

# AN INTREPID GUIDE TO ONTOLOGIES

Michael K. Bergman<sup>1</sup>, Coralville, Iowa USA

May 16, 2007

*AI3::Adaptive Information* blog

*Ontology* is one of the more daunting terms for those exposed for the first time to the semantic Web. Not only is the word long and without many common antecedents, but it is also a term that has widely divergent use and understanding within the community. It can be argued that this not-so-little word is one of the barriers to mainstream understanding of the semantic Web.

The root of the term is the Greek *ontos*, or *being* or the *nature of things*. Literally — and in classical philosophy — ontology was used in relation to the study of the nature of being or the world, [the nature of existence](#). [Tom Gruber](#), among others, made the term popular in relation to [computer science](#) and artificial intelligence [about 15 years ago](#) when he defined ontology as a “formal specification of a conceptualization.”

While there have been attempts to strap on more or less formal understandings or machinery around ontology, it still has very much the sense of a world view, a means of viewing and organizing and conceptualizing and defining a domain of interest. As is made clear below, I personally prefer a loose and embracing understanding of the term (consistent with Deborah McGuinness’s 2003 paper, [Ontologies Come of Age](#) [1]).

There has been a resurgence of interest in ontologies of late. Two reasons have been the emergence of Web 2.0 and tagging and folksonomies, as well as the nascent emergence of the *structured Web*. In fact, on April 23-24 one of the noted communities of practice around ontologies, [Ontolog](#), sponsored the [Ontology Summit 2007](#), “*Ontology, Taxonomy, Folksonomy: Understanding the Distinctions*.”

These events have sparked my preparing this guide to ontologies. I have to admit this is a somewhat intrepid endeavor given the wealth of material and diversity of opinions.

## Overview and Role of Ontologies

Of course, a fancy name is not sufficient alone to warrant an interest in ontologies. There are reasons why understanding, using and manipulating ontologies can bring practical benefit:

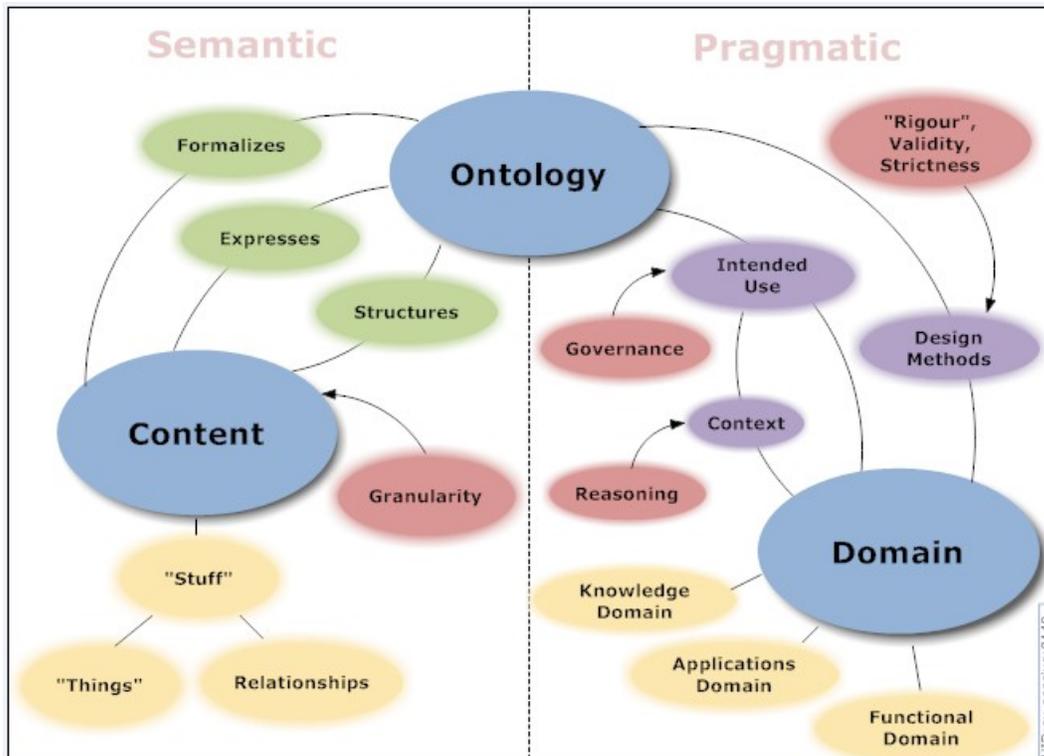
- Depending on their degree of *formalism* (an important dimension), ontologies help make explicit the scope, definition, and language and meaning (*semantics*) of a given domain or world view
- Ontologies may provide the power to generalize about their domains
- Ontologies, if hierarchically structured in part (and not all are), can provide the power of inheritance
- Ontologies provide guidance for how to correctly “place” information in relation to other information in that domain
- Ontologies may provide the basis to reason or infer over its domain (again as a function of its formalism)
- Ontologies can provide a more effective basis for information extraction or content clustering

---

<sup>1</sup>Email: [mike@mkbergman.com](mailto:mike@mkbergman.com)

- Ontologies, again depending on their formalism, may be a source of structure and controlled vocabularies helpful for disambiguating context; they can inform and provide structure to the “lexicons” in particular domains
- Ontologies can provide guiding structure for browsing or discovery within a domain, and
- Ontologies can help relate and “place” other ontologies or world views in relation to one another; in other words, ontologies can organize ontologies from the most specific to the most abstract.

Both structure and formalism are dimensions for classifying ontologies, which combined are often referred to as an ontology’s “expressiveness.” How one describes this structure and formality differs. One recent attempt is this figure from the [Ontology Summit 2007](#)’s wrap-up communique:



Note the bridging role that an ontology plays between a domain and its content. (By its nature, every ontology attempts to “define” and bound a domain.) Also note that the Summit’s 50 or so participants were focused on the trade-off between semantics v. pragmatic considerations. This was a result of the ongoing attempts within the community to understand, embrace and (possibly) legitimize “less formal” Web 2.0 efforts such as tagging and the folksonomies that can result from them.

There is an [M.C. Escher](#)-like recursion of the lizard eating its tail when one observes ontologists creating an ontology to describe the ontological domain. The above diagram, which itself would be different with a slight change in Summit participation or editorship, is, of course, but one representative view of the world. Indeed, a tremendous variety of scientific and research disciplines concern themselves with classifying and organizing the “nature of things.” Those disciplines go by such names as logicians, taxonomists, philosophers, information architects, computer scientists, librarians, operations researchers, systematicists, statisticians, historians, and so forth. (In short, given our *ontos*, every area of human endeavor has the urge to classify, to organize.) In each of these areas not only do their domains differ, but so do the adopted structures and classification schemes often used.

