A Foundation Ontology for Global City Indicators

Mark S. Fox

Professor of Industrial Engineering & Computer Science

Senior Fellow, Global Cities Institute

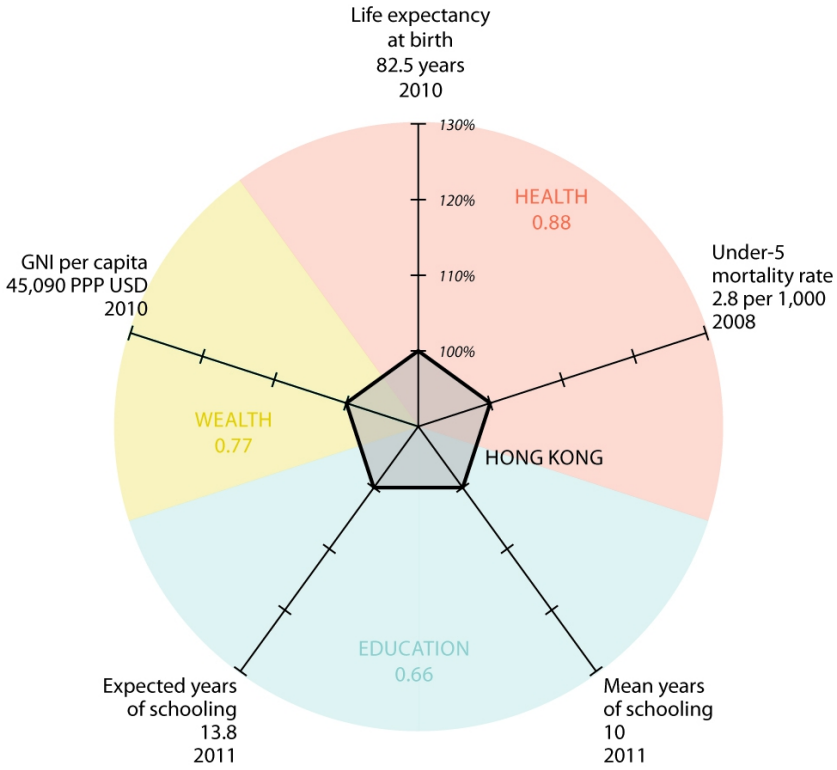
University of Toronto

msf@eil.utoronto.ca, www.eil.utoronto.ca

30 October 2013

Vision

You can't manage what you don't measure!



- 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.

Barriers

Indicator Gap

- World Bank funded a study, by the University of Toronto, of performance indicators at 9 cities.

- 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.

Student/Teacher Ratio

Student/teacher ratio

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

Student/teacher ratio

Numerator: Number of Students

Denominator: Number of Teachers

Global City Indicators Facility: Website User Guide. October 2012.

Aspects of Good City Indicators

Hoornweg, D., et al., (2006), "City Indicators: Now to Nanjing", Third World Urban Forum, Vancouver.

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

Global Cities Institute (U. of Toronto)

Over 100 indicators defined and submitted to ISO

City Services

- Education
- Finance
- Governance
- Recreation
- Transportation
- Wastewater
- Energy
- Fire and Emergency Services
- Health
- Safety
- Solid waste
- Urban Planning
- Water

Quality of Life

- Civic Engagement
- Economy
- Shelter
- Culture
- Environment
- Social Equity
- Technology and Innovation

ISO 37120 – Sustainable Development and Resilience of Communities – Indicators for City Services and Quality of Life (under TC268

Student/Teacher Ratio (STR)

Core Indicator Requirements: "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."

Data sources: "The number of full-time equivalent primary school classroom teachers and the number of enrolled primary school students should be collected from the local public school system or Ministry of Education."

Data interpretation: "The student/teacher ratio reflects teacher workload and the availability of teachers' services to their students. The lower the student/teacher ratio, the higher the availability of teacher services to students. The student/teacher ratio has implications not only for the cost of education, but also for the quality. Higher educational attainment is correlated with a lower student/ teacher ratio."

Semantic (Vocabulary) Gap

How do you compare data if they don't share a vocabulary?

Student/ Teacher Ratio	% solid waste burned	Number of Households	Life Expectancy
40	4.5	1.3	85

STR	Incinerated Waste	Family count	平均预期寿命
10	1.5	130	87

- Do these attributes match?
- What is the unit of measurement and scale of these numbers?
- When were they gathered? How? By whom?

Modeling Gap

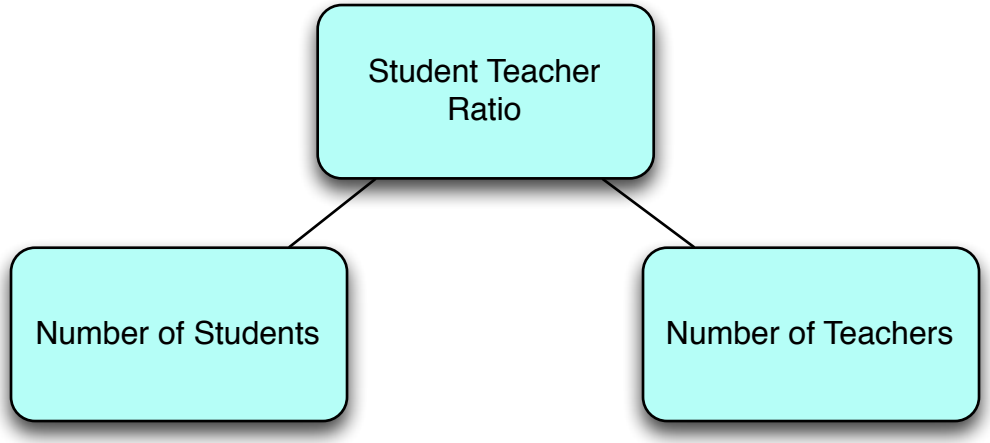
- What type of number is a STR?
- Unit of measurement? Meters? Grams?
- Ratio, Ordinal, Nominal?
- Scale? Kilos?

Student Teacher Ratio

How can we analyze indicators if the underlying data is not available?

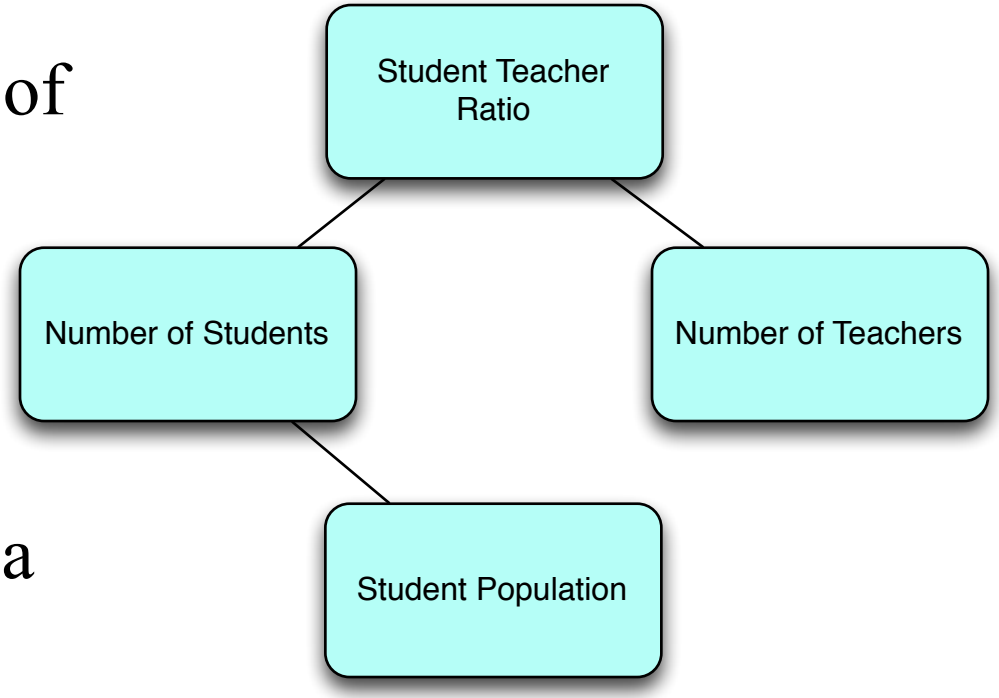
Modeling Gap

- What is an STR composed of?
- If a “division”, what is the numerator? denominator?
- What kind of numbers are these? Ordinal?
- Units? Kilo?



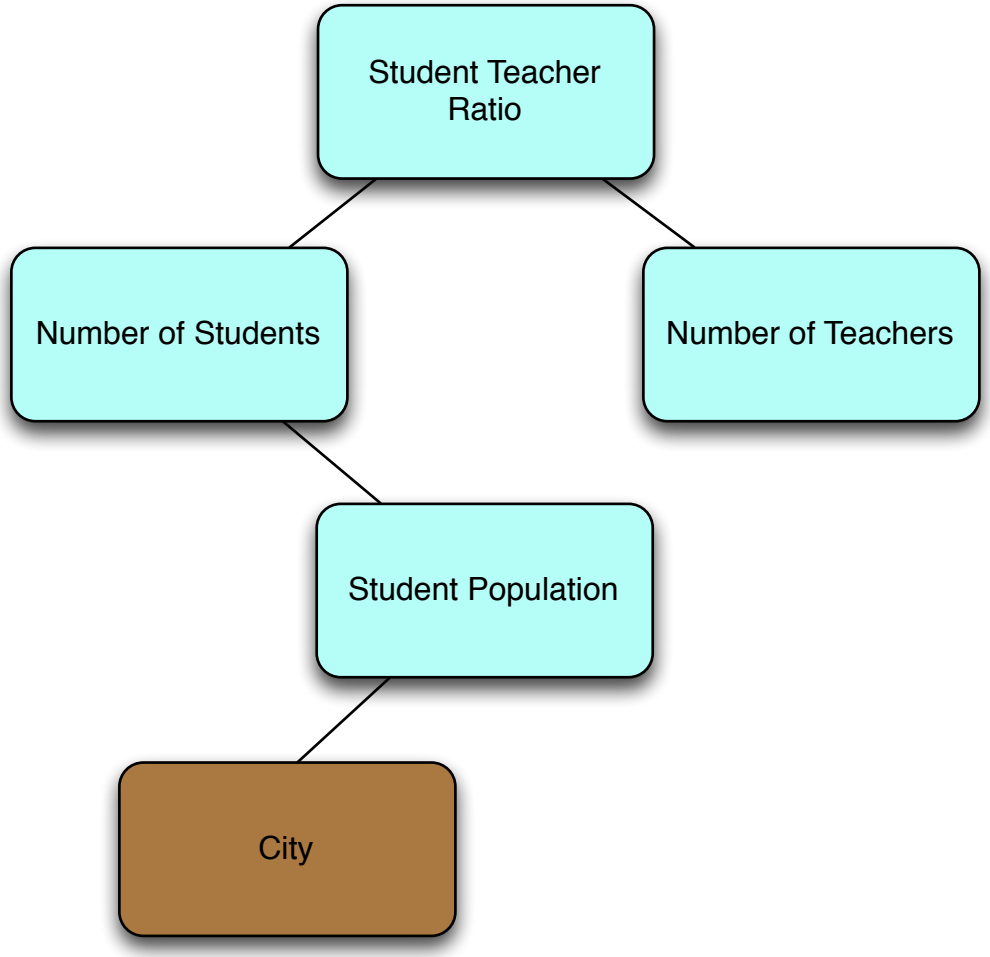
Modeling Gap

- What is the "number of students" representative of?
- How is it determined? A count of members of a population?



