

OOR For Public Sector Use Metadata Questions

Elisa Kendall

August 6, 2009

reuse challenges vary

- ∞ Many ontologies are developed for a specific purpose:
 - domain or application oriented
 - development assumptions that could impact reuse are not made explicit
- ∞ Research ontologies tend to be focused on demonstration-related content and are by nature incomplete, with varying coverage and levels of granularity due to funding limitations
- ∞ More recent ontologies are better documented, but many are also domain specific
 - <http://protege.stanford.edu/download/ontologies.html>)
 - similarly with the BioPortal (Open Biomedical Ontologies (OBO) Library), accessed via <http://www.bioontology.org/tools/portal/bioportal.html>
- ∞ Even with common metadata, specified via a registry framework such as ISO 11179, reuse is challenging without “design intent”

explicit policies for vocabulary/ontology management are key

- ∞ Linked data & mapping efforts show reuse greater for certain small-ish, fairly general vocabularies:
 - DOAP (Description of a Project) – <http://usefulinc.com/doap/>
 - Dublin Core – <http://www.dublincore.org/>
 - FOAF (Friend of a Friend) – <http://www.foaf-project.org/>
 - SKOS (Simple Knowledge Organization System) – <http://www.w3.org/2004/02/skos/>
 - SIOC (Semantically-Interlinked Online Communities) Ontology – <http://sioc-project.org/>
 - FinnONTO (National Semantic Web Ontology Project in Finland) – <http://www.seco.tkk.fi/projects/finnonto/>
- ∞ Critical factors for reuse appear to include:
 - Small development teams with larger user communities
 - Commitment to users and to continuous improvement
 - Publication of maintenance policies, URI naming conventions & policies, useful documentation
- ∞ Even well-used vocabularies receive mixed reviews for public sector applications, depending on application, metadata & provenance requirements

“good practices” for reusability

- ∞ Well-specified policies for vocabulary management, metadata, and provenance specification enable trust
- ∞ Commitment to forming, accommodating, serving, and working with a community of users is critical
- ∞ Portals such as NCOR’s BioPortal provide the library (repository), publish relevant metadata, manage versions, and provide web-based access to facilitate collaboration & reuse
- ∞ Minimal principles for vocabulary publication & management are provided in <http://www.w3.org/2006/07/SWD/Vocab/principles>
 - Use URIs for naming – publish not only the URI’s but policies for URI persistence, ownership, delegation of responsibility for specific vocabularies, etc.
 - Provide adequate readable documentation
 - Articulate maintenance policies that specify whether or not changes can be made, the process for doing so, a feedback loop so that the user community can comment on and be informed about changes
 - Identify versions – this is the minimum requirement; while ontology evolution is a research area, metadata recommendations are given in the document
 - Publish a formal schema in a recommended standard (*i.e.*, OWL, RDFS, CL)

lessons learned from ISO STEP

- ∞ Designing for reuse is critical, despite difficulties in specifying what that means
 - Results will include smaller clusters of models mapped to one another, or perhaps imported by one another to create larger federated models
 - Requires processes for determining how/when to split models or model groups as scope increases
 - Calls for tools that can manage and browse small groups of inter-related models
 - Requires a notion similar to a 'make file', for pulling smaller clusters together to create larger models, which themselves may be reusable in broader context

- ∞ Current STEP (STandard for the Exchange of Product Data) repository includes over 400 modules
 - Communities have built additional repositories around core STEP standards to add business-specific extension/content/user guides
 - There is a quality/integration review and signoff of everything that goes into the sharable repository, which frequently finds problems

** courtesy David Price, EuroSTEP*

essential metadata requirements

- ∞ Work on query answering & explanation, knowledge provenance infrastructure (Inference Web), and on a number of DoD projects indicates the critical nature of metadata (see www.ksl.stanford.edu/KSL_Abstracts/KSL-04-03.html for a number of requirements)
- ∞ Requirements range from understanding sources used, creation and revision dates, etc. at the ontology level to detailed provenance at the fact/individual level
- ∞ Reusability also depends on
 - understanding trustworthiness of sources
 - quality assessment metrics for the vocabulary & source materials
 - licensing, IP limitations
 - ease of integration with other relevant vocabularies
 - application specific requirements such as performance, security, maintainability
- ∞ A usable OOR must address at least some of these requirements to be useful from a public sector perspective
- ∞ More research is needed to determine which aspects are critical & how to approach design intent

metadata research & emerging standards

- ∞ Proof Mark-up Language (PML) 2.0 (InferenceWeb) – <http://iw.rpi.edu/documentation.html>
- ∞ OMV (Ontology Metadata Vocabulary) from AIFB/Karlsruhe – <http://ontoware.org/projects/omv/>
- ∞ ISO 11179-3 Metadata Registration & XMDR – <http://www.xmdr.org/>
- ∞ Dublin Core (<http://www.dublincore.org/>) & SKOS (Simple Knowledge Organization System), <http://www.w3.org/2006/07/SWD/>
- ∞ Research in micro-theories / micro-ontologies for version mapping, such as
 - <http://www.w3.org/2006/07/SWD/wiki/BestPracticeRecipesIssues/VersioningSnapshots>
 - <http://ontology.buffalo.edu/bfo/Versioning.pdf>
 - http://www3.lehigh.edu/images/userImages/jgs2/Page_3813/LU-CSE-06-026.pdf
 - <http://semweb4j.org/site/semversion/SemVersion>