There are at least 40 terms or concepts across these various disciplines, most related to Web and general knowledge content, that have organizational or classificatory aspects that — loosely defined — could be called an “ontology” framework or approach:

- [Tag cloud](#)
- [Controlled vocabulary](#)
- [Thesauri](#)
- [Collaborative tagging](#)
- [Folk taxonomy](#)
- [Directory](#)
- [Subject Map](#)
- [Semantic Web](#)
- [Cladistics](#)
- [Markup languages](#)
- [Social bookmarking](#)
- [Tags](#)
- [Tagging](#)
- [Taxonomy](#)
- [Folksonomy](#)
- [Classification](#)
- [Categorization](#)
- [RDF](#)
- [Metadata](#)
- [Systematics](#)
- [Ontology](#)
- [Microformats](#)
- [Data dictionary](#)
- [OPML](#)
- [XOXO](#)
- [OWL](#)
- [Subject Trees](#)
- [Information Architecture](#)
- [Data Reference Model](#)
- [Phylogeny](#)
- [Topic Maps](#)
- [Concept Maps](#)
- [Synsets](#)
- [Glossary](#)
- [WordNet](#)
- [Metadata](#)
- [Facets](#)
- [Structure](#)
- [Dublin Core](#)
- [Typology](#)

Actual domains or subject coverage are then mostly orthogonal to these approaches.

Loosely defined, the number of possible ontologies is therefore close to infinite: *domain X perspective X schema*. (Just kidding — sort of! In fact, UMBC’s [Swoogle](#) ontology search service claims 10,000 ontologies presently on the Web; the actual data from August 2006 ranges from about [16,000 to 92,000 ontologies](#), depending on how “formal” the definition. These counts are also limited to OWL-based ontologies.)

Many have misunderstood the semantic Web because of this diversity and the slipperiness of the concept of an ontology. This misunderstanding becomes flat wrong when people claim the semantic Web implies one single grand ontology or organizational schema, *One Ring to Rule Them All*. Human and domain diversities makes this viewpoint patently false.

## Diversity, ‘Naturalness’ and Change

The choice of an ontological approach to organize Web and structured content can be contentious. Publishers and authors perhaps have too many choices: from straight Atom or RSS feeds and feeds with tags to informal folksonomies and then Outline Processor Markup Language or microformats. From there, the formalism increases further to include the standard RDF ontologies such as SIOC (Semantically-Interlinked Online Communities), SKOS (Simple Knowledge Organizing System), DOAP (Description of a Project), and FOAF (Friend of a Friend) and the still greater formalism of OWL’s various dialects.

Arguing which of these is the theoretical best method is doomed to failure, except possibly in a bounded enterprise environment. We live in the real world, where multiple options will always have their advocates and their applications. All of us should welcome whatever structure we can add to our information base, no matter where it comes from or how it’s done. The sooner we can embrace content in any of these formats and convert it to a canonical form, we can then move on to needed developments in semantic mediation, the threshold condition for the semantic Web.

So, diversity is inevitable and should be accepted. But that observation need not also embrace chaos.

There are at least 40 concepts — loosely defined — that could be called an “ontology” framework or approach.

In my early training in biological systematics, [Ernst Haeckel’s recapitulation theory](#) that “ontogeny recapitulates phylogeny” (note the same *ontos* root, the difference from ontology being growth v. study) was losing favor fast. The theory was that the development of an organism through its embryological phases mirrors its evolutionary history. Today, modern biologists recognize numerous connections between ontogeny and phylogeny, explain them using [evolutionary theory](#), or view them as supporting evidence for that theory.

Yet, like the construction of phylogenetic trees, systematicists strive for their classifications of the relatedness of organisms to be “natural”, to reflect the true *nature* of the relationship. Thus, over time, that understanding of a “natural” system has progressed from appearance → embryology → embryology + detailed morphology → species and interbreeding → DNA. While details continue to be worked out, the degree of genetic relatedness is now widely accepted by biologists as a “natural” basis for organizing the [Tree of Life](#).

It is not unrealistic to also seek “naturalness” in the organization of other knowledge domains, to seek “naturalness” in the organization of their underlying ontologies. Like natural systems in biology, this naturalness should emerge from the shared understandings and perceptions of the domain’s participants. While subject matter expertise and general and domain knowledge are essential to this development, they are not the only factors. As tagging systems on the Web are showing, common usage and broad acceptance by the community at hand is important as well.

While it may appear that a domain such as the biological relatedness of organisms is more empirical than the concepts and ambiguous words in most domains of human endeavor, these attempts at naturalness are still not foolish. The phylogeny example shows that understanding changes over time as knowledge is gained. We now accept DNA over the recapitulation theory.

As the formal SKOS organizational schema for knowledge systems recognizes (see below), the ideas of narrower and broader concepts can be readily embraced, as well as concepts of relatedness and aliases (synonyms). These simple constructs, I would argue, plus the application of knowledge being gained in related domains, will enable tomorrow’s understandings to be more “natural” than today’s, no matter the particular domain at hand.

So, in seeking a “naturalness” within our organizational schema, we can also see that change is a constant. We also see that the tools and ideas underlying the seemingly abstract cause of merging and relating existing ontologies to one another will further a greater “naturalness” within our organizations of the world.

## A Spectrum of Formalisms

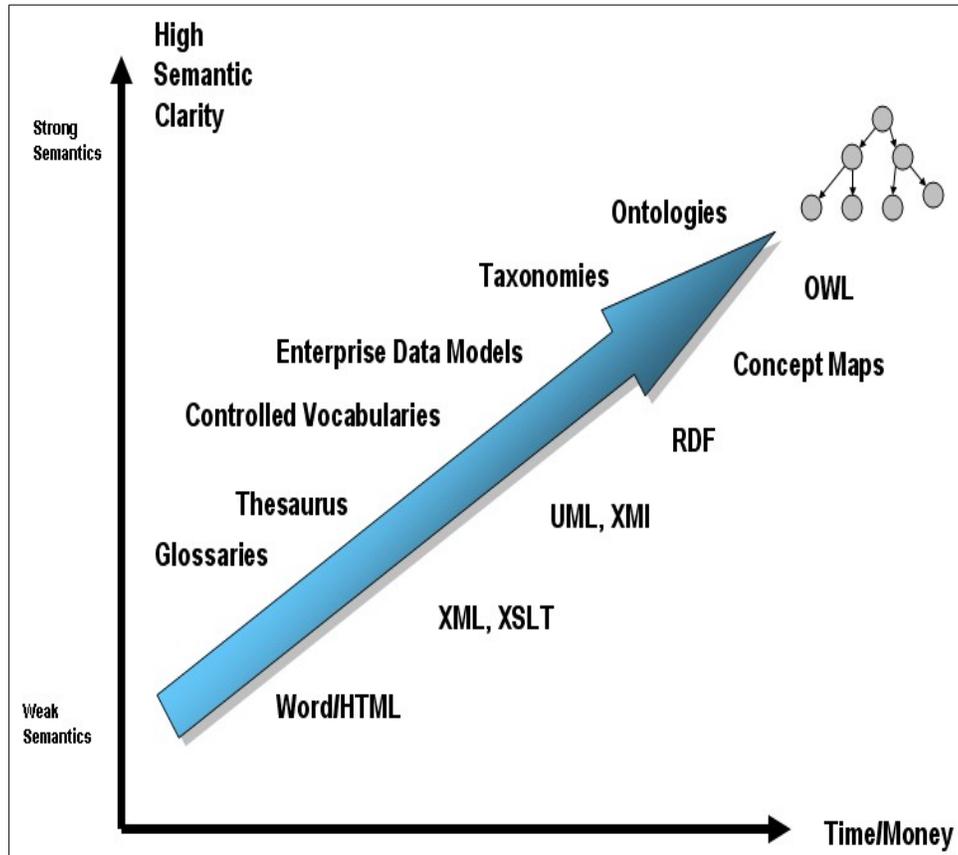
According to the Summit, *expressiveness* is the extent and ease by which an ontology can describe domain semantics. *Structure* they define as the degree of organization or hierarchical extent of the ontology. They further define *granularity* as the level of detail in the ontology. And, as the diagram above alludes, they define other dimensions of use, logical basis, purpose and so forth of an ontology.

