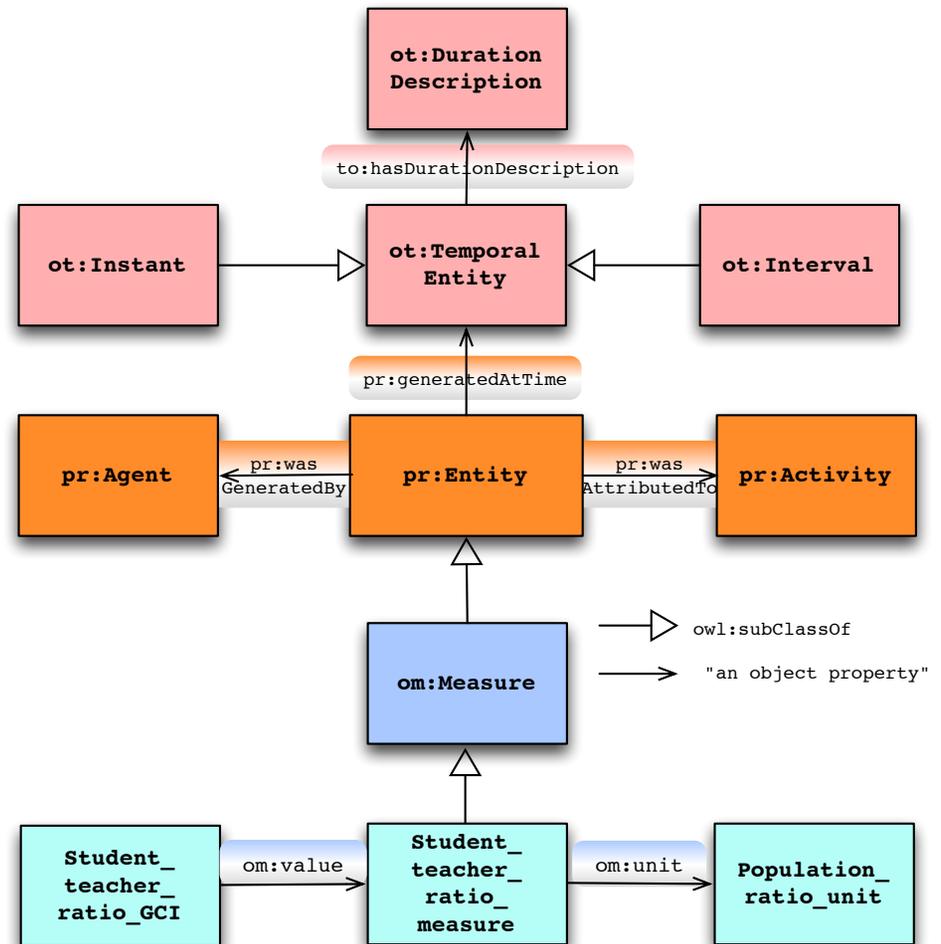


14.6

pr: <http://www.w3.org/ns/prov#>

# Time and Provenance

- To identify when entities are created, we link the provenance ontology to the time ontology via generatedAtTime property.

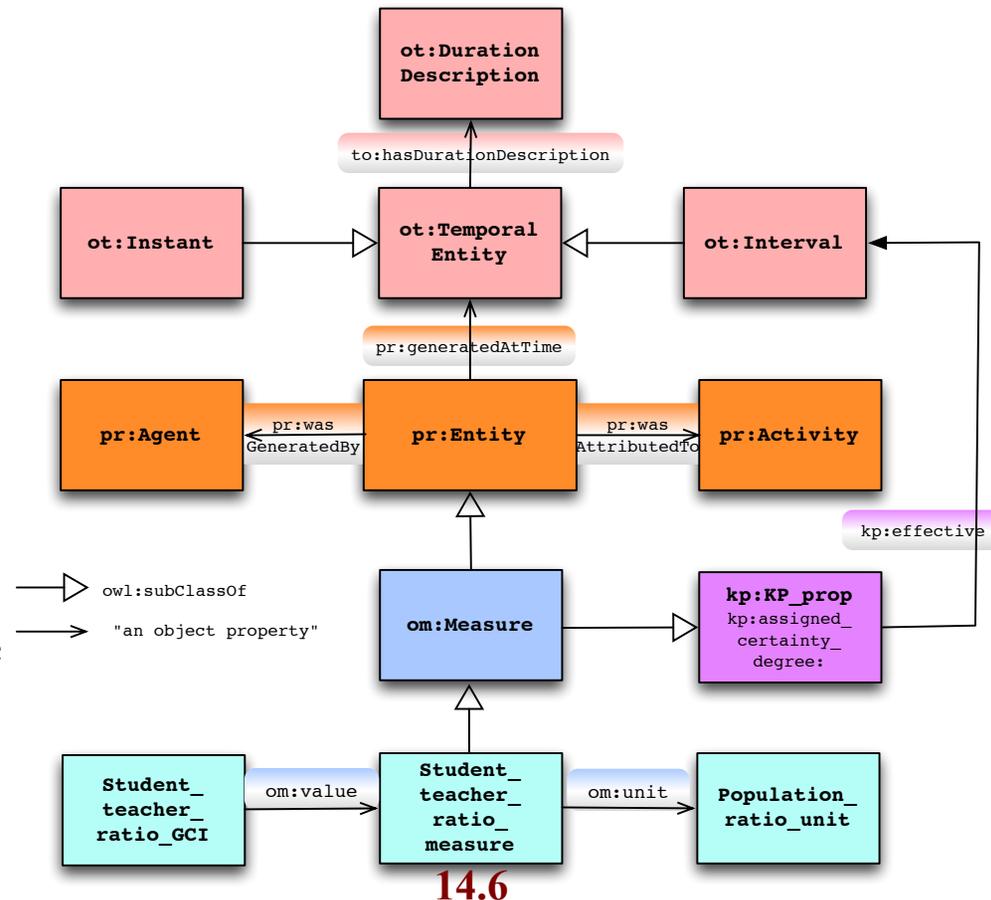


14.6

ot: <http://www.w3.org/2006/time>

# Specifying Validity

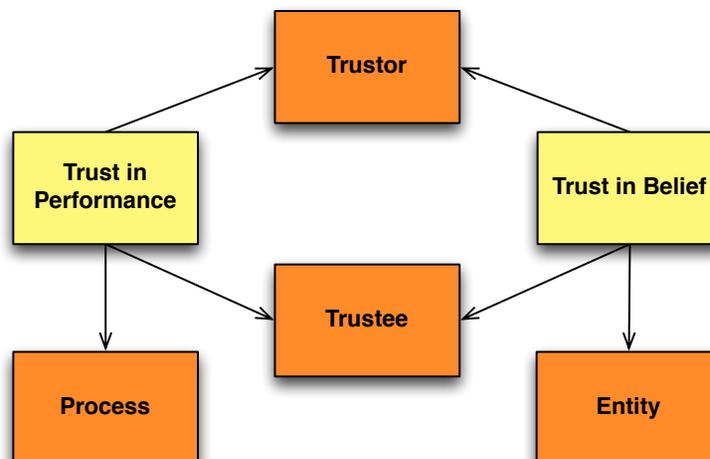
- Is the indicator true?
- Assign a degree of certainty that the indicator is valid.
- Kp\_prop is linked to a temporal entity by a kp:effective relation
- Relation defines the time period the degree of certainty holds.



Fox, M.S., and Huang, J., (2005a), "Knowledge Provenance in Enterprise Information", *International Journal of Production Research*, Vol. 43, No. 20., pp. 4471-4492.

