Today’s Session (Nov. 10, 2016)

Introduction to the Series on Improved Semantics for & with Domain Vocabularies

● Topic and Session Overview Gary Berg-Cross (Ontolog/RDA)

● Presenters:
  ● Mark Fox (University of Toronto)  An Ontology Design Pattern for Global City Indicators
  ● Boyan Brodaric (Research Scientist at Natural Resources Canada): What's a river? A foundational approach to a domain reference ontology for water

● Discussion
Domains have made good faith efforts to build CVs and register vocabularies in various ways. There may be implied community “agreement”. But extensive mapping of terms from various sources doesn’t solve all of the heterogeneity problems so they turn to the ontological community for a magic bullet or 2.


Gary Berg-Cross (Ontolog Forum & RDA US Advisory Group)
Help systematize the already large body of domain definition work on terms and their meaning using the rationalized, “consensus” knowledge of domain experts, especially as involved in RDA’s efforts.

Clarify alternate representation
Make implicit ideas more explicit & reasonably reflect the types of entities found in reality.
Ontology and Vocabulary Space

RDA

(Master) Vocabulary
("Standard") Terms, classification schemes, thesauri, lexica...

Tags/Annotations

Metadata

Past work (e.g. HC) demonstrates the value of controlled vocabularies as aiding "metadata" documentation to find, use & integrate data.

CVs are often used for indexing and retrieving data resources but early efforts were often arbitrary with little supporting conceptualizations or real standardization.

So there are some cases with dozens and dozens of "local" standards for data.

Examples include spatial locations for depiction on maps or water quality vocabularies that conflate multiple concepts but insert these into a single, compounded term.

E. G. observations mix substance with the medium (e.g. water) observed, along with the procedure used as part of the observation and the units used for measurement.
Semantic Integration Ontologies can mitigate variety in Big Data by aiding the annotation of data and its metadata.

- "Most of semantic web applications use ontologies as vocabularies to describe metadata and are aimed at semantic processing of them."

Data sets will differ in completeness of metadata, granularity and vocabulary used.

Ontologies can reduce some of this variety by normalizing terms and providing for absent metadata to ensure that there are no semantic mismatches.

RDF/S by itself is not a solution. RDF triples without ontological extensions may be underspecified bits of knowledge.

Triples can help with the vocabulary aspects of work, but better conceptualization & formalization with languages like OWL can more formally define and constrain meaning.
Two Views of Hydrographic Feature

Like many Earth Science domain(s) this is an extensive, flat, metadata vocabulary

USGS National Water Information System (NWIS) which has over 18000 codes for associated hydrologic variables.
Some Ontologies for this domain are just Lists of Features
http://hydro10.sdsc.edu/cinergi_ontology/GeographicFeatures.owl-

yagoGeoEntity

bank

cinergiGeoEntity
Naming schemes often suggest some implied semantics.

“Descriptions are more or less purposeful and theory-laden. Pharmacologists, for example, in their description of chemicals, emphasize the medical effects of chemicals, whereas "pure" chemists emphasis other things such as their structural properties.”

Semantics and Knowledge Organization, Birger Hjørland

There already exist lots of terms, but we need terms that are actually “Enriched Concepts”:
- With labels (names), definitions, & authorities
- “Naked terms” in an ontology / vocabulary not OK
  - E.g., what is “Litter” in the Sweet Ontology?
- Every term should be defined in natural language
- Every definition must have explicit source (authority)
  - Sources should be static reference to an unchanging authority (publication) or GUID – e.g. not a Wikipedia page URL
- Authority appropriate for science community
Loose talk (2015 Summit)— “Semantic Interfaces” or “Machine Learning”

• **Context** was IoT discussion of Semantic interoperability between heterogeneous information systems (service providers and service requestors)

  **Idea** - just develop comprehensive shared information models among the participant applications and businesses (like we always do)

• Usual problems
  - Differing standards & language about concepts which are rigid and inflexible when it comes to big data or processes
    • Hard to build semantic mediators (translators) to facilitate the needed conversion and conversations
    • Explosive complexity
    • What IoT devices have enough knowledge and smarts for what is needed?
What about Schema.org? - a Standardized/ing Vocabulary

“Schema.org has been tackling the formidable problem of:

- developing a generally accepted vocabulary that is now being used by over five million internet domains, and gradually introducing deeper semantics.”

Some Basics:

- Predecessor: data-vocabulary.org
- Adding structured, annotation information to web pages
  - Marks up contents and entities useful for search (products, broadcasting...)
- Community driven evolution and “deployed” on a large scale (~17% of all sites)
- Incorporates popular vocabularies but remains limited in coverage depth.

Ontolog Forum 2016, Domain Vocabulary Semantics
Another View: Building Ontologies using Non-Ontological Resources (NeOn)

Ontology Support Activities: Knowledge Acquisition (Elicitation); Documentation; Configuration Management; Evaluation (V&V); Assessment
Speakers

- Simon Scheider (Human Geography and Spatial Planning, Universiteit Utrecht) Ontological prerequisites for meaningful spatio-temporal analysis (maps, statistics)
- Olivier Bodenreider (NIH/NLM) Vocabulary semantics in the healthcare realm (SNOMED CT)
- Mike Bennett (Ontolog) Topic from FIBO Business Vocabulary
Some References:


Vocabulary and Mapping Tools Needed

Vocabulary mapping services are needed.

Large scale use of ontologies for the Internet and Big Data also require the use of tools to support ontology and vocabulary mapping and alignment....

Users and developers need to (naturally) use their own natural languages to both develop and use ontologies.

In many cases, the same ontologies will have to be mapped to multiple vocabularies (represented, for example, in SKOS), possibly each indistinct natural languages or used by distinct communities.

In addition, distinct ontologies, or modules of ontologies, will have to be mapped to other ontologies or otherwise aligned, to provide scalable semantics.

Tools and services to support vocabulary-to-ontology and ontology-to-ontology mapping are needed (see:Workshop on Ontology Matching (OM2013)

Many current ones employ terminology oriented approaches using things like SKOS.