Proceedings of the 2017 Industrial and Systems Engineering Conference
K. Coperich, E. Cudney, H. Nembhard, eds.

The PolisGnosis Project
Enabling the Computational Analysis of City Performance

Mark S. Fox
Enterprise Integration Laboratory, University of Toronto
msf@eil.utoronto.ca; www.eil.utoronto.ca

Abstract
Cities use a variety of metrics to evaluate and compare their performance. With the introduction of ISO 37120, which contains over 100 indicators for measuring a city's quality of life and sustainability, it is now possible to consistently measure and compare cities, assuming they adhere to the standard. The goal of this research is to develop theories, embodied in software, to perform longitudinal analysis (i.e., how and why a city's indicators change over time) and transversal analysis (i.e., how and why cities differ from each other), in order to discover the root causes of differences. The first phase of this project focuses on the creation of standard representations of city knowledge (i.e., Vocabularies and Ontologies) that can be used to represent indicators and their supporting data and publish them on the Semantic Web. The second phase focuses on the development of consistency axioms that automate the determination of whether a city's indicators and supporting data are consistent with the ISO 37120 definitions, and whether they are longitudinally and transversally consistent. The third phase focuses on the development of diagnostic algorithms that identify the root causes of longitudinal and transversal differences. Due to the heterogeneity of the supporting data, the applicability of classical diagnostic techniques is limited.

Keywords: City Indicators, city ontology, consistency analysis, root cause analysis.

1. Introduction
Open Data, part of the broader Open Government movement, is based on the belief that making data publicly available will lead to more effective public oversight. With more "eyes on the data", waste and inefficiencies can be detected and crowd-based solutions suggested. But there is a problem with this belief, it assumes that citizens have the ability to read, understand and analyse open data. But its sheer volume and complexity exceeds the typical citizen's abilities. Consequently, software applications need to be created that read, understand and analyze open data. The PolisGnosis project is exploring one such tool in the domain of city performance analysis.

Cities use a variety of metrics to evaluate their performance. With the introduction of ISO 37120 (2014), which contains over 100 indicators for measuring a city's quality of life and sustainability, it is now possible to consistently measure and compare cities, assuming they adhere to the standard and the data is openly published (Fox, 2015). But the volume of the data required to compute these indicators makes it impossible for citizens to analyse them.

The goal of this research is to construct an intelligent agent that can diagnose a city's performance. It will automate the longitudinal analysis, i.e., how and why a city's indicators change over time, and transversal analysis, i.e., how and why cities differ from each other, in order to discover the root causes of differences. In the following we describe the PolisGnosis vision and the progress to date.

2. PolisGnosis Agent Architecture
Our goal is to create a "universal" agent that is not tailored to specific indicators nor cities. Therefore the design of the PolisGnosis agent must satisfy the following requirements:

1. **Indicator Independence.** Since there are a vast number of indicators used by cities, beyond those defined in the ISO 37120 standard, and ISO standards evolve over time, we do not want our agent to have any knowledge of indicator definitions "hardwired" into its code. An indicator's definition must be an input to the agent.
2. **City Independence.** Cities openly publish vast amounts of data that that our agent would like to use. But the data lacks any standard models or vocabularies - every dataset differs in structure, attributes and

content. It would be practically impossible to construct an agent that can understand these datasets. Hence the agent will assume that cities will adopt or translate into a standard for representing the information used to derive its indicators.

3. **Analysis Independence.** Given the variety of indicators and the ensuing variety of data used to derive them, a variety of methods of analysis may also be required. Rather than hardwire these methods into PolisGnosis, it would be better if they too were inputs to the agent.

Due to these requirements, the PolisGnosis Agent will take as input:

- The definitions of city indicators,
- All information pertaining to how a city measured its indicators, and
- A set of analysis axioms.

Figure 1 depicts the architecture of PolisGnosis. The blue boxes focus on the representation of indicators and the data used to derive them. The goal is to transform existing city data into a standard representation that can be published on the Semantic Web. The green boxes focus on consistency analysis of published indicator data to verify that they are consistent with the definitions of the indicators, and that indicators for a city over time or comparing two cities are consistent. The orange boxes focus on theories to diagnose the root causes of longitudinal and transversal differences.

