

Publishing and Sharing Schemas: an overview

Rachel Heery

UKOLN, University of Bath

Schemas and ontologies workshop

16th May 2003



A centre of expertise in digital information management

www.ukoln.ac.uk



JISC

re:source

Contents

- Motivation
- Concepts and definitions
- MEG Registry
- Context in Digital Library world
- Context in wider world
- Connect to ontologies



Motivation for exploring registries

- Proliferation of metadata
 - Digital libraries
 - Open archives
 - Enterprise portals
 - eCommerce
 - eScience
 - Rights and conditions of use
 - Etc etc

An appropriate metadata element set is required for every implementation....



Managing metadata schema

Need to manage metadata efficiently and effectively to support this range of services

- Facilitate data exchange
- Share and re-use metadata
- Use common tools

Whilst allowing for

- Extensibility
- Localisation to meet local requirements

Tension between alignment and differentiation



Schema Registries: enabling discovery and re-use

Ambition: To promote re-use ; layer on added value services; support evolution of vocabularies

- To encourage ‘declaring and sharing’ of data elements, element sets and application profiles
- Enabling a common approach to extensibility and localisation

First step: building a ‘schema navigation service’ to provide information about existing element sets



Possible functions of Schema Registries

- To aid navigation
- Support evolution of schemas
- Provide mappings
- Provide annotations

For users who might be

- Software (metadata editors, metadata validators, conversion tools)
- Humans (developers, implementors, cataloguers)



Types of registry: collection policy

- Sector or Domain specific
- Enterprise based
- Standards based (DCMI)
- Application profiles
- Functional (rights, resource discovery)



Design choices...

- Thin registry (links) or thick registry (database)
- Creating schemas Centralised v. distributed
- Accessible by humans v. software

- Registry as a shared collaborative Web tool ? Or as a Web Service?



Definition of Schema

A schema is a structured representation that defines and identifies the data elements in an element set

- Provides authoritative declaration of semantics of data elements
- Indicates semantic relationship between data elements
- Supports unique identification of data elements
- Schema can be expressed formally in RDF or other vocabulary description language



Extract from RDF schema

```
<rdf:Property ID="legalStatus">  
<rdfs:label>Legal Status</rdfs:label>  
<rdfs:comment>A statement of the legal status of the  
collection.</rdfs:comment>  
<rdfs:subPropertyOf  
  rdf:resource="http://purl.org/dc/elements/1.1/description"  
  />  
</rdf:Property>
```



Element

A unit of data for which the definition, identification, representation, and permissible values can be specified that is used to describe a characteristic or attribute of a resource

- Elements are ‘properties’ of things
- Elements can be documented i.e.defined, given unique identifiers
- Elements can be ‘represented’ by different syntaxes
- Rules for the value of data elements can be specified



Examples of data elements

- Title
- Author
- Creator
- Abstract
-



Definition of Element Set

An element set is coherent bounded set of elements (terms) formulated as a basis for metadata creation

- Designed for particular purpose e.g. domain-specific, resource description, rights
- Identifies designated authority



Examples of element sets

- DCMI v1.1
- MARC
- IEEE LOM
- GILS
- vCARD



Definition of Application Profile

An application profile is an element set optimised for a particular local application. It may include data elements (terms) drawn from one or more element sets. It may specify rules for values of elements.

- Application profiles are declarations of usage
- Application profile *reuse* data elements already defined elsewhere; or *use* data elements from a local element set



