



# The SIMILE Project: Extending the Metadata Capabilities of the DSpace System

Robert Tansley  
Hewlett-Packard Laboratories



- SIMILE: Semantic Interoperability of Metadata and Information in unLike Environments
- 3-year collaboration started February 2003
- MIT Lab for Computer Science
  - Haystack Project/David Karger
- Worldwide Web Consortium
  - Semantic Web Advanced Development Activity/Eric Miller
- MIT Libraries
  - MacKenzie Smith
- HP Labs
  - Palo Alto CA, Cambridge MA (USA); Bristol (UK)

- DSpace open source digital repository system developed during 2-year collaboration between HP and MIT Libraries
- Metadata support is currently limited to Dublin Core and bespoke relational database (METS to follow soon)
- Different communities at MIT and elsewhere require richer metadata
  - Open CourseWare (IMS learning object metadata)
  - Image metadata (VRA Core)
  - Biomedical images (various schema)
  - MARC
  - Controlled vocabularies

# Approach



- Extend DSpace's metadata support using Semantic Web techniques
  - RDF
  - RDF Schema
  - DAML + OIL
  - OWL
- Explore primary use case
- Demonstrator by end of 2003
- Explore further use cases
- Explore architectures, metadata stores/query languages, user interface considerations
- Implement atop DSpace

# Primary Use Case: Heterogeneous Schemas and Instance Data Support



- Potential scenarios:
  - Community has no schema or metadata
  - Community has metadata/schema in non-RDF form
  - Community has metadata/schema in RDF form
- Need schema in RDF form, validation information, UI information
  - May need to convert existing instance data/schema to RDF
- Support adding and editing instance data
- Navigate and search instance data
- Add relationships with other schemas

# Simplifying Assumptions



- This use case will not consider distribution
- All relevant, canonical data and schemas are held locally

# Specific 'Sub-use Cases'



- Visual Images Support (VRA Core, controlled vocabularies)
- Learning object support (IMS metadata)
- Schema and Vocabulary Registry, Visualization, and Search
- Biomedical Images

# First Year Demonstrator



- Load images with VRA Core metadata
- Load images with IMS metadata
- Demonstrate results of queries of these 'islands'
- Introduce mappings
- Demonstrate results of queries with these mappings



# Other Use Cases



- Lifecycle Management of Digital Objects
- OCLC Authority Control Service
- Audit System
- Distributed Collections
- Mining from unstructured information

# For More Information



- <http://www.dspace.org/>
- <http://web.mit.edu/simile/>



**i n v e n t**