Brigham Young University
BYU ScholarsArchive

International Congress on Environmental Modelling and Software

3rd International Congress on Environmental Modelling and Software - Burlington, Vermont, USA - July 2006

Jul 1st, 12:00 AM

Information Integration through Ontology and Metadata for Sustainability Analysis

V. Brilhante

A. Ferreira

J. Marinho

J. S. Pereira

Follow this and additional works at: https://scholarsarchive.byu.edu/iemssconference

Brilhante, V.; Ferreira, A.; Marinho, J.; and Pereira, J. S., "Information Integration through Ontology and Metadata for Sustainability Analysis" (2006). *International Congress on Environmental Modelling and Software*. 191.
https://scholarsarchive.byu.edu/iemssconference/2006/all/191

This Event is brought to you for free and open access by the Civil and Environmental Engineering at BYU ScholarsArchive. It has been accepted for inclusion in International Congress on Environmental Modelling and Software by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.

Information Integration through Ontology and Metadata for Sustainability Analysis

V. Brilhante [a], A. Ferreira[b], J. Marinho[b] and J. S. Pereira [a]

[a]Computing Science Department, Federal University of Amazonas – UFAM, Manaus, Brazil
{virginia, jssp}@dcc.ufam.edu.br

[b]Federal Centre of Amazonas for Technological Education – CEFET-AM, Manaus, Brazil
{a.ferreira, joy_marinho}@cefetam.edu.br

Abstract:
Indicators of Sustainable Development (ISDs) make up a sophisticated set of information that can be employed as a means to study many aspects of a system's sustainability. Methods for producing ISDs, as well as environmental models in general, typically are strongly backed by data, so as to heighten their trustworthiness. This work proposes a sustainability analysis framework where higher level forms of data description, namely, a domain ontology and metadata, are employed to harness the data and to provide relevant elicited information to a software tool which is ultimately responsible for yielding levels of sustainability of systems under analysis. Having the framework equipped with a metadata standard and a formal ontology of the domain which are specified in standard and W3C (World Wide Web Consortium) recommended representational formalisms (XML and OWL) fosters information exchange and systems interoperability on the WWW platform.

Keywords: Sustainability Analysis; Indicators of Sustainable Development; Domain Ontology; Metadata

1 Introduction

A system is deemed sustainable if it is capable of functioning in such a way that its survival, development and perpetuated existence are guaranteed within its environment. Indicators of Sustainable Development (ISDs) are a means for assessment of such capability, lending themselves to a wide range of applications such as software tools to support sustainability analysis leading to decision making. We present a sustainability analysis framework, called MOeMA-IS[1], that is able to accommodate different theoretical approaches for analysis, including Bossel's methodology which is illustrative of them in the paper. This choice has been consolidated by applying the methodology to a sustainability analysis of the State of Amazonas, Brazil.

Formal domain ontologies are knowledge engineering artifacts that describe an agreed interpretation of a knowledge domain, making explicit the terms that represent pertinent concepts and their intended meaning [Uschold and Gruninger, 1996]. In this way, they are a means of specification that enhances appropriate use and sharing of the domain knowledge in different contexts. Ontologies are to be machine interpretable, so the range of such knowledge users includes humans as well as software systems. In the scope of a single software system, an ontology can act as an underlying, prescriptive specification for consistent domain information handling; on top of that, it can be shared by different systems which leads to interoperability between them.

Our framework includes an ontology of the ISDs domain called ISD-Economics. Such an ontology can serve a system concerned with ISDs various purposes, for example: it can provide an appropriate vocabulary for the user interface and a hierarchical organisation of the domain (e.g., Trade is a subtheme of the Economic Structure theme) so that the system can be structured accordingly; the system can perform operations on indicators in line with their properties and relationships as prescribed in the ontology (e.g., an input for the Balance of Trade in Goods and Services indicator is the value of exported goods in US dollars; the Gross Domes-

[1]MOeMA-IS stands for the equivalent in Portuguese of Metadata, Ontologies and Sustainability Indicators integrated to Environmental Modelling.

tic Product per Capita indicator is linked to population growth); etc.

The remainder of the paper unfolds as follows. The theoretical foundations of the sustainability analysis framework are presented in Section 2. Section 3 addresses the use of data, metadata and XML in the production and analysis of ISDs. The ISD-Economics ontology is presented in Section 4. Section 5 gives an integrated view of the framework, followed by conclusions in Section 6. Section 7 closes the paper with future work and outlook.

