

# A Basic Guide to Ontologies

From OSF Wiki

*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.<sup>[1]</sup>

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 (<http://en.wikipedia.org/wiki/Ontology>). Tom Gruber (<http://tomgruber.org/>), among others, made the term popular in relation to computer science ([http://en.wikipedia.org/wiki/Ontology\\_\(computer\\_science\)](http://en.wikipedia.org/wiki/Ontology_(computer_science))) and artificial intelligence about 15 years ago (<http://tomgruber.org/writing/ontologia-kaj-1993.htm>) 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, we prefer a loose and embracing understanding of the term (consistent with Deborah McGuinness’s 2003 paper, *Ontologies Come of Age* ([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))<sup>[2]</sup>

## Contents

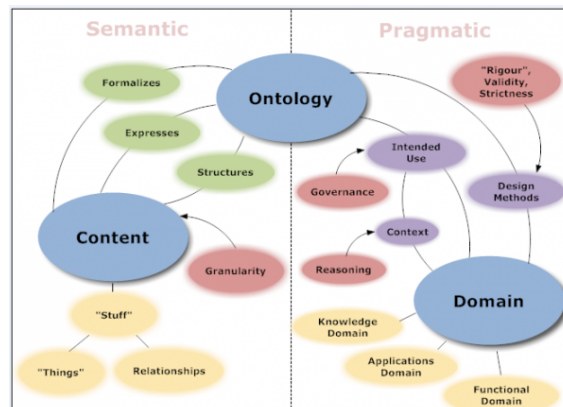
- 1 Overview and Role of Ontologies
- 2 Diversity, ‘Naturalness’ and Change
- 3 A Spectrum of Formalisms
- 4 Ontology Approaches on the Web
- 5 Still-Another “Level” of Ontologies
- 6 The SUMO Example
- 7 Ontology Binding and Integration Mechanisms
  - 7.1 Centralized Approaches
  - 7.2 Federated Approaches
- 8 The Role of SKOS – the Simple Knowledge Organization System
- 9 Conclusions
- 10 Endnotes

## 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
- 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 (<http://ontology.cim3.net/cgi-bin/wiki.pl?OntologySummit2007>)’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 ([http://en.wikipedia.org/wiki/M.\\_C.\\_Escher](http://en.wikipedia.org/wiki/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<br>( <a href="http://en.wikipedia.org/wiki/Tag_cloud">http://en.wikipedia.org/wiki/Tag_cloud</a> ) | ■ Social bookmarking<br>( <a href="http://en.wikipedia.org/wiki/Social_bookmarking">http://en.wikipedia.org/wiki/Social_bookmarking</a> ) | ■ Ontology ( <a href="http://en.wikipedia.org/wiki/Ontology_(computer_science)">http://en.wikipedia.org/wiki/Ontology_(computer_science)</a> ) | ■ Topic Maps ( <a href="http://en.wikipedia.org/wiki/Topic_map">http://en.wikipedia.org/wiki/Topic_map</a> )          |
| ■ Controlled vocabulary<br>( <a href="http://en.wikipedia.org/">http://en.wikipedia.org/</a> )                 | ■ Tags ( <a href="http://en.wikipedia.org/wiki/Tag_(metadata)">http://en.wikipedia.org/wiki/Tag_(metadata)</a> )                          | ■ Microformats<br>( <a href="http://en.wikipedia.org/wiki/Microformats">http://en.wikipedia.org/wiki/Microformats</a> )                        | ■ Concept Maps<br>( <a href="http://en.wikipedia.org/wiki/Concept_map">http://en.wikipedia.org/wiki/Concept_map</a> ) |