# Specifying Trust

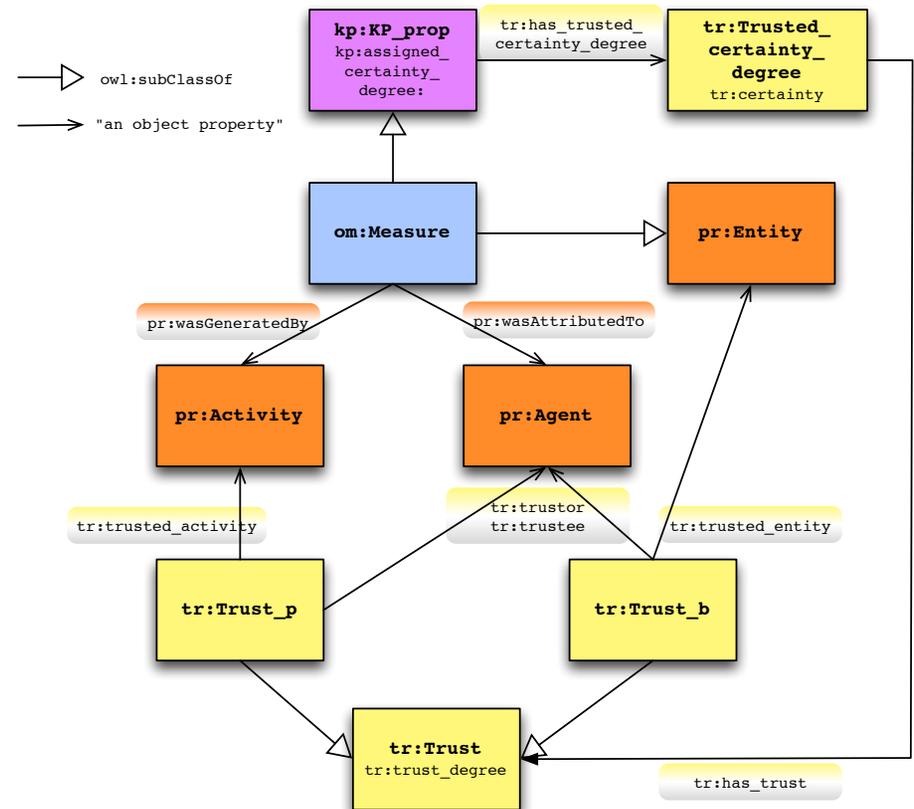
- Do you trust the creator of the GCI?
  - Do you trust the process used to create the CGI?
  - Does Joe trust it?
  - Does Frank trust it?
- What does it mean to trust an indicator?
    - Trustor's trust in the value that the trustee specifies?
      - I.e., trust in belief
    - Trustor's trust in the trustee's process that produces it?
      - I.e., trust in performance



Trust is uniquely defined by the trustor for a specific trustee.

# Trust

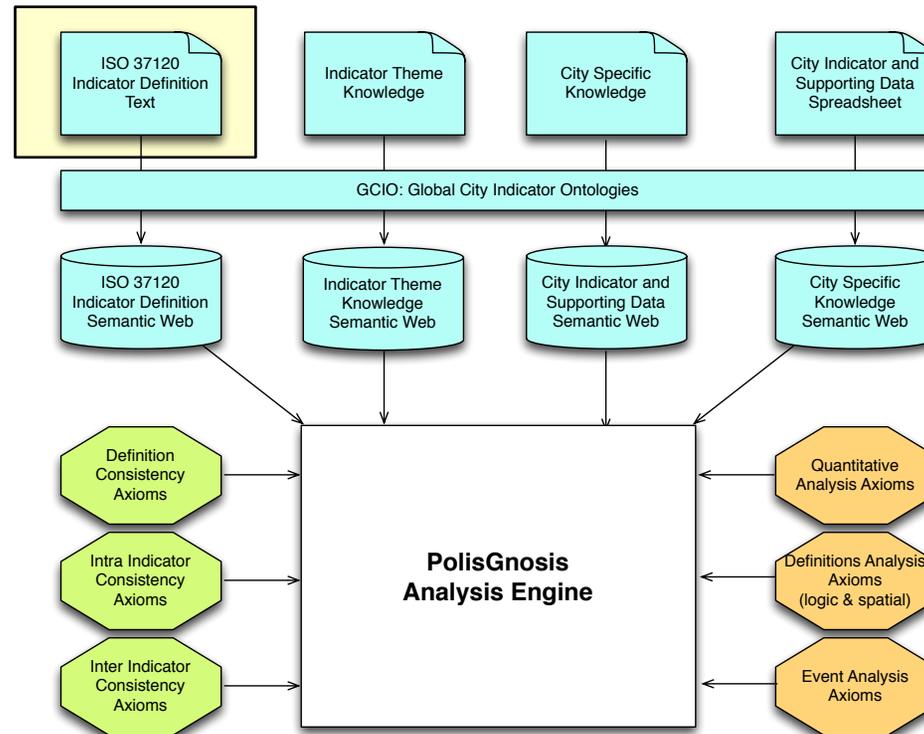
- Every GCI inherits a property that points to a trusted certainty degree specific to the trustor/trustee pair and derived from the trustee's original asserted certainty.
- There can be many trusted certainty degrees linked to a GCI.



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



# Representing Indicator Definitions, Instances and Supporting Data





# Primary 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 **preschool** teachers and staff **shall not be included**.
- The number of teachers shall be counted in fifth time increments, for example, a teacher working one day per week should be counted as 0.2 teachers, and a teacher working three days per week should be counted as 0.6 teachers.”
- **Grades, courses, Catholic school, ...**



# Heterogeneous Representation

Indicator unit of measure  
(*measurement ontology*)

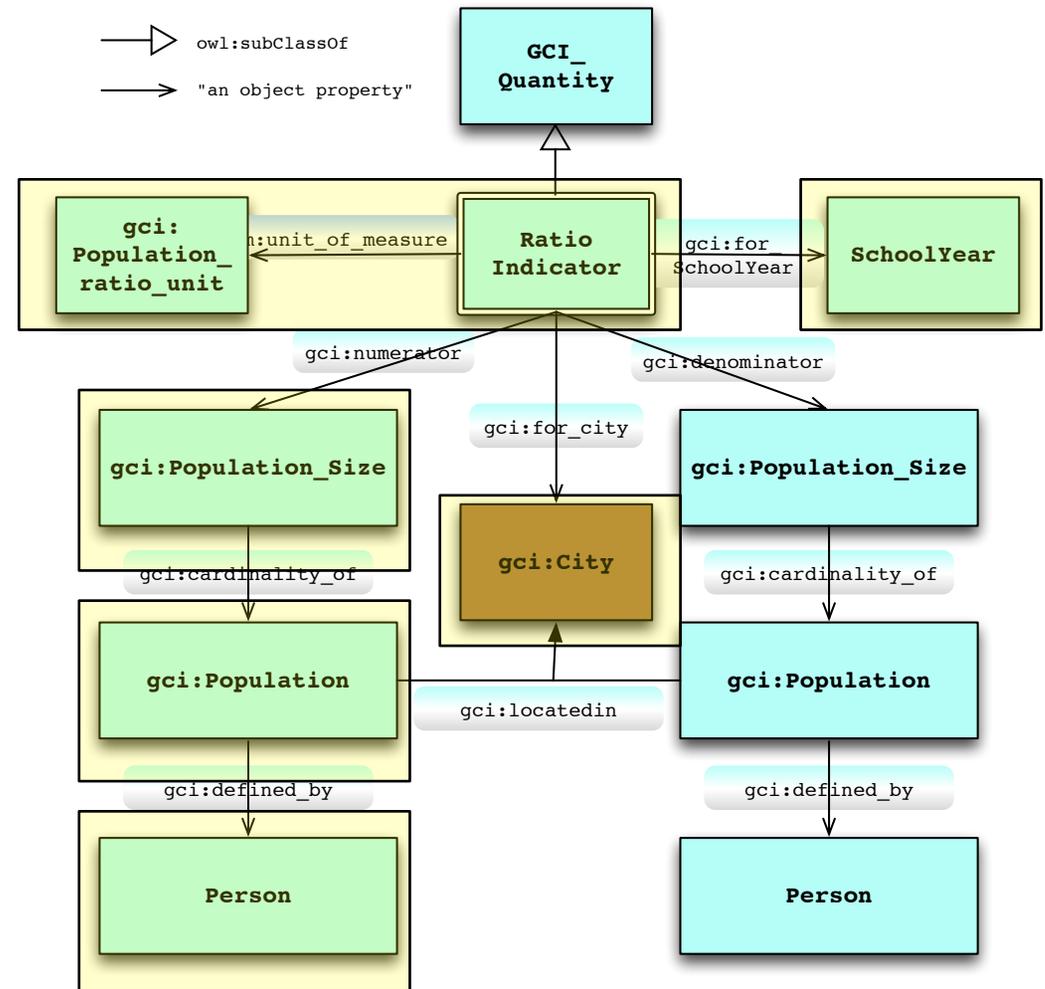
Year of measurement  
(*temporal ontology*)