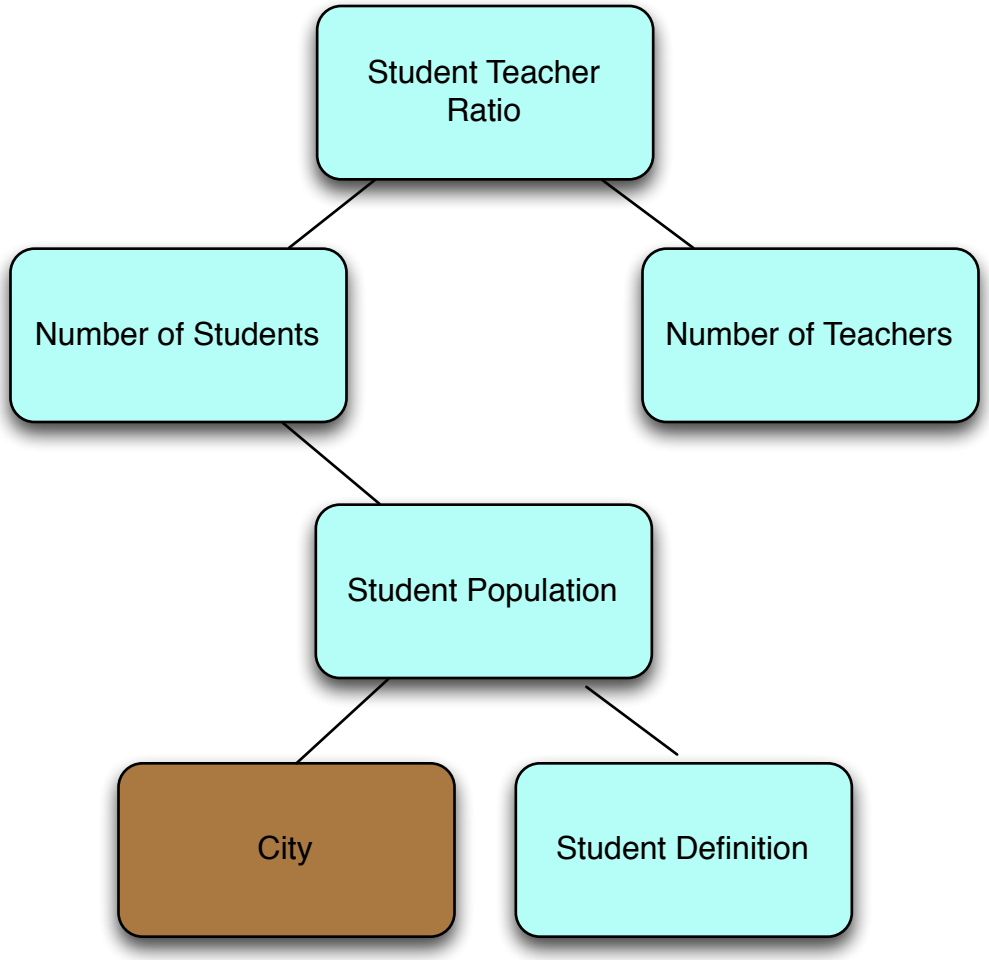
Modeling Gap

- Where is the population drawn from? Toronto? Toronto OH?



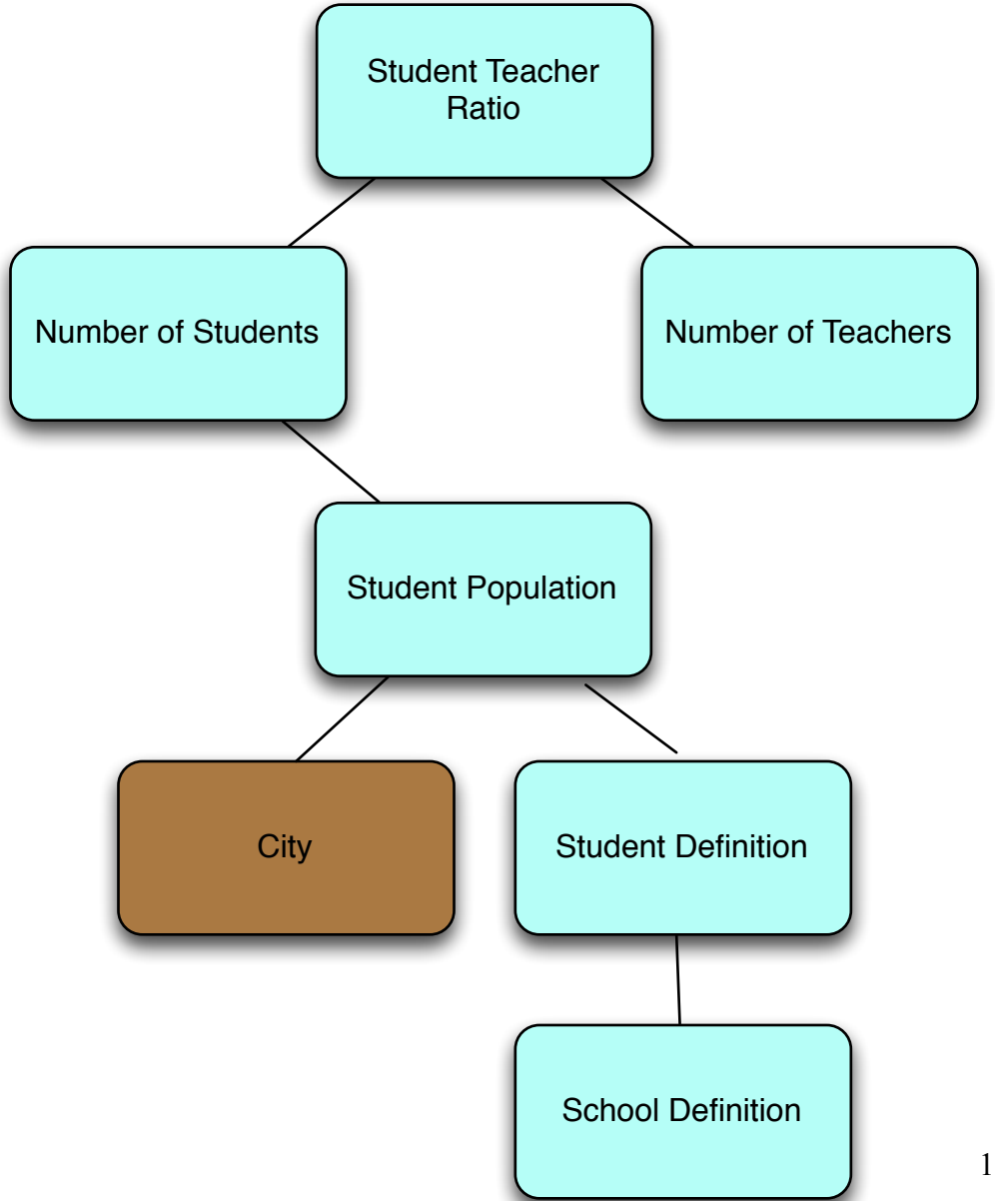
Modeling Gap

- How do you determine a member of the population?
- What is a student? Full or part time? Regular or special? Primary or secondary grades?

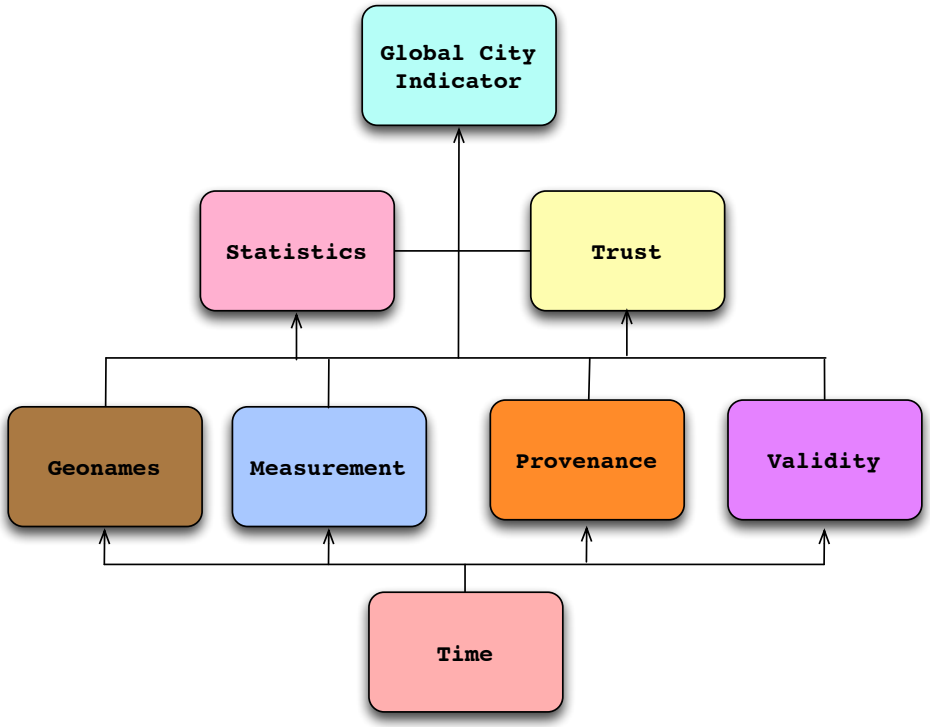


Modeling Gap

- What schools are included in defining students? Public, Private?
- What are the primary grades?



Observation



- A single indicator requires the integration of many types of representations.

Indicator Ontology: Open, Linked, Standard

1. **Develop an ontology for city indicators that builds on existing foundation ontologies by integrating and extending them, and publish on the Semantic Web.**

Ontology

"explicit representation of a shared understanding"

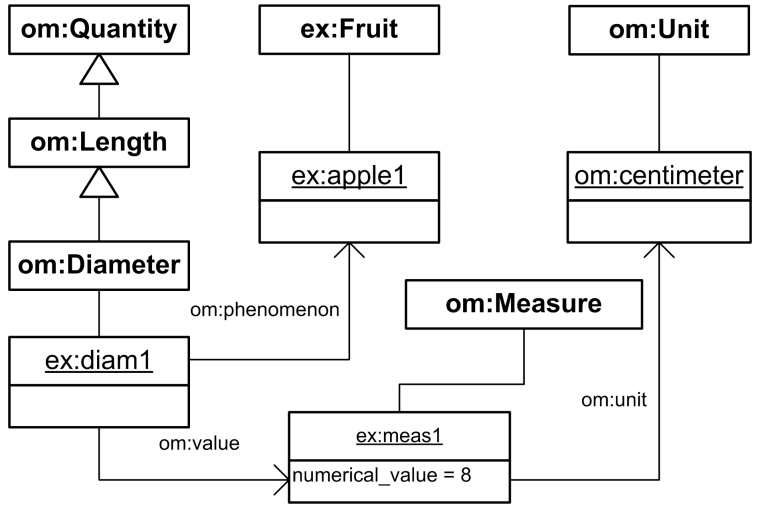
(Gruber, 1993)

- A **Formal Ontology** is the specification of:
  - a **Terminology**, that specify the classes, properties and data types of the domain, and
  - **Axioms**, that define and constrain the interpretation of the terminology (in FOL, OWL).

How do we Measure things?

