Open Ontology Repository Initiative

Frank Olken

Lawrence Berkeley National Laboratory
National Science Foundation
folken@nsf.gov

presented to
CENDI/NKOS Workshop
World Bank
Sept. 11, 2008

Version 6.0

DISCLAIMER

- Remarks, opinions, etc. about the Open Ontology Repository (OOR) Initiative are solely my own and do not reflect the position(s) of either Lawrence Berkeley National Laboratory or the National Science Foundation.

Joint Work

This is joint work by many persons: Peter Yim, Leo Obrst, Mike Dean, Michelle Raymond, Mark Musen, Barry Smith, Fabian Neuhaus, Michael Gruninger, Pat Hayes, Steve Ray, Ravi Sharma, Natasha Noy, Deborah McGuiness, Pat Cassidy, Elisa Kendall, Evan Wallace, John Sowa, myself (Frank Olken), et al.

Many of these slides have been adapted from the OOR Communique without attribution.

What is the OOR Initiative?

- **Open** = accessible, minimal intellectual property encumbrances on the ontologies, preferably open source code for the repository
- **Ontology** = formal conceptualization (degree of formalization may vary: frames, graphs (RDF), logic (OWL-DL, Common Logic, ...))
- **Repository** = collection of ontologies, related materials, support for storage, retrieval, integration, etc.
- **Initiative =** group/effort to create OOR

Definition of Ontology Repository

"An ontology repository is a facility where ontologies and related information artifacts can be stored, retrieved and managed."

Goals of OOR Initiative

- To promote global use and sharing of ontologies by:
  - Establishing a hosted registry-repository;
  - Enabling open, federated, collaborative ontology repositories, and
  - Establishing best practices for expressing interoperable ontologies and taxonomies in repositories.

Why OOR?

- To provide access to good quality ontologies for use in a variety of software, and for education
- To provide a place for persons to contribute new ontologies
- To provide a means for ontology queries, ontology integration
- To facilitate development of new, composite ontologies

Use Cases

- Store a new ontology about time.
- Find me a good ontology about time.
- Find me a concept “duration” in ontology X
- What is relationship between “event” in ontology X and “occurrence” in ontology Y?
- Find all “subclasses” of “event” in ontology Z
- Find immediate superclasses of “event” in ontology W
- Find a mapping between SIC and NAIC classifications

Ontology Formality?

- Degree and nature of formalization of ontologies will vary:
  - Frames, SKOS, RDF, Description Logics, First Order Logic, Higher Order Logics
- Formalizable:
  - KR language has known syntax
- Formal:
  - KR language with well specified syntax
  - KR language with well specified semantics (e.g., a model theory)

Macro-level Operations

- Macro-level = level of entire ontologies
- Support storage, retrieval, query, annotation at the ontology level
- Relationships among ontologies:
  - Extends, specializes, revises, ...

Micro-level Operations

- Micro-level:
  - Partitions of ontologies
  - Individual concepts, axioms, ...
- Micro-level Operations:
  - Storage, querying, retrieval, updates, annotation
  - Inference
  - Mapping between concepts in different ontologies
  - Composition, integration of ontologies, portions of ontologies

Ontology Admission Criteria

- Required attributes of an ontology for inclusion in OOR:
  - in a publicly described language and format.
  - read accessible.
  - in a formal language with a well-defined syntax.
  - has the required metadata.
  - has a clearly specified and clearly delineated scope.
  - versions of the ontology are clearly identified.
  - appropriately named.

Quality Control

- Build a repository of “good quality” ontologies
- There will be an editorial function by which ontologies will be reviewed to assess their “quality”
- Ontologies will be annotated with “quality assessments”.
- Quality assessments are intend as a guide to users, ontology developers, integrators
- Details yet to be settled, cf. ISO 11179

Metadata Uses

- determine whether an ontology is suitable for a user purpose;
- capture the design rationales that underlie the ontology;
- find information about author, author credentials, and source of ontology reference material
- retrieve ontologies for use in domain applications;
- retrieve ontologies to be integrated with other ontologies;
- retrieve ontologies that will be extended to create new ontologies;

Types of Metadata

- Knowledge Representation Language
- Modularity (module structure)
- Relationships among ontologies
- Provenance
- Version
- Existing applications of the ontology (e.g. interoperability, search, decision support)
- Domain-specificity

Provenance Metadata

- Many ontologies are based on definitions taken from government regulations, legislation, court decisions, standards, ...
- Need fine grained provenance, citations for individual concepts, axioms which cite authoritative sources for each item in ontology
- This is especially important for ontologies used for administrative or regulatory applications in federal agencies (e.g., EPA, FDA, OSHA, IRS, ...)