Place of measurement  
(*placename and geospatial ontologies*)

Statistic being measured  
(*Statistics ontology*)

Population being measured  
(*Population ontology*)

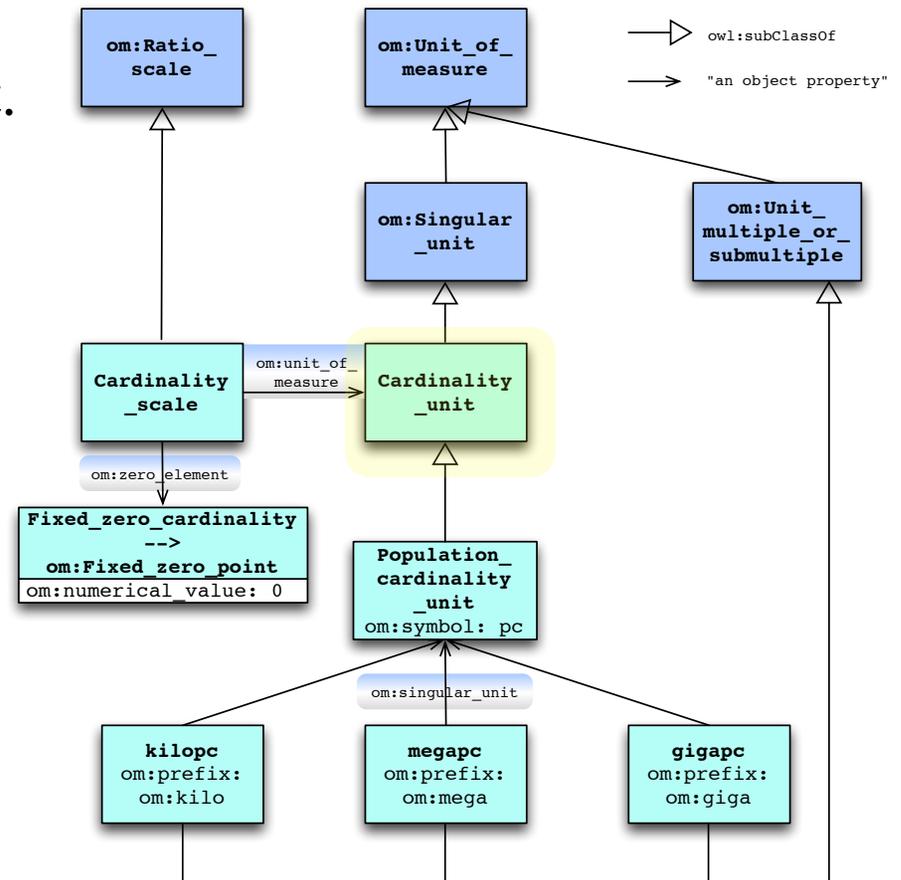
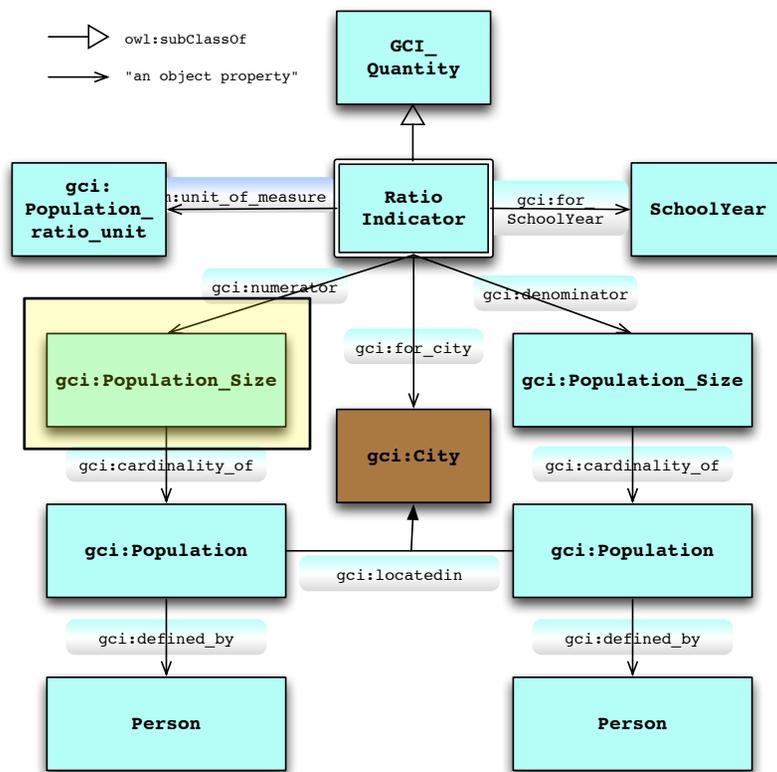
Person being counted  
(*Theme ontology*)





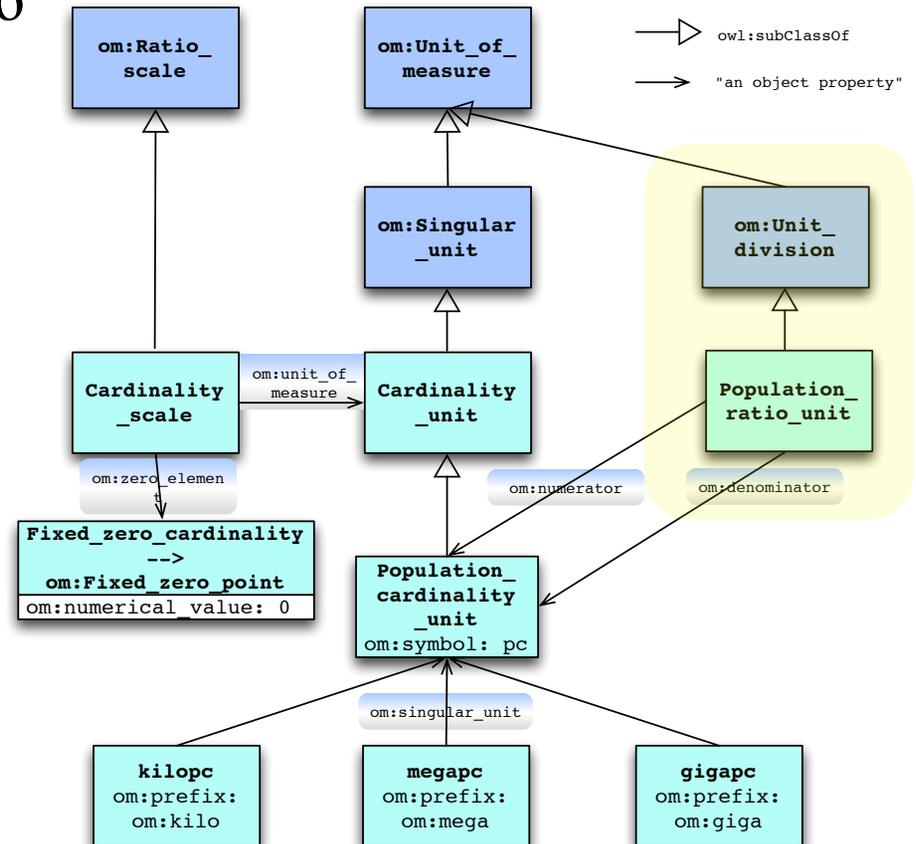
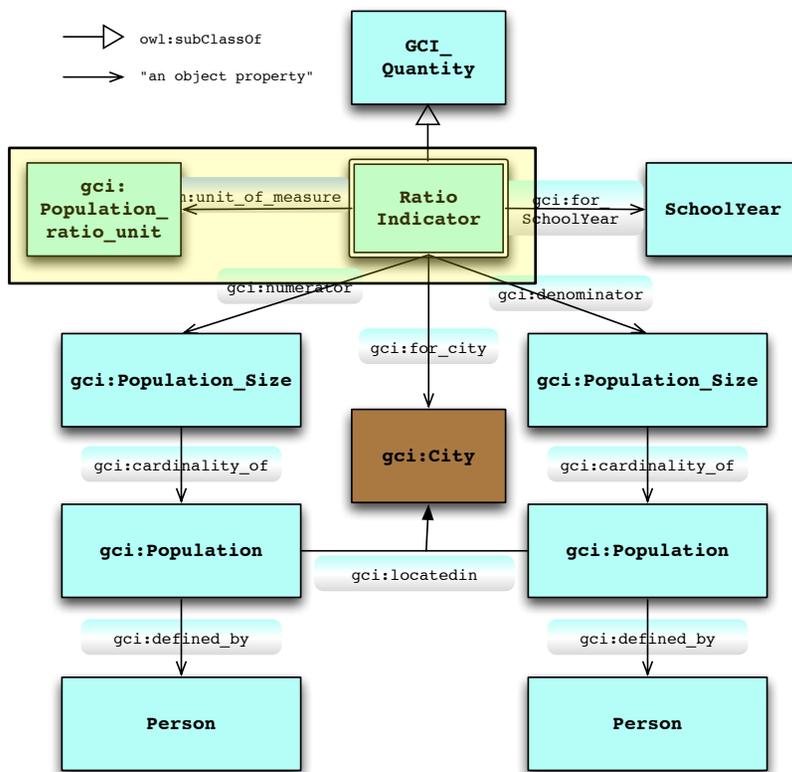
# Population Unit of Measure