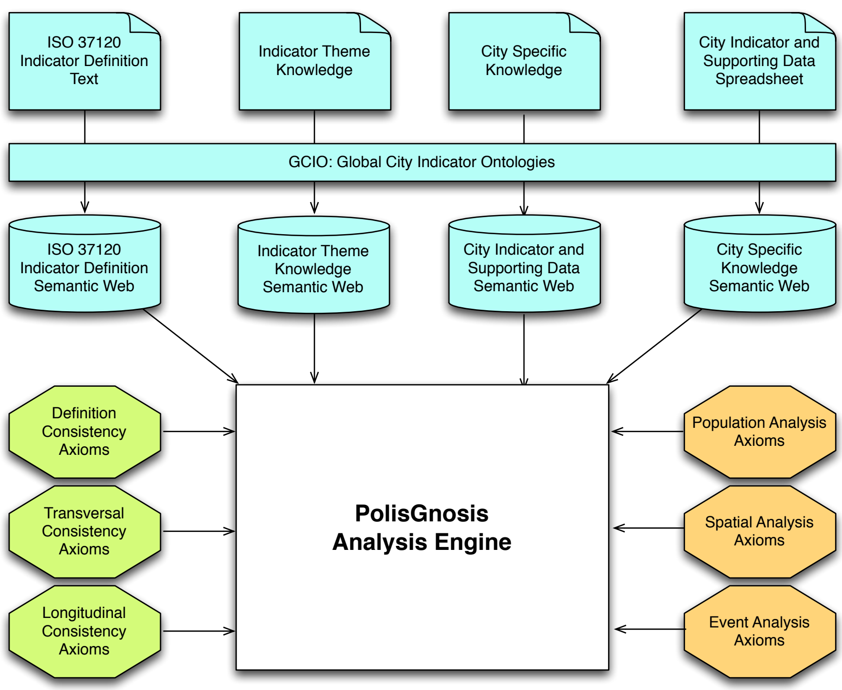

Figure 1: PolisGnosis Architecture

In the next sections we summarize these three areas. We use as an example indicator 6.4 in the ISO 37120 standard: Primary School Student Teacher Ratio. It measures the ratio of students who attend primary public schools to teachers who actually teach at primary public schools.

3. GCIO: Global City Indicators Ontologies

The goal of our ontology research is to create a standard with which cities can openly publish their data on the Semantic Web, enabling intelligent agents to read and analyse it. Figure 2 identifies four types of information that need to be published and for which standards need to be developed. First, the agent needs to understand the definition of the indicator. Cities may be using ISO 37120, other emerging standard indicators, and/or their own. Hence we need to translate the indicator definitions into a computer understandable representation – this requires an ontology. Second, the agent needs to understand a certain amount of theme specific "common sense" knowledge in order to interpret indicator definitions and analyse the data properly. For example, if dealing with primary school student teacher ratio, it needs to understand the concepts of school, primary vs secondary, public vs private, student, teacher, grade, etc. – this too requires an ontology. Third, in order to analyze specific cities, the agent needs to understand how the common sense theme knowledge is interpreted for 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, do they exclude England's public schools)? This too requires an ontology. Finally, the agent needs to understand a city's specific indicator values and the data used to derive them. This information may be available in PDF files or spreadsheets but needs to be translated into a computer understandable representation – this too requires an ontology.

The focus of our ontology research has been on ISO 37120. This standard is divided into 17 themes such as Health, Transportation, Education, Finance, etc. Figure 2 depicts the organization of ontologies used to define the ISO 37120 indicators. At the highest level, i.e., ISO 37120 Ontology level, the ISO 37120 module contains the globally unique identifier (IRI) for each ISO 37120 indicator. For example, for "6.4 Primary School Student Teacher Ratio" indicator, the IRI is: "http://ontology.eil.utoronto.ca/ISO37120.owl#6.4".

