The Open Ontology Repository Initiative: Requirements and Research Challenges

Ken Baclawski
Todd Schneider

OOR Charter

Promote the global use and sharing of ontologies by:

- establishing a hosted registry-repository;
- enabling and facilitating open, federated, collaborative ontology repositories, and
- establishing best practices for expressing interoperable ontology and taxonomy work in registry-repositories.

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

Why?

- Isn't the Semantic Web notion of distributed islands of semantics sufficient as a de facto repository?
- If you put it out there, will they come?
- If you build it better and put it out there, will they prefer yours?
- History does not show this laissez faire "field of dreams" is good reality

Okay, but why a new infrastructure?

- The “clickable” web has been very successful in employing a “lazy strategy” for HTML documents.
- However the use and content of the Semantic Web has different characteristics (next slide) that make it far less tolerant of the change and frequent errors which are commonplace on the clickable web.

SemWeb Distinguishing Characteristics

- Machines are the primary consumers
  - Humans can handle errors and noisy content.
  - This is likely to be fatal for machine processing.
- Import dependencies
  - owl:imports introduces a strong transitive dependency between ontology documents.
  - Changes in an imported document have serious consequences.
    - Can cause inconsistencies.
    - Can change meaning significantly.

Open Ontology Repository

- *"An ontology repository is a facility where ontologies and related information artefacts can be stored, retrieved and managed"*
- The persistent store
- The registry
- Value-added services

Open Ontology Repository

All types of artefacts on the ontology spectrum

Open Ontology Repository

- Folksonomies
- Terminologies
- Controlled vocabularies
- Taxonomies
- Thesauri
- Data schemata

Open Ontology Repository

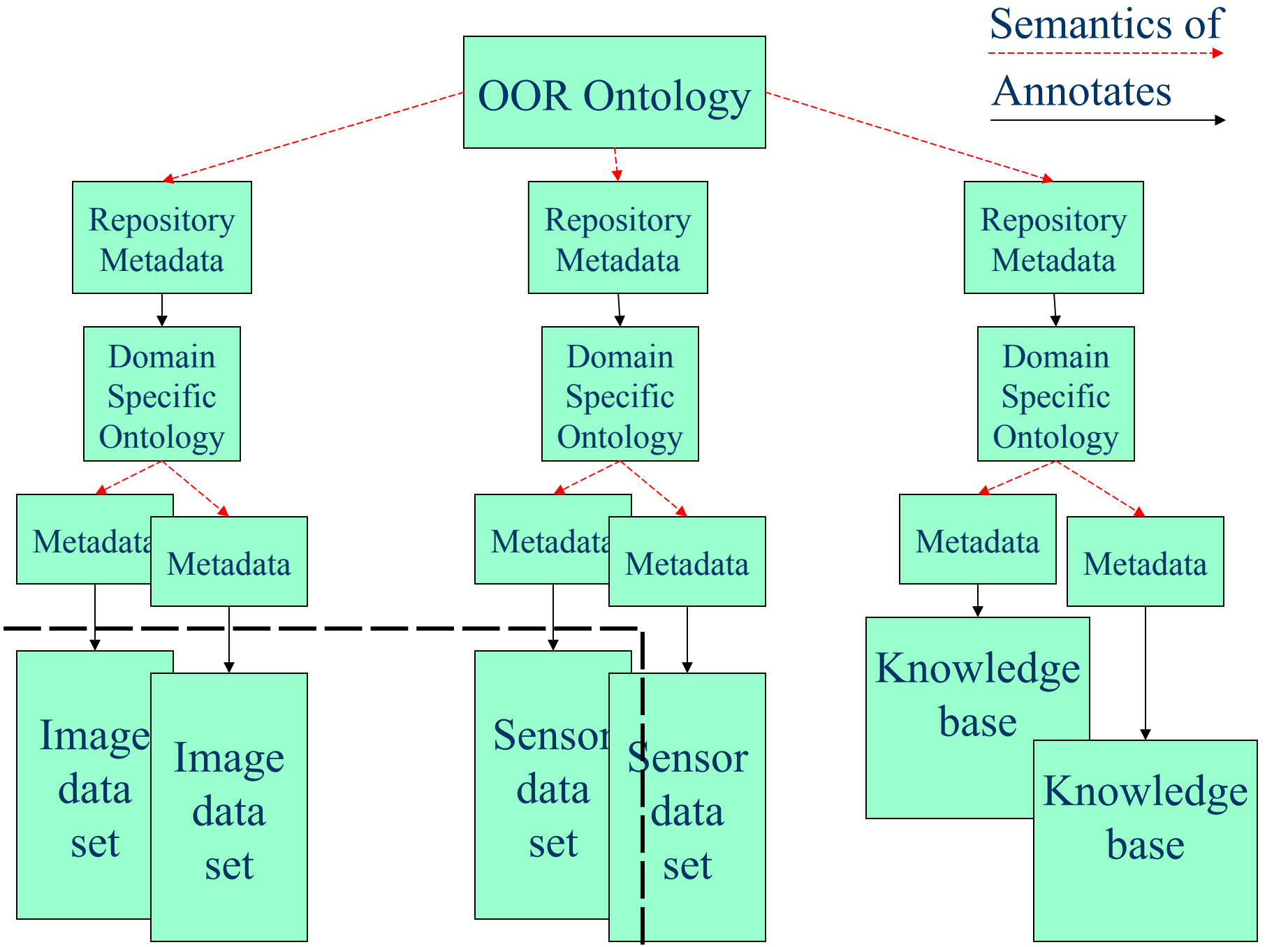
- Open access
  - compliance with open standards
  - open technology (with open source)
  - open knowledge (open content)
  - open collaboration (transparent community process)
- Open to integration with “non-open” repositories via an open interface