- Define a measure of population size: introduce the notion of measuring the cardinality of a set.



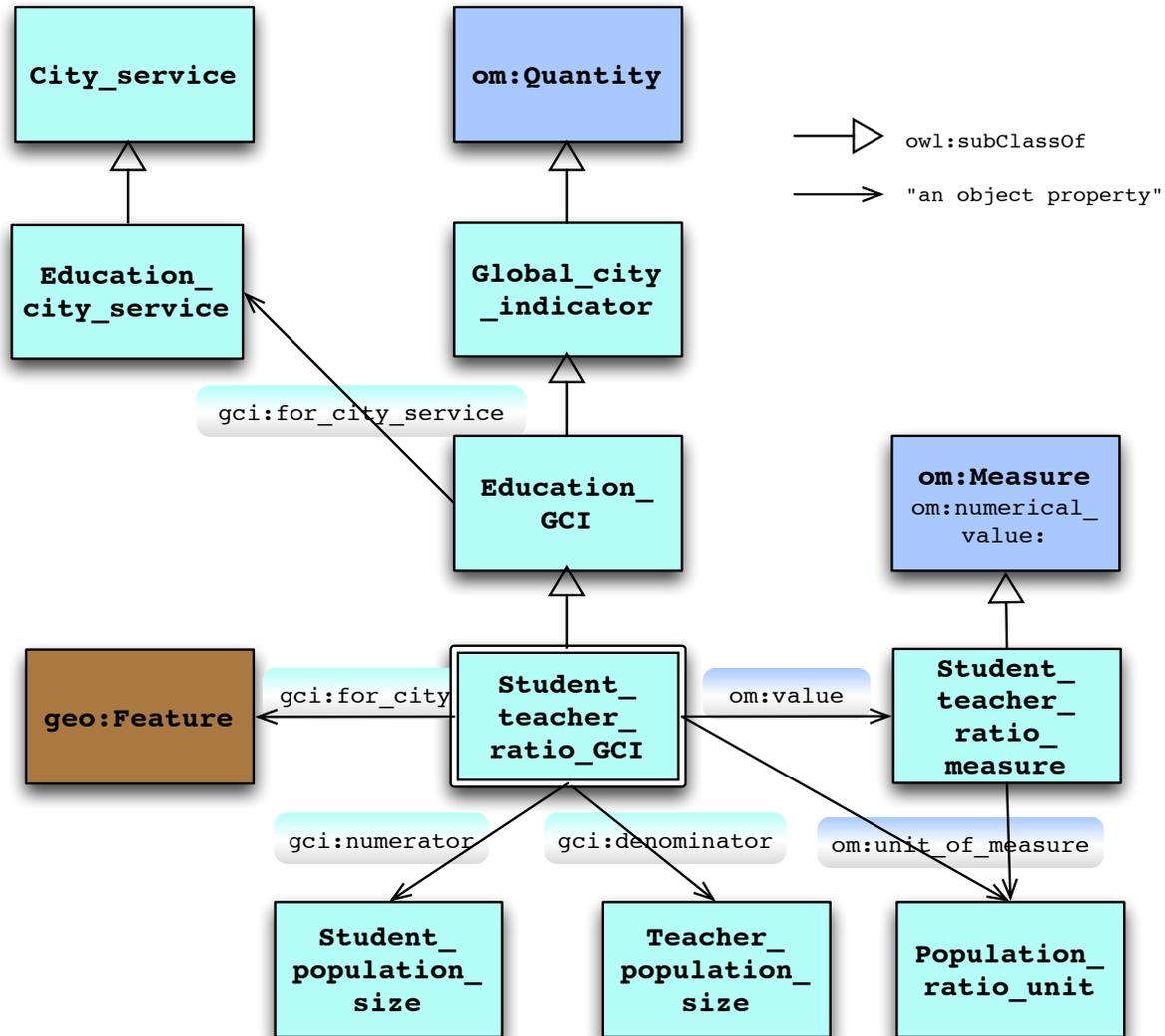
# Ratios of Populations

- Introduce `Population_ratio_unit` which is a measure of the ratio of two populations.



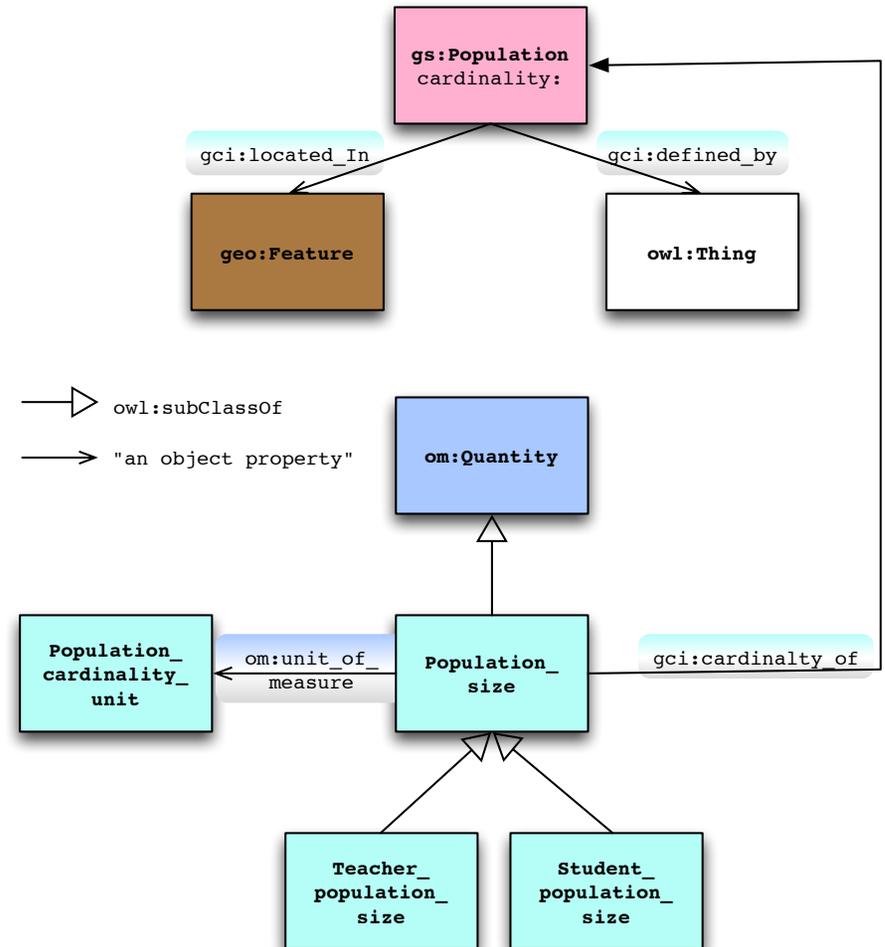
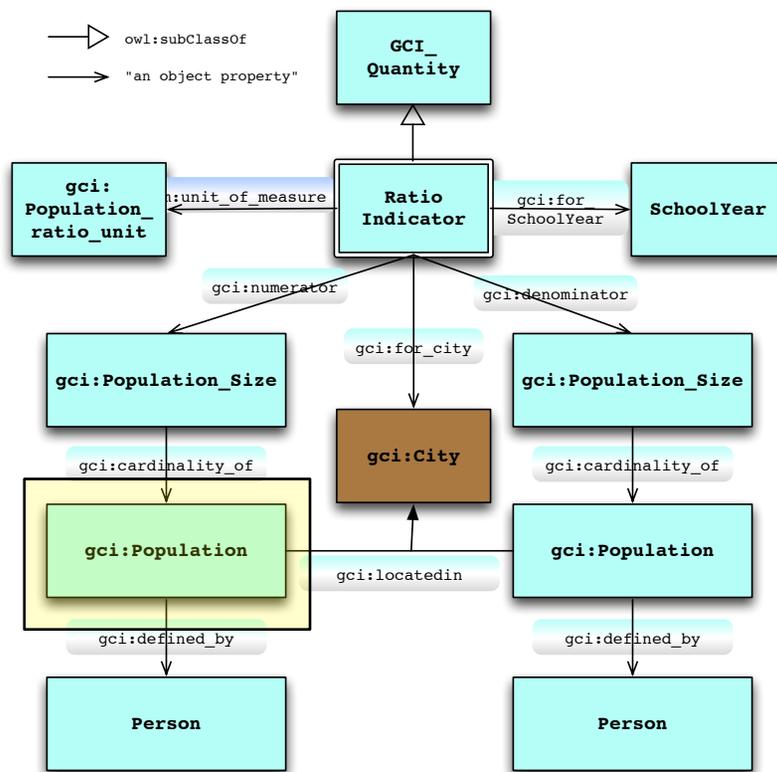