The over fifty respondents from 42 communities submitted some 70 different ontologies under about 40 terms to a survey that was used by the Summit to construct their diagram. These submissions included:

*... formal ontologies (e.g., [BFO](#), [DOLCE](#), [SUMO](#)), biomedical ontologies (e.g., *Gene Ontology*, *SNOMED CT*, *UMLS*, *ICD*), thesauri (e.g., [MeSH](#), *National Agricultural Library Thesaurus*), folksonomies (e.g., *Social bookmarking tags*), general ontologies ([WordNet](#), [OpenCyc](#)) and specific ontologies (e.g., *Process Specification Language*). The list also includes markup languages (e.g., [NeuroML](#)), representation formalisms (e.g., *Entity-Relation model*, *OWL*, *WSDL-S*) and various ISO standards (e.g., *ISO 11179*). This [Ontolog] sample clearly illustrates the diversity of artifacts collected under “ontology”.*

I think the simplest spectrum for such distinctions is the **formalism** of the ontology and its approach (and language or syntax, not further discussed here). More formal ontologies have greater expressiveness and structure and inferential power, less formal ones the opposite. Constructing more formal ontologies is more demanding, and takes more effort and rigor, resulting in an approach that is more powerful but also more rigid and less flexible. Like anything else, there are always trade-offs.

Based on work by Leo Obrst of Mitre as interpreted by Dan McCreary, we can view this as a trade-off as one of semantic clarity v. the time and money required to construct the formalism [[12](#), [13](#)]:



[Click on image for full-size pop-up]

Note this diagram reflects the more conventional, practitioner’s view of the “formal” ontology, which does not include taxonomies or controlled vocabularies (for example) in the definition. This represents the more “closely defined” end of the [ontology \(semantic\) spectrum](#).

However, since we are speaking here of ontologies and the *structured Web* or the semantic Web, I believe we need to embrace a concept of ontology aligned to actual practice. Not all content providers can or want to employ ontology engineers to enable formal inferencing of their content. Yet, on the other hand, their content in its various forms does have some meaningful structure, some organization. The trick is to extract this structure for more meaningful use such as data exchange or data merging.

### Ontology Approaches on the Web

Under such “loosely defined” bases we can thus see a spectrum of ontology approaches on the Web, proceeding from less structure and formalism to **more so**:

Type or Schema	Examples	Comments on Structure and Formalism
Standard Web Page	entire Web	General metadata fields in the and internal HTML codes and tags provide minimal, but useful sources of structure; other HTTP and retrieval data can also contribute
Blog / Wiki Page	examples from <a href="#">Technorati</a> , <a href="#">Bloglines</a> , <a href="#">Wikipedia</a>	Provides still greater formalism for the organization and characterization of content (subjects, categories, posts, comments, date/time stamps, etc.). Importantly, with the addition of plug-ins, some of the basic software may also provide other structured characterizations or output (SIOC,

Type or Schema	Examples	Comments on Structure and Formalism
RSS / Atom feeds	most blogs and most news feeds	FOAF, etc.; highly varied and site-specific given the diversity of publishing systems and plug-ins) RSS extends basic XML schema for more robust syndication of content with a tightly controlled vocabulary for feed concepts and their relationships. Because of its ubiquity, this is becoming a useful baseline of structure and formalism; also, the nature of adoption shows much about how ontological structure is an <i>artifact</i> , not <i>driver</i> , for use
RSS / Atom feeds with tags or OPML	<a href="#">Grazr</a> , most newsfeed aggregators can import and export OPML lists of RSS feeds	The OPML specification defines an outline as a hierarchical, ordered list of arbitrary elements. The specification is fairly open which makes it suitable for many types of list data. See also <a href="#">OML</a> and <a href="#">XOXO</a>
<a href="#">Hierarchical Faceted Metadata</a>	<a href="#">XFML</a> , <a href="#">Flamenco</a>	These and related efforts from the information architecture (IA) community are geared more to library science. However, they directly contribute to faceted browsing, which is one of the first practical instantiations of the semantic Web
<a href="#">Folksonomies</a>	<a href="#">Flickr</a> , <a href="#">del.icio.us</a>	Based on user-generated tags and informal organizations of the same; not linked to any “standard” Web protocols. Both tags and hierarchical structure are arbitrary, but some researchers now believe over large enough participant sets that structural consensus and value does emerge
<a href="#">Microformats</a>	Example formats include <a href="#">hAtom</a> , <a href="#">hCalendar</a> , <a href="#">hReview</a> , <a href="#">hResume</a> , <a href="#">rel-directory</a> , <a href="#">xFolk</a> , <a href="#">XFN</a> and <a href="#">XOXO</a>	A microformat is HTML mark up to express semantics with strictly controlled vocabularies. This markup is embedded using specific HTML attributes such as class, rel, and rev. This method is easy to implement and understand, but is not free-form
Embedded RDF	<a href="#">RDFa</a> , eRDF	An embedded format, like microformats, but free-form, and not subject to the approval strictures associated with microformats
<a href="#">Topic Maps</a>	<a href="#">Infoloom</a> , <a href="#">Topic Maps Search Engine</a>	A topic map can represent information using topics (representing any concept, from people, countries, and organizations to software modules, individual files, and events), associations (which represent the relationships between them), and occurrences (which represent relationships between topics and information resources relevant to them)
<a href="#">RDF</a>	Many; <a href="#">DBpedia</a> , etc.	RDF has become the canonical data model since it represents a “universal” conversion format
<a href="#">RDF Schema</a>	<a href="#">SKOS</a> , <a href="#">SIOC</a> , <a href="#">DOAP</a> , <a href="#">FOAF</a>	RDFS or RDF Schema is an extensible knowledge representation language, providing basic elements for the description of ontologies, otherwise called RDF vocabularies, intended to structure RDF resources. This becomes the canonical ontology common meeting ground
<a href="#">OWL Lite</a> <a href="#">OWL DL</a>	Here are some existing <a href="#">OWL ontologies</a> ; also see <a href="#">Swoogle</a> for OWL search facilities	The Web Ontology Language (OWL) is a language for defining and instantiating Web ontologies. An OWL ontology may include descriptions of classes, along with their related properties and instances. OWL is designed for use by applications that need to process the content of information instead of just presenting information to humans. It facilitates greater machine interpretability of Web content than that supported by XML, RDF, and RDF Schema (RDF-S) by providing additional vocabulary along with a formal semantics. The three language versions are in order of increasing expressiveness
<a href="#">OWL Full</a>		
Higher-order “formal”	<a href="#">SUMO</a> , <a href="#">DOLCE</a> , <a href="#">PROTON</a> , <a href="#">BFO</a> , <a href="#">Cyc</a>	These provide comprehensive ontologies and often related knowledge bases, with the goal of enabling AI applications to perform human-like