Purpose and Scope of the OOR Initiative

- Limited to providing an infrastructure that enables ontology and metadata management
- Formal architecture
  - Interfaces
  - Required services
  - Enable interoperability among OOR instances
- Reference implementation
  - Basic services
  - Basic metadata lifecycle

Metadata management

- The next slide shows an organization suitable for data sharing and interoperability within domains.
- Cross-domain interoperability requires relationships between ontologies
  - Import relationships
  - Ontology mapping/mediation relationships


Semantics of
Annotates
OOR Ontology
Repository Metadata
Repository Metadata
Repository Metadata
Domain Specific Ontology
Domain Specific Ontology
Domain Specific Ontology
Metadata
Metadata
Metadata
Metadata
Metadata
Metadata
Image data set
Image data set
Sensor data set
Sensor data set
Knowledge base
Knowledge base

OOR Value Added: Part 1

- The OOR is reliably available
- The OOR is persistent and sustainable, so you can be confident when committing to its use
- The OOR has information about when, why, and how an ontology has changed, so you can be aware of changes that may effect its usability
- Ontologies and metadata can easily be found.

OOR Value Added: Part 2

- Ontologies and metadata are registered, so you know who built them
- Metadata provides the ontology purpose, KR language, user group, content subject area, etc.
- The OOR includes mappings, so you can connect ontologies to other ontologies
- The OOR content has quality and value, as gauged by recognized criteria

OOR Value Added: Part 3

- The OOR enables services, so that ontologies can map and be mapped, find and be found, can review/certify and be reviewed/certified
  - Use services that others have developed, or
  - Plug in your own services
- Ontologies can reuse or extend other ontologies, including common middle and upper ontologies
- The OOR can be extended

Top Requirements

- A well-maintained persistent store (with high availability and performance) where ontological work can be stored, shared and accessed
- Properly registering and “governing” ontologies, with provenance and versioning support, made available (logically) in one place so that they can be browsed, discovered, queried, analysed, validated and reused
- Allow ontologies to be “open” and unencumbered by IPR constraints, in terms of access and reuse

Top Requirements

- Providing services across disparate ontological artefacts to support cross-domain interoperability, mapping, application and making inferences.
- Registering semantic services to support peer OORs
- OOR Use Case development is ongoing.

Research Challenges

- Computational Complexity
- Inadequacies of current repositories
- Representation languages
- Policies and best practices
- Outreach and education

Computational Complexity

- Repositories permit the creation of large ontologies by import and mediation.
- Complexity of logical inference increases more rapidly than traditional database query complexity

Current repositories

- Repositories already have many important features: registration/upload, browsing/search, metrics/statistics,...
- Ontologies are generally treated as independent entities, although this is now changing.
- Lack of repository standards
- No federation ability

Representation languages

- The diversity of metadata and ontology languages is large and growing.
- Languages for representing relationships are emerging/diverging.
- Transformations between representations don't preserve semantics.
- A CL repository could address this issue.

Policies and best practices

- Intellectual property concerns
- Policies and procedures
  - Naming conventions and policies
  - Maintenance policies
  - Documentation
- Provenance to enable trust

Outreach and Education

- Integration with other semantic tools
  - Semantic wikis
  - Others?
- Education
  - Documentation is not enough
- Community outreach
  - Meet with representatives of the community
  - Assist in transition to ontology repositories