# STR Indicator Partial Definition



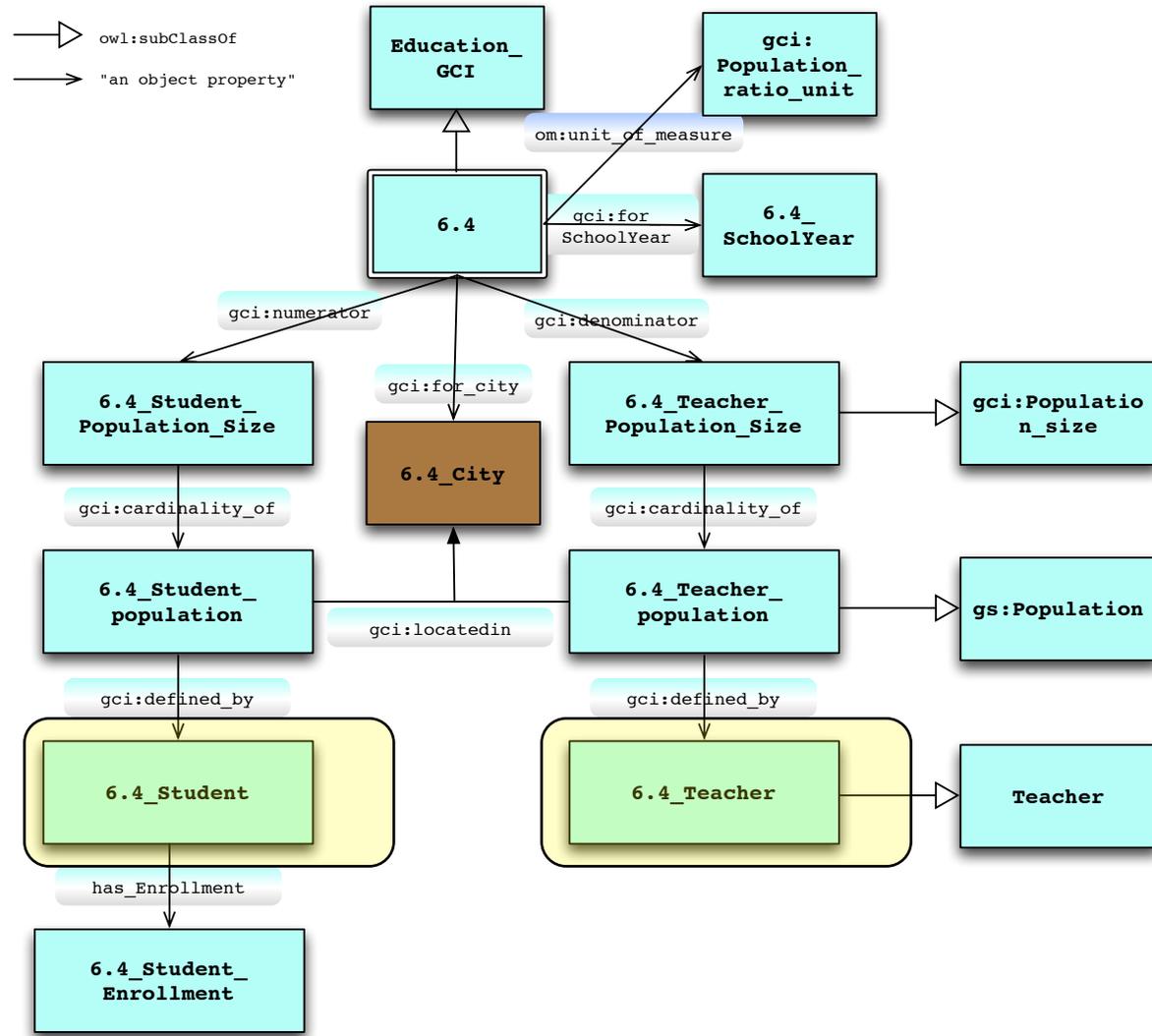
# What is the Population?

- Introduce Population as defined by a city and person.





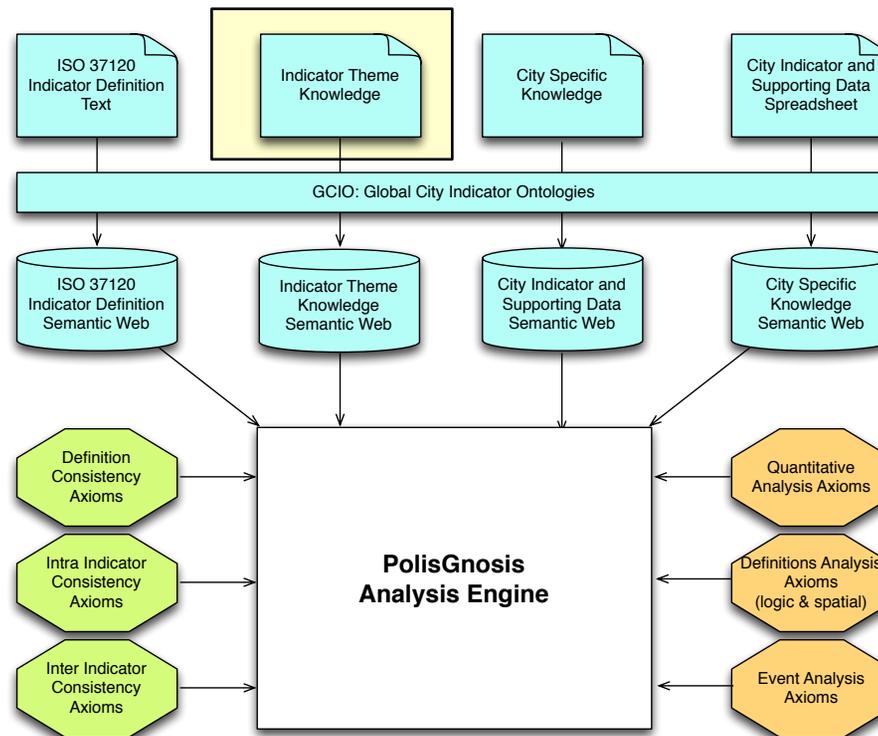
# 6.4: Primary Education Student/Teacher Ratio





# Enabling Definitions with Theme Specific Knowledge

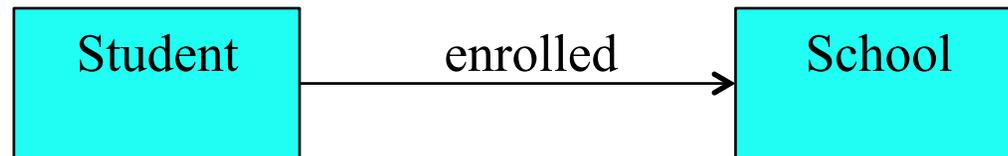
## (Education Ontology)