- [/wiki/Controlled\\_vocabulary\)](http://en.wikipedia.org/wiki/Controlled_vocabulary)
- Thesauri (<http://en.wikipedia.org/wiki/Thesauri>)
- Collaborative tagging ([http://en.wikipedia.org/wiki/Collaborative\\_tagging](http://en.wikipedia.org/wiki/Collaborative_tagging))
- Folk taxonomy ([http://en.wikipedia.org/wiki/Folk\\_taxonomy](http://en.wikipedia.org/wiki/Folk_taxonomy))
- Directory (<http://en.wikipedia.org/wiki/Directory>)
- Subject Map (<http://www.mulberrytech.com/Extreme/Proceedings/html/2005/Newcomb01/EML2005Newcomb01.html>)
- Semantic Web ([http://en.wikipedia.org/wiki/Semantic\\_Web](http://en.wikipedia.org/wiki/Semantic_Web))
- Cladistics (<http://en.wikipedia.org/wiki/Cladistics>)
- Markup languages ([http://en.wikipedia.org/wiki/Markup\\_language](http://en.wikipedia.org/wiki/Markup_language))
- Tagging (<http://en.wikipedia.org/wiki/Tagging>)
- Taxonomy (<http://en.wikipedia.org/wiki/Taxonomy>)
- Folksonomy (<http://en.wikipedia.org/wiki/Folksonomy>)
- Classification ([http://en.wikipedia.org/wiki/Library\\_classification](http://en.wikipedia.org/wiki/Library_classification))
- Categorization (<http://en.wikipedia.org/wiki/Categorization>)
- RDF ([http://en.wikipedia.org/wiki/Resource\\_Description\\_Framework](http://en.wikipedia.org/wiki/Resource_Description_Framework))
- Metadata (<http://en.wikipedia.org/wiki/Metadata>)
- Systematics (<http://en.wikipedia.org/wiki/Systematics>)
- Data dictionary ([http://en.wikipedia.org/wiki/Data\\_dictionary](http://en.wikipedia.org/wiki/Data_dictionary))
- OPML (<http://en.wikipedia.org/wiki/OPML>)
- XOXO (<http://en.wikipedia.org/wiki/XOXO>)
- OWL ([http://en.wikipedia.org/wiki/Web\\_Ontology\\_Language](http://en.wikipedia.org/wiki/Web_Ontology_Language))
- Subject Trees ([http://en.wikipedia.org/wiki/Tree\\_structure](http://en.wikipedia.org/wiki/Tree_structure))
- Information Architecture ([http://en.wikipedia.org/wiki/Information\\_architecture](http://en.wikipedia.org/wiki/Information_architecture))
- Data Reference Model ([http://en.wikipedia.org/wiki/Data\\_Reference\\_Model](http://en.wikipedia.org/wiki/Data_Reference_Model))
- Phylogeny (<http://en.wikipedia.org/wiki/Phylogeny>)
- Synsets (<http://en.wikipedia.org/wiki/Synset>)
- Glossary (<http://en.wikipedia.org/wiki/Glossary>)
- WordNet (<http://en.wikipedia.org/wiki/WordNet>)
- Metadata (<http://en.wikipedia.org/wiki/Metadata>)
- Facets ([http://en.wikipedia.org/wiki/Faceted\\_classification](http://en.wikipedia.org/wiki/Faceted_classification))
- Structure ([http://en.wikipedia.org/wiki/Structure\\_\(mathematical\\_logic\)](http://en.wikipedia.org/wiki/Structure_(mathematical_logic)))
- Dublin Core ([http://en.wikipedia.org/wiki/Dublin\\_Core](http://en.wikipedia.org/wiki/Dublin_Core))
- Typology (<http://en.wikipedia.org/wiki/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*. In fact, UMBC's Swoogle (<http://swoogle.umbc.edu/>) 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 (<http://ebiquity.umbc.edu/blogger/untangling-ontologies-on-the-semantic-web/>), 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* ([http://en.wikipedia.org/wiki/One\\_Ring](http://en.wikipedia.org/wiki/One_Ring)). 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.

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 → 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 ([http://en.wikipedia.org/wiki/Tree\\_of\\_life\\_\(Science\)](http://en.wikipedia.org/wiki/Tree_of_life_(Science))).

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, 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 noted above, *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.

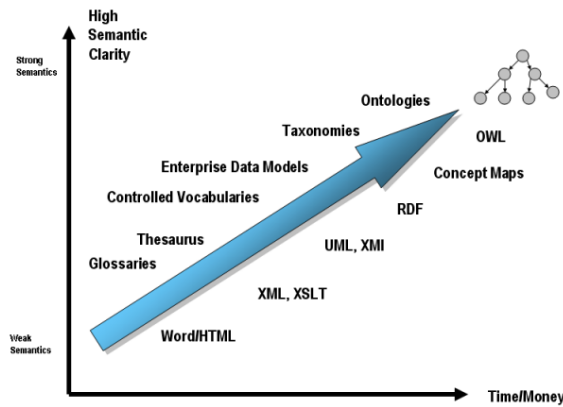
The over fifty respondents from 42 communities submitted some 70 different ontologies under about 40 terms to a survey that was used by the Summit to construct their diagram. These submissions included:

