An Integrated Knowledge Base for Sustainability Indicators

Lida Ghahremanloo
School of Computer Science and Information Technology, RMIT University
GPO Box 2476, Melbourne 3001, Victoria, Australia
lida.ghahremanloo@rmit.edu.au
+64-3-99252758
(PhD Enrolment Date: 31/08/2010)

Keywords
Sustainability Indicators, Urban Sustainability, Ontology Engineering, Ontology Matching

ABSTRACT
The objective of this research is to develop a knowledge base to represent and integrate data schemas for sustainability indicators. Over the past 30 years, sustainability and sustainable development - not only in environmental, but also in economic and social areas - has become a key theme in both academic and popular literature. The focus in this research is on how to represent the important features of sustainability indicators (defined and used by different organisations) that can be reused for different purposes, such as reasoning, reapplying and querying. Selection of sustainability indicators forms a key part of current organisational planning and reporting activities. In particular, the importance of measuring sustainability in urban environments has received considerable attention in the past couple of decades. One of the technical difficulties for stakeholders in cities is what sustainability indicators should be chosen in the urban context. We propose here that presenting multiple indicator sets in a commensurate way can be solved by use of an ontological approach. In order to address the task of choosing the most suitable ontology for urban indicator sets, first, we have reviewed and compared the indicators used by the cities of Vancouver and Melbourne. We have then looked to compare the combined list of indicators against the GRI (company-oriented) and OECD (country-oriented) indicator sets, to identify broadly their degree of conceptual overlap. Next, we have applied the ROMEO methodology to evaluate the degree of fit between both GRI and OECD indicators, and those used by existing cities in practice. One early outcome at this stage is that both GRI and OECD do not provide full coverage of candidate city categories. A subordinate result from this small sample also indicates that they miss some key social categories.

1. INTRODUCTION
A common definition of the term sustainability is to ensure our use of resources is sufficient to meet the needs of the present without compromising the ability of future generations to meet their own needs [2]. Over the past 30 years, sustainability and sustainable development - not only in environmental, but also in economic and social areas - has become a key theme in both academic and popular literature. It is often stated that in order to manage something, it needs to be measured. Accordingly, managing complex economic, environmental and social systems required robust and comprehensive systems of measurement indicators. Two aspects of sustainability indicators are particularly important. First, indicators must have well-defined and understood meanings; and second, they must be able to be applied regularly [1]. Many indicator systems have been developed and are in use today: standards bodies such as, the Global Reporting Initiative (GRI)\(^1\), Organisation for Economic Co-operation and Development (OECD)\(^2\) and International Organisation for Standardization (ISO)\(^3\) have developed indicator sets as parts of reporting standards, while companies (such as Fuji Xerox Australia\(^4\)) and governments (such as City of Melbourne) and international organisation (such as the United Nations Statistics Division\(^5\)) have been adopted or adapted these indicator sets as part of their reporting practices.

1.1 Background
A major challenge in sustainability is that researchers still debate over its structuring knowledge. According to Kuma\(z\,\text{a}\,\text{a}\,\text{wa}\,\text{[15]}\), the structuring knowledge should include three perspectives: whenever, whatever, and whoever. Whenever applies to reusability of the knowledge. Whatever refers to versatility, that is a knowledge base should be compatible to as many different domains as possible not just a specific domain. Whoever suggests using a knowledge base should be the same for every user tracing the same structuring process. Moreover, the dynamic nature of the problem domains in sustainability science requires a knowledge structure to adjust redefinition of the problem domains. Thus, extensibility appears as the third requirement for a well-designed knowledge base [15]. In this research, we have designed and maintained a knowledge base for sustainability indicator sets that includes the aforementioned characteristics: reusability, versatility and extensibility.

Another aspect of this research is knowledge representation in reference to representing the key features of sustainability indicators. Knowledge representation is a division of science that suggests the interdisciplinary approaches to extract and reproduce the important information from content. One of the methodological approaches for knowledge representation

\(^1\)http://www.globalreporting.org/Home
\(^2\)http://www.oecd.org/home/
\(^3\)http://www.iso.org/iso/home.html
\(^5\)http://unstats.un.org/unsd/default.htm
in artificial intelligence is defining an *ontology*. In computing disciplines, an ontology is referred to an explicit model of a domain of knowledge consisting of a set of *concepts, properties* and *individuals* [6]. The primary concern in any ontology development task is defining a finite list of concepts and the relationships among them. These are written in a formal language, e.g. Web Ontology Language that is understandable by computers [8].