Repository Architecture

- Federated
- Service Oriented Architecture (SOA)
- Scalable.
- Optimized for sharing, collaboration and reuse.
- Supporting ontologies in multiple formats and levels of formalism.
- Distributed repositories.
- Explicit machine usable/accessible formal semantics for the meta-model of the repository.

Repository Architecture (cont.)

- A mechanism to address intellectual property and related legal issues/problems.
- Support for adding, searching and mapping across ontologies and data related to the stored ontologies.
- Support additional services both directly within the province of the repository and as external services.
- Support all phases of the ontology life cycle.

Repository Services

- Ontology creation tools
- Ontology editors
- Ontology differencing tools
- Ontology modularization tools (clustering, etc.)
- Ontology export
- Ontology visualization (e.g., graph visualization)
- Version management
- Access control
- Ontology syntax checking
- Ontology consistency checking – no cycles in taxonomies

Status of OOR Initiative

- **Ontology Summit in April of 2008**
- **Ongoing teleconference discussions of:**
  - **Integration of component technologies, gathering use cases & requirements, OOR architecture, open source code development, knowledge representation language, specification of ontology mappings, outreach to early adopters, ...**
- **Early discussions about:**
  - **collaboration between contributing institutions, preparation of OOR research proposal(s) to funding agencies**
  - **Producing a joint OOR-Ontolog panel series**
- **See: http://ontolog.cim3.net/cgi-bin/wiki.pl?OpenOntologyRepository**

How to participate?

- **Read the OOR Wiki:**
  - **http://ontolog.cim3.net/cgi-bin/wiki.pl?OpenOntologyRepository**
- **Look at Ontology Summit 2008**
  - **http://ontolog.cim3.net/cgi-bin/wiki.pl?OntologySummit2008**
- **Listen to previous teleconferences, read mailing list archive**
- **Contact peter.yim@cim3.com to join mailing list**
- **Join the teleconferences**
- **Read / Contribute to email discussion**
- **Help write the OOR proposal(s)**

Supplementary Material

Repository Goals

- Supporting the Open Ontology Repository (OOR) Initiative that will promote the global use of ontologies, their instance bases, rules, and services, and mappings among these.
- Enabling and facilitating open, federated, collaborative ontology repositories.
- Establishing best practices for expressing interoperable ontology work in open registries/repositories.
- Enabling and facilitating the development of common services to support the repository and to extend the capabilities available to providers, users, and developers who use the repository.

Repository Management

- enforce access policies
- enforce submission policies
- enforce governance policies
- enforce change management policies
- control user and administrator access

Registry vs. Repository

- Registry:
  - Contains a list of ontologies and some related metadata about the ontologies
- Repository:
  - Contains actual ontologies, not just a list of ontologies
  - Supports storage, retrieval, updates, integration of ontologies

Repository Functions

- create usage reports
- validate syntax
- check logical consistency
- automatically categorize a submission

Repository Search Criteria

- domain
- author/creator/source
- version
- language
- terminology and controlled vocabularies
- quality
- mapping
- inference

Related Work

- XMDR project – metadata repositories (LBNL, UC Berkeley, et al.) ISO 11179 prototype
- Bioportal, Protege [medical ontologies, ontology editor, etc.](Stanford)
- Swoogle (UMBC) ontology search engine
- OBO [biological ontologies, frame based]
- Ecocyc, etc. [microbial ontologies, frame based] (SRI)
- OMG ODM (Ontology Definition Metamodel)

Knowledge Representation Languages

- Resource Description Framework (RDF)
- OWL-DL (description logic)
- OWL-Full (first order logic)
- Common Logic (first order logic)
- OBO (frame based)

Additional Resources

- http://ontolog.cim3.net/cgi-bin/wiki.pl?OntologySummit2008_Communique
- http://ontolog.cim3.net/cgi-bin/wiki.pl?OntologySummit2008
- http://ontolog.cim3.net/cgi-bin/wiki.pl?OpenOntologyRepository

Contact Info

- **Frank Olken**
  - folken@nsf.gov
  - **703-292-7350**
  - **National Science Foundation**
    - Computer and Information Science and Engineering Directorate
    - Intelligent Information Systems Division
    - Information Integration and Informatics Program
    - 4201 Wilson Blvd., Suite 1125, Arlington, VA 22230

OOR Initiative Contacts

- **Peter Yim**
  - **peter.yim@cim3.com**
- **Leo Obrst (MITRE)**
  - **lobrst@mitre.org**
- **Michelle Raymond (Honeywell)**
  - **michellearaymond@gmail.com**
- **Mark Musen (Stanford)**
  - **musen@stanford.edu**
- **Mike Dean (BBN)**
  - **mdean@bbn.com**

Acknowledgments

- My portion of this work has been supported by the National Science Foundation under the Individual Research Plan, as a part of my IPA appointment to NSF from Lawrence Berkeley National Laboratory.