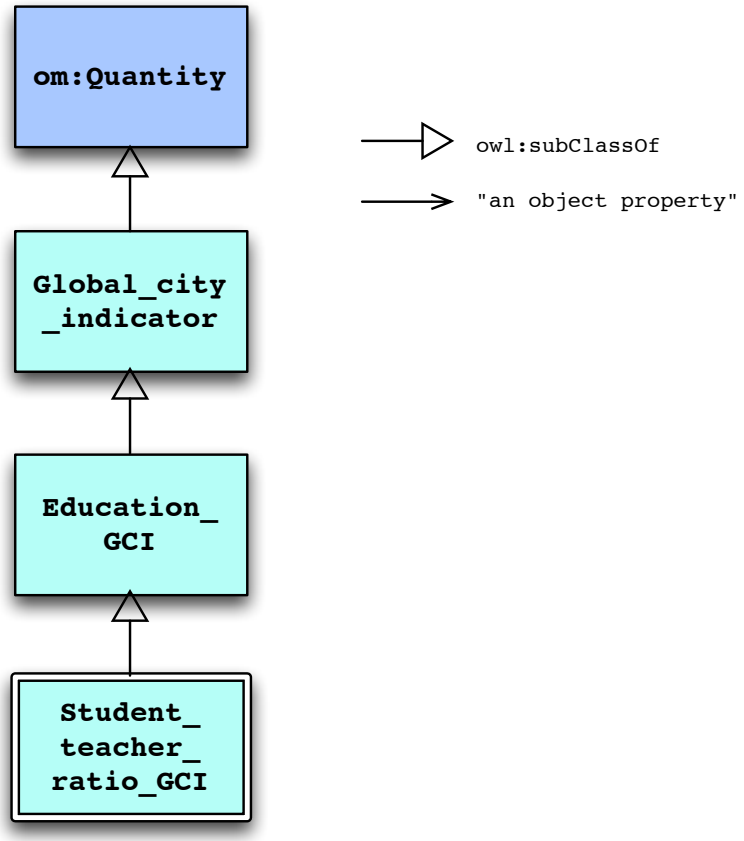
- What is the semantics of an indicator?
- What is its scale?
  - Can one indicator be 2x another?
- What are its units?
  - Mega, kilo?
- Can they be compared?

Measurement Theory provides the basic concepts.



STR is a Quantity

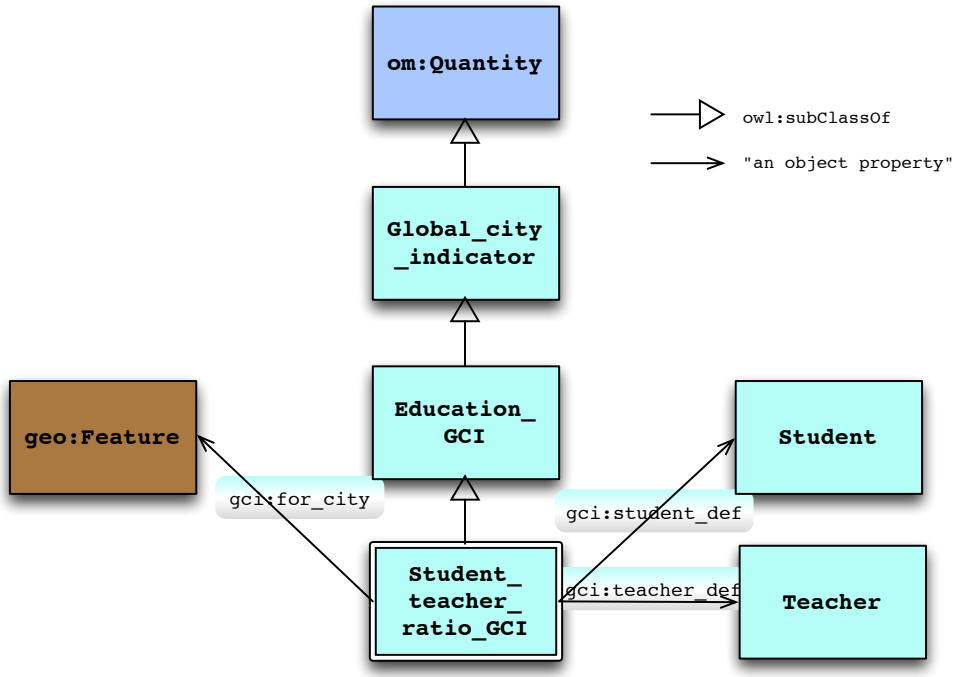
- The actual representation of a STR is as a quantity in the Measurement Ontology.



om: http://www.wurvoc.org/vocabularies/om-1.8/

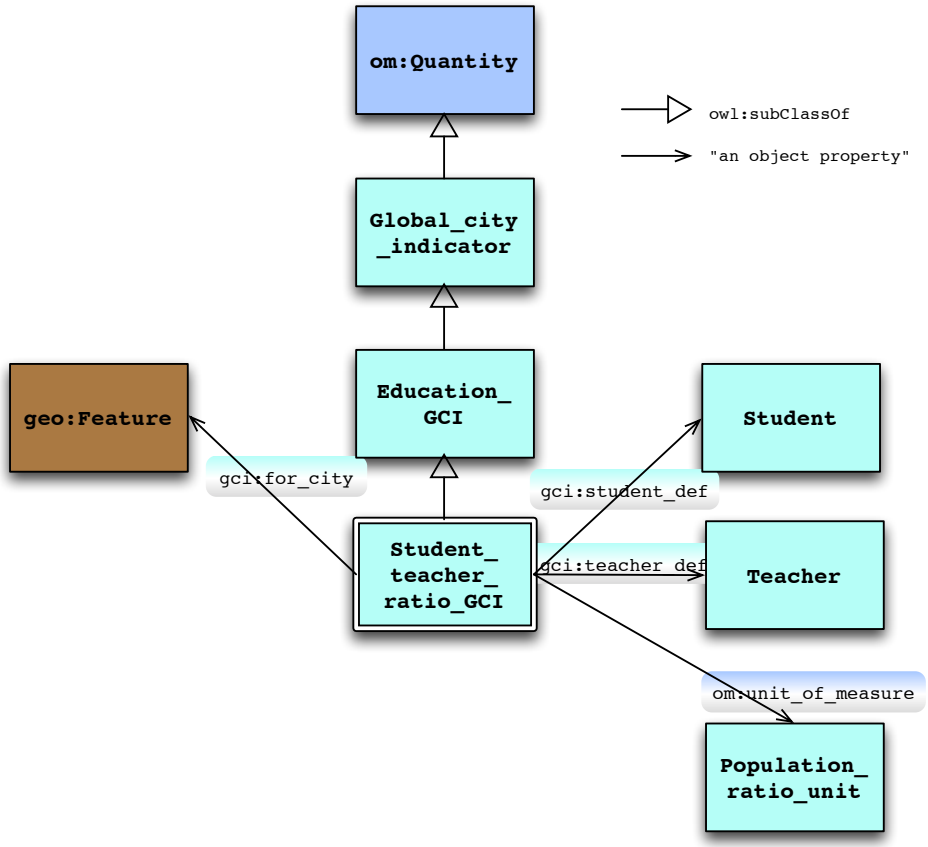
STR Specifies Key Components

- STR links to the definition of Student, Teacher and the city it is for.



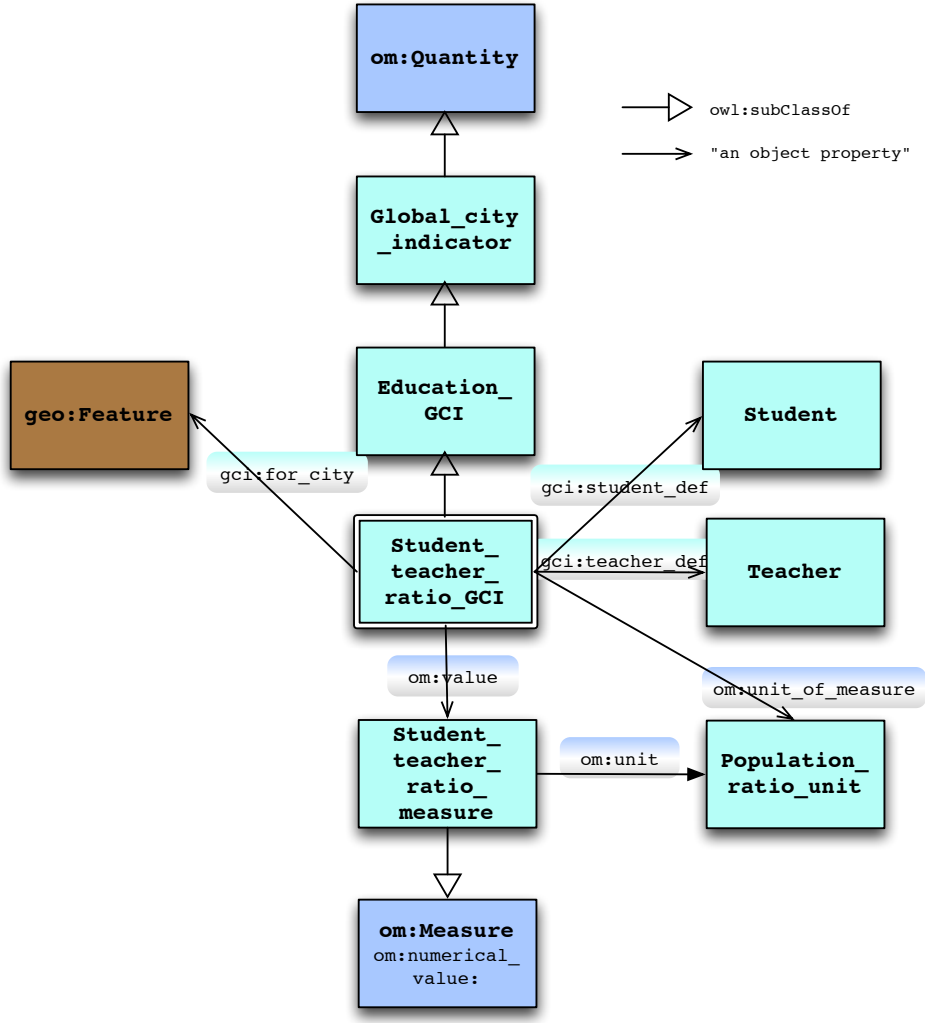
STR's Unit of Measure

- STR also links to its unit of measure that provides its semantics.
- *Population ratio unit will be defined soon.*



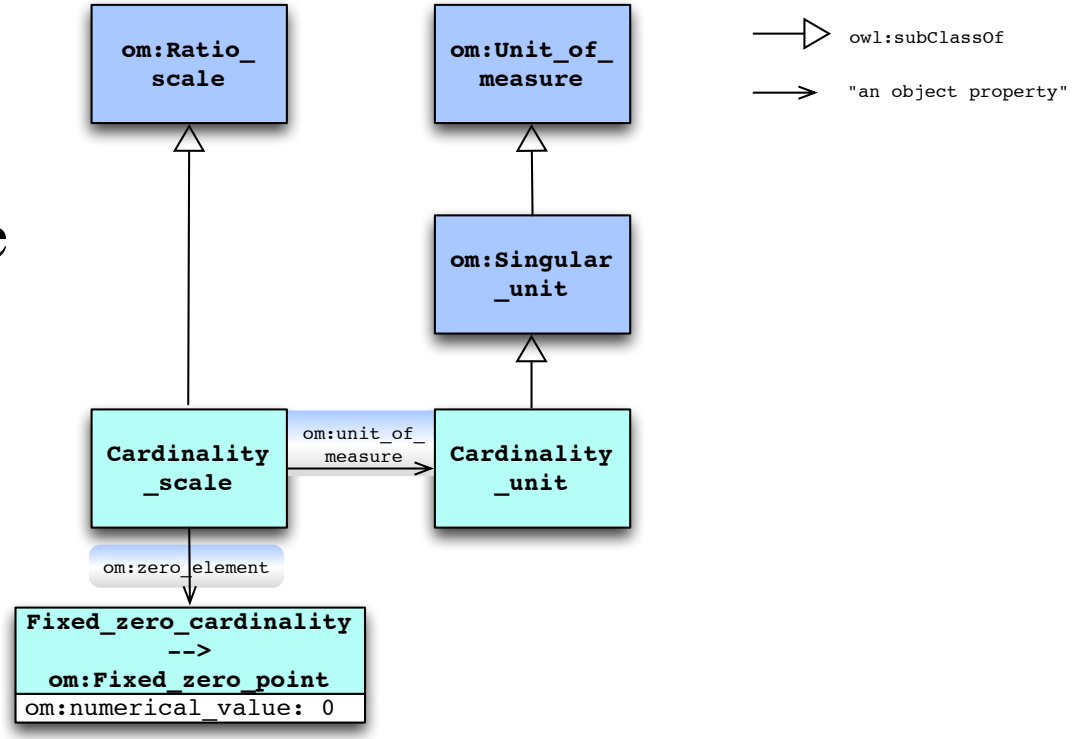
STR's Value

- We can now point to the actual value of the quantity which links to the unit of measure.