2 SUSTAINABILITY INDICATORS AND ANALYSIS

The present generations shall make use of available resources without hindering the capacity of future generations to do the same. This is the notion of intergenerational solidarity which became central to discussions on sustainable development since the United Nations (UN) Conference on Environment and Development held in Rio de Janeiro in 1992. Numerous studies try and establish increasingly elaborate criteria, taking into account social, economic, environmental and institutional dimensions, for determining levels of sustainability of systems, be they geographical regions, countries, organisations, etc. To carry out a study such as this, firstly, it is necessary to understand the system's characteristics, subsystems, environment, internal and external interactions that sustain the system as such. Secondly, indicators must be carefully selected so as to convey information needed to verify the system's behaviour status and thus support analyses of its development sustainability levels.

2.1 Bossel's Methodology

The outcomes of the first formal workshop on sustainability indicators, held in the Netherlands in April 1996, have had influence on the methodology proposed by Bossel [Bossel, 1999]. It is a methodology for correctly defining and applying sets of indicators to sustainability analyses, having as principle the satisfaction of basic guiding factors (called *orientors*) for a system, e.g., adaptability, effectiveness, etc., within the perspective of General Systems Theory. According to the methodology, the trustworthiness of an analysis results is dependent on how representative of the subsystems in question the applied indicators are.

We have put Bossel's methodology to the test through a study of sustainability levels of the Amazonas State in the north of Brazil. The Amazonas State system was analysed, using data relative to the 90s time period, through indicators of the Human (e.g., life expectancy at birth, adult literacy rate), Support (e.g., gross domestic product per capita, debt to gross national product ratio) and Natural (e.g., forest fires, use of agricultural pesticides) subsystems. To each of these three subsystems, seven indicators were used (one indicator per orientor) amounting to 21 indicators considered in the analysis, which resulted in the orientors star in Figure 1. The shaded area represents the extent of the Amazonas State sustainability in relation to the seven orientors proposed by the methodology. A detailed critique of this result together with all the indicators used in the study and the associated data, regression equations and evaluation functions can be found in [Ferreira, 2004; Ferreira et al., 2005].

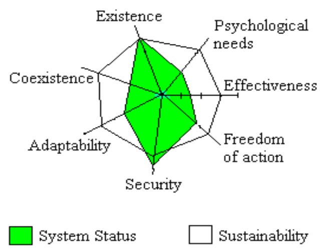

Figure 1: Orientors star of the Amazonas State in the 90s.

3 DATA, METADATA AND XML IN ISDS PRODUCTION AND ANALYSIS

A source of indicators for the sustainability analysis framework can be institutions such as Eurostat of the European Commission, the Brazilian Institute of Geography and Statistics (IBGE), etc., which maintain, or give access to, databases of quantitative ISDs with respect to specific systems and time periods. ISDs can also be produced by the open source software system under implementation in connection with the framework, which includes a comprehensive base of sustainability variables and methods (henceforth SVM-DB) that has been designed to suit sustainability indicators-based studies in general. SVM-DB has a metadata specification which follows the FGDC (Federal Geographic Data Committee)[2] metadata standard – adopted in our project, for the most part, due to its customisation facilities – and is encoded in XML (eXtensible Markup Language), a data markup standard.

3.1 ISDs Production

The algorithm in the MOeMA-IS system that produces, or calculates, an indicator also works in a

[2] www.fgdc.gov

generic manner to allow for easy inclusion of new indicators, given that the set of indicators to be used in a system's sustainability analysis can vary depending on technical and political criteria. Each indicator is associated with a stack data structure, stored in SVM-DB, containing variables and operators used in the indicator's measurement method. For an indicator to be calculated, values for the input variables are retrieved, also from SVM-DB, and the ISD-Economics ontology is consulted for units-of-measure consistency. The input values are then operated in a way similar to that of a reverse polish calculator. For the system to include new indicators it is only necessary for its corresponding stack and input variable values to be added to SVM-DB. An example of the workings of the algorithm, in connection with the ISD-Economics ontology, is presented in Section 5.

