Citizen-centric access to statistics

Bo Sundgren

May 2011

Summary

- The project will investigate the possibilities to define an easy-to-use, powerful, and flexible standard interface for communication of statistical data and metadata between the dissemination databases of statistics producers (statistical agencies) and citizen-centric applications, developed by entrepreneurs or by citizens themselves.
- A consortium consisting of members from Estonia, Iceland, and Sweden will run the project.
- Interested entrepreneurs and other actors from government agencies and private business, even outside the Nordic-Baltic countries, are welcome to participate in an open network associated with the project.
- Three major seminars will be organised in Estonia, Iceland, and Sweden during 2011.
- In addition to producing its own conclusions concerning the problems, opportunities, and possible solutions of citizen-centric access to statistics, the project will prepare for full-scale development and implementation projects.
- More information about the project: https://sites.google.com/site/accesstostatistics/
  There is also a blog: http://stataccess.blogspot.com/

Background

- There is a rapidly growing supply of statistics, and everyone is using statistics in all sectors of society, private and public: researchers, analysts, business decision makers, politicians, journalists, and ordinary citizens – all use statistics to try to understand an ever more complex world, to make decisions for themselves and others, to participate in democratic processes, in arguments and negotiations, etc.
- National statistical offices have the role in the government of a modern, democratic society to provide objective statistics of good quality to all those mentioned above, including the ordinary citizens, and to make those statistics available to everyone, often as a public good, free of charge, financed by the taxpayers.
- Traditionally national statistical agencies and other producers of statistics have made statistics available through printed publications. Since the advent of computers, and especially since the general breakthrough of the Internet in the mid 1990's, statistical agencies have made statistics available via websites. Nowadays the web-based dissemination of statistics is offered for free by most national and international statistical agencies – with OECD still as a notable exception.
- Many different forms of dissemination are offered: from electronic versions of printed publications to multi-dimensional databases, giving the user very flexible access to subsets of statistics, time series as well as cross-sectional data, selected by the user herself, and presented in a variety of forms, ranging from traditional statistical tables to graphs and animations. The statistical data are accompanied by metadata in order to enable the user to interpret and understand the presented statistics.

But

- It is still the producer statisticians in the statistical offices who by and large determine which statistics should be made available, and how they should be made available.
- Of course many statistics producers have an ongoing, constructive communication with their users and make their best to fulfil their needs and demands within the budget limits.
- Users of statistics themselves, or innovative entrepreneurs acting in cooperation with them, come up with new ideas about how to make statistics more understandable and useful. Example: Hans Rosling.
- Innovate initiatives by entrepreneurs or by the citizens themselves are often hampered by the lack of simple, easy-to-use standards for downloading statistical data and metadata to the users' own systems and to new citizen-centric services developed by entrepreneurs.
- The EU Public Sector Information (PSI) Directive deals with these problems and urges member states to implement legislation supporting initiatives by entrepreneurs to develop and market new, innovative services based on available public sector data.
- However, many governmental agencies, including national statistical offices, have still not properly understood the intentions of the PSI directive. They often still prefer to develop and implement new services themselves, rather than leaving this to the entrepreneurs, working in free competition and on their own risk, without the budget limitations and other restrictions existing for government agencies.
- Another complication: entrepreneurs and citizens often want to combine data from many different producers, using different standards and procedures.