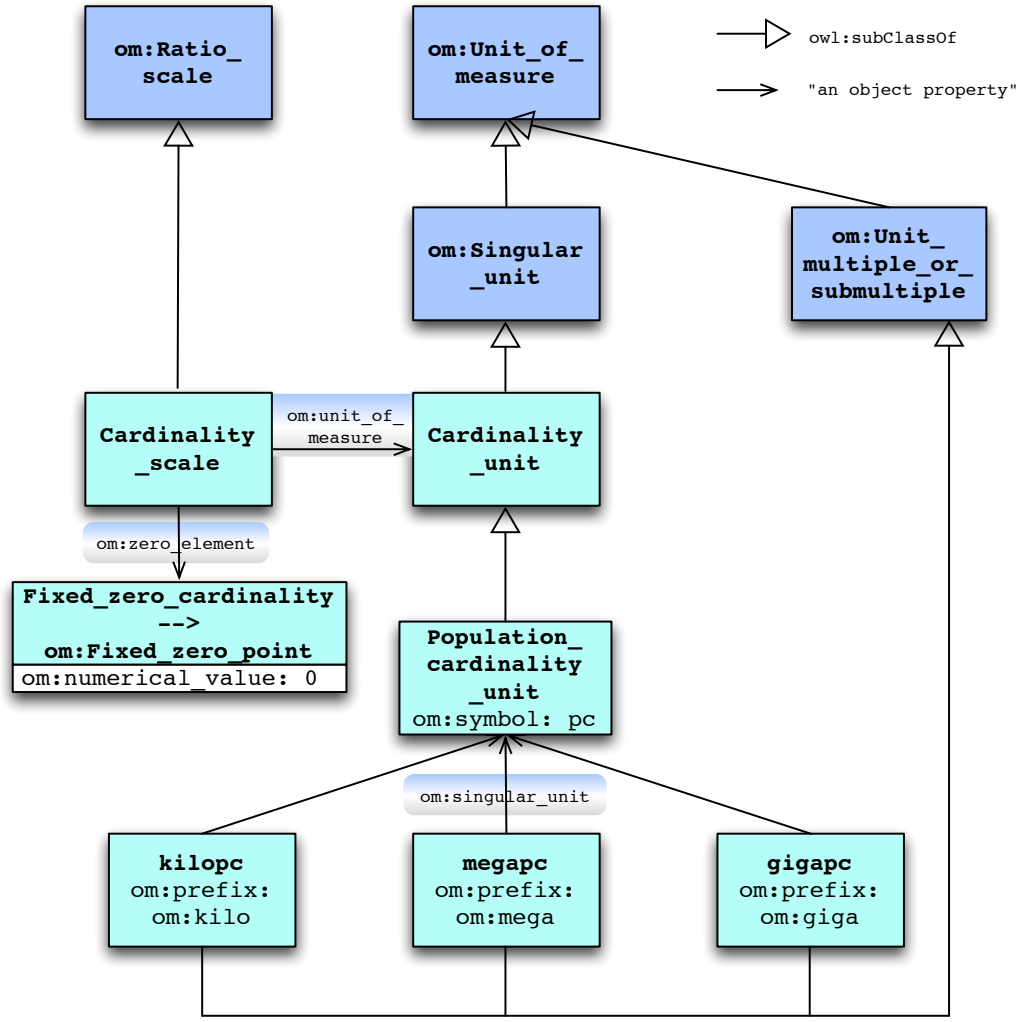
Population Unit of Measure

- For STR, we have to first define a measure of population size.
- Have to first introduce the notion of measuring the cardinality of a set.



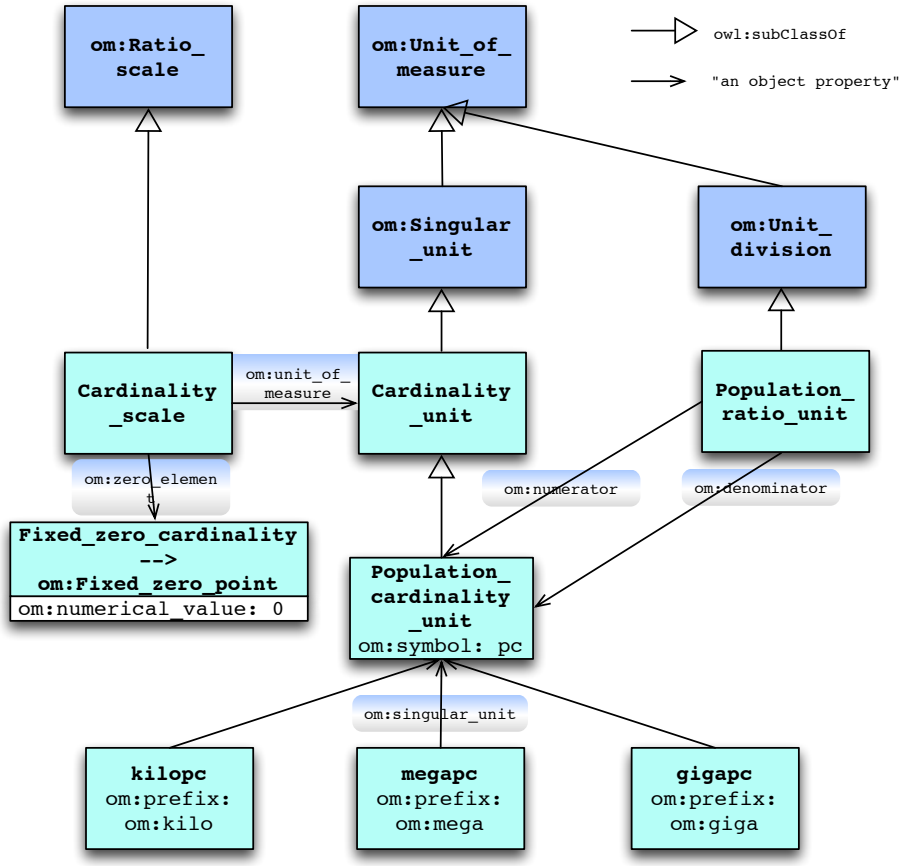
Population Unit of Measure

- Introduce a unit of measure “pc”: Population_cardinality _unit.
- Introduce scaled versions of the unit.



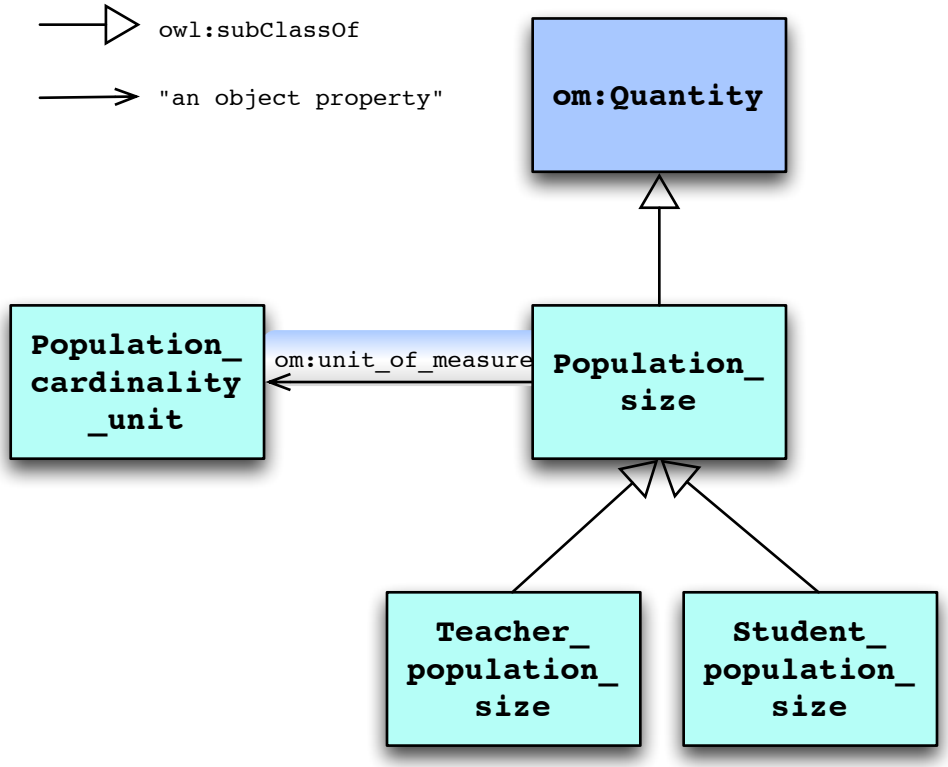
Population Unit of Measure

- Introduce Population_ratio_unit which is a measure of the ratio of two populations.
- It is a division unit, defined to be composed of a numerator and denominator both which are population cardinality units.



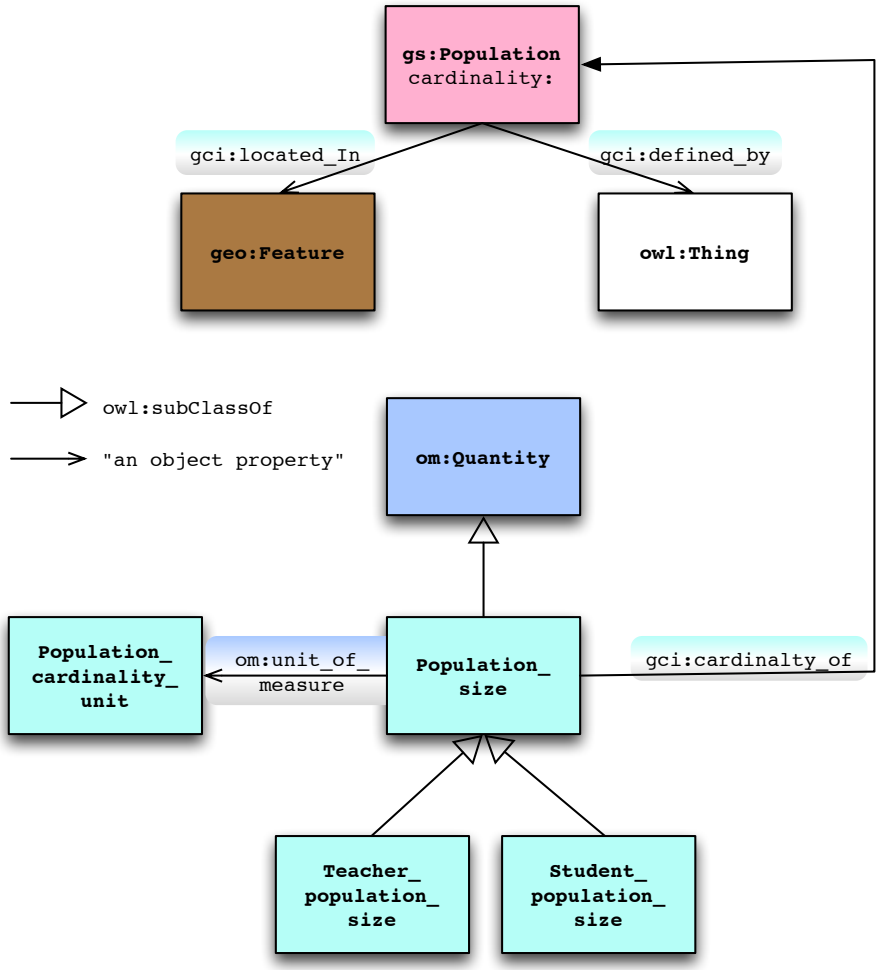
Population Size Quantity

- With the Population_cardinality_unit, we can now define a Quantity Population_size which is the cardinality of a Population (set).