# What Defines a Student?

---



- Does "enrolled" satisfy the definition of "Primary Student"?
  - What grade? Fulltime or part time? What year?
- **Intermediation** is the process of expanding the property into a class with additional information

# Student and Enrollment

## (Description Logic: Manchester Syntax)



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

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

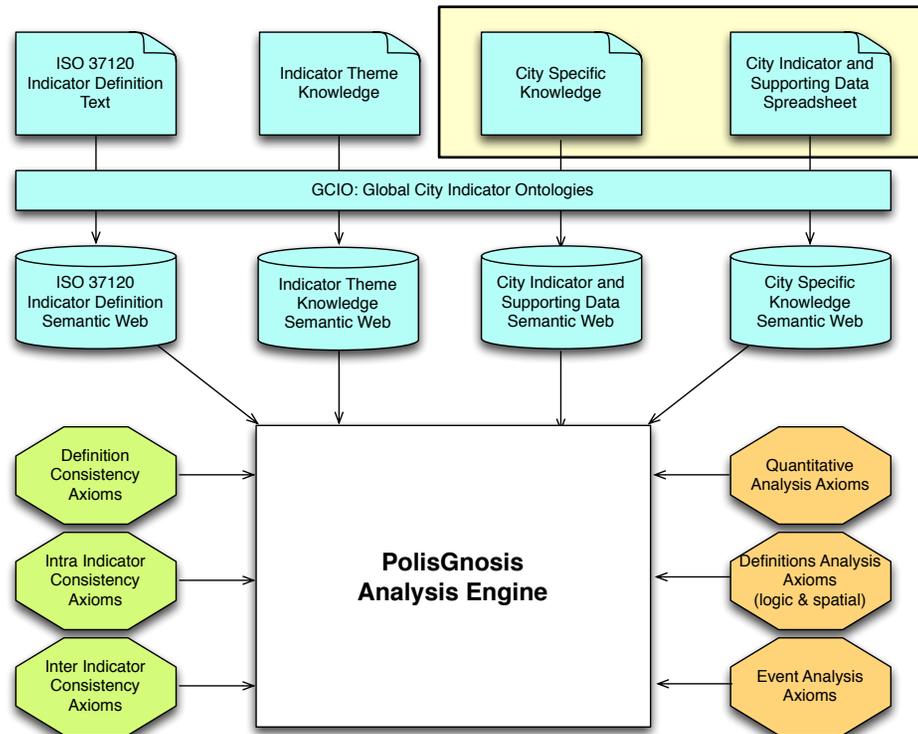
# Program



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



# Instantiating City Indicator Values



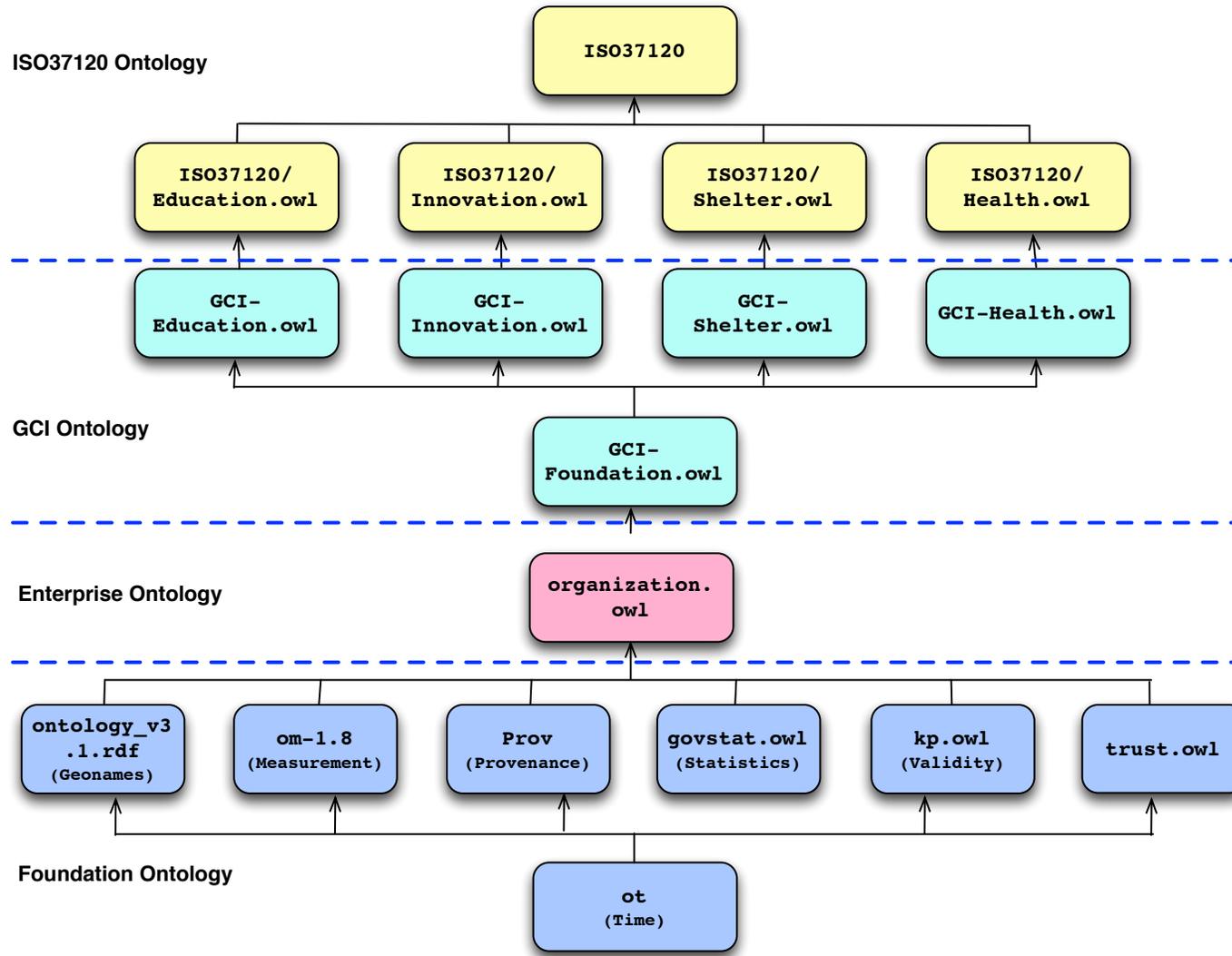


# Toronto Primary Education

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



# ISO37120 Ontologies Structure





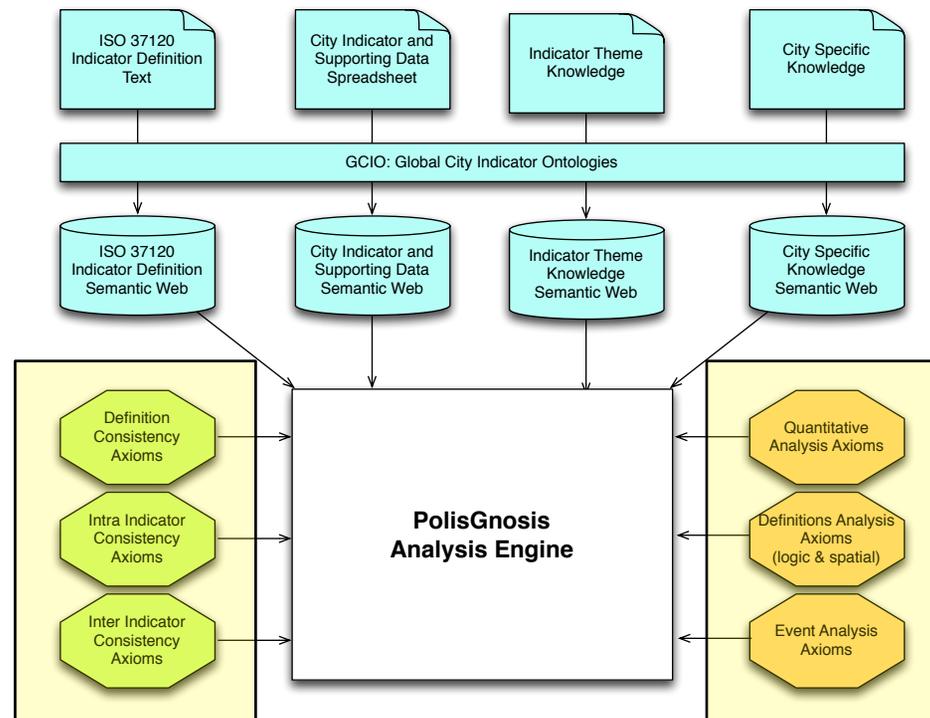
# Status

---

- Foundation, Education, Innovation, Shelter and Health theme ontologies and indicator definitions complete.
- Environment, Finance, Safety and Transportation under development.
- City of Toronto has begun to publish its ISO 37120 indicators using our ontologies.
- City Protocols has adopted our ontologies for the representation of their indicators, which include the ISO 37120 indicators.