Project proposal

- Investigate the possibilities to define an easy-to-use, powerful, and flexible standard interface for communication of statistical data and metadata between the dissemination databases of statistics producers (statistical agencies) and citizen-centric applications, developed by entrepreneurs or by citizens themselves.
- The interface should fulfil the requirements stated above, in particular the requirements implied by the EU Public Sector Information (PSI) Directive.
- The investigations would be based upon, *inter alia*:
  - a preliminary investigation made by Björgvin Sigurdsson, Statistics Iceland, presented at a conference in Copenhagen in August 2010 (see reference)
  - a preliminary investigation made by Statistics Sweden and Gapminder Foundation in cooperation (see reference to come)
  - the PSI directive and related legislation in EU member countries (see reference)
  - requirements by citizen-centric systems and applications, as expressed, for example, by Hans Rosling and other entrepreneurs and creators of citzen-centric services
  - relevant existing and/or emerging standards, such as SDMX, www.sdmx.org, developed by a consortium of seven international organisations; however, these standards have not been developed with citizens and entrepreneurs in focus, and they have to be further developed for these purposes, especially as regards simplicity and richness of user-oriented metadata

A handshake at the Seminar on innovative approcaches to turning statistics into knowledge, organised in Stockholm 26-27 May 2008 by OECD and Statistics Sweden

www.oecd.org/oecdworldforum/statknowledge

Kjell Jansson, Director General of Statistics Sweden, 2006-2008:
"I agree with you, Hans. We will fulfil your wishes."

Professor Hans Rosling, Founder and Chairman of Gapminder Foundation, www.gapminder.org:
"We want all your data, objective, well documented, and with your quality stamp on them – in raw form – for us to analyse and present in ways that make them interesting for citizens, students, and decision-makers."

Other entrepreneurs and innovators associated with the project:
The factlab: Peter & Mike Andersson, http://www.thefactlab.com/
ncomva: Mikael Jern & collaborators, http://www.ncomva.com/
Google: Google Public Data Explorer, http://www.google.com/publicdata/admin
Gov2u: Government to you, an NGO, http://www.gov2u.org/
Izwe: Share opinions, make a difference, http://www.izwe.com/
Prognoz: Business intelligence – made simple, http://www.prognoz.com/en/index.php

Good intentions – so what is the problem?

- What is it – more precisely – that the entrepreneurs request: how should the data be delivered, which standards should be followed as regards contents and for, which metadata should accomany them, how should they be quality-declared?
- How should all these requirements be communicated and discussed between all those concerned, how should the contents-oriented and technical problems be solved, and how should an agreement be achieved and implemented?
- How should the producers be motivated to do their part of the job?

Motivation and communication

GOVERNMENT

ENTREPRENEURS

CITIZENS

MOTIVATION & COMMUNICATION

Top bureaucrats

"HACKERS"