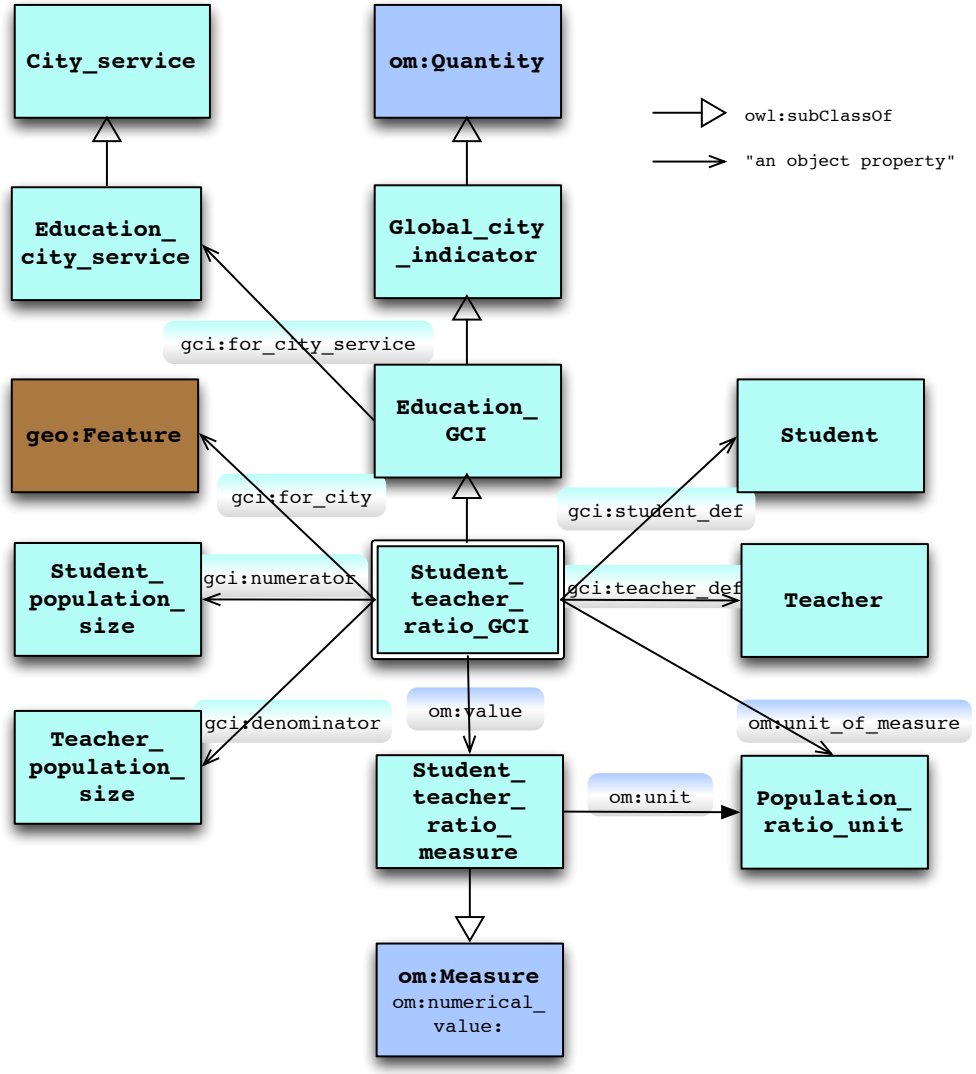
What is the Population?

- Introduce Population as what is being measured linked to the population size quantity.
- Population is defined as being located in a placename and a defined by a class description like a student in grades 1 thru 12, full and part time, …



Student/Teacher Ratio

- Final pieces are in place to define STR as being the division of a student population size by the teacher population size.



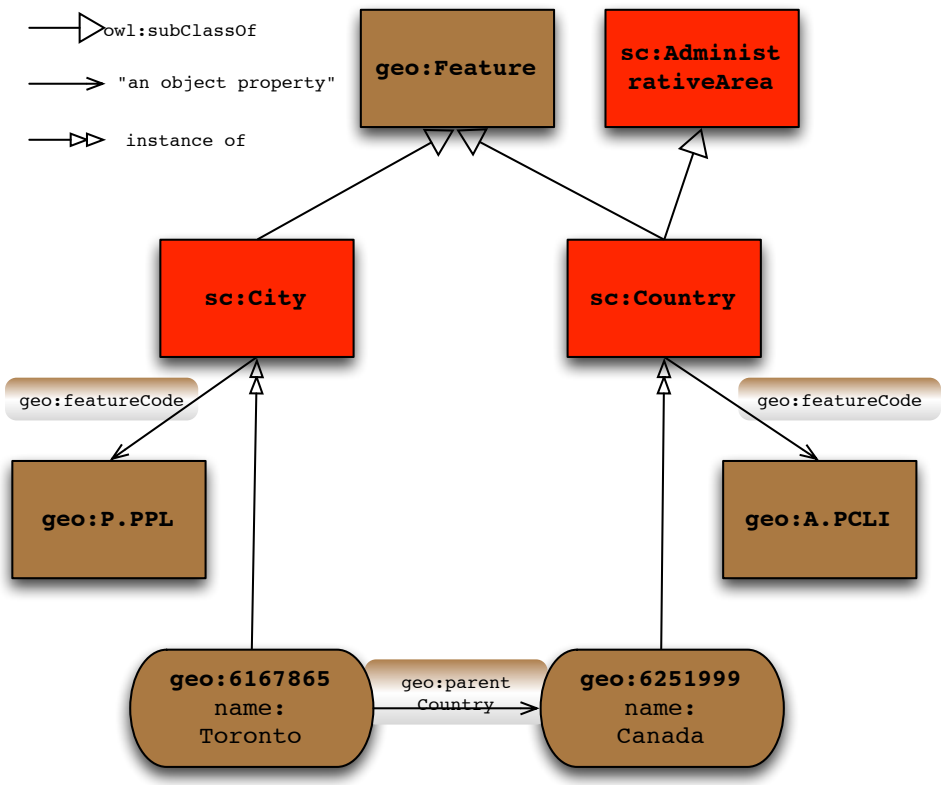
Identifying Place

- What is the city being measured?
  - Toronto pre-merger
  - Toronto post-merger
  - GTA
- 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: city, country
  - Linkedgeodata.org: hospitals, valleys, prisons, rivers, …

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.

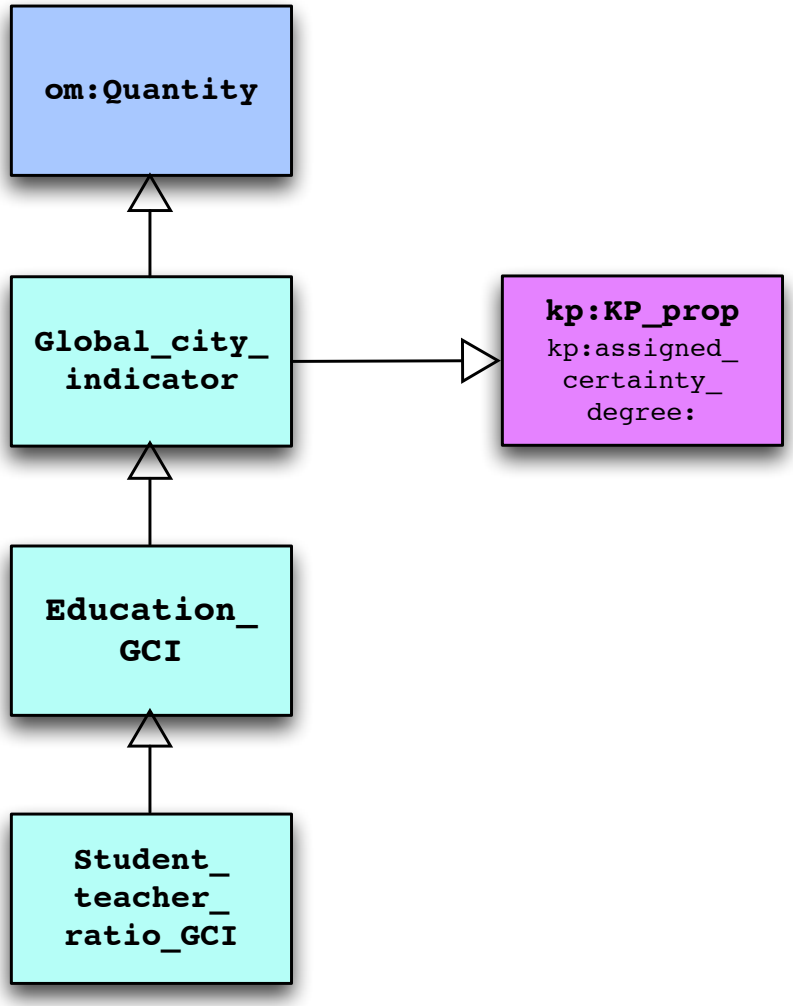


sc: http://schema.org/ ,
geo: http://www.geonames.org/ontology/ontology_v3.1.rdf#

Meta-Information

Is the Indicator Valid? Accurate?

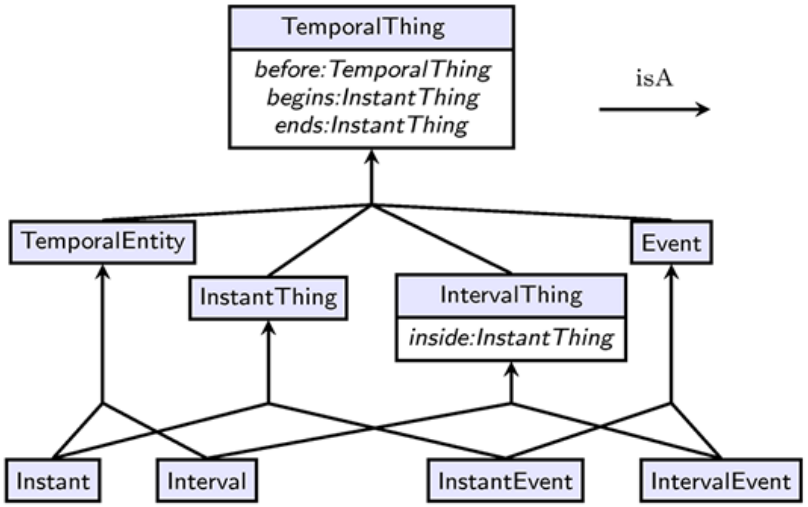
- Is the indicator true?
- Assign a degree of certainty that the indicator is valid.