... formal ontologies (e.g., BFO (<http://www.ifomis.uni-saarland.de/bfo/>), DOLCE (<http://www.loa-cnr.it/DOLCE.html>), SUMO (<http://www.ontologyportal.org/>)), biomedical ontologies (e.g., Gene Ontology, SNOMED CT, UMLS, ICD), thesauri (e.g., MeSH (<http://www.nlm.nih.gov/mesh/termscon.html>), National Agricultural Library Thesaurus), folksonomies (e.g., Social bookmarking tags), general ontologies (WordNet (<http://www.cogsci.princeton.edu/~wn>), OpenCyc (<http://www.cyc.com/cyc/encyc/overview>)) and specific ontologies (e.g., Process Specification Language). The list also includes markup languages (e.g., NeuroML (<http://www.neuroml.org/>)), 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".

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<sup>[3], [4]</sup>.

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



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 ([http://en.wikipedia.org/wiki/Semantic\\_spectrum](http://en.wikipedia.org/wiki/Semantic_spectrum)).

However, since we are speaking here of ontologies and the *structured Web* or the semantic Web, 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 Technorati ( <a href="http://technorati.com/tag/">http://technorati.com/tag/</a> ), Bloglines ( <a href="http://www.bloglines.com/">http://www.bloglines.com/</a> ), Wikipedia ( <a href="http://en.wikipedia.org/wiki/Main_Page">http://en.wikipedia.org/wiki/Main_Page</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, FOAF, etc.; highly varied and site-specific given the diversity of publishing systems and plug-ins)
RSS ( <a href="http://en.wikipedia.org/wiki/Rss">http://en.wikipedia.org/wiki/Rss</a> ) / Atom ( <a href="http://en.wikipedia.org/wiki/Atom_(standard)">http://en.wikipedia.org/wiki/Atom_(standard)</a> ) feeds	most blogs and most news feeds	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 ( <a href="http://en.wikipedia.org/wiki/Tag_(metadata)">http://en.wikipedia.org/wiki/Tag_(metadata)</a> ) or OPML ( <a href="http://en.wikipedia.org/wiki/OPML">http://en.wikipedia.org/wiki/OPML</a> )	Grazr ( <a href="http://www.grazr.com/">http://www.grazr.com/</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 OML ( <a href="http://en.wikipedia.org/wiki/OML">http://en.wikipedia.org/wiki/OML</a> ) and XOXO ( <a href="http://en.wikipedia.org/wiki/XOXO">http://en.wikipedia.org/wiki/XOXO</a> )
Hierarchical Faceted Metadata ( <a href="http://en.wikipedia.org/wiki/Faceted_classification">http://en.wikipedia.org/wiki/Faceted_classification</a> )	XFML ( <a href="http://xfml.org/">http://xfml.org/</a> ), Flamenco ( <a href="http://flamenco.berkeley.edu/index.html">http://flamenco.berkeley.edu/index.html</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
Folksonomies ( <a href="http://en.wikipedia.org/wiki/Folksonomy">http://en.wikipedia.org/wiki/Folksonomy</a> )	Flickr ( <a href="http://en.wikipedia.org/wiki/Flickr">http://en.wikipedia.org/wiki/Flickr</a> ), del.icio.us ( <a href="http://en.wikipedia.org/wiki/Del.icio.us">http://en.wikipedia.org/wiki/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
Microformats ( <a href="http://en.wikipedia.org/wiki/Microformats">http://en.wikipedia.org/wiki/Microformats</a> )	Example formats include hAtom ( <a href="http://en.wikipedia.org/wiki/HAtom">http://en.wikipedia.org/wiki/HAtom</a> ), hCalendar ( <a href="http://en.wikipedia.org/wiki/HCalendar">http://en.wikipedia.org/wiki/HCalendar</a> ), hCard ( <a href="http://en.wikipedia.org/wiki/HCard">http://en.wikipedia.org/wiki/HCard</a> ), hReview ( <a href="http://en.wikipedia.org/w/index.php?title=HReview&amp;action=edit">http://en.wikipedia.org/w/index.php?title=HReview&amp;action=edit</a> ), hResume ( <a href="http://en.wikipedia.org/wiki/HResume">http://en.wikipedia.org/wiki/HResume</a> ), rel-directory ( <a href="http://en.wikipedia.org/w/index.php?title=Rel-directory&amp;action=edit">http://en.wikipedia.org/w/index.php?title=Rel-directory&amp;action=edit</a> ), xFolk ( <a href="http://en.wikipedia.org/w/index.php?title=xFolk&amp;action=edit">http://en.wikipedia.org/w/index.php?title=xFolk&amp;action=edit</a> ), XFN ( <a href="http://en.wikipedia.org/wiki/XHTML_Friends_Network">http://en.wikipedia.org/wiki/XHTML_Friends_Network</a> ) and XOXO ( <a href="http://en.wikipedia.org/wiki/XOXO">http://en.wikipedia.org/wiki/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	RDFa ( <a href="http://en.wikipedia.org/wiki/RDFa">http://en.wikipedia.org/wiki/RDFa</a> ), eRDF	An embedded format, like microformats, but free-form, and not subject to the approval