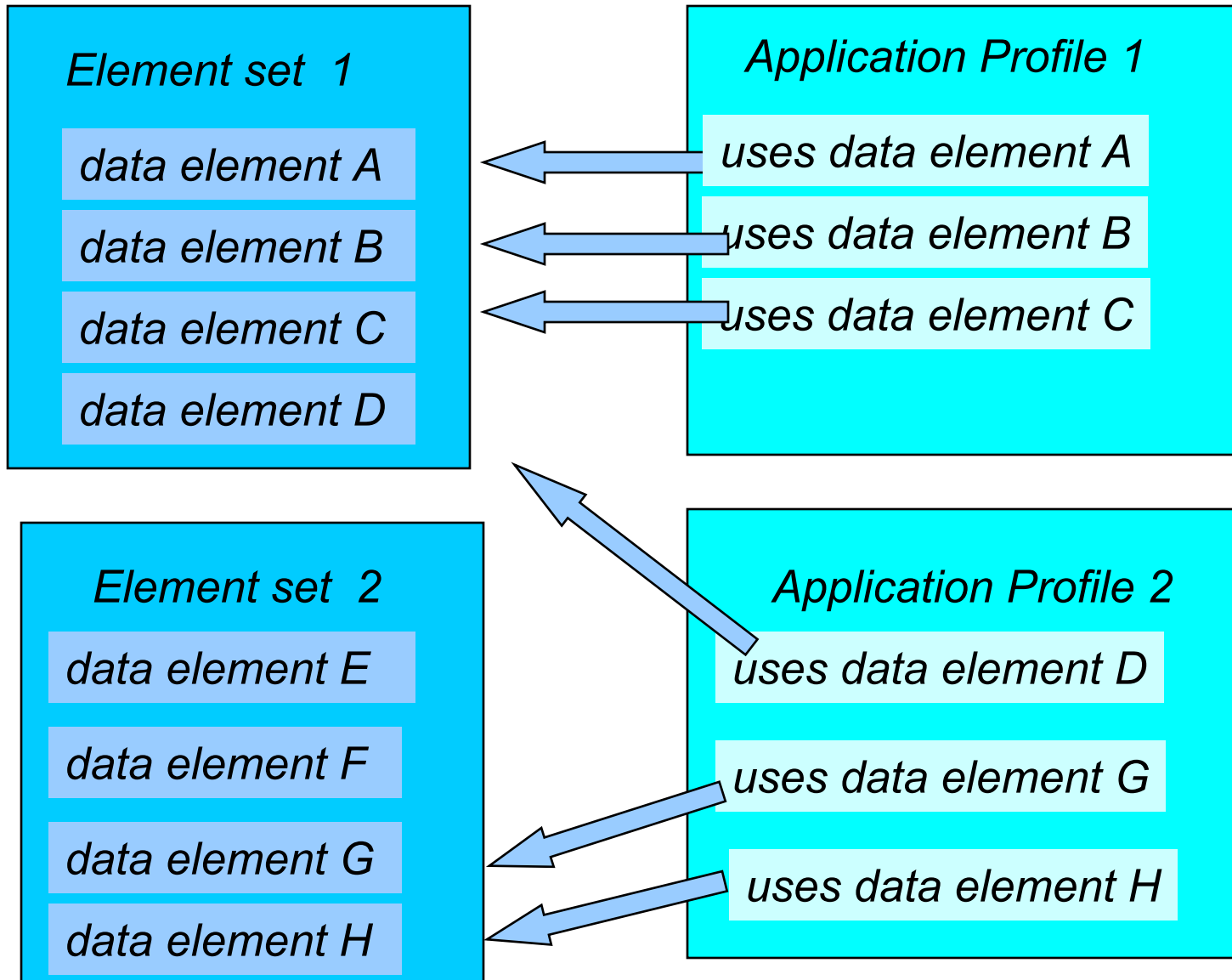
Examples of Application Profiles

- Renardus European Subject Gateway
- Australian Government Locator Service
- Food and Agricultural Organisation
- European Environment Agency

- Various UK educational initiatives
 - Qualifications and Curriculum Authority
 - Virtual Teacher Centre
- DCMI Application Profiles
 - Libraries, Government



Application Profile model



Context: within Digital Library world

- DESIRE Registry (SCHEMAS Registry)
 - Sharing schemas and application profiles (and investigating RDFS)
<http://www.ukoln.ac.uk>
- MetaForm
 - database of Dublin Core Crosswalks, crosscuts and mappings maintained at the State and University Library in Goettingen
<http://www2.sub.uni-goettingen.de/metaform/>
- DCMI Registry
 - Authoritative specification of current element sets
 - Access to multilingual definitions of data elements
<http://dublincore.org/dcregistry/index.html>



.....Trial creation and loading of schemas



Metadata for Education Group

- Facilitated by Interoperability Focus (UKOLN)
- Forum for discussing provision of educational resources at **all** levels across UK
- Encourages consistency of approaches to description
- Initially explored (re-)use of DESIRE registry (2000 – 2001)