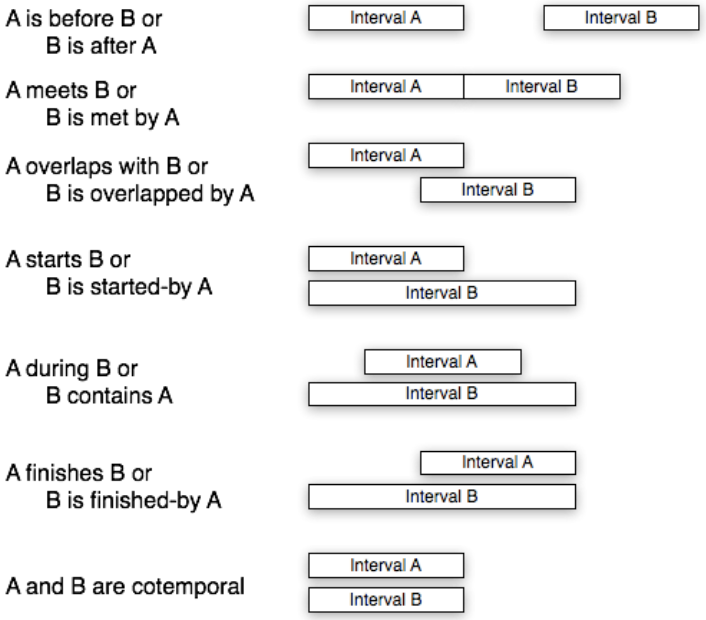


kp: http://ontology.eil.utoronto.ca/kp#

Temporal Validity

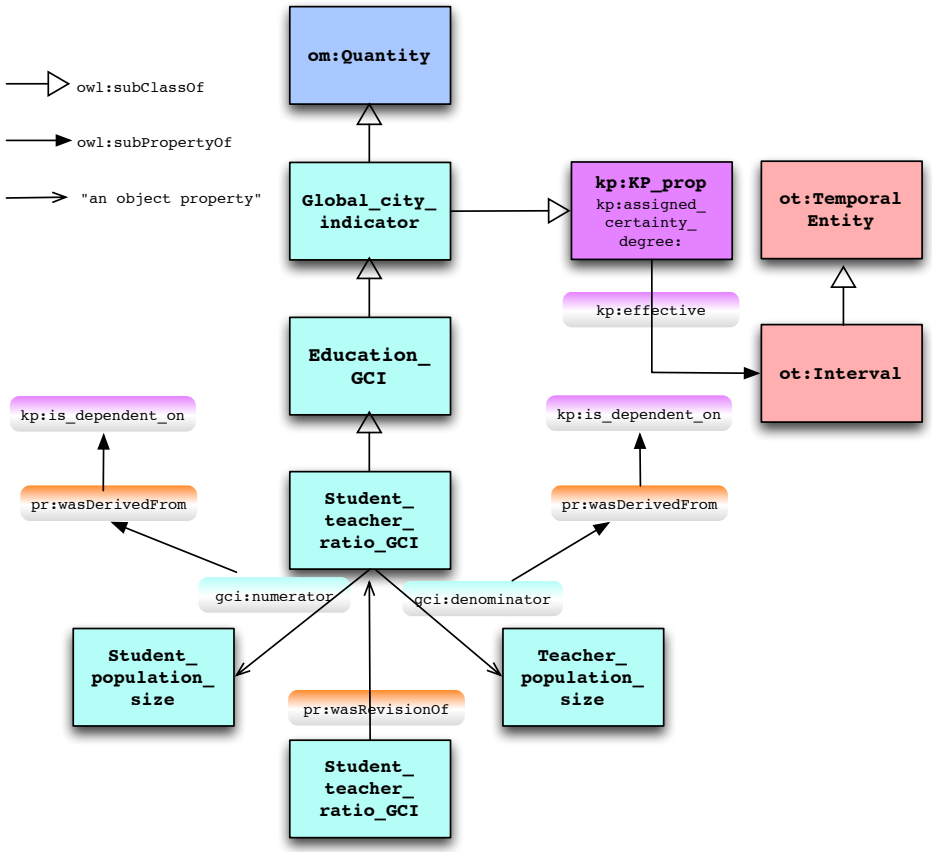
- Over what period of time is the indicator valid?





Temporal Validity

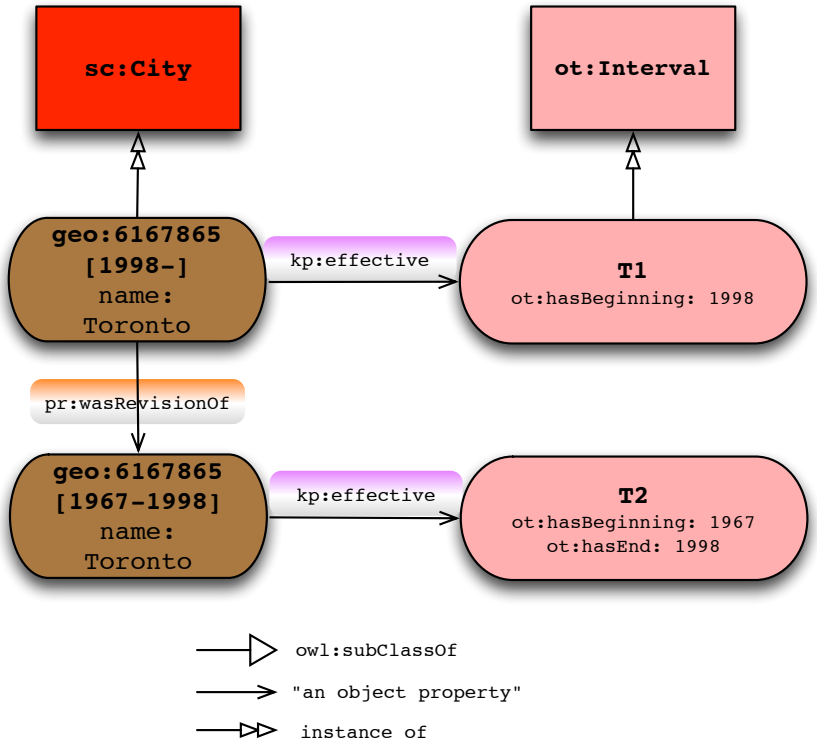
- Kp_prop is linked to a temporal entity by a kp:effective relation
- Relation defines the time period the degree of certainty holds.