Topic Maps ( <a href="http://en.wikipedia.org/wiki/Topic_map">http://en.wikipedia.org/wiki/Topic_map</a> )	Infoloom ( <a href="http://www.topicmaps.net/">http://www.topicmaps.net/</a> ), Topic Maps Search Engine ( <a href="http://search.topicmaps.de/">http://search.topicmaps.de/</a> )	<p>strictures associated with microformats</p> <p>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)</p>
RDF ( <a href="http://en.wikipedia.org/wiki/Resource_Description_Framework">http://en.wikipedia.org/wiki/Resource_Description_Framework</a> )	Many; DBpedia ( <a href="http://dbpedia.org/">http://dbpedia.org/</a> ), etc.	RDF has become the canonical data model since it represents a “universal” conversion format
RDF Schema ( <a href="http://en.wikipedia.org/wiki/RDF_Schema">http://en.wikipedia.org/wiki/RDF_Schema</a> )	SKOS ( <a href="http://en.wikipedia.org/wiki/SKOS">http://en.wikipedia.org/wiki/SKOS</a> ), SIOC ( <a href="http://en.wikipedia.org/wiki/SIOC">http://en.wikipedia.org/wiki/SIOC</a> ), DOAP ( <a href="http://en.wikipedia.org/wiki/DOAP">http://en.wikipedia.org/wiki/DOAP</a> ), FOAF ( <a href="http://en.wikipedia.org/wiki/FOAF_(software)">http://en.wikipedia.org/wiki/FOAF_(software)</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
OWL ( <a href="http://en.wikipedia.org/wiki/Web_Ontology_Language">http://en.wikipedia.org/wiki/Web_Ontology_Language</a> ) Lite	Here are some existing OWL ontologies ( <a href="http://protege.cim3.net/cgi-bin/wiki.pl?ProtegeOntologiesLibrary#nid81S">http://protege.cim3.net/cgi-bin/wiki.pl?ProtegeOntologiesLibrary#nid81S</a> ); also see Swoogle ( <a href="http://swoogle.umbc.edu/">http://swoogle.umbc.edu/</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 (RDFS) by providing additional vocabulary along with a formal semantics. The three language versions are in order of increasing expressiveness
OWL ( <a href="http://en.wikipedia.org/wiki/Web_Ontology_Language">http://en.wikipedia.org/wiki/Web_Ontology_Language</a> ) DL		
OWL ( <a href="http://en.wikipedia.org/wiki/Web_Ontology_Language">http://en.wikipedia.org/wiki/Web_Ontology_Language</a> ) Full		
Higher-order “formal” and “upper-level” ontologies	SUMO ( <a href="http://en.wikipedia.org/wiki/Suggested_Upper_Merged_Ontology">http://en.wikipedia.org/wiki/Suggested_Upper_Merged_Ontology</a> ), DOLCE ( <a href="http://wonderweb.semanticweb.org/deliverables/documents/D18.pdf">http://wonderweb.semanticweb.org/deliverables/documents/D18.pdf</a> ), PROTON ( <a href="http://proton.semanticweb.org/D1_8_1.pdf">http://proton.semanticweb.org/D1_8_1.pdf</a> ), BFO ( <a href="http://www.ifomis.org/bfo">http://www.ifomis.org/bfo</a> ), Cyc ( <a href="http://en.wikipedia.org/wiki/Cyc">http://en.wikipedia.org/wiki/Cyc</a> ), OpenCyc	These provide comprehensive ontologies and often related knowledge bases, with the goal of enabling AI applications to perform human-like 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.