3.2 ISDs Analysis

The framework encompasses two other XML specifications. One of them consists of evaluation functions which map quantitative values of indicators, once produced or retrieved from an external ISDs base, into qualitative ones. Let us take, for example, the Adult Literacy Index indicator for Brazil. A quantitative value of 90% to 95% could be mapped into a 'good' level for the indicator. For other, more developed countries, values in this range could be considered 'insufficient', say. In our project, such evaluation functions have been designed taking into account UN recommendations and a comparative analysis of data and indicators concerning both developing and developed countries. The other XML specification consists of Bossel's methodology, with subsystems to consider and the corresponding sets of indicators through which satisfaction of the methodology's basic orientors can be verified, as discussed in Section 2.1.

4 The ISD-Economics Ontology

In order to promote studies on sustainable development of countries worldwide, the UN have put together a work programme on ISDs involving governments and a number of contributing organisations and experts. The programme delivered a report which provides countries wishing to undertake sustainable development programmes with a core set of indicators and associated guidelines and methodologies [United Nations, 2001]. ISD-Economics is an ontology on indicators of sustainable development with emphasis on the economic dimension, which is being developed on the basis of this UN report and expert consultation.

The tool we have been using to build ISD-Economics is Protégé coupled with a plugin for OWL (Web Ontology Language), a language founded on Description Logics (DL) whose semantics draws on set theory[3]. Figure 2 shows an excerpt of ISD-Economics hierarchy in Protégé. The classes that compose this part of the hierarchy are discussed in the sections that follow. Their formal descriptions are given using DL quantifiers and constructors. For ease of understanding, a reading of each formula is provided. Some descriptions refer to classes of the OWL version of SUMO (Suggested Upper Merged Ontology)[4], an upper level ontology developed by IEEE which provides definitions of concepts that are not specific to our domain.

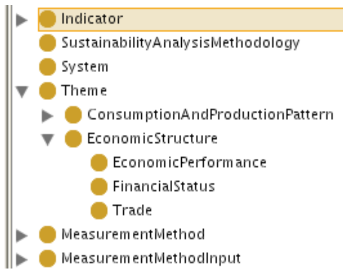

Figure 2: A section of ISD-Economics hierarchy.

4.1 Economic Themes, Subthemes and Indicators

The ontology includes a two-level hierarchy of themes in which the economic dimension of ISDs is structured according to [United Nations, 2001]. There are two themes, namely, `ConsumptionAndProductionPattern` and `EconomicStructure`, which are further classified into a few subthemes. Material and energy consumption are examples of subthemes in `ConsumptionAndProductionPattern`, which are deemed to be related to levels of depletion of natural resources and deterioration of the global environment. `EconomicStruture` comprises subthemes that are relevant for economic growth and sustainable development, *viz*, `EconomicPerformance`, `Trade` and `FinancialStatus`.

The specification of ISDs in general in ISD-Economics is as follows.

`Indicator` *is a subclass of the class of individuals that refer to a* `System` *and to a period of time*:

$$\mathtt{Indicator} \sqsubseteq \exists\, \mathtt{refersTo} \,.\, \mathtt{System} \sqcap \exists\, \mathtt{refersTo} \,.\, \mathtt{sumo:TimeInterval} \tag{1}$$

[3] The Protégé environment together with associated plugins and documentation are available from http://protege.stanford.edu.
[4] SUMO and associated documentation are available from http://www.ontologyportal.org/.

System is specified as a subclass of sumo : Agent which includes geopolitical areas, organisations, organisms, etc. And sumo : TimeInterval represents the class of time intervals, having both an extent and a location on the universal time line.

In the sequel, we will illustrate definitions of classes of indicators in ISD-Economics (which are subclasses of the Indicator class) using as examples indicators of the EconomicStruture theme.

Gross Domestic Product (GDP) per Capita Indicator. GDP per Capita is a basic economic growth indicator which measures the level and extent of total economic output, reflecting changes in the total production of goods and services. Measurements for the indicator are obtained by dividing annual or period GDP at current market prices by population.

GDPperCapita *is defined as the class of individuals which are* Indicators *and belong to and only to the subtheme of* EconomicPerformance *and have at least the* GDPperCapitaMethod *as measurement method and can only have* UnitedStatesDollar *as unit of measure:*

$$\begin{aligned}&\texttt{GDPperCapita} \equiv \\ &\quad \texttt{Indicator} \sqcap \\ &\quad \exists,\forall\, \texttt{belongsToSubtheme} . \texttt{EconomicPerformance} \sqcap \\ &\quad \exists\, \texttt{hasMeasurementMethod} . \texttt{GDPperCapitaMethod} \sqcap \\ &\quad \exists,\forall\, \texttt{hasUnitOfMeasure} . \texttt{sumo}:\texttt{UnitedStatesDollar}\end{aligned} \tag{2}$$

Table 1 shows another example of an ISD definition.