Type or Schema	Examples	Comments on Structure and Formalism
----------------	----------	-------------------------------------

“upper-level” ontologies	OpenCyc	reasoning. Their reasoning languages often use higher-order logics
--------------------------	---------	--

As a rule of thumb, items that are less “formal” can be converted to a more formal expression, but the most formal forms can generally not be expressed in less formal forms.

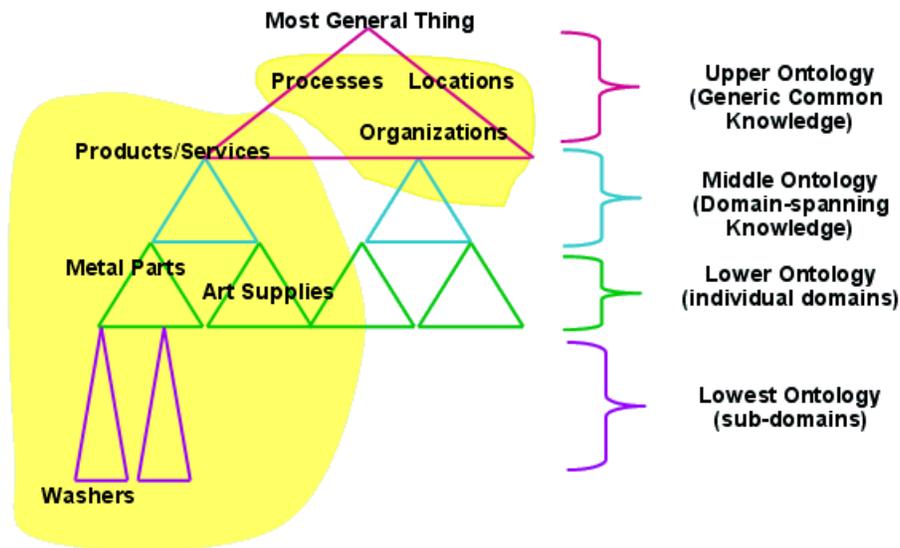
As latter sections elaborate, I see RDF as the universal data model for representing this structure into a common, canonical format, with RDF Schema (specifically SKOS, but also supplemented by FOAF, DOAP and SIOC) as the organizing ontology [knowledge representation language](#) (KRL).

This is not to say that the various dialects of OWL should be neglected. In bounded environments, they can provide superior reasoning power and are warranted if they can be sufficiently mandated or enforced. But the RDF and RDF-S systems represent the most tractable “meeting place” or “middle ground,” IMHO.

### Still-Another “Level” of Ontologies

As if the formalism dimension were not complicated enough, there is also the practice within the ontology community to characterize ontologies by “levels”, specifically *upper*, *middle* and *lower* levels. For example, chances are that you have heard particularly of [“upper-level” ontologies](#).

The following figure helps illustrate this “level” dimension. This diagram is also from Leo Obrst of Mitre [12], and was also used in another 2006 talk by [Jack Park and Patrick Durusau](#) (discussed further below for other reasons):



Examples of upper-level ontologies include the Suggested Upper Merged Ontology ([SUMO](#)), the Descriptive Ontology for Linguistic and Cognitive Engineering ([DOLCE](#)), [PROTON](#), [Cyc](#) and [BFO](#) (Basic Formal Ontology). Most of the content in their upper-levels is akin to broad, abstract relations or concepts (similar to the primary classes, for example, in a [Roget’s Thesaurus](#) — that is, real *ontos* stuff) than to “generic common knowledge.” Most all of them have both a hierarchical and networked structure, though their actual subject structure relating to concrete things is generally pretty weak [2].

The above diagram conveys a sense of how multiple ontologies can relate to one another both in terms of narrower and broader topic matter and at the same “levels” of generalization. Such “meta-structure” (if you will) can provide a reference structure for relating multiple ontologies to one another.

It resides exactly in such bindings or relationships that we can foresee the promise of querying and relating multiple endpoints on the Web with accurate semantics in order to connect dots and combine knowledge bases. Thus, the understanding of the relationships and mappings amongst ontologies becomes a critical infrastructural component of the semantic Web.

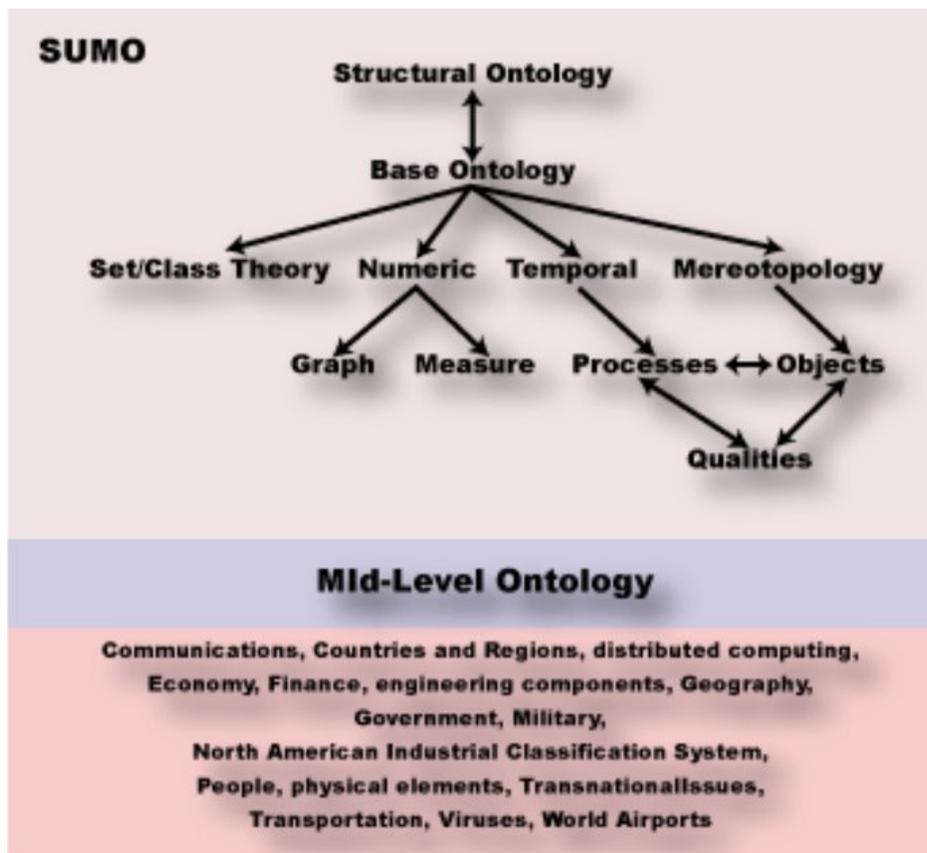
The relationships and mappings amongst ontologies is a critical infrastructure component of the semantic Web.

