Semantic Metadata

A Tale of Two Types of Vocabularies
What is the semantic web?

- Making content web-accessible in a format that can be read and used by automated tools, so that people and machines can find, share and integrate information more easily.

- Some current examples, especially if they use semantics as the basis for that integration.
  - **Named Entity Services** – Named entity links to “authority files” are rendered by browsers.
  - **Dynamic Web Pages** – Content changes in response to different contexts or conditions.
  - **Personalization** – Tailoring to a user based on personal details or characteristics they provide (or are inferred based on other information).
  - **Mashups** – Combining data from more than one source into an integrated application.
Negotiations With Iran Over Nuclear Program May Resume

By STEVEN LEE MYERS and RICK GLADSTONE
Published: February 17, 2012

WASHINGTON — The United States and the European Union signaled on Friday that negotiations with Iran over its nuclear program could soon resume for the first time in nearly a year, even as a telecommunications network vital to the global banking industry prepared to expel Iranian banks.

While senior American and European officials stopped short of declaring a diplomatic breakthrough, Iran dropped previously unacceptable preconditions for talks in a letter this week from its senior nuclear negotiator, Saeed Jalili, who declared his country’s “readiness for dialogue” at “the earliest possibility.”

After weeks of official bluster, ominous threats of military retaliation, and efforts by Israel and others to stoke public anxiety, Iran’s new offer could pave the way for discussions that eluded American and European diplomats last year when the negotiations became too contentious to continue.

More news and information about Iran.
Google’s new right rail – another form of linked data data
Oracle events mashup with Google maps
Linked data cloud
Most widely used vocabularies in the linked data cloud (as of 9/11/2011)

<table>
<thead>
<tr>
<th>Vocabulary prefix</th>
<th>Vocabulary link</th>
<th>Number of usages in data sets</th>
<th>Data sets that use the vocabulary</th>
</tr>
</thead>
<tbody>
<tr>
<td>dc</td>
<td><a href="http://purl.org/dc/elements/1.1/">http://purl.org/dc/elements/1.1/</a></td>
<td>92 (31.19 %)</td>
<td>Data sets that use dc</td>
</tr>
<tr>
<td>foaf</td>
<td><a href="http://xmlns.com/foaf/0.1/">http://xmlns.com/foaf/0.1/</a></td>
<td>81 (27.46 %)</td>
<td>Data sets that use foaf</td>
</tr>
<tr>
<td>skos</td>
<td><a href="http://www.w3.org/2004/02/skos/core#">http://www.w3.org/2004/02/skos/core#</a></td>
<td>58 (19.66 %)</td>
<td>Data sets that use skos</td>
</tr>
<tr>
<td>geo</td>
<td><a href="http://www.w3.org/2003/01/geo/wgs84_pos#">http://www.w3.org/2003/01/geo/wgs84_pos#</a></td>
<td>25 (8.47 %)</td>
<td>Data sets that use geo</td>
</tr>
<tr>
<td>xhtml</td>
<td><a href="http://www.w3.org/1999/xhtml/vocab#">http://www.w3.org/1999/xhtml/vocab#</a></td>
<td>19 (6.44 %)</td>
<td>Data sets that use xhtml</td>
</tr>
<tr>
<td>akt</td>
<td><a href="http://www.aktors.org/ontology/vocab#">http://www.aktors.org/ontology/vocab#</a></td>
<td>17 (5.76 %)</td>
<td>Data sets that use akt</td>
</tr>
<tr>
<td>bibo</td>
<td><a href="http://purl.org/ontology/bibo/">http://purl.org/ontology/bibo/</a></td>
<td>14 (4.75 %)</td>
<td>Data sets that use bibo</td>
</tr>
<tr>
<td>mo</td>
<td><a href="http://purl.org/ontology/mo/">http://purl.org/ontology/mo/</a></td>
<td>13 (4.41 %)</td>
<td>Data sets that use mo</td>
</tr>
<tr>
<td>vcard</td>
<td><a href="http://www.w3.org/2006/vcard/ns#">http://www.w3.org/2006/vcard/ns#</a></td>
<td>10 (3.39 %)</td>
<td>Data sets that use vcard</td>
</tr>
<tr>
<td>sioc</td>
<td><a href="http://rdfs.org/sioc/ns#">http://rdfs.org/sioc/ns#</a></td>
<td>10 (3.39 %)</td>
<td>Data sets that use sioc</td>
</tr>
<tr>
<td>cc</td>
<td><a href="http://creativecommons.org/ns#">http://creativecommons.org/ns#</a></td>
<td>8 (2.71 %)</td>
<td>Data sets that use cc</td>
</tr>
<tr>
<td>geonames</td>
<td><a href="http://www.geonames.org/ontology#">http://www.geonames.org/ontology#</a></td>
<td>6 (2.03 %)</td>
<td>Data sets that use geonames</td>
</tr>
<tr>
<td>frbr</td>
<td><a href="http://purl.org/vocab/frbr/core#">http://purl.org/vocab/frbr/core#</a></td>
<td>6 (2.03 %)</td>
<td>Data sets that use frbr</td>
</tr>
<tr>
<td>xsd</td>
<td><a href="http://www.w3.org/2001/XMLSchema#">http://www.w3.org/2001/XMLSchema#</a></td>
<td>6 (2.03 %)</td>
<td>Data sets that use xsd</td>
</tr>
<tr>
<td>time</td>
<td><a href="http://www.w3.org/2006/time#">http://www.w3.org/2006/time#</a></td>
<td>5 (1.69 %)</td>
<td>Data sets that use time</td>
</tr>
<tr>
<td>event</td>
<td><a href="http://purl.org/NET/c4dm/event.owl#">http://purl.org/NET/c4dm/event.owl#</a></td>
<td>5 (1.69 %)</td>
<td>Data sets that use event</td>
</tr>
<tr>
<td>dbpedia</td>
<td><a href="http://dbpedia.org/resource/">http://dbpedia.org/resource/</a></td>
<td>5 (1.69 %)</td>
<td>Data sets that use dbpedia</td>
</tr>
<tr>
<td>gr</td>
<td><a href="http://purl.org/goodrelations/v1#">http://purl.org/goodrelations/v1#</a></td>
<td>4 (1.36 %)</td>
<td>Data sets that use gr</td>
</tr>
</tbody>
</table>