### 1.2 Motivation

The United Nation Global Compact Cities Programme has developed a set of indicator sets under the category of *The Circles of Sustainability* in four domains: Economics, Ecology, Politics and Culture [13]. We contribute to this project using those indicator sets.

One contribution of the current research is representing the knowledge about sustainability indicators. Although there are some existing works that have shown the key concepts about indicators in a specific format, such as, GRI sustainability report represented in XBRL\(^6\), a systematic approach to this is missing. The issues related to which components of sustainability indicators are important to represent in a knowledge base and how to express them in an unique structure are still unanswered. The outcome of this research is expected to address these questions.

Another contribution of this research is integrating existing data schemas into our knowledge base. This mapping connects our local knowledge base to other global indicator sets.

### 1.3 Research Questions

The objective of the proposed research is developing a knowledge base for representing multiple sets of sustainability indicators that can be reused later on for different purposes of reasoning, applying and querying the indicators. In particular, we have focused on the research questions and appropriate methodologies as follows:

**RQ1: How to design an ontology to systematically represent knowledge about sustainability indicators?**

Research question one focuses on representing the knowledge about sustainability indicators following a structured design. The first step to address this is understanding sustainability indicators and their main features. The second step consists of designing a systematic approach for representing sustainability indicator key information e.g. the use of ontologies. The steps of building ontologies is referred to *ontology engineering*. The process involves the construction, refinement and usage of a given ontology for an application (more detail in Section 2.1).

The final step is evaluating the ontology after the design phase. There are three ontology evaluation approaches: *Gold standard*, *Criteria-based* and *Task-based*. The first approach compares an ontology with a benchmark ontology. Maedche and Staab [12] proposed a gold standard approach to empirically measure similarities between ontologies from different views such as lexical and conceptual aspects. These measures determine the accuracy of discovered relations generated from the proposed ontology compared with an existing ontology, but the main disadvantage is that they are not so useful outside the domain of the proposed ontology. Criteria-based approach — as the name suggests — evaluates the ontology based on the criteria such as *consistency, completeness*, *coherence, expandability* and *sensitivity* that are proposed in different studies [4, 5, 6]. These criteria focus on the characteristics of the ontology in isolation from the application. So, the main drawback with this approach is that while ontology criteria maybe met, it may not satisfy all the needs of the application. The task-based approach evaluates an ontology based on the competency of the ontology in completing tasks [20]. This approach judges whether an ontology is suitable for the application or task in a quantitative manner by measuring its performance within the context of the application. The disadvantage of this approach is that an evaluation for one application or task may not be comparable with another task. Hence, evaluations need to be taken for each task being considered.

For the specific purpose of this research, selecting a suitable ontology for urban indicator sets, we have chosen an ontology evaluation approach namely *ROMEO* — given in section 2.2 — that involves validation experiments using task-based performance comparisons on ontologies.

**RQ2: What technologies and algorithms are the most appropriate to build, maintain and access a reusable, versatile and extensible knowledge base for sustainability indicators?**

Research question two focuses on constructing, maintaining and querying a knowledge base with the specific characteristics. In order to build a knowledge base choosing a suitable language to represent the ontology that formulates the fundamental rules is the first task. A knowledge representation language must have four essential features: “vocabulary”, “syntax”, “semantics” and “rules of inference” [10]. We have represented the ontology in RDF/OWL. We have then used SPARQL\(^7\) to query the knowledge base. Another stage to address this question is identifying crucial criteria for evaluating our ontology such as, “Clarity”, “Coherence”, “Extendibility”, etc. Two suggested evaluation mechanisms are: *OntoClean* [7] and *OntoMetric* [11]. The former evaluates the *correctness* status of the ontology. It specifies whether the ontology modelled entities and properties correctly reflected entities in the application being modelled. The latter focuses in Metric Methodology that uses an adapted version of the Analytical Hierarchy Process (AHP) method [14] to improve the decision-making on multiple criteria.

