

The Enterprise Ontology

by Dave McComb

At the time of this writing almost no enterprises in North America have a formal enterprise ontology. Yet we believe that within a few years this will become one of the foundational pieces to most information system work within major enterprises. In this paper, we will explain just what an enterprise ontology is, and more importantly, what you can expect to use it for and what you should be looking for, to distinguish a good ontology from a merely adequate one.

What is an ontology?

An ontology is a “specification of a conceptualization.” This definition is a mouthful but bear with me, it's actually pretty useful. In general terms, an ontology is an organization of a body of knowledge or, at least, an organization of a set of terms related to a body of knowledge. However, unlike a glossary or dictionary, which takes terms and provides definitions for them, an ontology works in the other direction. An ontology starts with a concept. We first have to find a concept that is important to the enterprise; and having found the concept, we need to express it in as precise a manner as possible and in a manner that can be interpreted and used by other computer systems. One of the differences between a dictionary or a glossary and ontology is, as we know, dictionary definitions are not really processable by computer systems. But the other difference is that by starting with the concept and specifying it as rigorously as possible, we get definitive meaning that is largely independent of language or terminology. Then the definition states that an ontology is a “specification of a conceptualization.” That is what we just described. In addition, of course, we then attach terms to these concepts, because in order for us humans to use the ontology we need to associate the terms that we commonly use.

Why is this useful to an enterprise?

Enterprises process great amounts of information. Some of this information is structured in databases, some of it is unstructured in documents or semi structured in content management systems. However, almost all of it is “local knowledge” in that its meaning is agreed within a relatively small, local context. Usually, that context is an individual application, which may have been purchased or may have been built in-house.

One of the most time- and money-consuming activities that enterprise information professionals perform is to integrate information from disparate applications. The reason this typically costs a lot of money and takes a lot of time is not because the information is on different platforms or in different formats – these are very easy to accommodate. The expense is because of subtle, semantic differences between the applications. In some cases, the differences are simple: the same thing is given different names in different systems. However, in many cases, the differences are much more subtle. The customer in one system may have an 80 or 90% overlap with the definition of a customer in another system, but it's the 10 or 20% where the definition is not the same that causes most of the confusion; and there are many, many terms that are far harder to reconcile than “customer.”

So the intent of the enterprise ontology is to provide a “lingua franca” to allow, initially, all the systems within an enterprise to talk to each other and, eventually, for the enterprise to talk to its trading partners and the rest of the world.

Isn't this just a corporate data dictionary or consortia of data standards?

The enterprise ontology does have many similarities in scope to both a corporate data dictionary and consortia data standard. The similarity is primarily in the scope of the effort: both of those initiatives, as well as enterprise ontologies, aim to define the shared terms that an enterprise uses. The difference is in the approach and the tools. With both a corporate data dictionary and a consortia data standard the interpretation and use of the definitions is strictly by humans, primarily system designers. Within an enterprise ontology, the expression of the ontology is such that tools are able to interpret and make inferences on the information when the system is running.

How to build an enterprise ontology

The task of building an enterprise ontology is relatively straightforward. You would be greatly aided by purchasing a good ontology editor, although reasonable ontology editors are available for free. The analytical work is similar to building a conceptual enterprise data model and involves many of the same skills: the ability to form good abstractions, to elicit information from users through interviews, as well as to find informational clues through existing documentation and data. One of the interesting differences is that as the ontology is

being built it can be used in connection with data profiling to see whether the information that is currently being stored in information systems does in fact comply with the rules that the ontology would suggest.

What to look for in an enterprise ontology

What distinguishes a good or great enterprise ontology from a merely adequate one are several characteristics that will mostly be exercised later in the lifecycle of the actual use of the ontology. Of course, they are important to consider at the time you're building the ontology.

Expressiveness

The ontology needs to be expressive enough to describe all the distinctions that an enterprise makes. Most enterprises of any size at all have tens of thousands to hundreds of thousands of distinctions that they use in their information systems. Not only is each piece of schemata in all of their databases a distinction but so are many of the codes they have in code tables as well as decisions that are called out either in code or in procedure manuals. The sum total of all these distinctions is the operating ontology of the enterprise. However, they are not formally expressed in one place. The structure as well as the base concepts used need to be rich enough that when a new concept is uncovered it can be expressed in the ontology.

Elegance

At the same time, we need to strive for an elegant representation. It would be simple but perhaps simplistic to take all the distinctions in all the current systems and put them in a simple repository and call them an ontology. This misses some of the great strengths of an ontology. We want to use our ontology not only to document and describe distinctions but also to find similarities. In these days of Sarbanes-Oxley regulations it would be incredibly helpful to know which distinctions and which parts of which schemas deal with financial commitments and "material transactions."

Inclusion and exclusion criteria

Essentially, the ontology is describing distinctions amongst "types." In many cases, what we would like to know is whether a given instance is of a particular type. Let's say it's a record in a product table, therefore it's a type "product." But in another system we may have inventory

and we would like to know whether this instance is also compatible with the type that we've defined as inventory. In order to do this, we need in the ontology a way to describe inclusion and exclusion criteria: what other clues we would use if we or another system were evaluating a particular instance to determine whether it was, in fact, of a particular type. For instance, if inventory were defined as being physical goods held for resale, one inclusion criteria might be weight because weight is an indicator of a physical good. Clearly, there would be many more, as well as criteria for excluding. But this gives you an idea.

Cross referencing capability

Another criterion that is very important is the ability to keep track of where the distinction was found; that is, which system currently implements and uses this particular distinction. This is very important for producing any type of where-used information because as we change our distinctions it might have side effects on other systems.

Inferencing

Inferencing is the ability to find or infer additional information based on the information we have. For instance, if we know that an entity is a person we can infer that the person has a birthday, whether we know it or not, and we can also infer that the person is less than 150 years old. While this sounds simple at this level, the power in an ontology is when the inference chains become long and complex and we can use the inferencing engine itself to make many of these conclusions on-the-fly.

Foreign-language support

As we described earlier, the ontology is a specification of a conceptualization that we attach terms to. It doesn't take much to add the ability to add foreign language terms.. This adds a great deal of power for developers who wish to present the same information, and the same screens, in multiple languages, as we are really just manipulating the concepts and attaching the appropriate language at runtime.

Some of these characteristics are aided by the existence of tools or infrastructures, but many of them are produced by the skill of the ontologist.

Summary

We believe that the enterprise ontology will become a cornerstone in many information systems in the future. It will become a primary part of the systems integration infrastructure as one application will be translated into the ontology and we will very rapidly know what the corresponding schema and terms are and what transformations are needed to get to another application. It will become part of the corporate search strategy as search moves beyond mere keywords into actually searching for meaning. It will become part of business intelligence and data warehousing systems as naïve users can be led to similar terms in the warehouse repository and aid their manual search and query construction.

Many more tools and infrastructures will become available over the next few years that will make use of the ontology, but the prudent information manager will not wait. He or she will recognize that there is a fair lead time to learn and implement something like this, and any implementation will be better than none because this particular technology promises to greatly leverage all the rest of the system technologies.