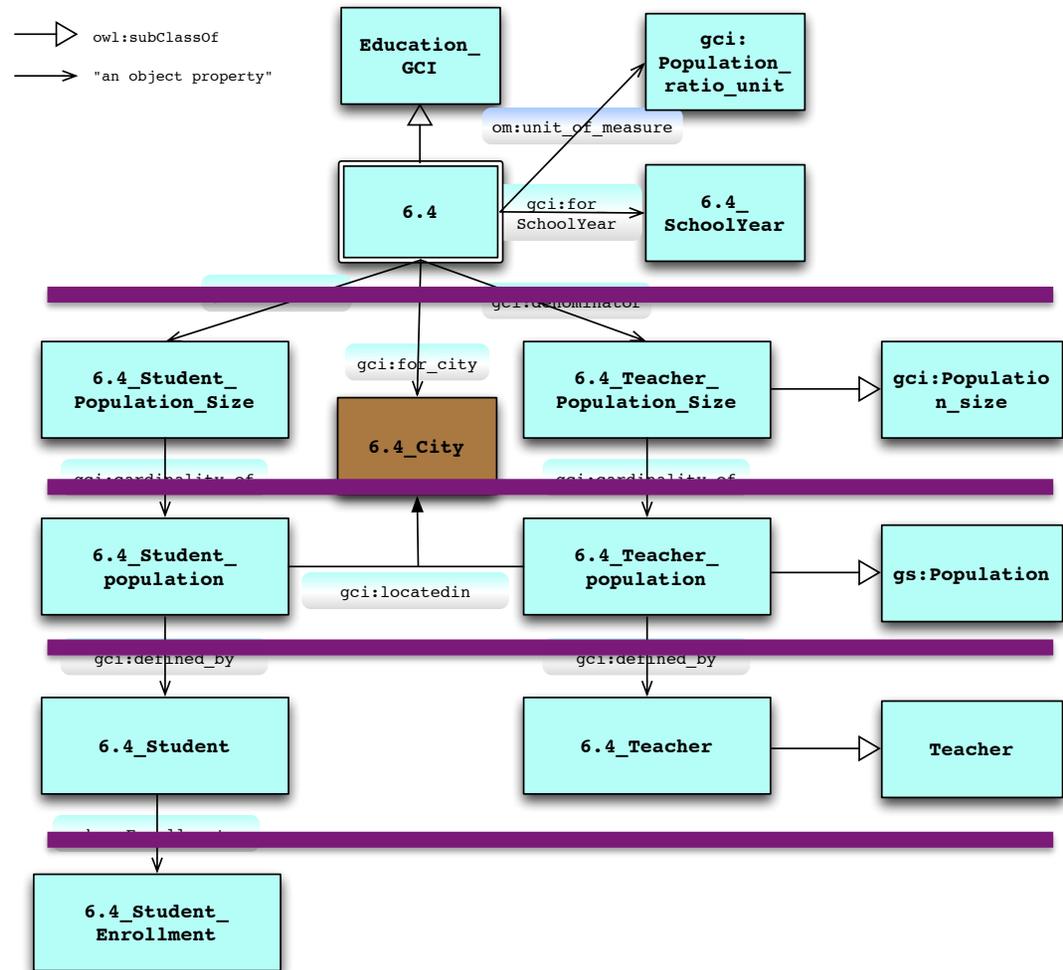
# Indicator Analysis



# Open City Data Completeness Metric

- Measuring the degree to which a city's open data satisfies the information requirements of computing indicators.

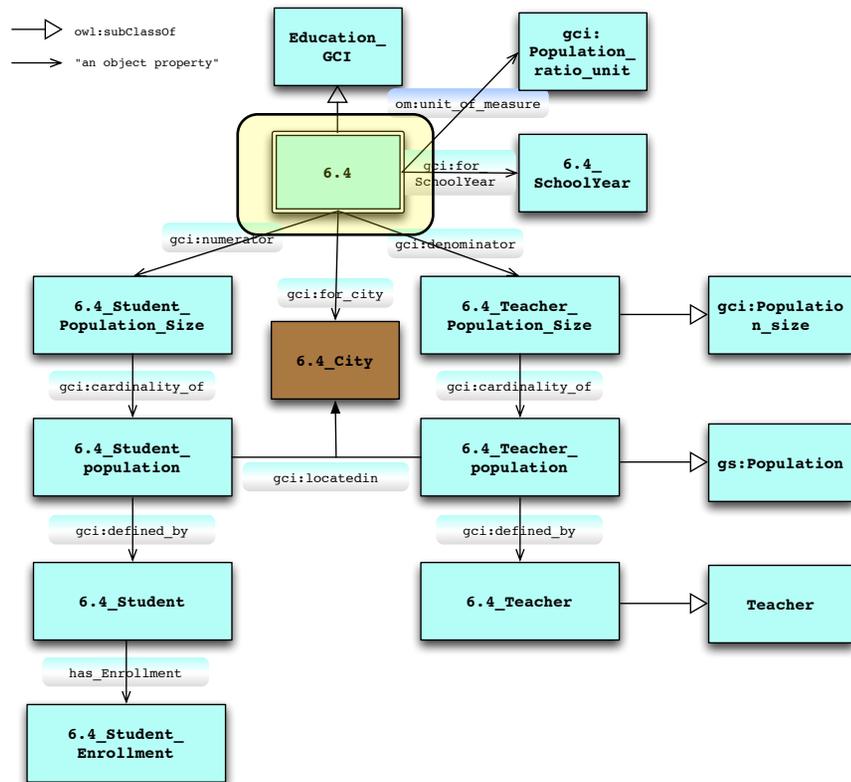
## Student Teacher Ratio



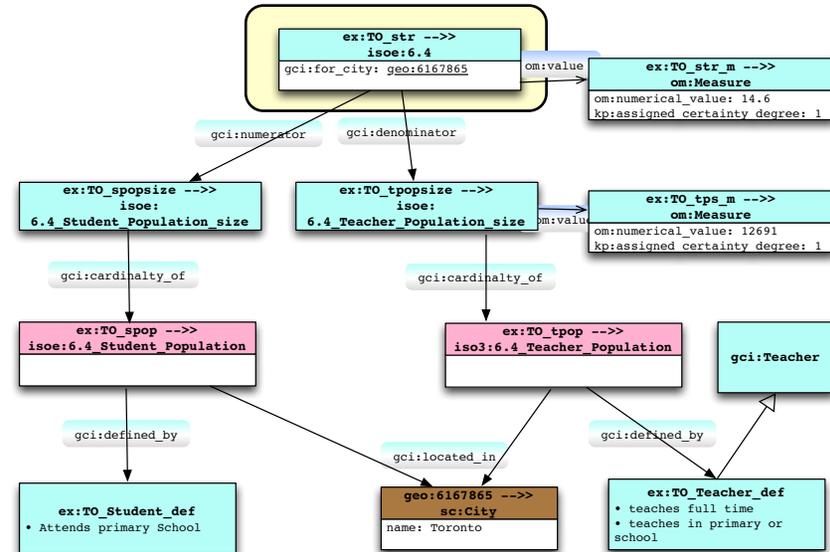
Fox, M.S., and Pettit, C.J., (2015), "On the Completeness of Open City Data for Measuring City Indicators", *IEEE Conference on Smart Cities*, Guadalajara MX, to appear.