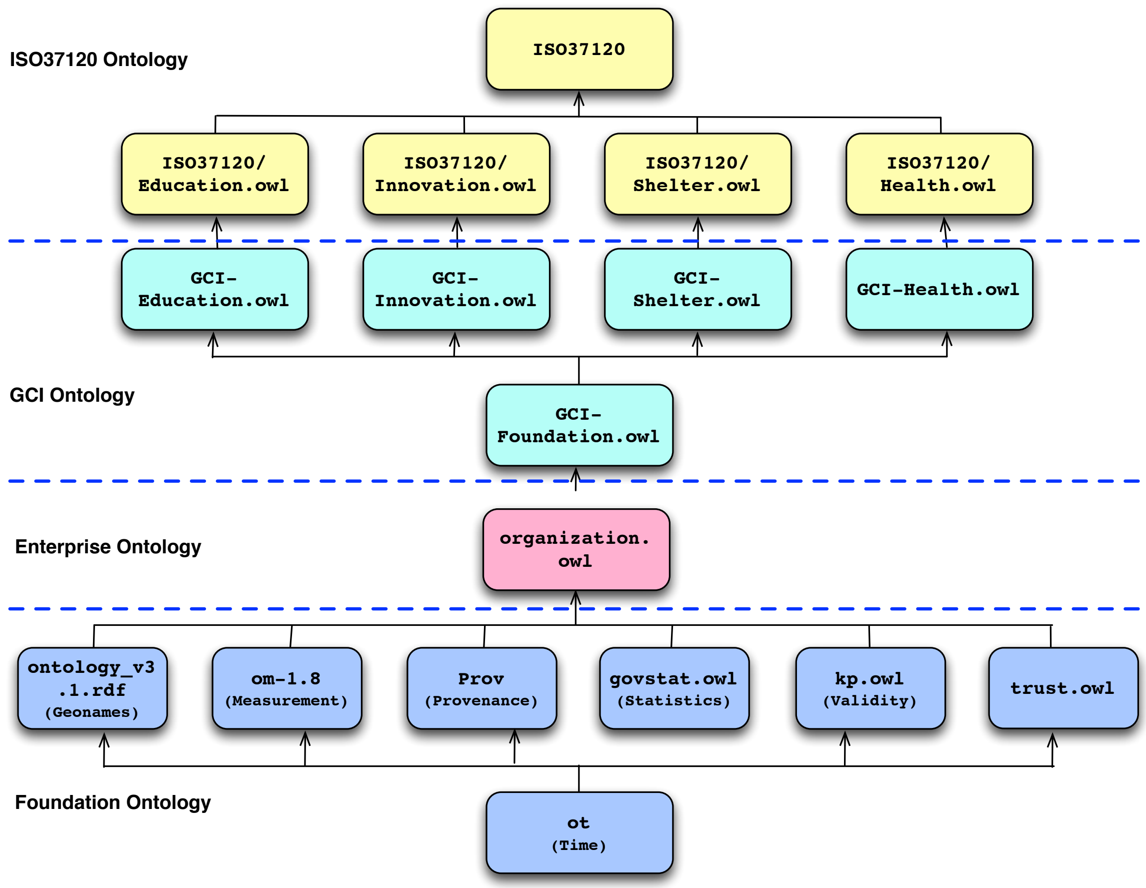

Figure 2: GCIO Ontology Hierarchy

For each indicator theme in the ISO 37120 specification, for the Education theme, there is a separate file that provides the definition of each indicator in that theme. For example, ISO37120/Education.owl provides a complete OWL definition for the education theme indicators in the ISO 37120 specification. The GCI Ontology level provides the generic, theme-specific ontologies required to define each theme's indicators. For example, to define the ISO 37120 Education indicators, we need an ontology covering common sense education concepts such as schools, grades, students, teachers, etc. GCI-Education.owl provides the classes used by Education.owl.

All of the theme specific indicator ontologies rely about the GCI Foundation ontology (Fox, 2013; 2015) for more generic concepts including: Populations and their specifications, and extensions to OM measurement ontology (Rijgersberg et al., 2011) to include population-related metrics.

The Enterprise Ontology level is based on the TOVE Enterprise Modelling ontologies (Fox & Gruninger, 1998). In this figure we only show the Organization Ontology file (Fox et al., 1998). Finally, the Foundation Ontology level provides very basic ontologies that were selected as the foundation for the GCI-Foundation.owl ontology. This includes the following ontologies: Time (Hobbs & Pan, 2006), Measurement (Rijgersberg et al., 2011), Statistics (Pattuelli, 2009), Validity (Fox & Huang, 2005), Trust (Huang & Fox, 2006), and Placenames (www.geonames.org).

To date, we have developed ontologies for representing the common sense knowledge of and indicator definitions in the following ISO 37120 themes:

- Education, including schools, teachers, students (Fox, 2013),
- Environment, including pollution and sensors (Dahleh & Fox, 2016),
- Finance, including debt, asset, liability, revenue, expenditures and tax (Wang & Fox, 2016),
- Telecommunication and Innovation, including residency and service, (Forde & Fox, 2015), and
- Shelter, including households, homeless and shelters (Wang & Fox, 2015).

In addition, ontologies are under development for the following indicator themes: Transportation, Energy, Health, Recreation, Governance, Fire & Emergency, and Safety. All of these ontologies are accessible at: http://ontology.eil.utoronto.ca.

4. Consistency Research

We believe that the majority of problems with city data arise from inconsistencies. Being able to detect inconsistencies is a necessary pre-curser to diagnosis.