DebtToGNPRatio *is defined as the class of individuals which are* Indicators *and belong to and only to the subtheme of* FinancialStatus *and have at least the* DebtToGNPRatioMethod *as measurement method and are expressed as a percentage:*

$$\begin{aligned}&\texttt{DebtToGNPRatio} \equiv \\ &\quad \texttt{Indicator} \sqcap \\ &\quad \exists,\forall\, \texttt{belongsToSubtheme} . \texttt{FinancialStatus} \sqcap \\ &\quad \exists\, \texttt{hasMeasurementMethod} . \texttt{DebtToGNPRatioMethod} \sqcap \\ &\quad \texttt{isExpressedAs} \ni \text{'\%'}\end{aligned} \tag{3}$$

Table 1: Definition of the Debt to Gross National Product (GNP) Ratio indicator in the ISD-Economics ontology.

4.2 Indicators' Measurement Methods

As seen in Section 4.1, indicator definitions include the property of having a measurement method, that is, how input variables are operated together to produce indicators values. One such method can be structured in several levels, i.e., an input variable used in an indicator's method can, in turn, also have a measurement method that require other input variables, and so on and so forth. ISD-Economics includes a class of measurement methods of which the methods in the definitions of particular indicators are subclasses. Recall the definition of the GDPperCapita indicator (Formula 2). It has GDPperCapitaMethod as measurement method, specified as:

GDPperCapitaMethod *is a subclass of the class of individuals which are a* MeasurementMethod *and have* GDP *and* Population *as inputs*:

$$\begin{aligned}&\texttt{GDPperCapitaMethod} \sqsubseteq \\ &\quad \texttt{MeasurementMethod} \sqcap \\ &\quad \exists\, \texttt{hasInput} . \texttt{GDP} \sqcap \exists\, \texttt{hasInput} . \texttt{Population}\end{aligned} \tag{4}$$

GDP, in turn, is specified in ISD-Economics as:

GDP *is a subclass of the class of individuals which are a* MeasurementMethodInput *and are measured at* CurrentMarketPrices *with respect to a year or to other period of time and can only have* UnitedStatesDollar *as unit of measure:*

$$\begin{aligned}&\texttt{GDP} \sqsubseteq \texttt{MeasurementMethodInput} \sqcap \\ &\quad \exists\, \texttt{atPrice} . \texttt{CurrentMarketPrice} \sqcap \\ &\quad \exists\, \texttt{refersTo} . (\texttt{sumo}:\texttt{Year} \sqcup \texttt{sumo}:\texttt{TimeInterval}) \sqcap \\ &\quad \exists,\forall\, \texttt{hasUnitOfMeasure} . \texttt{sumo}:\texttt{UnitedStatesDollar}\end{aligned} \tag{5}$$

And Population is specified as:

Population *is a subclass of the class of individuals which are a* MeasurementMethodInput *and are a positive integer number:*

$$\begin{aligned}&\texttt{Population} \sqsubseteq \texttt{MeasurementMethodInput} \sqcap \\ &\qquad \texttt{sumo} : \texttt{PositiveInteger}\end{aligned} \tag{6}$$

4.3 Sustainability Analysis Methodology

The sustainability analysis concept discussed in Section 2 is modelled in ISD-Economics as the following class.

SustainabilityAnalysisMethodology *is a subclass of the class of individuals that are applied to systems of the* System *class and use indicators of the* Indicator *class*:

$$\begin{aligned}&\texttt{SustainabilityAnalysisMethodology} \sqsubseteq \\ &\quad \exists,\forall\, \texttt{isAppliedToSystem} . \texttt{System} \sqcap \\ &\quad \exists,\forall\, \texttt{usesIndicator} . \texttt{Indicator}\end{aligned} \tag{7}$$

One of such individuals is BosselMethodology which has an annotation property rdfs:isDefinedBy[5] in ISD-Economics pointing to the URI where an XML specification of the methodology (Section 3.2) is placed.