# Definitional Consistency

## Definition (isoe)

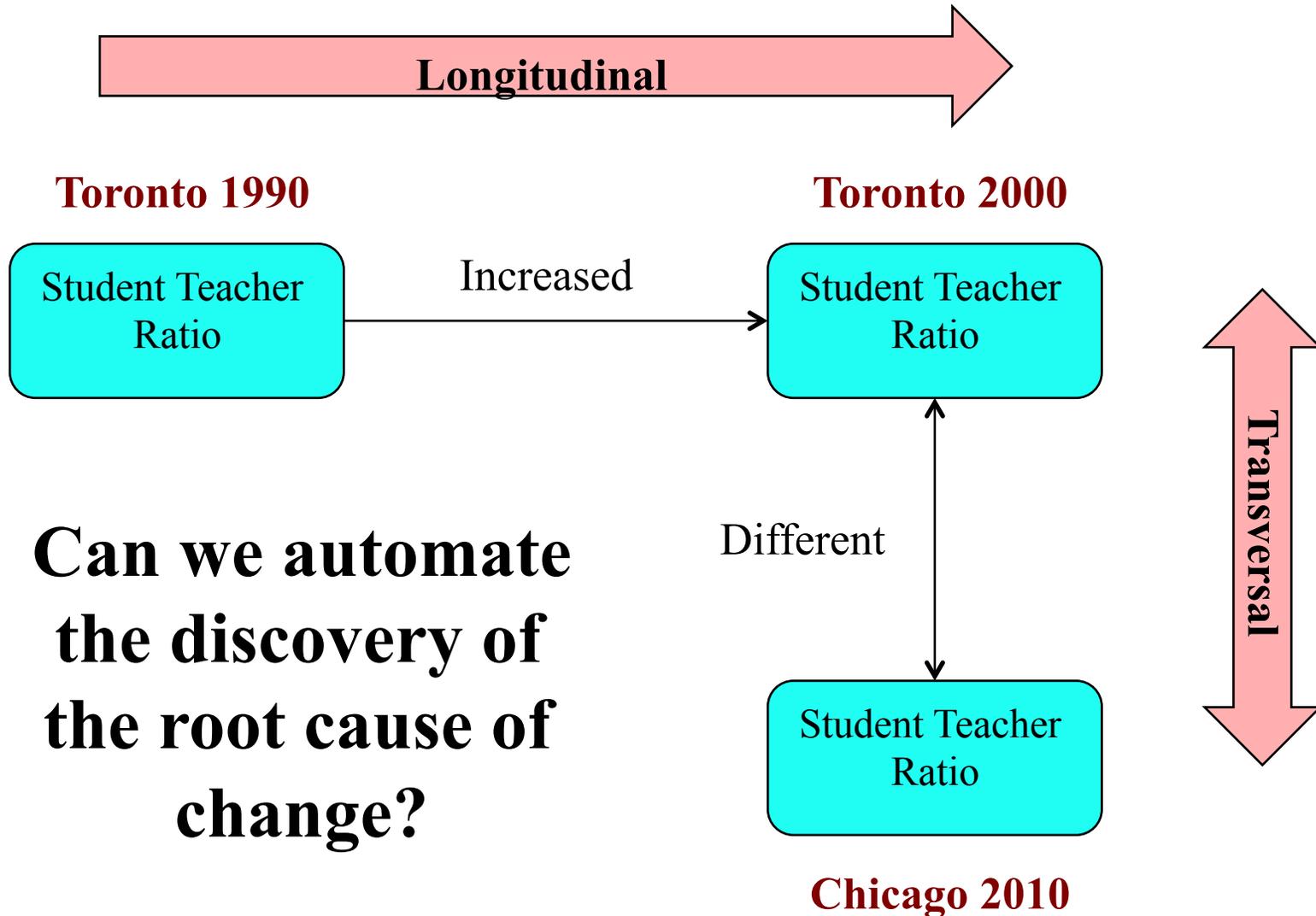


## Reported Indicator





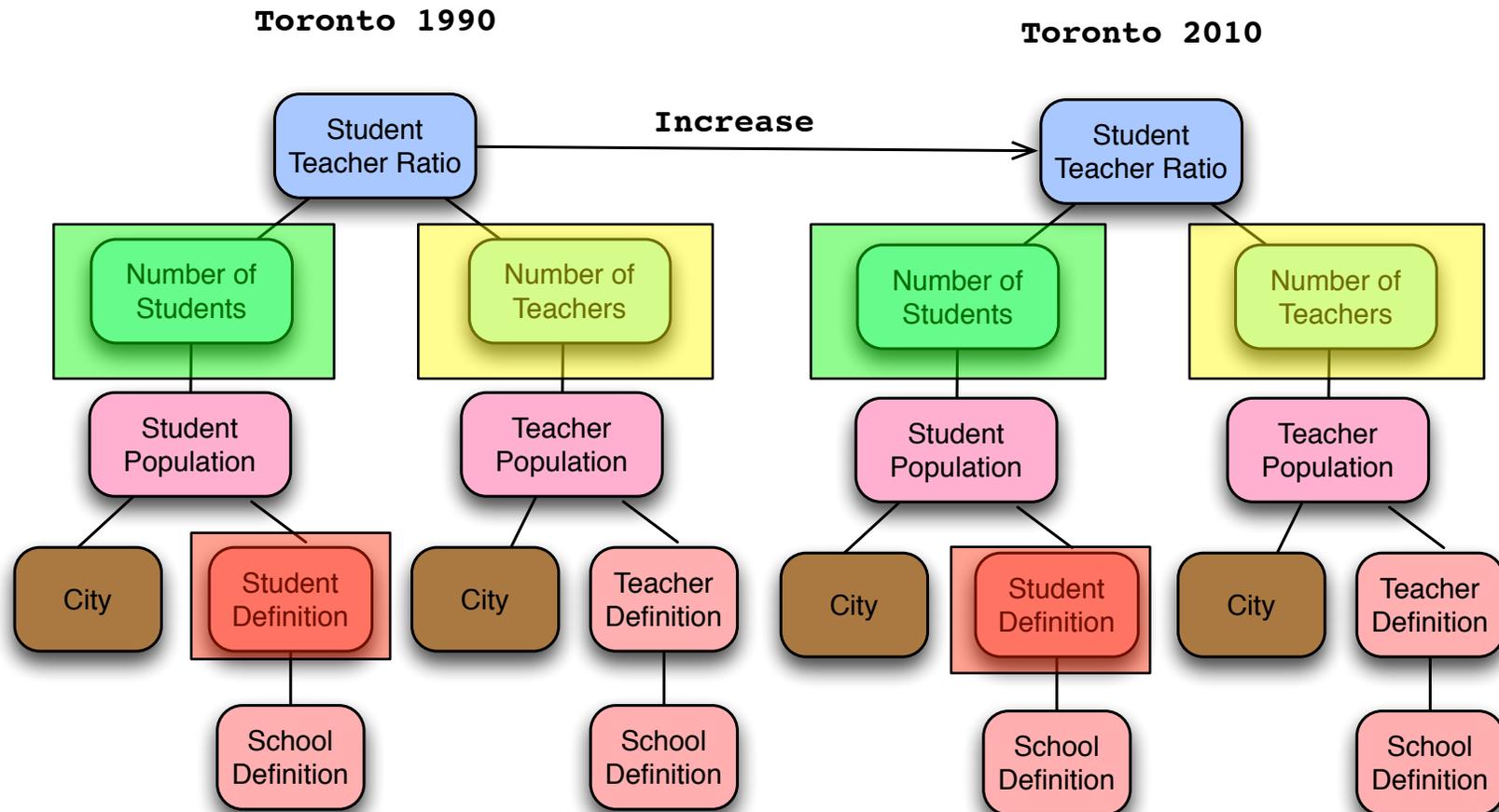
# Bridge the Analysis Gap





# Automated Analysis

## Longitudinal Analysis



### Heterogeneous Models

Each knowledge type requires its own micro-theory to determine the root cause of differences.



# Conclusion

---

- ISO37120 and subsequent standards have created an opportunity for the measurement of city performance.
- But the Open Data movement is not yet aligned, creating a morass of data of limited usefulness.
- PolisGnosis solves two major problems:
  1. What is the target representation for city knowledge to be published in the open data portals, and
  2. How can we effectively analyze this data?