

Metadata Support for OMG's Emerging Ontology & Vocabulary Management Initiative

Elisa Kendall Sandpiper Software

April 10, 2008



Management Information RFI

- ∞ Specific areas of knowledge requested:
 - Application experience
 - Tooling
 - Tool interoperability
 - Querying and accessibility
 - Knowledge management and mapping
 - Standards of practice
 - Related or competing standards activities
 - Example repositories
- ∞ Managed ontologies/vocabularies may include metadata supporting a number of these areas



- ∞ http://www.omg.org/cgi-bin/doc?ontology/08-03-02
- ∞ Anyone who wishes may respond
- ∞ Emphasis on 3 issues in content management
 - provenance where the information comes from
 - effectivity at what time, location, and/or use is the content applicable or valid
 - evolution how we track change



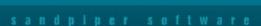
policies for vocabulary management are key

- ∞ Recent 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 success 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



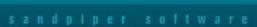
"good practices" for reusability

- ∞ Emerging 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., ODM/XMI, OWL, RDFS, SBVR, & perhaps mappings between them)



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 interrelated models; metadata must facilitate this
 - 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; again, metadata is key
- - 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



essential metadata requirements

- Our 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



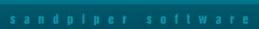
metadata research & emerging standards

- ∞ OMV (Ontology Metadata Vocabulary) from AIFB/Karlsruhe http://ontoware.org/projects/omv/
- 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/ServingSnapshots
 - http://ontology.buffalo.edu/bfo/Versioning.pdf,
 - http://www3.lehigh.edu/images/userlmages/jgs2/Page_3813/L U-CSE-06-026.pdf
 - http://semweb4j.org/site/semversion/ SemVersion



issues we see

- ∞ Content assessment
- ∞ Criteria & procedures for acceptance
- ∞ Freshness management
- ∞ Finding desired content
 - Annotations
 - Naming
 - Query support
- ∞ Persistence
 - Coherent organization
 - Version management strategies & support



concrete steps to address them

- Design a repository structure, version strategy, & naming conventions
- ∞ Determine metrics for content assessment / evaluation
- ∞ Create rules & procedures for content acceptance
- Determine mechanisms for content annotation / classification & querying
- ∞ Create a strategy/schedule for deployment at OMG



Contacts

- ∞ Evan K. Wallace -- co-Chair Ontology PSIG
 - ewallace@nist.gov
- ∞ Elisa Kendall -- co-Chair Ontology PSIG
 - ekendall@sandsoft.com
- Mark Linehan -- Chair Date-Time RFP submission team
 - mlinehan@us.ibm.com