## The SUMO Example

We can better understand these mapping and inter-relationship concepts by using a concrete example with a formal ontology. We’ll choose to use the [Suggested Upper Merged Ontology](#) simply because it is one of the best known. We could have also selected another upper-level system such as PROTON [3] or Cyc [4] or one of the many with narrower concept or subject coverage.

SUMO is one of the formal ontologies that has been mapped to the [WordNet](#) lexicon, which adds to its semantic richness. SUMO is written in the [SUO-KIF](#) language. SUMO is free and owned by the [IEEE](#). The ontologies that extend SUMO are available under [GNU General Public License](#).

The abstract, conceptual organization of SUMO is shown by this diagram, which also points to its related [MILO \(Mid-Level Ontology\)](#), which is being developed as a bridge between the abstract content of the SUMO and the richer detail of various domain ontologies:





These views may be a bit misleading because the actual underlying structure, while it has hierarchical aspects as shown here, really is in the form of a directed acyclic graph (showing other relatedness options, not just hierarchical ones). So, alternate visualizations include traditional network graphs.

The other thing to note is that the “things” covered in the ontology, the entities, are also fairly abstract. That is because the intention of a standard “upper-level” ontology is to cover all relevant knowledge aspects of each entity’s domain. This approach results in a subject and topic coverage that feels less “concrete” than the coverage in, say, an encyclopedia, directory or card catalog.

## Ontology Binding and Integration Mechanisms

According to Park and Durusau, upper ontologies are diverse, middle ontologies are even more diverse, and lower ontologies are more diverse still. A key observation is that ontological diversity is a given and increases as we approach real user interaction levels. Moreover, because of the “loose” nature of ontologies on the Web (now and into the future), diversity of approach is a further key factor.

Recall the initial discussion on the role and objectives of ontologies. About half of those roles involve effectively accessing or querying more than one ontology. The objective of “upper-level” ontologies, many with their own binding layers, is also expressly geared to ontology integration or federation. So, what are the possible mechanisms for such binding or integration?

A fundamental distinction within mechanisms to combine ontologies is whether it is a unified or centralized approach (often imposed or required by some party) or whether it is a schema mapping or binding approach. We can term this distinction centralized v. federated.

### Centralized Approaches

Centralized approaches can take a number of forms. At the most extreme, adherence to a centralized approach can be contractual. At the other end are reference models or standards. For example, illustrative reference models include:

- the [Data Reference Model \(DRM\)](#), one of the five reference models of the Federal Enterprise Architecture (FEA)
- [UDEF](#) (Unified Data Element Framework), an approach toward a unified description framework, or
- the eXtended MetaData Registry ([XMDR](#)) project.

Though I have argued that *One Ring to Rule them All* is not appropriate to the general Web, there may be cases within certain enterprises or where through funding clout (such as government contracts), some form of centralized approach could be imposed [5]. And, frankly, even where compliance can not be assured, there are advantages in economy, efficiency and interoperability to attempt central ontologies. Certain industries — notably pharmaceuticals and petrochemicals — and certain disciplines — such as some areas of biology among others — have through trade associations or community consensus done admirable jobs in adopting centralized approaches.

### Federated Approaches

However, combining ontologies in the context of the broader Internet is more likely through federated approaches. (Though federated approaches can also be improved when there are consensual standards within specific communities.) The key aspect of a federated approach is to acknowledge that multiple schema need to be brought together, and that each contributing data set and its schema will not be altered directly and will likely remain in place.

Thus, the key distinctions within this category are the mechanisms by which those linkages may take place. An important goal in any federated approach is to achieve interoperability at the data or instance level without unacceptable loss of information or corruption of the semantics. Numerous specific approaches are possible, but three example areas in RDF-topic map interoperability, the use of “subject maps”, and binding layers can illustrate some of the issues at hand.

In 2006 the W3C set up a working group to look at the issue of RDF and topic maps interoperability. Topic maps have been embraced by the library and information architecture community for some time, and have standards that have been adopted under ISO. Somewhat later but also in parallel was the development of the RDF standard by W3C. The interesting thing was that the conceptual underpinnings and objectives between these two efforts were quite similar. Also, because of the substantive thrust of topic maps and the substantive needs of its community, quite a few topic maps had been developed and implemented.

One of the first efforts of the W3C work group was to evaluate and compare five or six extant proposals for how to relate RDF and topic maps [6]. That report is very interesting reading for any one desirous of learning more about specific issues in combining ontologies and their interoperability. The result of that evaluation then led to some guidelines for best practices in how to complete this mapping [7]. Evaluations such as these provide confidence that interoperability can be achieved between relatively formal schema definitions without unacceptable loss in meaning.

A different, “looser” approach, but one which also grew out of the topic map community, is the idea of “subject maps.” This effort, backed by Park and Durusau noted above, but also with the support of other topic map experts such as Steve Newcomb and Robert Barta via their proposed [Topic Maps Reference Model](#) (ISO 13250-5), seems to be one of the best attempts I’ve seen that both respects the reality of the actual Web while proposing a workable, effective scheme for federation.

The basic idea of a subject map is built around a set of subject “proxies.” A subject proxy is a computer representation of a subject that can be implemented as an object, must have an identity, and must be addressable (this point provides the URI connector to RDF). Each contributing schema thus defines its own subjects, with the mappings becoming meta-objects. These, in turn, would benefit from having some accepted subject reference schema (not specifically addressed by the proponents) to reduce the breadth of the ultimate mapped proxy “space.”

I don’t have the expertise to judge further the specifics, but I find the presentation and papers by Park and Durusau, [Avoiding Hobson’s Choice In Choosing An Ontology](#) and [Towards Subject-centric Merging of Ontologies](#) to be worthwhile reading in any case. I highly recommend these papers for further background and clarity.

As the third example, “binding layers” are a comparatively newer concept. Leading upper-level ontologies such as SUMO or PROTON propose their own binding protocols to their “lower” domains, but that approach takes place within the construct of the parent upper ontology and language. Such designs are not yet generalized solutions. By far the most promising generalized binding solution is the [SKOS](#) (Simple Knowledge Organization System). Because of its importance, the next section is devoted to it.