http://www4.wiwiss.fu-berlin.de/lodcloud/state/#structure
Linked data cloud characteristics (as of 9/11/2011)

Distribution of triples by domain

Distribution of links by domain

http://www4.wiwiss.fu-berlin.de/lodcloud/state/#structure
Types of Vocabularies

- In the linked data cloud, there are two types of vocabularies:
  - Concept schemes – metadata schemes like Dublin Core
  - Semantic schemes – value vocabularies like taxonomies, thesauri, ontologies, etc.
Why Dublin Core? According to Todd Stephens …

- Dublin Core is a de-facto standard across many other systems and standards
  - RSS (1.0), OAI (Open Archives Initiative), SEMI E36, etc.
  - Inside organizations – ECMS, SharePoint, etc.
- Mapping to DC elements from most existing schemes is simple.
- Metadata already exists in enterprise applications
  - Windchill, OpenText, MarkLogic, SAP, Documentum, MS Office, SharePoint, Drupal, etc.
MDM model that integrates taxonomy and metadata

- Taxonomies, Vocabularies, Ontologies
- Dublin Core

Source: Todd Stephens, BellSouth

Per-Source Data Types, Access Controls, etc.
Semantic Schemes: Simple to Complex

- **Synonym Ring**: A set of words/phrases that can be used interchangeably for searching. E.g., Hypertension, High blood pressure.

- **Controlled Vocabulary**: A list of preferred and variant terms.

- **Taxonomy**: A system for identifying and naming things, and arranging them into a classification according to a set of rules.

- **Classification Scheme**: An arrangement of knowledge usually enumerated, that does not follow taxonomy rules. E.g., Dewey Decimal Classification.

- **Thesaurus**: A tool that controls synonyms and identifies the semantic relationships among terms.

- **Ontology**: A faceted taxonomy but uses richer semantic relationships among terms and attributes and strict specification rules.

**Relationships**

- **Equivalence**: A set of preferred and variant terms.

- **Hierarchy**: An arrangement of knowledge usually enumerated, that does not follow taxonomy rules. E.g., Dewey Decimal Classification.

- **Associative**: A tool that controls synonyms and identifies the semantic relationships among terms.