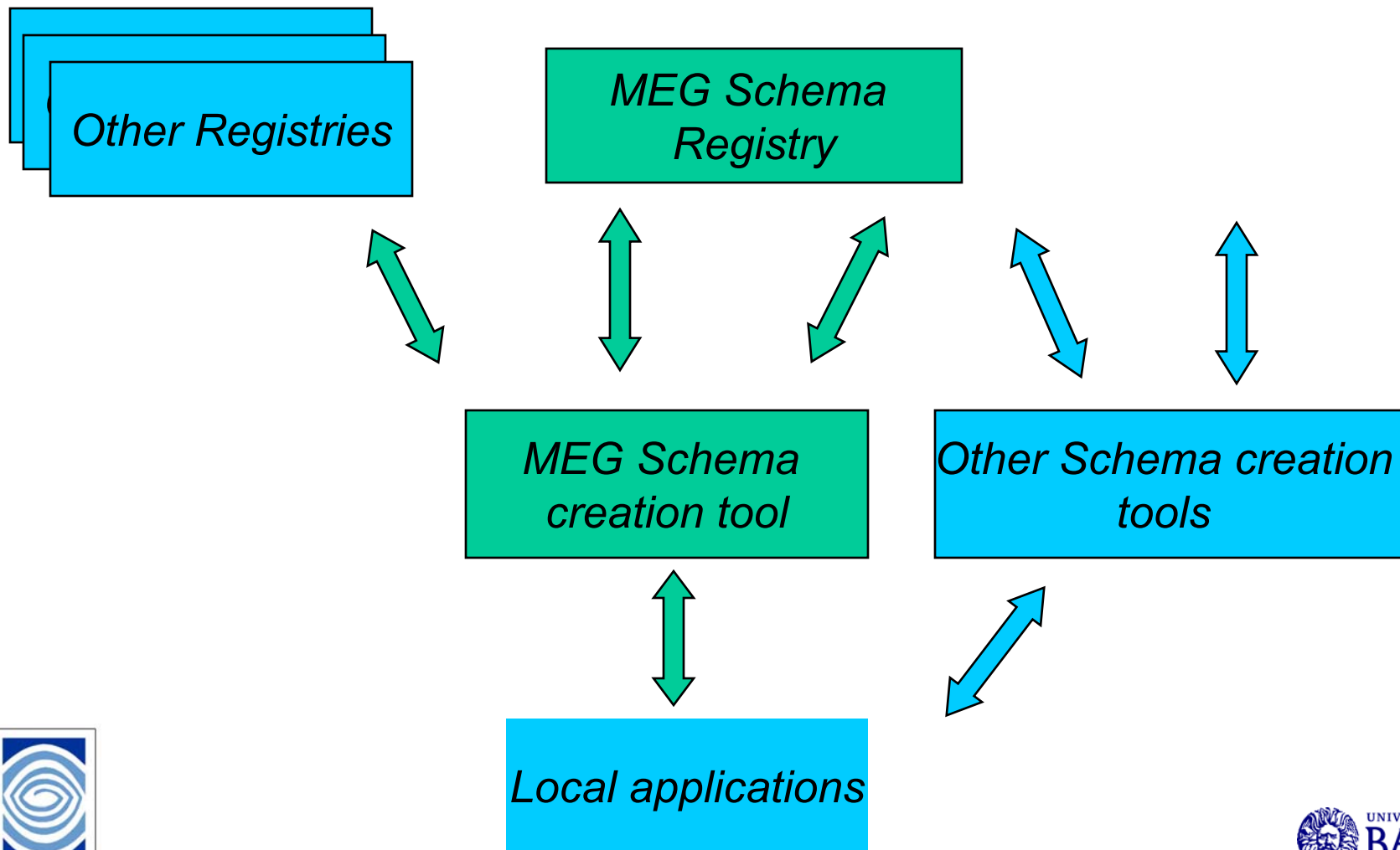


Aims of MEG Registry project

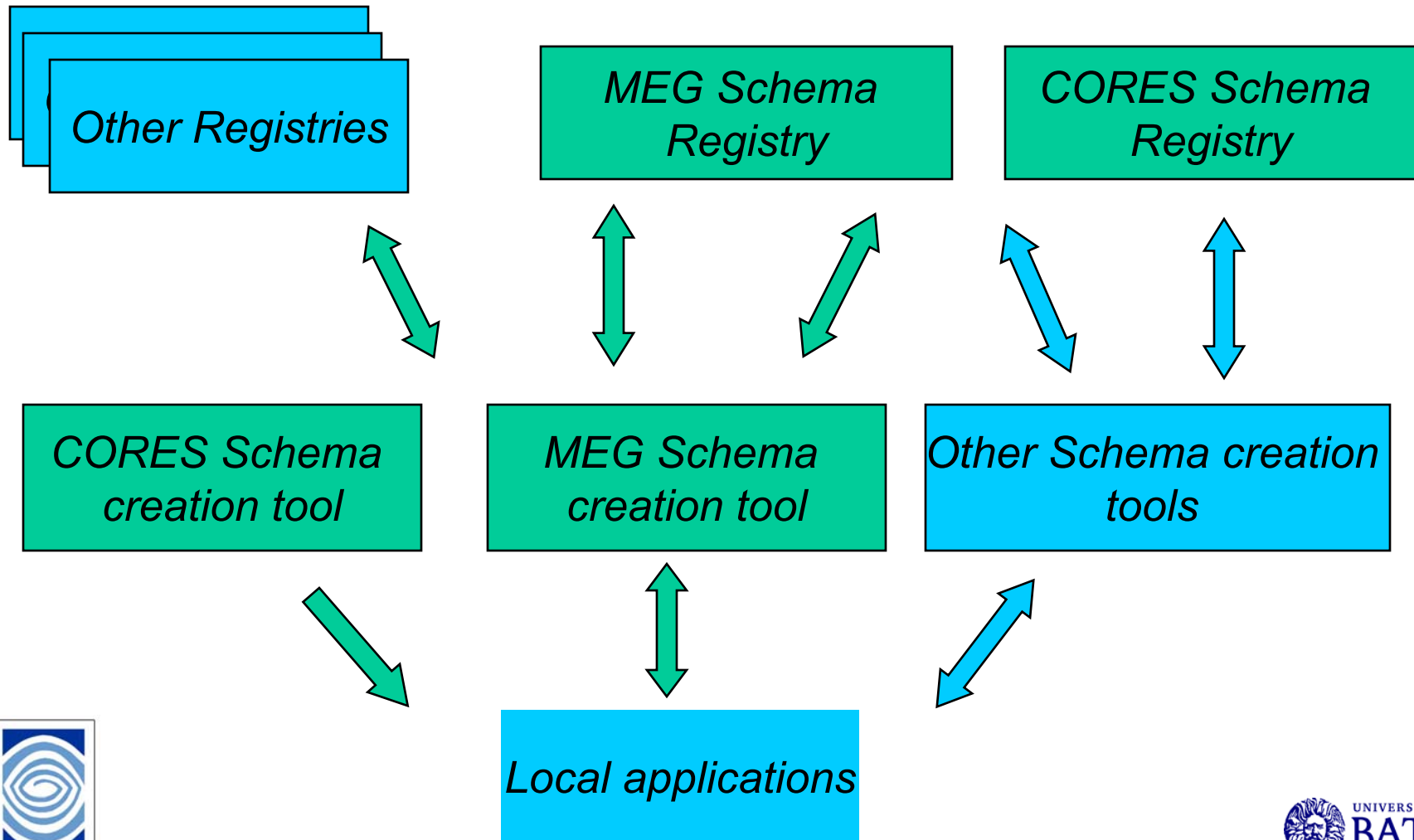
- Funded by JISC/BECTa (2002-3)
- Provide interactive environment for schema creation
 - Complementary creation tool and registry
- Enable schema creators to declare their own schemas
 - To create machine readable schemas
 - Users require no knowledge of complex schema language
 - Scalable solution
- Explore use of RDFS



Distributed schema creation and registration



Distributed schema creation and registration



Context: wider world of registries

Coverage of the Open Forum on Metadata Registries, 2003

<http://metadata-stds.org/openforum2003/>

- ISO/IEC 11179 Metadata Registries
- Universal Description, Discovery, and Integration (UDDI)
- XML Registries/Repositories
- Database Catalogs (e.g. relational DBMS/SQL)
- CASE Tool Repositories
- Software Component Registries
- Ontology Registries
- DCMI Registry



Context: ISO/IEC 11179 data registries

Specification and standardisation of data elements

- National Health Information Knowledgebase hosted by the Australian Institute of Health and Welfare <http://www.aihw.gov.au/knowledgebase/index.html>
- Environmental Data Registry hosted by the US Environmental Protection Agency <http://www.epa.gov/edr>



Environmental Data Registry

System of registries

- New System of Registries is a virtual gateway for accessing various EPA's existing Environmental Data Registry (EDR) and the Substance Registry System