There are three categories of consistency analysis we are pursuing. The first category, definitional consistency, determines whether a city's published indicator data is consistent with the ISO 37120 definition. Consider the primary student teacher ratio example. Is the data published by the city consistent with the definition of "public school"? Are the grades included consistent with "primary grades" definition? Are teachers included consistent with the ISO definition of teachers, e.g., they cannot be administrators? This category of consistency analysis has to have access to the data from which a city's indicator value is derived, and not just the aggregate numbers.

The second category of consistency, intra-indicator consistency, determines whether the data a city used to derive an indicator is internally consistent? One type intra-indicator consistency is temporal consistency. For example, was the count of the number of students for the same time period as the count of the number of teachers? Another type is spatial consistency. For example, were the counts for both students and teachers for the same geographic areas? A third type is measurement consistency. For example, are the units for both the count of students and teachers of the same units? Could one be in 100s and the other 1000s?

The third category of consistency, inter-indicator consistency, determines whether the published indicator data for two cities being compared are consistent. If we assume that each city's indicator data are definitional and intra-indicator consistent, what could be the types of inconsistency that arise? Consider longitudinal consistency analysis (comparing the same indicator for the same city at two different times). Temporal-spatial consistency problems arise when the geographic boundaries of the city change in the intervening time between the measuring of the indicators, e.g., merger of suburbs into the city. Temporal-definitional consistency problems arise when the definition of entities change over time, e.g., definition of student changes to include special needs students. Event consistency problems arise when a significant event occurs during the intervening time. For example, climatic events such as a hurricane may have an effect.

To date we have developed and tested a system for all three categories of consistency called City Indicator Consistency Checker: CICC (Wang & Fox, 2017). Given an indicator definition, a city's indicator value and its supporting data, all represented using the GCIO, CICC is able to detect all three categories of inconsistencies (Fig 3).

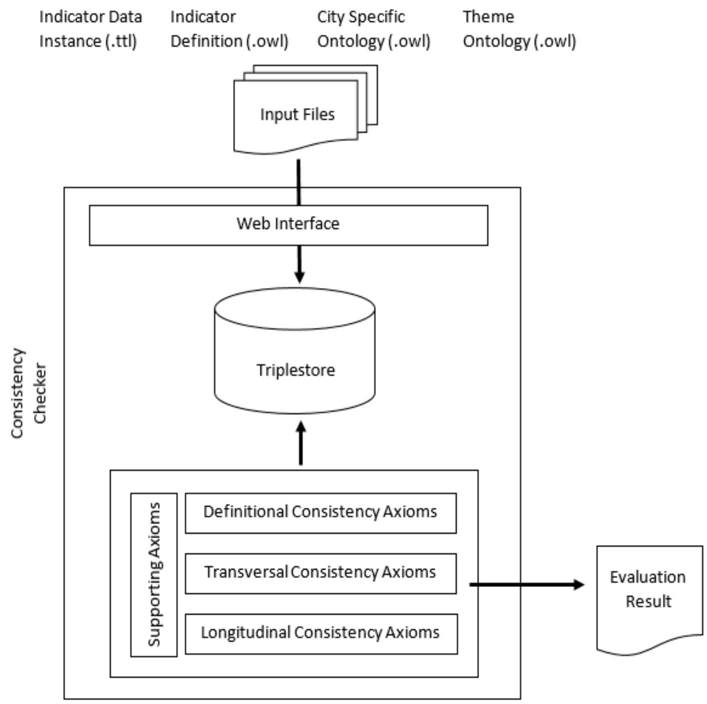

Figure 3: City Indicator Consistency Checker

5. Diagnosis Research

The final goal of our research is to diagnose the root causes of a city's performance, as measured by indicators. It is easy to detect a change or difference in an indicator's value, but the issue is what has led to the change? Diagnosing the root cause of differences is a challenge because the data is composed of many types, including: quantitative (e.g., counts and ratios), statistical (e.g., Population sampling), logical (e.g., definitions of students, teachers), spatial (e.g., city boundaries) and temporal. Sources of change can occur anywhere.

Luckily, consistency analysis provides insight into root causes, such as definitional consistency (e.g., adding special needs students to the student population), and longitudinal spatial changes (e.g., city boundary changes over time). We can use consistency analysis to identify the possible sources longitudinal and transversal changes. They are essentially hypotheses as to the root cause. What remains is to determine whether these proposed root causes are significant sources of change. Secondly, there may be other root causes of change not addressed by inconsistency analysis. For example, significant events can play a major role in changes in indicator values. For example, hurricanes can affect economic indicators, hosting the Pan American games can affect recreation indicators, and immigration can impact shelter indicators. This research remains to be done.