Finally, with respect to federated approaches, there are quite a few software tools that have been developed to aid or promote some of these specific methods. For, example, about twenty of the software applications in my [Sweet Tools](#) listing of 500+ semantic Web and -related tools could be interpreted as aiding ontology mapping or creation. You may want to check out some of these specific tools depending on your preferred approach [8].

## **The Role of SKOS – the Simple Knowledge Organization System**

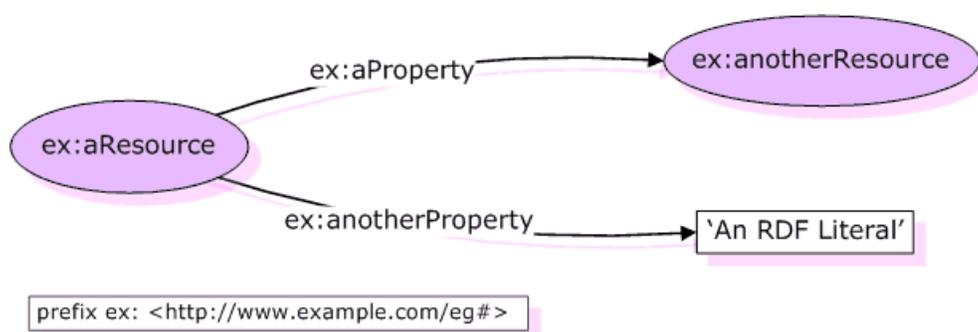
SKOS, or the Simple Knowledge Organization System, is a formal language and schema designed to represent such structured information domains as [thesauri](#), [classification schemes](#), [taxonomies](#), [subject-heading](#)

[systems](#), [controlled vocabularies](#), or others; in short, most all of the “loosely defined” ontology approaches discussed herein. It is a [W3C](#) initiative more fully defined in its [SKOS Core Guide](#) [9].

SKOS is built upon the [RDF](#) data model of the *subject–predicate–object* “triple.” The *subjects* and *objects* are akin to nouns, the *predicate* a verb, in a simple Dick-sees-Jane sentence. *Subjects* and *predicates* by convention are related to a [URI](#) that provides the definitive reference to the item. *Objects* may be either a URI resource or a literal (in which case it might be some indexed text, an actual image, number to be used in a calculation, etc.).

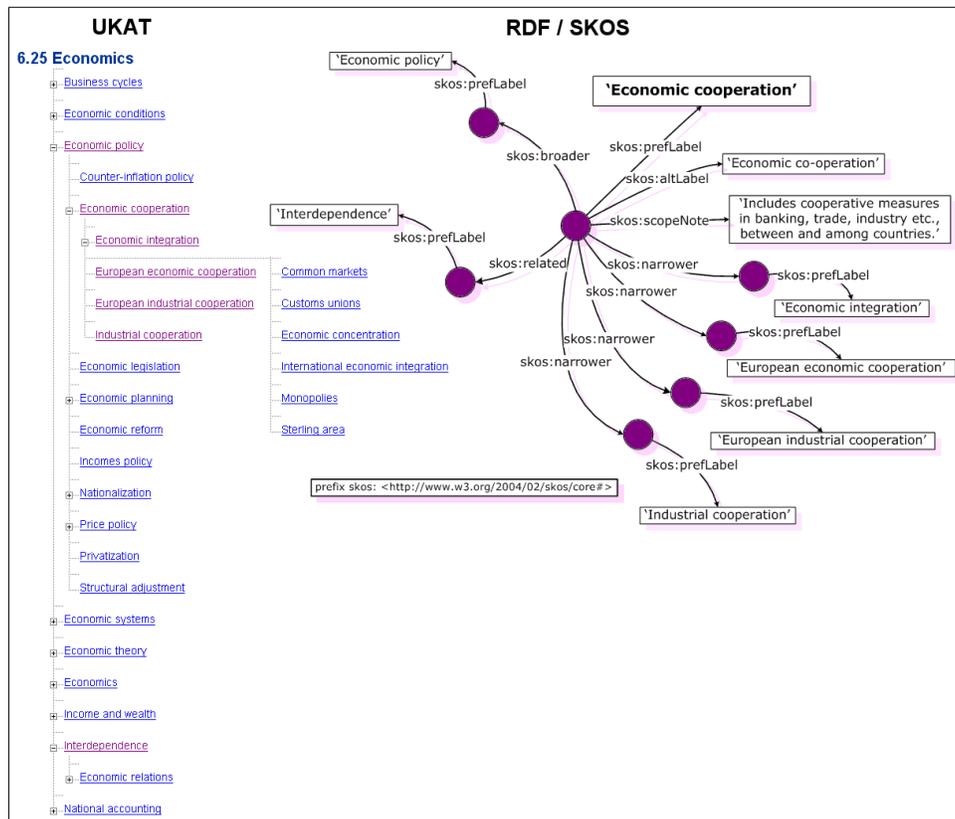
Being an [RDF Schema](#) simply means that SKOS adds some language and defined relationships to this RDF baseline. This is a bit of recursive understanding, since RDFS is itself defined in RDF by virtue of adding some controlled vocabulary and relations. The power, though, is that these schema additions are also easily expressed and referenced.

This RDFS combination can thus be shown as a standard RDF triple graph, but with the addition of the extended vocabulary and relations:



The power of the approach arises from the ability of the triple to express virtually any concept, further extended via the RDFS language defined for SKOS. SKOS includes concepts such as “broader” and “narrower”, which enable hierarchical relations to be modeled, as well as “related” and “member” to support networks and arrays, respectively [9].

We can visualize this transforming power by looking at how an “ontology” in a totally foreign scheme can be related to the canonical SKOS scheme. In the figure below the left-hand portion shows the native hierarchical taxonomy structure of the UK Archival Thesaurus ([UKAT](#)), next as converted to SKOS on the right (with the overlap of categories shown in dark purple). Note the hierarchical relationships visualize better via a taxonomy, but that the RDF graph model used by SKOS allows a richer set of additional relationships including related and alternative names:



[Click on image for full-size pop-up]

SKOS also has a rich set of annotation and labeling properties to enhance human readability of schema developed in it [9]. There is also a useful draft schema that the W3C's SWEO (Semantic Web Education and Outreach) group is developing to organize semantic Web-related information [10].

Combined, these constructs provide powerful mechanisms for giving contributory ontologies a common conceptualization. When added to other sibling RDF schema such as FOAF or SIOC or DOAP, still additional concepts can be collated.

## Conclusions

While not addressed directly in this piece, it is obviously of first importance to have content with structure before the questions of connecting that information can even arise. Then, that structure must also be available in a form suitable for merging or connection.

At that point, the subjects of this posting come into play.

We are stubbing our toes on the rocks while we gaze at the heavens.

We see that the daily Web has a diversity of schema or ontologies “loosely defined” for representing the structure of the content. These representations can be transferred to more complex schema, but not in the opposite direction. Moreover, the semantic basis for how to make these mappings also needs some common referents.