[5] One of the pre-defined RDF (Resource Description Framework) Schema primitives that OWL makes available for annotating the ontology with additional information, such as comments, URIs (Uniform Resource Identifiers), etc.

5 THE INTEGRATED MOEMA-IS FRAMEWORK

Throughout the paper we have discussed several parts of the MOeMA-IS sustainability analysis framework. This section focuses on how these parts are interconnected to compose the framework, as depicted in Figure 3.

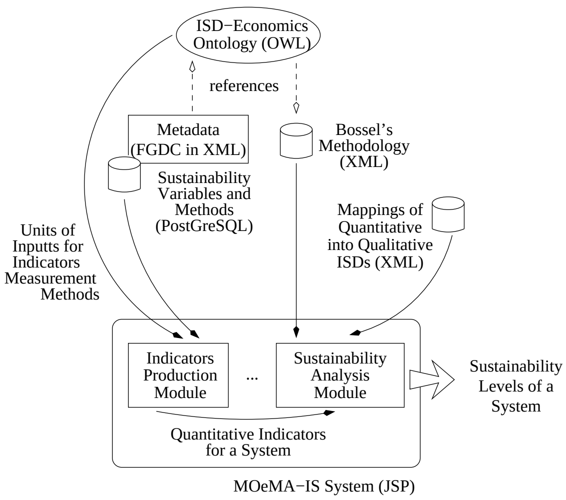

Figure 3: The MOeMA-IS framework.

As discussed in Section 3, the framework includes SVM-DB with a metadata specification which documents the data, reflecting its structure. Correspondences that exist between entities or attributes in the metadata and ISD-Economics are part of the metadata specification too. A tag is added to an entity or attribute by way of which the concerned data is linked up to the ontological concept that describes it. For example, the entities Country, State, Region, etc. are marked up with the tag <Entity_Ontology_Spec_Reference> with content moemais.ufam.edu.br/ontologias/isd-e.owl#System, which is the URI of the class System in ISD-Economics.

Within the framework, one working, practical use of the ontology is in keeping units of measure consistent in the production of indicators. This is done coupled with a calculator implementation because OWL does not lend itself to representing and handling arithmetic expressions. As discussed in Section 3.1, SVM-DB furnishes the Indicators Production Module with a stack of operands (inputs) and operators needed to calculate an indicator. Indicators can have specificities that may cause different values for an input to be retrieved from SVM-DB; for example, the GDP input for GDP per Capita may refer to a year or to some other period of time (see Formula 5). This will be determined by the user interacting with the MOeMA-IS system through an interface that presents options of indicators and their inputs according to the ISD-Economics definitions.

XML specifications of a sustainability analysis methodology and of mappings of quantitative to qualitative ISDs (Section 3.2) inform the Sustainability Analysis Module in the MOeMA-IS system[6], which also receives quantitative indicators from the Indicators Production Module[7]. The framework is designed so as to allow other of such specifications to be fitted into it. ISD-Economics can easily be made to refer to other analysis methodologies, as it currently does to Bossel's (Section 4.3).

An Example: Sustainability Analysis of the Amazonas State and the GDP per Capita Indicator. GDP per Capita, whose related ontological descriptions have been detailed in Section 4, is one of the 21 indicators used in the sustainability analysis of the Amazonas State system (Section 2.1). GDP per Capita was selected, specifically, for the Effectiveness orientor of the Support subsystem. Associated with the indicator, there is the $[\texttt{GDPperCapita GDP Population }'/']$ stack in SVM-DB, containing the parameters for the Indicators Production Module to execute $\texttt{GDPperCapita} = \texttt{GDP}/\texttt{Population}$, after values for the GDP and Population inputs are retrieved from the database and the ontology consulted for units consistency. Recall that Formula 4 specifies the inputs to produce a GDP per Capita indicator, and that Formulae 5 and 6 specify the unit of measure, or otherwise, in which the input is to be expressed (Population is expressed as a positive integer and a unit of measure does not apply). The values for the inputs retrieved from SVM-DB can be converted to the units prescribed by the ontology, if necessary. It is worth noting that in the ISD-Economics definitions for indicators, such as Formula 2, the unit of measure defined for an indicator is consistent with its respective inputs and measurement method – dividing GDP given in US dollars by Population given as a positive integer, results in GDPperCapita given in US dollars.