After: Amy Warner. *Metadata and Taxonomies for a More Flexible Information Architecture*
Q: How do you share a vocabulary across (and outside of) the enterprise?
A: With standards

- **ISO 2788:1986** Guidelines for the Establishment and Development of Monolingual Thesauri
- **ISO 5964:1985** Guidelines for the Establishment and Development of Multilingual Thesauri
- **ISO 25964** (combines 2788 and 5964) Thesauri and Interoperability with other Vocabularies
- **Zthes** specifications for thesaurus representation, access and navigation
- **W3C SKOS** Simple Knowledge Organization System
Why SKOS? According to Alistair Miles …

- **Ease of combination** with other standards
  - Vocabularies are used in great variety of contexts.
    - E.g., databases, faceted navigation, website browsing, linked open data, spellcheckers, etc.
  - Vocabularies are re-used in combination with other vocabularies.
    - E.g., [ISO3166 country codes](https://en.wikipedia.org/wiki/ISO_3166) + [USAID regions](https://www.usaid.gov); [USPS zip codes](https://www.usps.com) + [US Congressional districts](https://www.congress.gov); [USPS states](https://www.usps.com) + [EPA regions](https://www.epa.gov), etc.

- **Flexibility and extensibility** to cope with variations in structure and style
  - Variations between types of vocabularies
    - E.g., list vs. classification scheme
  - Variations within types of vocabularies
    - E.g., [Z39.19-2005](https://www.loc.gov/policy/z39/) monolingual controlled vocabularies and the [NASA Taxonomy](https://www.nasa.gov/topics/multimedia/unifiedtaxonomystandard.html)
Why SKOS? (2)

- **Publish managed vocabularies** so they can readily be consumed by applications
  - Identify the concepts
    - What are the named entities?
  - Describe the relationships
    - Labels, definitions and other properties
  - Publish the data
    - Convert data structure to standard format
    - Put files on an http server (or load statements into an RDF server)

- **Ease of integration** with external applications
  - Use web services to use or link to a published concept, or to one or more entire vocabularies.
    - E.g., Google maps API, NY Times article search API, Linked open data

- **A W3C standard** like HTML, CSS, XML... and RDF, RDFS, and OWL
# Semantic relationships

<table>
<thead>
<tr>
<th>Relationship</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concept</td>
<td>A unit of thought, an idea, meaning, or category of objects or events. A Concept is independent of the terms used to label it.</td>
</tr>
<tr>
<td>Preferred Label</td>
<td>A preferred lexical label for the resource such as a term used in a digital asset management system.</td>
</tr>
<tr>
<td>Alternate Label</td>
<td>An alternative label for the resource such as a synonym or quasi-synonym.</td>
</tr>
<tr>
<td>Broader Concept</td>
<td>Hierarchical link between two Concepts where one Concept is more general than the other.</td>
</tr>
<tr>
<td>Narrower Concept</td>
<td>Hierarchical link between two Concepts where one Concept is more specific than the other.</td>
</tr>
<tr>
<td>Related Concept</td>
<td>Link between two Concepts where the two are inherently &quot;related&quot;, but that one is not in any way more general than the other.</td>
</tr>
</tbody>
</table>
Some semantic relationships for IBM

<table>
<thead>
<tr>
<th>Subject</th>
<th>Predicate</th>
<th>Object</th>
</tr>
</thead>
<tbody>
<tr>
<td>lc:n79142877</td>
<td>skos:prefLabel</td>
<td>International Business Machines Corporation</td>
</tr>
<tr>
<td>lc:n79142877</td>
<td>skos:altLabel</td>
<td>IBM</td>
</tr>
<tr>
<td>lc:n79142877</td>
<td>skos:altLabel</td>
<td>I.B.M.</td>
</tr>
</tbody>
</table>
My company sells IBM’s XIV product

A high-end disk storage server designed to provide high performance, scalability, and availability in disk storage.

Originally developed by Israeli company XIV, which was acquired by IBM in 2007.
QUESTIONS?

Joseph A Busch, Principal

jbusch@taxonomystrategies.com

twitter.com/joebusch

415-377-7912
Semantic Metadata: A Tale of Two Types of Vocabularies

- Semantic metadata is metadata that is expressed using a standard syntax that can be commonly processed by applications and tools. There is always an implied statement in any description or "classification" of an object, for example, <News Item><Topic><US Presidential Election 2012>. This is a subject-predicate-object triple, or more specifically, a class-attribute-value triple. The first two elements of the triple – class, attribute – are metadata elements with a defined semantic relationship. The third element is a value, from a controlled vocabulary. This talk will focus on: The two types of vocabularies involved with semantic metadata, the class-attribute vocabulary, and the value vocabulary. Examples of standard metadata vocabularies such as Dublin Core and FOAF, and canonical lists of named entities (people, organizations, places, events and things) especially well-branded names such as products and services will be shown.