RDF provides the canonical data model for the data transfers and representations. RDFS, especially in the form of SKOS, appears to form one basis for the syntax and language for these transformations. And SKOS,

with other schema, also appears to offer much of the appropriate “middle ground” for data relationships mapping.

However, lacking in this story is a referential structure for subject relationships [11]. (Also lacking are the ultimately critical domain specifics required for actual implementation.)

Abstract concepts of interest to philosophers and deep thinkers have been given much attention. Sadly, to date, concrete subject structures in which tangible things and tangible actions can be shared, is still very, very weak. We are stubbing our toes on the rocks while we gaze at the heavens.

Yet, despite this, simple and powerful infrastructures are well in-hand to address all foreseeable syntactic and semantic issues. There appear to be no substantive limits to needed next steps.

Lastly, many valuable resources for further reading and learning may be found within the [Ontolog Community](#), [W3C](#), [TagCommons](#) and [Topics Maps](#) groups. Enjoy! And be wary of *ontology* no longer.

## Acknowledgements

This article was originally posted on the *AI3::Adaptive Information* Web site at <http://www.mkbergman.com/374/an-intrepid-guide-to-ontologies/>. This version has been edited and reformatted slightly for PDF distribution. We thank Cognonto Corporation for making this content freely available.

- 
- [1] Deborah L. McGuinness. “Ontologies Come of Age”. In Dieter Fensel, Jim Hendler, Henry Lieberman, and Wolfgang Wahlster, editors. *Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential*. MIT Press, 2003. See [http://www.ksl.stanford.edu/people/dlm/papers/ontologies-come-of-age-mit-press-\(with-citation\).htm](http://www.ksl.stanford.edu/people/dlm/papers/ontologies-come-of-age-mit-press-(with-citation).htm)
- [2] I think it would be much clearer to refer to “upper level” ontologies as *abstract* or *conceptual*, “mid levels” as *mapping* or *binding*, and “lower levels” as *domain* (without any hierarchical distinctions such as *lower* or *lowest* or *sub-domain*), but current practice is probably too entrenched to change now.
- [3] There are many aspects that make [PROTON](#) one of the more attractive reference ontologies. The PROTON ontology (PROTo ONtology), developed within the scope of the [SEKT project](#), is attractive because of its understandability, relatively small size, modular architecture and a simple subsumption hierarchy. It is available in an OWL Lite form and is easy to adopt and extend. On the face of it, the Topic class within PROTON, which is meant to serve as a bridge between different ontologies, may also provide a binding layer to specific subject topics as sub-classes or class instances.
- [4] See my [earlier post on Cyc](#).
- [5] Even with such clout, it is questionable to get rather complete adherence, as [Ada](#) showed within the Federal government. However, where circumstances allow it, central schema and ontologies may be worth pursuing because of improved interoperability and lower costs, even where some portions do not adhere and are more chaotic like the standard Web.
- [6] See, [A Survey of RDF/Topic Maps Interoperability Proposals](#), W3C Working Group Note 10 February 2006, Pepper, Vitali, Garshol, Gessa, Presutti (eds.)
- [7] See, [Guidelines for RDF/Topic Maps Interoperability](#), W3C Editor’s Draft 30 June 2006, Pepper, Presutti, Garshol, Vitali (eds.)
- [8] Here are some [Sweet Tools](#) that may have a usefulness to ontology federation and creation:
- [Adaptiva](#) — is a user-centered ontology building environment, based on using multiple strategies to construct an ontology, minimising user input by using adaptive information extraction
  - [Altova SemanticWorks](#) — is a visual RDF and OWL editor that auto-generates RDF/XML or nTriples based on visual ontology design
  - [CMS](#) — the CROSI Mapping System is a structure matching system that capitalizes on the rich semantics of the OWL constructs found in source ontologies and on its modular architecture that allows the system to consult external linguistic resources
  - [ConceptTool](#) — is a system to model, analyze, verify, validate, share, combine, and reuse domain knowledge bases and ontologies, reasoning about their implication
  - [ConRef](#) — is a service discovery system which uses ontology mapping techniques to support different user vocabularies

- [FOAM](#) — is the Framework for Ontology Alignment and Mapping. It is based on heuristics (similarity) of the individual entities (concepts, relations, and instances)
- [hMAFRA](#) (Harmonize Mapping Framework) — is a set of tools supporting semantic mapping definition and data reconciliation between ontologies. The targeted formats are XSD, RDFS and KAON
- [IF-Map](#) — is an Information Flow based ontology mapping method. It is based on the theoretical grounds of logic of distributed systems and provides an automated streamlined process for generating mappings between ontologies of the same domain
- [IODT](#) — is IBM’s toolkit for ontology-driven development. The toolkit includes EMF Ontology Definition Metamodel (EODM), EODM workbench, and an OWL Ontology Repository (named Minerva)
- [KAON](#) — is an open-source ontology management infrastructure targeted for business applications. It includes a comprehensive tool suite allowing easy ontology creation and management and provides a framework for building ontology-based applications. An important focus of KAON is scalable and efficient reasoning with ontologies
- [LinKFactory](#) — is Language & Computing’s ontology management tool. It provides an effective and user-friendly way to create, maintain and extend extensive multilingual terminology systems and ontologies (English, Spanish, French, etc.). It is designed to build, manage and maintain large, complex, language independent ontologies
- [M3t4.Studio Semantic Toolkit](#) — is Metatomix’s free set of Eclipse plug-ins to allow developers to create and manage OWL ontologies and RDF documents
- [MAFRA Toolkit](#) — the Ontology MAPPING FRamework Toolkit allows to create semantic relations between two (source and target) ontologies, and apply such relations in translating source ontology instances into target ontology instances
- [OntoEngine](#) — is a step toward allowing agents to communicate even though they use different formal languages (i.e., different ontologies). It translates data from a “source” ontology to a “target.”
- [OntoPortal](#) — enables the authoring and navigation of large semantically-powered portals
- [OWLS-MX](#) — the hybrid semantic Web service matchmaker OWLS-MX 1.0 utilizes both description logic reasoning, and token based IR similarity measures. It applies different filters to retrieve OWL-S services that are most relevant to a given query
- [pOWL](#) — is a semantic Web development platform for ontologies in PHP. pOWL consists of a number of components, including RAP
- [Protege](#) — is an open source visual ontology editor written in Java with many plug-in tools
- [Semantic Net Generator](#) — is a utility for generating topic maps automatically from different data sources by using rules definitions specified with Jelly XML syntax. This Java library provides Jelly tags to access and modify data sources (also RDF) to create a semantic network
- [SOFA](#) — is a Java API for modeling ontologies and Knowledge Bases in ontology and Semantic Web applications. It provides a simple, abstract and language neutral ontology object model, inferencing mechanism and representation of the model with OWL, DAML+OIL and RDFS languages
- [Terminator](#) — is a tool for creating term to ontology resource mappings (documentation in Finnish)
- [WebOnto](#) — supports the browsing, creation and editing of ontologies through coarse grained and fine grained visualizations and direct manipulation.