**RQ3: What are the most effective algorithms and techniques for integrating and evaluating existing indicator sets into our knowledge base?**

Research question three focuses on an integrating and mapping process from existing indicator sets to our knowledge base. The first phase requires us to compare various taxonomies represented by different organisations. The sec-

\(^6\)http://www.globalreporting.org/reportingframework/

\(^7\)http://www.w3.org/TR/rdf-sparql-query/
ond phase to address the integration task is applying ontology matching techniques and algorithms to our ontology. One approach is manual matching techniques using XSLT, SAX and other programmatic approaches for translating data instances from other formats into the desired template of our ontology. One of the early outcomes of this step is that we have successfully loaded GRI indicator sets into our knowledge base translated from XBRL into RDF using SAX technology.

The final step is overall evaluation of our knowledge base from the practical value perspective. This can be performed by consulting experts in sustainability indicator representations and by interviewing end users of our system and evaluating the ontology with reference to their feedback. This technique is called User-Oriented in which final users are directly involved in verifying ontology requirements and development cycles [17].

1.4 Problem Definition
Selection of sustainability indicators forms a key part of current organisational planning and reporting activities. The importance of measuring sustainability in urban environments has received considerable attention in the past couple of decades, following the publication of United Nations Agenda 21. Studying current literature indicates that one of the technical difficulties is what sustainability indicators should be chosen in the urban context. In fact, developing an adequate indicator set for cities is extraordinarily difficult. In this paper, we propose that presenting multiple indicator sets in a commensurate way can be solved by use of ontological approach. Particularly, we have narrowed the problem into the question below:

Task: Which ontology is the most appropriate to report on urban sustainability indicators?

As a further refinement of our approach, we have considered here the challenge of building robust urban indicator sets from an ontology engineering perspective. We have assumed that a city endeavouring to construct a robust ontological model of indicators will need to do so with reference to the sources mentioned in Section 2.1: existing, though non-standardised, city indicators, as well as standardised indicators used by companies and countries.

2. METHODOLOGY
In order to address the task of choosing the most suitable ontology for urban indicator sets, we have performed the steps below: First, we have reviewed and compared the indicators used by the cities of Vancouver and Melbourne. These two cities are selected because they rank first and second against the Liveability Index, and also produce substantial sustainability reports themselves. They can therefore be taken as good exemplars of the kinds of sustainability reporting requirements other cities are likely to have, and therefore their categories and indicators can be used as a useful guide for evaluating the GRI and OECD standards. We then look to compare the combined list of indicators against the GRI, company-oriented and OECD, country-oriented indicator sets, to identify broadly their degree of conceptual overlap. Next, we apply the ROMEO methodology steps - given in Section 2.2 - that could be used to evaluate the degree of fit between both GRI and OECD indicators, and those used by existing cities in practice.

Figure 1: A Tree Representation of GRI Indicators

2.1 The GRI and OECD ontologies
In the first step to address the selection of a suitable ontology for urban indicators, we have investigated the indicator sets of cities of Melbourne and Vancouver. Table 1 illustrates the number of nodes in each category level for each candidate city and two resource organisations as well as the total indicators and by interviewing end users of our system and evaluating the ontology with reference to their feedback. This technique is called User-Oriented in which final users are directly involved in verifying ontology requirements and development cycles.

Table 1: The number of indicators & nodes per category

<table>
<thead>
<tr>
<th>Category</th>
<th>1st Level</th>
<th>2nd Level</th>
<th>3rd Level</th>
<th>Indicators</th>
</tr>
</thead>
<tbody>
<tr>
<td>Melbourne:</td>
<td>6</td>
<td>33</td>
<td>151</td>
<td>82</td>
</tr>
<tr>
<td>Vancouver:</td>
<td>9</td>
<td>-</td>
<td>-</td>
<td>54</td>
</tr>
<tr>
<td>GRI:</td>
<td>6</td>
<td>33</td>
<td>-</td>
<td>78</td>
</tr>
<tr>
<td>OECD:</td>
<td>5</td>
<td>16</td>
<td>28</td>
<td>47</td>
</tr>
</tbody>
</table>

The second phase is building the ontology for GRI and OECD indicator sets. Figure 1 illustrates a snapshot of GRI indicator set. It captures the tree category structured in the middle layers and indicators at the lowest bottom level. In order to build the ontology for the two aforementioned organisations, we have followed ontology engineering principles proposed in [19] as listed below:

1. Identify purpose and scope: Sustainability Indicators
2. Building the ontology:
   - ontology capture
3. Evaluation: ROMEO Methodology

As the list above indicates, two main concepts and three relations have been assumed for GRI and OECD ontology. It is also summarised the language and the tools that we have used for editing the ontologies.