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

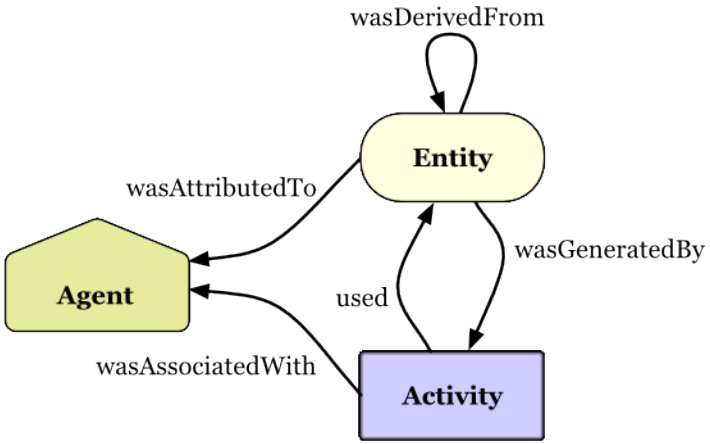
Dynamic Placenames

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

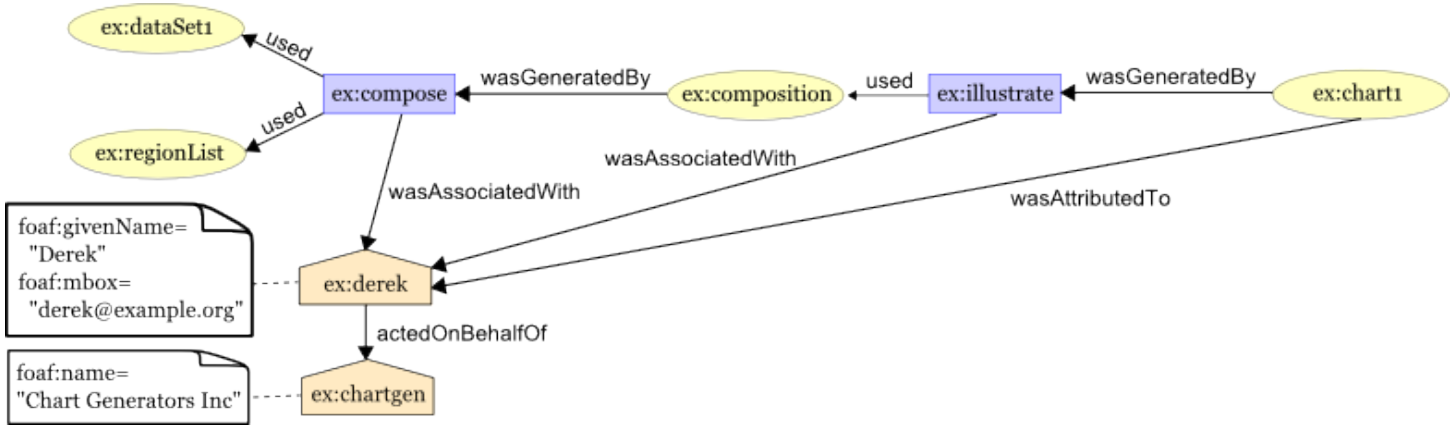


Provenance

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

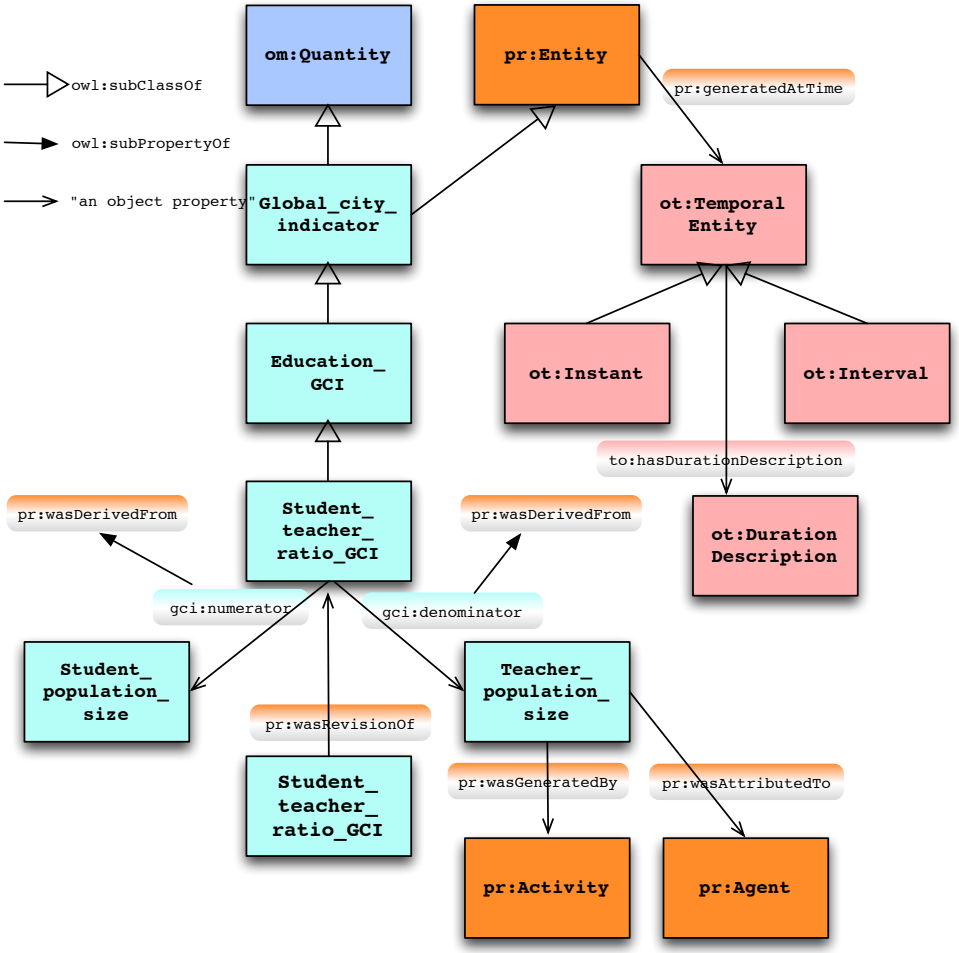


- *PROV:*



Provenance

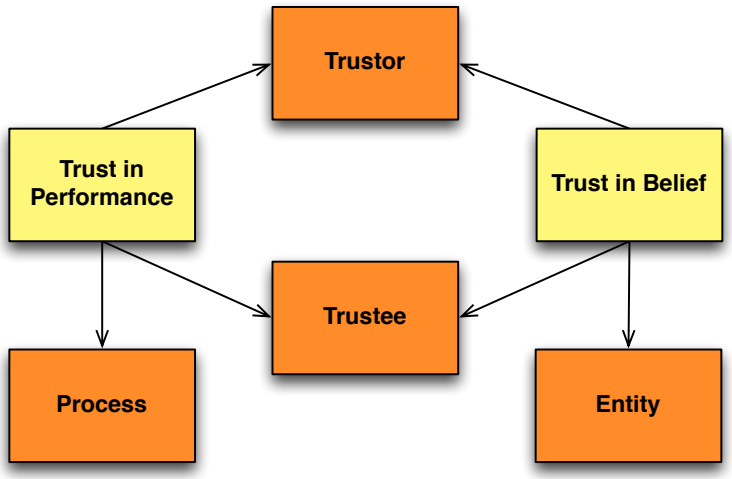
- By making a Global_city_indicator a subclass of a provenance Entity, it inherits all of the attributes of it.
- Can then link indicators and source numbers to activities, agents, time, etc.



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

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?

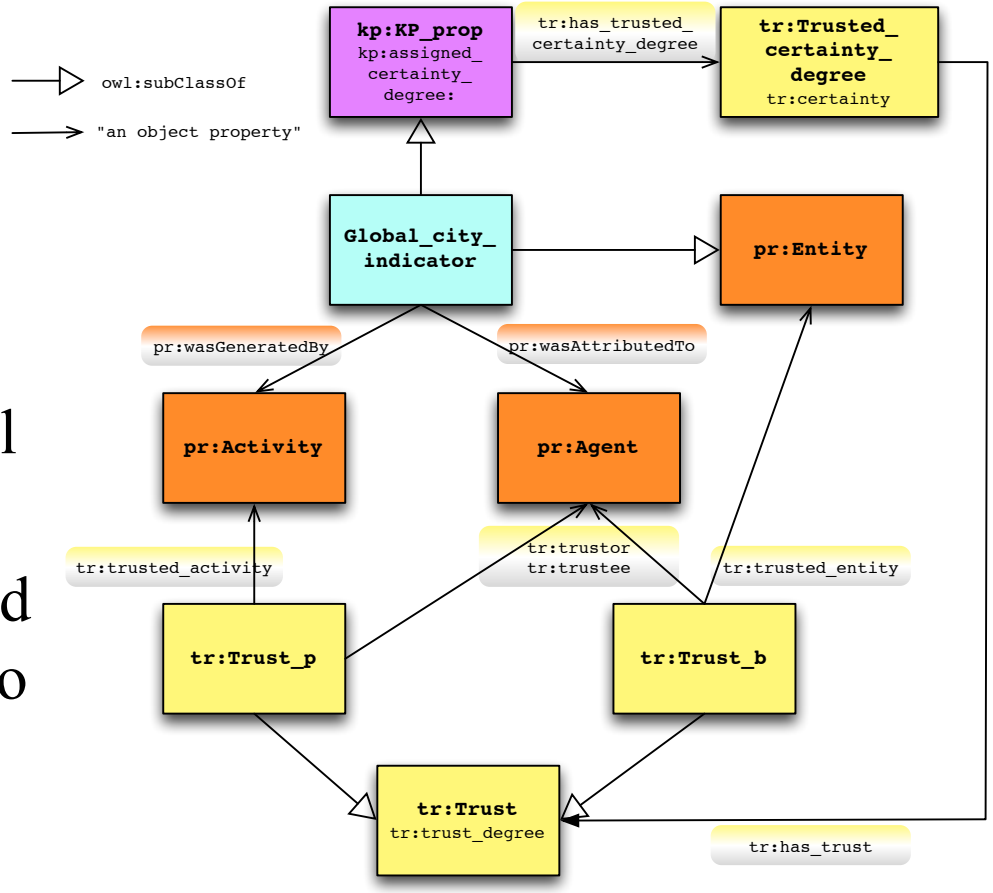


- What does it mean to trust an indicator?
  - Trustor's trust in the value that the trustee specifies?
    - I.e., trust in belief
  - Trustor's trust in the trustee's process that produces it?
    - I.e., trust in performance

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.



tr: Trust Ontology

Global City Indicators Foundation Ontology

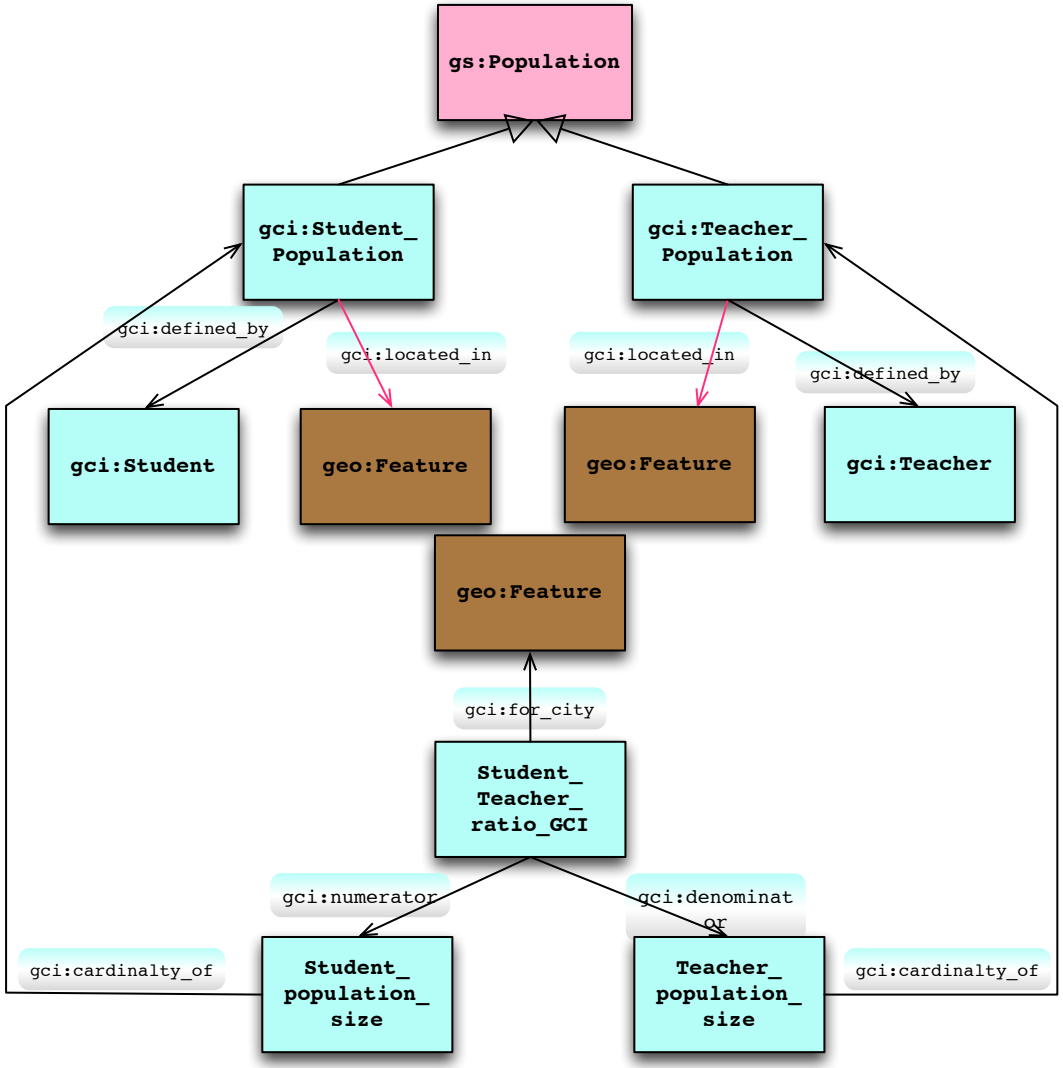
1. Develop an ontology for city indicators that builds on existing foundation ontologies by integrating and extending them, and publish on the Semantic Web.
2. **Develop bridge axioms that detect inconsistencies among the integrated ontologies.**

Bridge Axioms: Consistency

- With the grounding of an indicator requiring the integration of a multiple ontologies, we need to guarantee that the instantiation of their classes are consistent.