[9] The SKOS language has the following classes:

- [CollectableProperty](#) — A property which can be used with a skos:Collection
- [Collection](#) — A meaningful collection of concepts
- [Concept](#) — An abstract idea or notion; a unit of thought
- [ConceptScheme](#) — A set of concepts, optionally including statements about semantic relationships between those concepts. Thesauri, classification schemes, subject heading lists, taxonomies, ‘folksonomies’, and other types of controlled vocabulary are all examples of concept schemes. Concept schemes are also embedded in glossaries and terminologies.
- [OrderedCollection](#) — An ordered collection of concepts, where both the grouping and the ordering are meaningful . . . and the following properties:
- [altLabel](#) — An alternative lexical label for a resource. Acronyms, abbreviations, spelling variants, and irregular plural/singular forms may be included among the alternative labels for a concept
- [altSymbol](#) — An alternative symbolic label for a resource
- [broader](#) — A concept that is more general in meaning. Broader concepts are typically rendered as parents in a concept hierarchy (tree)
- [changeNote](#) — A note about a modification to a concept

- [definition](#) — A statement or formal explanation of the meaning of a concept
- [editorialNote](#) — A note for an editor, translator or maintainer of the vocabulary
- [example](#) — An example of the use of a concept
- [hasTopConcept](#) — A top level concept in the concept scheme
- [hiddenLabel](#) — A lexical label for a resource that should be hidden when generating visual displays of the resource, but should still be accessible to free text search operations
- [historyNote](#) — A note about the past state/use/meaning of a concept
- [inScheme](#) — A concept scheme in which the concept is included. A concept may be a member of more than one concept scheme
- [isPrimarySubjectOf](#) — A resource for which the concept is the primary subject
- [isSubjectOf](#) — A resource for which the concept is a subject
- [member](#) — A member of a collection
- [memberList](#) — An RDF list containing the members of an ordered collection
- [narrower](#) — A concept that is more specific in meaning. Narrower concepts are typically rendered as children in a concept hierarchy (tree)
- [note](#) — A general note, for any purpose. The other human-readable properties of definition, scopeNote, example, historyNote, editorialNote and changeNote are all sub-properties of note
- [prefLabel](#) — The preferred lexical label for a resource, in a given language. No two concepts in the same concept scheme may have the same value for skos:prefLabel in a given language
- [prefSymbol](#) — The preferred symbolic label for a resource
- [primarySubject](#) — A concept that is the primary subject of the resource. A resource may have only one primary subject per concept scheme
- [related](#) — A concept with which there is an associative semantic relationship
- [scopeNote](#) — A note that helps to clarify the meaning of a concept
- [semanticRelation](#) — A concept related by meaning. This property should not be used directly, but as a super-property for all properties denoting a relationship of meaning between concepts
- [subject](#) — A concept that is a subject of the resource
- [subjectIndicator](#) — A subject indicator for a concept. [The notion of ‘subject indicator’ is defined here with reference to the latest definition endorsed by the OASIS Published Subjects Technical Committee]
- [symbol](#) — An image that is a symbolic label for the resource. This property is roughly analagous to rdfs:label, but for labelling resources with images that have retrievable representations, rather than RDF literals. Symbolic labelling means labelling a concept with an image.

[10] The [SWEO classification ontology](#) is still under active development and has these draft classes. Note, however, the relative lack of actual subject or topic matter:

Classes are currently defined as:

- article – magazine article
- blog – blog discussing SW topics
- book – indicates a textbook, applies to the book’s home page, review or listing in Amazon or such
- casestudy – Article on a business case
- conference/event – conferences or events where you can learn about the Semantic Web
- demo/demonstration – interactive SW demo
- forum – a forum on semantic web or related topics
- presentation – Powerpoint or similar slide show
- person – If this is a person’s home page or blog, see below
- publication – a scientific publication
- ontology – a formalisation of a shared conceptualization using OWL, RDFS, SKOS or something else based on RDF
- organization – If the page is the home page of an organization, research, vendor etc, see below
- portal – a portal website Semantic Web or related topics, usually hosting information items, mailinglists, community tools

- project – a research (for example EU-IST) or other project that addresses Semantic Web issues
- mailinglist – a mailinglist on semantic Web or related topics
- person – ideally a person that is well known regarding the Semantic Web (people who can do keynote speakers), may also be any related person
- press – a press release by a company or an article about Semantic Web
- recommended – If the resource is seen to be in the top 10 of its kind
- specification – a Semantic Web specification (RDF, RDF/S, OWL, etc)
- categories – (perhaps using tags or other free form annotation)
- successstory – Article that can contain advertisement and clearly shows the benefit of semantic web
- tutorial – a tutorial teaching some aspect of semantic web, an example
- vocabulary – a RDF vocabulary
- software project/tool – For product/project home pages

If the page describes an organization, it can be tagged as:

- vendor
- research
- enduser

If the page is a person's home page or blog or similar, it could be:

- opinionleader
- researcher
- journalist
- executive
- geek

The type of audience can also be tagged, for example:

- general public
- beginners
- technicians
- researchers.

[11] The [OASIS](#) Topic Maps Published Subjects Technical Committee was formed a number of years back to promote Topic Maps interoperability through the use of Published Subjects Indicators (PSIs). Their [resulting report](#) was a very interesting effort that unfortunately did not lead to wide adoption, perhaps because the effort was a bit ahead of its time or it was in advance of the broader acceptance of RDF. This general topic is the subject of a [later posting by me](#).

[12] See further, Leo Obrst, "The Semantic Spectrum & Semantic Models," a Powerpoint presentation ([http://ontolog.cim3.net/file/resource/presentation/LeoObrst\\_20060112/OntologySpectrumSemanticModels-LeoObrst\\_20060112.ppt](http://ontolog.cim3.net/file/resource/presentation/LeoObrst_20060112/OntologySpectrumSemanticModels-LeoObrst_20060112.ppt)) made as part of an Ontolog Forum (<http://ontolog.cim3.net/>) presentation in two parts, "What is an Ontology? – A Briefing on the Range of Semantic Models" (see [http://ontolog.cim3.net/cgi-bin/wiki.pl?ConferenceCall\\_2006\\_01\\_12](http://ontolog.cim3.net/cgi-bin/wiki.pl?ConferenceCall_2006_01_12)), in January 2006. Leo Obrst is a principal artificial intelligence scientist at MITRE's (<http://www.mitre.org>) Center for Innovative Computing and Informatics and a co-convener of the Ontolog Forum. His presentation is a rich source of practical overview information on ontologies.

[13] The actual diagram is an unattributed modification by Dan McCreary (see [http://www.danmccreary.com/presentations/sem\\_int/sem\\_int.ppt](http://www.danmccreary.com/presentations/sem_int/sem_int.ppt)) based on Obrst's material in [12].