The third step is applying ROMEO methodology to evaluate the designed ontologies that is given in Section 2.2.

Table 2: Requirement 1 for Urban Ontology

<table>
<thead>
<tr>
<th>Requirement 1:</th>
<th>Adequate level of coverage the structured categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analyse:</td>
<td>GRI and OECD Indicator sets</td>
</tr>
<tr>
<td>For the purpose of:</td>
<td>Selection of suitable ontology for a city</td>
</tr>
<tr>
<td>With respect to:</td>
<td>Matching structured categories between city reports, and GRI and OECD standards</td>
</tr>
<tr>
<td>From the view point of view:</td>
<td>Stakeholders of the cities</td>
</tr>
<tr>
<td>In the context of:</td>
<td>Matching categories</td>
</tr>
<tr>
<td>Motivation:</td>
<td>The ontology of the candidate organisations (GRI and OECD) are compared to measure the coverage of the structured hierarchy of sustainability indicators for our example cities (City of Melbourne and City of Vancouver).</td>
</tr>
</tbody>
</table>

2.2 ROMEO

Requirements Oriented Methodology for Evaluating Ontologies (ROMEO) [20] addresses the general concern with ontologies that is the lack of appropriate mappings of ontology requirements with related measures. In order to confront this problem, the ROMEO methodology focuses on a “requirement-oriented” approach. This approach links the requirements to known ontology evaluation measures through a set of questions. Yu, et al. [20] describe the components involved in ROMEO: First, the set of roles of the ontology ought to be defined to clarify the needs of that ontology. Next, these roles are linked to a corresponding set of ontology requirements that follows with the set of related questions. Ontology requirements must reflect a set of needs from a suitable ontology for the given application. Some of the aspects of the requirements are “competency”, “capability”, “functionality” and “standardised” in our case. Ultimately, the questions are used on the basis of ontology evaluation measures to determine how the questions have been answered and hence if a given requirement has been met.

As the third step of our approach, we consider two major requirements for evaluating GRI and OECD ontologies within cities indicator sets with respect to two key concepts in our ontology − category and indicator −. Yu, et al. [20] suggest breaking down the process into smaller phases, as we show in relation to our requirements in Table 2 and 3. The requirements are first, the level of overlap of indicator categories, and second, the level of overlap of indicators themselves.

3. PRELIMINARY RESULTS

Figure 2 illustrates a small portion of our preliminary results (two levels) that matches between indicator sets for the city of Melbourne within the GRI and OECD. The original table has 190 rows that the first column features the city of Melbourne indicator set structure, the second and third columns show the most similar categories that we could match among the GRI and OECD indicator sets respectively. Similar experiment and results has been obtained for City of Vancouver. One early outcome at this stage is that both GRI and OECD do not provide full coverage of candidate city categories. A subordinate result from this small sample also indicates that they miss some key social categories, such as “A sense of community” and “Collaborative community”.

The final step to complete the table is an overall evaluation of our matches that is performed by consulting an expert panel (consisting of social scientists and urban practitioners). This technique is called User-Oriented in which final users are directly involved in verifying ontology requirements and development cycles [17]. However, we have confronted some challenges to analyse the results. First, considering the
general ambiguity of what an indicator is, we are required to revise our result with an expert panel as well as applying other approaches such as, automatic matching. This approach needs to improve in regards to processing larger indicator sets. Second, we have not yet found adequate measures for our findings. It requires us to study existing literature, in particular, some of the case studies in biology which has a well-developed ontology cluster [18, 9]. Third, two proposed requirements in this paper might not be sufficient to investigate indicator sets of other cities that have different structure than Melbourne and Vancouver, such as Mumbai or Tehran. Therefore, more cities and requirements are ought to be considered.

4. FUTURE WORK
The future experiment for this task is in relation to three challenges mentioned in the previous section. Questions to be addressed are: How to improve manual matching process? One possible answer could be applying automatic matching approaches, such as schema-based [16] and semantic schema matching algorithms (S-Match) [3]. What measures are suitable to analyse the results? Furthermore, how to determine whether the evaluation of our work meets the real world requirements?

5. ACKNOWLEDGEMENT
This research is a part of an Australian Research Council (ARC) funded linkage project on Accounting for Sustainability: Developing an Integrated Approach for Sustainability Assessments.

6. REFERENCES