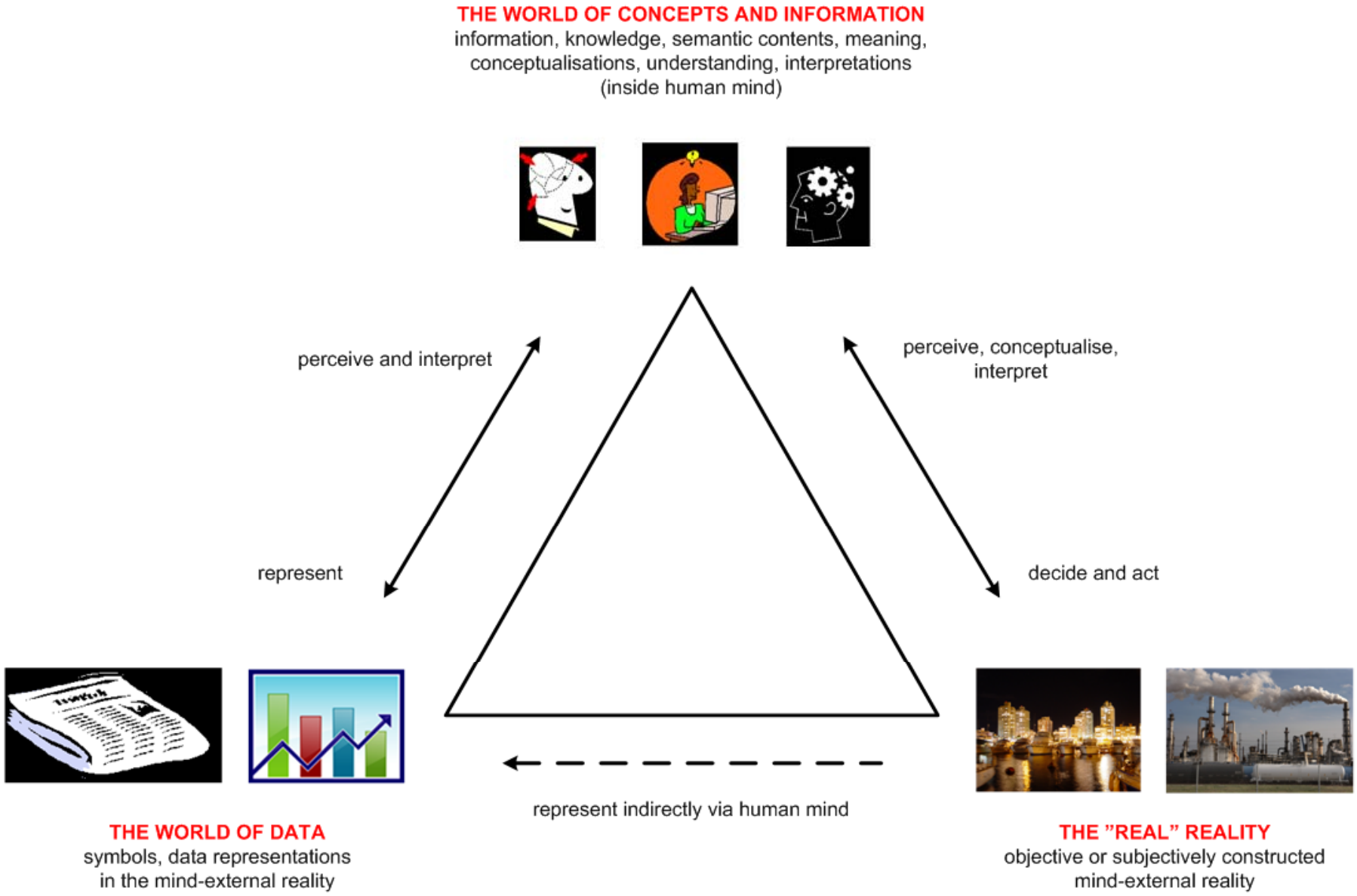
Professionals

IT specialists

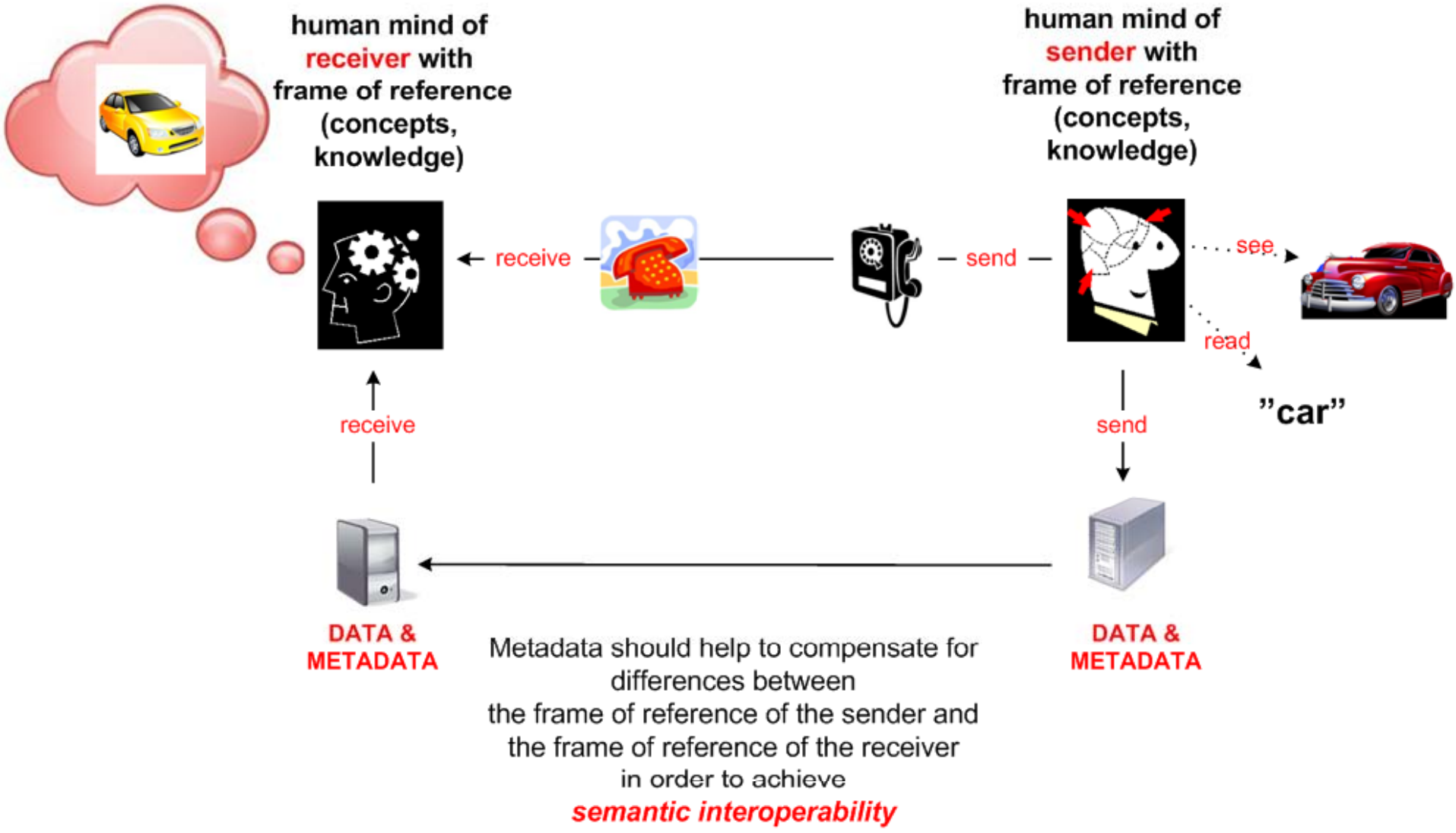
Bureaucrats

AGENCY

Information – data - reality



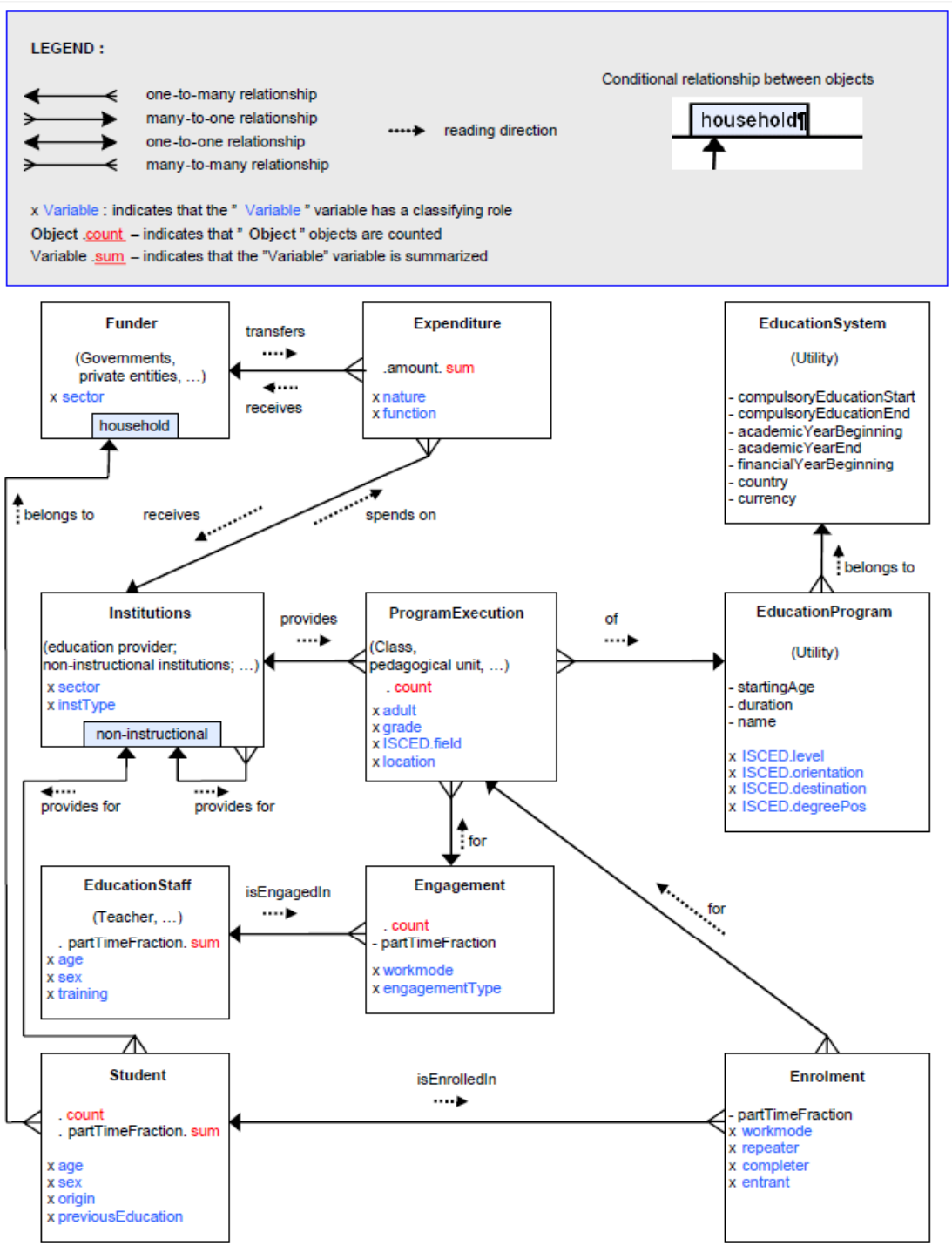
Semantic interoperability



Ontologies and conceptual models

- Short definition: “An ontology is a specification of a conceptualization” (Tom Gruber, 1992)
- An ontology defines a set of representational primitives with which to model a domain of knowledge or discourse. The representational primitives are typically classes (or sets), attributes (or properties), and relationships (or relations among class members).
- The definitions of the representational primitives include information about their meaning and constraints on their logically consistent application.
- In the context of database systems, ontology can be viewed as a level of abstraction of data models, analogous to hierarchical and relational models, but intended for modeling knowledge about individuals, their attributes, and their relationships to other individuals.
- Ontologies are typically specified in languages that allow abstraction away from data structures and implementation strategies; in practice, the languages of ontologies are closer in expressive power to first-order logic than languages used to model databases. For this reason, ontologies are said to be at the "semantic" level, whereas database schema are models of data at the "logical" or "physical" level.
- Due to their independence from lower level data models, ontologies are used for integrating heterogeneous databases, enabling interoperability among disparate systems, and specifying interfaces to independent, knowledge-based services.
- In the technology stack of the Semantic Web standards, ontologies are called out as an explicit layer.
- There are now standard languages and a variety of commercial and open source tools for creating and working with ontologies.

  (Tom Gruber, 2009)

Domain of discourse: Education (UNESCO)

ACTOR	COMPLEX OBJECT	UTILITY
• person • organisation	• event • transaction • relationship • activity	• thing • asset • resource • good / evil