6. Conclusion

This paper presents our vision, architecture and status of the PolisGnosis System, that analyses the longitudinal and transversal performance of cities based on open data. In order to achieve this vision, ontologies are required to define a common standard for publishing open data on the semantic web, and city and indicator independent algorithms are required to analyse the data. Progress has been made on the development of ontologies for each of the 17 themes of the ISO 37120 standard, i.e., the Global City Indicator Ontology. These ontologies can be used to represent the definitions of the ISO 37120 indicators, and to represent and publish on the Semantic Web the data used by cities to derive their indicator values. Secondly, we have developed the City Indicator Inconsistency Checker that is able to identify definitional, transversal and longitudinal inconsistencies in the published data, which make comparisons invalid.

Never the less, there are two major hurdles we face in the practical application of this research: 1) how can we take the results of this and other related research, and develop global standards for representing and publishing city data, and 2) how can we convince cities to adopt these standards and publish the information necessary for an intelligent agent to perform comparative analyses. Addressing the former hurdle, two new ISO projects have been initiated: ISO 21893 which focuses on creating a standard ontology for representing city meta data, and ISO 21972 which focuses on creating a standard ontology for representing city indicators. With respect to the latter hurdle, there remains a long way to go evidence by the lack of publishing of city data on city open data web sites (Fox & Pettit, 2015).

Acknowledgements

This research is supported in part by the Natural Science and Engineering Research Council of Canada. The PolisGnosis team includes Dalia Dahleh, Jide Falodi, Andrew Forde, Jingwei Huang, Kourosh Khazei, Alanna Komisar, Ping Lou, Iris Wang, and Yetian Wang.

References

Dahleh, D., and Fox, M.S., (2016), "An Environment Ontology for Global City Indicators (ISO 37120)", Working Paper, Enterprise Integration Laboratory, University of Toronto, http://eil.mie.utoronto.ca/wp-content/uploads/2015/06/An-Environmental-Ontology-for-Global-City-Indicators-12oct2016.pdf

Falodi, J., and Fox, M.S., (2016), "A Health Ontology for Global City Indicators (ISO 37120)", Working Paper, Enterprise Integration Laboratory, University of Toronto, to appear.

Forde, A., and Fox, M.S., (2015), "An Innovation Ontology for Global City Indicators (ISO 37120)", Working Paper, Enterprise Integration Laboratory, University of Toronto, 1 April 2015. http://eil.utoronto.ca/smartcities/papers/forde-eilwp15.pdf.

Fox, M.S., (2013), "A Foundation Ontology for Global City Indicators", Working Paper 3, Global Cities Institute, University of Toronto. http://eil.utoronto.ca/smartcities/papers/GCI-Foundation-Ontology.pdf.

Fox, M.S., (2014), "An Education Ontology for Global City Indicators (ISO 37120)", Working Paper, Enterprise Integration Laboratory, University of Toronto, 14 October 2014. http://eil.utoronto.ca/smartcities/papers/GCI-Education.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.

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

Fox, M.S., and Grüninger, M., (1998), "Enterprise Modelling", *AI Magazine,* AAAI Press, Fall 1998, pp. 109-121.

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

Fox, M.S., and 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.

Huang, J., and Fox, M.S, (2006), "An Ontology of Trust – Formal Semantics and Transitivity," Proceedings of the International Conference on Electronic Commerce, pp. 259-270. http://www.eil.utoronto.ca/km/papers/huang-ec06.pdf

ISO37120, (2014), "ISO 37120: Sustainable Development of Communities – Indicators for City Services and Quality of Life", International Organization for Standardization, First Edition, 2014-05-15, ISO37120:2014(E)

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.

Wang, Y., and Fox, M.S., (2015), "A Shelter Ontology for Global City Indicators", Working Paper, Enterprise Integration Laboratory, University of Toronto, 1 April 2015.

Wang, Y., (2017), "Consistency Analysis of City Performance" M.A.Sc, Enterprise Integration Laboratory, University of Toronto, in preparation.

Wang, Y., and Fox, M.S., (2017), "Consistency Analysis of City Indicator Data", *Planning Support Systems for Smarter Urban Futures*, Springer, to appear.

Wang, Z., and Fox, M.S., (2016), "Expenditure, Revenue and Taxes - Towards a Standard for Representing City Finance Open Data", *Proceedings of the IEEE Smart Cities Conference*, pp. 1-6, Trento Italy.