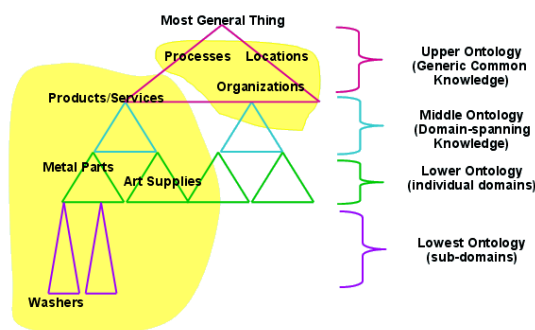
As latter sections elaborate, RDF may be viewed as a universal data model for representing this structure into a common, canonical format, with RDF Schema (specifically SKOS, but also supplemented by FOAF, DOAP and SIOC) as the organizing ontology knowledge representation language ([http://en.wikipedia.org/wiki/Knowledge\\_representation](http://en.wikipedia.org/wiki/Knowledge_representation)) (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.

### 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 ([http://en.wikipedia.org/wiki/Upper\\_ontology\\_\(computer\\_science\)](http://en.wikipedia.org/wiki/Upper_ontology_(computer_science))).

The following figure helps illustrate this “level” dimension. This diagram is also from Leo Obrst of Mitre [12 (<http://www.mkbergman.com/index.php?p=374#onto12>)], and was also used in another 2006 talk by Jack Park and Patrick Durusau ([http://ontolog.cim3.net/file/resource/presentation/JackPark-PatrickDurusau\\_20060427](http://ontolog.cim3.net/file/resource/presentation/JackPark-PatrickDurusau_20060427)) (Avoiding\_Hobson-s\_Choice\_In\_Choosing\_An\_Ontology--JackPark-PatrickDurusau\_20060427.ppt) (discussed further below for other reasons):



Examples of upper-level ontologies include the Suggested Upper Merged Ontology (SUMO ([http://en.wikipedia.org/wiki/Suggested\\_Upper\\_Merged\\_Ontology](http://en.wikipedia.org/wiki/Suggested_Upper_Merged_Ontology))), the Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE (<http://wonderweb.semanticweb.org/deliverables/documents/D18.pdf>)), PROTON ([http://proton.semanticweb.org/D1\\_8\\_1.pdf](http://proton.semanticweb.org/D1_8_1.pdf)), UMBEL (<http://umbel.org>), Cyc (<http://en.wikipedia.org/wiki/Cyc>) and BFO (<http://www.ifomis.org/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 ([http://en.wikipedia.org/wiki/Roget's\\_Thesaurus](http://en.wikipedia.org/wiki/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.<sup>[5]</sup>

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.

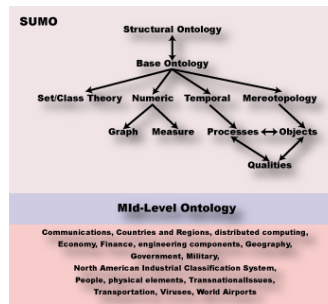
### The SUMO Example

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

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 (<http://www.ontologyportal.org/>) simply because it is one of the best known. We could have also selected another upper-level system such as PROTON<sup>[6]</sup> or Cyc<sup>[7]</sup> or one of the many with narrower concept or subject coverage.