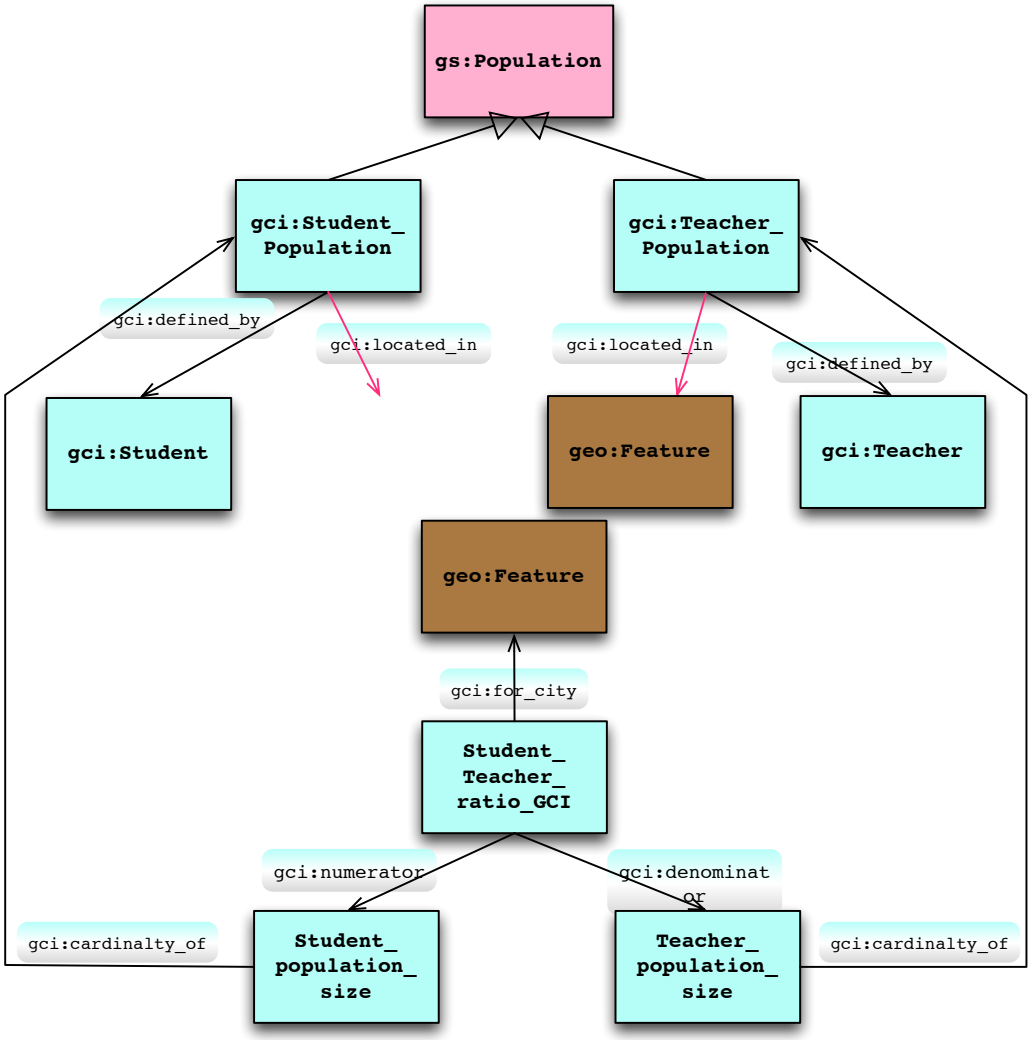
Placename Rules

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



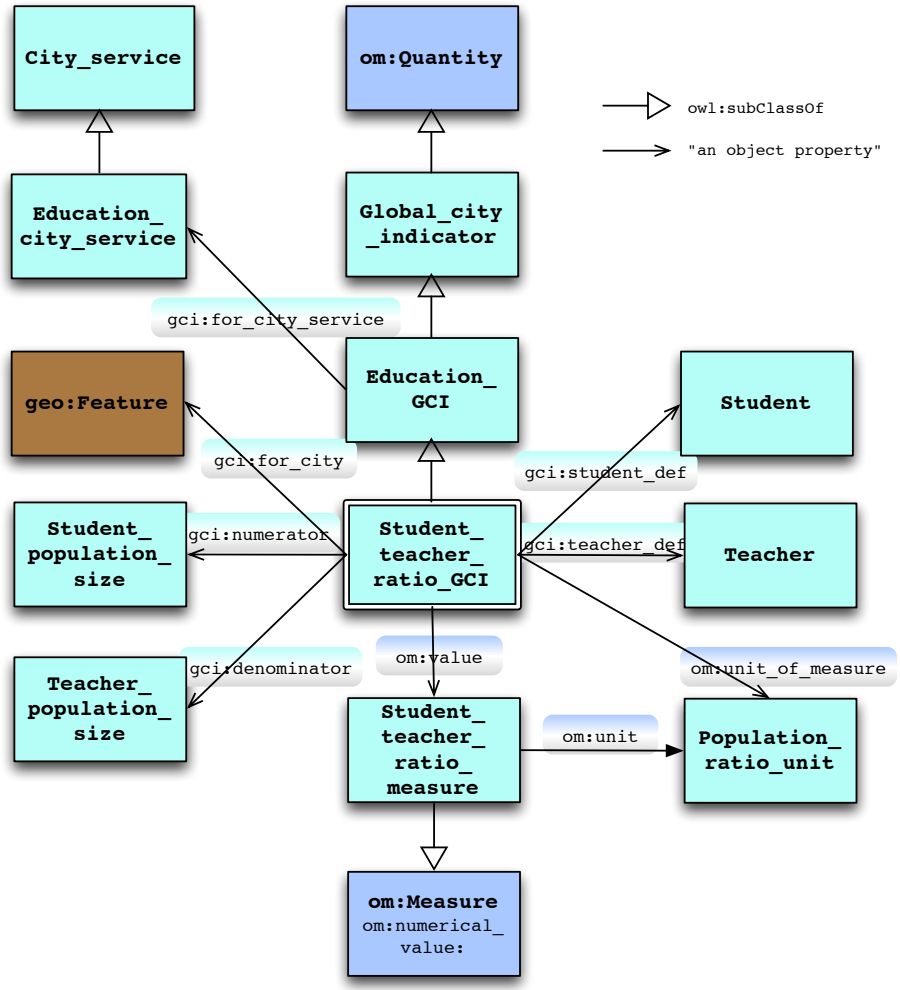
Bridge Axioms: Hypothesize

- We can use bridge axioms to hypothesize missing information.



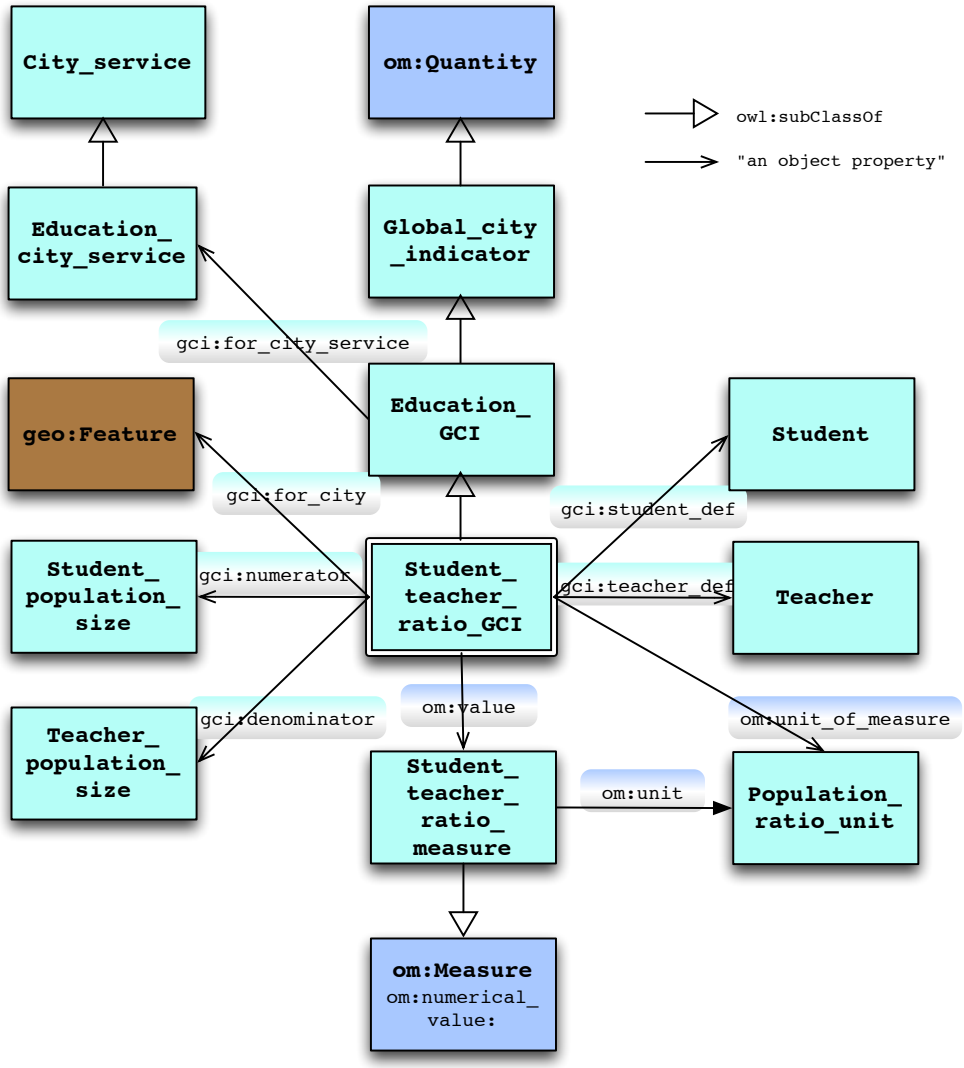
Measurement Rules

- **Rule M1:** The numerator and denominator of a gci:Student_teacher_ratio_GCI are the correct type (as specified by the GCI).
- **Rule M2:** The numerator and denominator of the gci:Student_teacher_ratio_GCI are consistent with the numerator and denominator of its unit of measure.
- **Rule M3:** If the numerator and denominator of a gci:Student_teacher_ratio_GCI are the same type, then they should have the same units (scale).
- **Rule M4:** The units of the actual measurement are the same as defined by GCI it is a measure of.
- **Rule M5:** The value of the gci:Student_teacher_ratio_measure is equal to the value of the gci:Student_teacher_ratio_GCI numerator divided by the denominator.



Population Rules

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



Meta Information Rules

- **Validity**
  - **Rule V1:** The effective time period for which an indicator is valid is contained within the effective time periods of its numerator and denominator.
  - **Rule V2**: The effective period for an indicator is after the time the indicator was generated.
- **Provenance**
  - **Rule P1:** If two versions of the same indicator exist, then they are inconsistent with each other if different methods were used to generate them.
- **Trust**
  - **Rule T1:** The trustee in a trust relationship is the same as the pr:wasAttributedTo Agent for an indicator.
  - **Rule T2:** The trusted certainty degree of an indicator is less than or equal to the indicator's certainty assigned by its creator.

Ongoing Research

1. Develop an ontology for city indicators that builds on existing foundation ontologies by integrating and extending them, and publish on the Semantic Web.
2. Develop bridge axioms that detect inconsistencies among the integrated ontologies.
3. Develop bridge axioms that identify inconsistencies between measures separated by time (longitudinal) or location (transversal).
4. Develop axioms that identify root causes of longitudinal and transversal differences.
5. Develop a set of micro-theories for estimating missing data by mining the web.

Systems Perspective

1. Develop an ontology for **system metrics** that builds on existing foundation ontologies by integrating and extending them, and publish on the Semantic Web.
2. Develop bridge axioms that detect inconsistencies among the integrated ontologies.
3. Develop bridge axioms that identify inconsistencies between measures separated by time (longitudinal) or location (transversal) **within heterogeneous models**.
4. Develop axioms that identify root causes of longitudinal and transversal differences **within heterogeneous models**.
5. Develop a set of micro-theories for estimating missing data by mining the web.

Summary

- **Objective**: clear, well defined, precise and unambiguous, simple to understand.
- **Measurable and replicable**: easily quantifiable, systematically observable.
- **Auditable**: valid, subject to third-party verification, quality controlled data (legitimacy across users).
- **Statistically representative** at the city level.
- **Comparable/ Standardized** longitudinally (over time) and transversally (across cities).
- **Flexible:** can accommodate continuous improvements to what is measured and how. Have a formal mechanism for all cities and interested parties to comment on.
- **Interrelated**: indicators should be constructed in an interconnected fashion (social, environmental and economics).
- **Consistent and sustainable over time**: frequently presented and independent of external capacity and funding support.

References

- 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.
- Fox, M.S., (2013), "A Foundation Ontology for Global City Indicators", Global City Institute Working Paper #3.
- Fox, M.S., and Huang, J., (2005), "Knowledge Provenance in Enterprise Information", *International Journal of Production Research,* Vol. 43, No. 20., pp. 4471-4492.
- Hoornweg, D., et al., (2006), "City Indicators: Now to Nanjing", Third World Urban Forum, Vancouver.
- 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.
- Hobbs, J.R., and Pan, F., (2006), "Time Ontology in OWL", http:www.w3.org/TR/owl-time.
- Pattuelli, M.C., (2003), "The GovStat Ontology: Technical Report". The GovStat Project, Integration Design Laboratory, School of Information and Library Science, University of North Carolina at Chapel Hill, http://ils.unc.edu/govstat/papers/govstatontology.doc.
- Rijgersberg, H., Wigham, M., and Top, J.L., (2011), "How Semantics can Improve Engineering Processes: A Case of Units of Measure and Quantities", *Advanced Engineering Informatics*, Vol. 25, pp. 276-287.