The quantitative value for the indicator, expressed in the ontology-compliant unit-of-measure, is then mapped into a qualitative value. Finally, this value is combined, according to Bossel's methodology,

[6]We use SAX (Simple API (Application Program Interface) for XML) to implement data exchange between the XML specifications and the modules of the MOeMA-IS system implemented with JSP (JavaServer Pages) technology. ISD-Economics source code is in OWL which is XML compatible.

[7]The MOeMA-IS system comprises several other modules and datasets which are not within the scope of this paper.

with the qualitative values of the other 20 indicators, produced in a similar fashion, to compose the levels of sustainability of the Amazonas State represented by the orientors star in Figure 1.

6 CONCLUSIONS

This work has presented a framework, in connection with a software implementation, for analysis of systems sustainability which can be customised to operate with any analysis methodology that makes use of quantitative and/or qualitative sustainability indicators. The framework has been designed to be versatile enough to suit a broad range of indicators and related data. It comprises, in sum, an ontology of ISDs of the economic dimension, a data base of sustainability variables and methods with a metadata specification, specifications of a sustainability analysis methodology and of mappings of quantitative to qualitative indicators, and software system modules for production and analysis of indicators. The ontology provides a formal description of the domain of interest which could not be suitably captured by a metadata standard alone and, so far, is applied at the implementation level to maintain the consistency of units of measure in the production of ISDs by the MOeMA-IS system.

7 FUTURE WORK AND OUTLOOK

The development of ISD-Economics is not yet complete. The current version of the ontology specifies for each indicator only its basic definition, not including variant definitions it may have. Moreover, only specifications of inputs for measurement methods of indicators are in place. They have not been extended to the inputs themselves. As discussed in Section 4.2, a measurement method may require several levels of inputs. For instance, the inputs for the GDP per Capita measurement method are GDP and Population. The latter is a count of people in a certain area which does not take other inputs into account. GDP, on the other hand, can involve product prices, currency conversion factors, etc. – it is a highly aggregated composite measure with several possible definitions [United Nations, 2001], which a more comprehensive version of ISD-Economics should include.

ISD-Economics makes explicit an interpretation of the ISDs domain using OWL which is a W3C standard. If other systems adopt the ontology, direct interoperability can occur between them and MOeMA-IS on the Web platform as the systems would share the same interpretation of the domain. In a setting where an interacting system does not adopt ISD-Economics, interoperability can still be facilitated since the interpretation of the domain under which MOeMA-IS functions can be accessed by the other system. Within our own framework and extended MOeMA-IS system, we aim at using, in practice, more of the ontology and inference capabilities to ensure other kinds of consistency besides that of units of measure.

The Indicators Production Module in the framework assumes that SVM-DB's provision of data is adequate and complete. The MOeMA-IS system is not equipped with resources for statistical treatment of the data. Finally, developments are under way to enrich the framework with more environmental data and metadata together with our ontology for such domain, called Ecolingua, so as to allow the exchange of ISDs or related data between the MOeMA-IS system and simulation environmental models.

ACKNOWLEDGMENTS

This work is funded by FAPEAM (Fundação de Amparo a Pesquisa do Estado do Amazonas) through the research project Metadata, Ontologies and Sustainability Indicators integrated to Environmental Modelling.

REFERENCES

Bossel, H. Indicators for sustainable development: theory, method, applications. Technical Report, International Institute for Sustainable Development. Winnipeg, Canada, 1999.

Ferreira, A. *Construccion de una Herramienta para el Análisis de la Sostenibilidad a través del Uso de Indicadores y de la Teoria General de Sistemas*. PhD thesis, Universitat de les Illes Balears, Spain, 2004.

Ferreira, A., A. Riera, and J. Machado. Scenes analysis for sustainable development supported by multi-criteria decision analysis. In *Proceedings of the 2nd Intl. ICSC Symposium on Information Technologies in Environmental Engineering (ITEE'2005)*, pages 186–201, Magdeburg, Germany, 2005.

United Nations. Indicators of Sustainable Development: Guidelines and Methodologies. http://www.un.org/esa/sustdev/publications/indisd-mg2001.pdf, 2001. Last accessed on 29 Apr 2006.

Uschold, M. and M. Gruninger. Ontologies: principles, methods and applications. *The Knowledge Engineering Review*, 11(2):93–136, 1996.