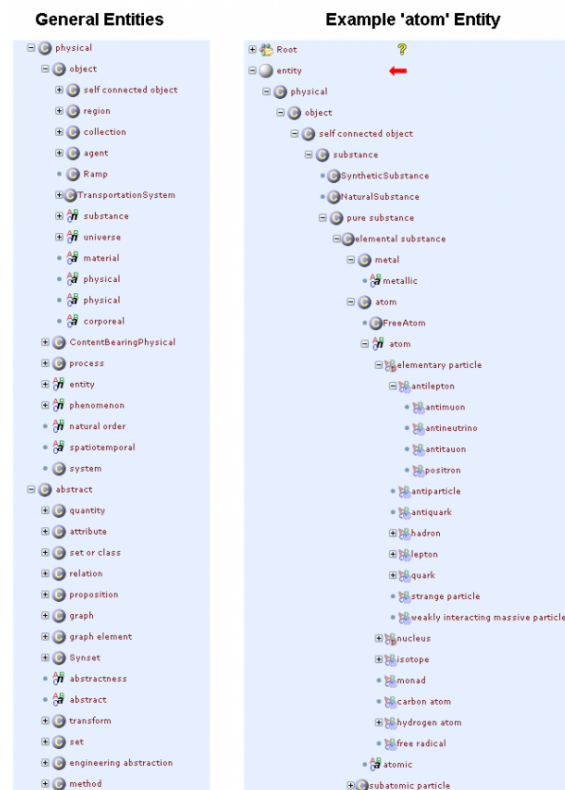
SUMO is one of the formal ontologies that has been mapped to the WordNet (<http://www.cogsci.princeton.edu/~wn/>) lexicon, which adds to its semantic richness. SUMO is written in the SUO-KIF ([http://sigmakee.cvs.sourceforge.net/\\*checkout\\*/sigmakee/sigma/suo-kif.pdf](http://sigmakee.cvs.sourceforge.net/*checkout*/sigmakee/sigma/suo-kif.pdf)) language. SUMO is free and owned by the IEEE (<http://www.ieee.org/portal/site>). The ontologies that extend SUMO are available under GNU General Public License (<http://www.gnu.org/copyleft/gpl.html>).

The abstract, conceptual organization of SUMO is shown by this diagram, which also points to its related MILO (Mid-Level Ontology) ([http://sigmakee.cvs.sourceforge.net/\\*checkout\\*/sigmakee/KBs/Mid-level-ontology.kif](http://sigmakee.cvs.sourceforge.net/*checkout*/sigmakee/KBs/Mid-level-ontology.kif)), which is being developed as a bridge between the abstract content of the SUMO and the richer detail of various domain ontologies:



At this level, the structure is quite abstract. But one can easily browse the SUMO structure. A nifty tool to do so is the KSMSA (Knowledge Support for Modeling and Simulation) ontology browser (<http://virtual.cvt.cz/ksmsaWeb/browser/title>). Using a hierarchical tree representation, you can navigate through SUMO, MILO, WordNet (<http://www.cogsci.princeton.edu/~wn/>), and (with the locally installed version) Wikipedia.

The figure below shows the upper-level *entity* concept on the left; the right-hand panel shows a drill-down into the example *atom* entity:



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 ([http://en.wikipedia.org/wiki/Data\\_Reference\\_Model](http://en.wikipedia.org/wiki/Data_Reference_Model)) (DRM ([http://www.whitehouse.gov/omb/egov/documents/DRM\\_2\\_0\\_Final.pdf](http://www.whitehouse.gov/omb/egov/documents/DRM_2_0_Final.pdf))), one of the five reference models of the Federal Enterprise Architecture (FEA)
- UDEF (<http://en.wikipedia.org/wiki/UDEF>) (Unified Data Element Framework), an approach toward a unified description framework, or
- the eXtended MetaData Registry (XMDR (<http://xmdr.org/>)) project.

Though *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.<sup>[8]</sup> 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.<sup>[9]</sup> 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.<sup>[10]</sup> 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 (<http://www.jtc1sc34.org/repository/0710.pdf>) (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.”

The presentation and papers by Park and Durusau, *Avoiding Hobson’s Choice In Choosing An Ontology* ([http://ontolog.cim3.net/file/resource/presentation/JackPark-PatrickDurusau\\_20060427/Avoiding\\_Hobson-s\\_Choice\\_In\\_Choosing\\_An\\_Ontology-JackPark-PatrickDurusau\\_20060427.ppt](http://ontolog.cim3.net/file/resource/presentation/JackPark-PatrickDurusau_20060427/Avoiding_Hobson-s_Choice_In_Choosing_An_Ontology-JackPark-PatrickDurusau_20060427.ppt)) and *Towards Subject-centric Merging of Ontologies* ([http://protege.stanford.edu/conference/2006/submissions/abstracts/3.3\\_Park\\_SubjectCentricProtege.pdf](http://protege.stanford.edu/conference/2006/submissions/abstracts/3.3_Park_SubjectCentricProtege.pdf)) can provide additional reading on these topics.

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 (<http://en.wikipedia.org/wiki/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, we separately list about a hundred ontology tools [xxxx] useful to these topics.

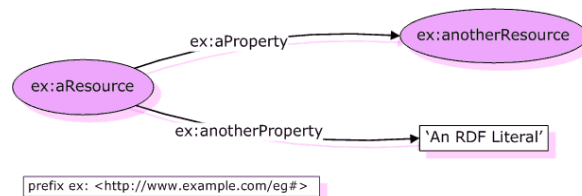
### The Role of SKOS - the Simple Knowledge Organization System

SKOS, or the Simple Knowledge Organization System, is a formal language and schema designed to represent such structured information domains as thesauri (<http://en.wikipedia.org/wiki/Thesaurus>), classification schemes ([http://en.wikipedia.org/wiki/Classification\\_scheme](http://en.wikipedia.org/wiki/Classification_scheme)), taxonomies (<http://en.wikipedia.org/wiki/Taxonomy>), subject-heading systems ([http://en.wikipedia.org/wiki/Authority\\_control](http://en.wikipedia.org/wiki/Authority_control)), controlled vocabularies ([http://en.wikipedia.org/wiki/Controlled\\_vocabulary](http://en.wikipedia.org/wiki/Controlled_vocabulary)), or others; in short, most all of the “loosely defined” ontology approaches discussed herein. It is a W3C ([http://en.wikipedia.org/wiki/World\\_Wide\\_Web\\_Consortium](http://en.wikipedia.org/wiki/World_Wide_Web_Consortium)) initiative more fully defined in its SKOS Core Guide (<http://www.w3.org/TR/swbp-skos-core-guide>).<sup>[11]</sup>

SKOS is built upon the RDF ([http://en.wikipedia.org/wiki/Resource\\_Description\\_Framework](http://en.wikipedia.org/wiki/Resource_Description_Framework)) 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 ([http://en.wikipedia.org/wiki/Uniform\\_Resource\\_Identifier](http://en.wikipedia.org/wiki/Uniform_Resource_Identifier)) 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 ([http://en.wikipedia.org/wiki/RDF\\_Schema](http://en.wikipedia.org/wiki/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.

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 (<http://www.ukat.org.uk/thesaurus/hierarchy.php?m=625>)), 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:





- 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.

13. ↑ The OASIS (<http://www.oasis-open.org/home/index.php>) 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 (<http://www.oasis-open.org/committees/download.php/3050/pubsobj-pt1-1.02-cs.pdf>) 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 alater posting by me (<http://www.mkbergman.com/?p=375>).

Retrieved from "http://wiki.opensemanticframework.org/index.php?title=A\_Basic\_Guide\_to\_Ontologies&oldid=6747"

Categories: Background Documents | Ontology Introductions | ZWiki

- This page was last modified on 6 January 2014, at 22:18.
- This page has been accessed 9,843 times.
- Content is available under Creative Commons Attribution unless otherwise noted.



Resources

Structured Dynamics is an innovative development and consulting company dedicated to semantic technologies.



- Overview
- About SD
- Open Source Codes

## Software and Semantic Technologies

- Open Semantic Framework, a complete stack of open source semantic technologies
- OSF for Drupal, leveraging the Drupal 7 CMS as a front-end to OSF
- OSF Web Services are a platform-independent series of about 30 Web services for RDF data
- OSF Tagger, an information extractor and tagger using domain-specific vocabularies
- OSF Widgets, visualization and display components for OSF records
- OSF Management Tools, a variety of command-line tools for managing OSF data and ontologies

## Sandboxes and Demos

- Citizen Dan is an open source system for aggregating local, community data
- UMBEL, an upper ontology of 28,000 reference concepts for the Web

## Technical Documentation and Methodologies

- OSF Wiki, ~ 500 free, technical articles on SD's technologies, mostly for OSF
- Open SEAS, an open source methodology for transitioning to a semantic enterprise
- Sweet Tools, an online database of 1000 semantic tools, most open source

## Ontologies and Vocabularies

- UMBEL has 28,000 reference concepts for categorizing Web data and a vocabulary for designing domain ontologies
- The Bibliographic Ontology provides main concepts and properties for describing citations and bibliographic references
- irON is a specification and vocabulary for specifying RDF triples and schema in JSON, XML or CSV formats
- MUNI Ontology, a general ontology for cities and local communities
- Music Ontology, a leading vocabulary about musical artists, albums and tracks

## Founder Blogs

- Mike Bergman's AI3::Adaptive Information blog is one of the most popular in the semantic Web and linked data space
- Fred Giasson's blog has been a source of technical posts and oft-cited innovations since 2005