EDR Data Element Comparison Matrix

- Compares elements by: name, definition, XML tag, type, identifier, status, range, example. resource acronym, values



xml.org

- The xml.org directory of XML document specifications
- Facilitates re-use of Document Type Definition (DTD)
- Hosted by the Organization for the Advancement of Structured Information Standards (OASIS)

<http://www.xml.org/xml/registry.jsp>



RDF Dictionaries

- LexML, an open source development of a multi-lingual and multi-jurisdictional RDF Dictionary for the legal world

<http://home.snafu.de/mmuller/lexmlde/rdf.htm>

- The Semantic Web Agreement Group (SWAG) Dictionary, a database of terms for the Semantic Web that can be referred to by humans and software agents

<http://groups.yahoo.com/group/swag-dev/>



How does this relate to ontologies?



Definitions of ontology

An ontology defines the terms used to describe and represent an area of knowledge. ...The word ontology has been used to describe artifacts with different degrees of structure. These range from simple taxonomies (such as the Yahoo hierarchy), to metadata schemas (such as the Dublin Core), to logical theories.

Jeff Heflin. Web Ontology Language (OWL) Use Cases and Requirements, W3C, February 2003, <http://www.w3.org/TR/webont-req/>

In this paper, we will require the following properties to hold in order to consider something an ontology. Specifications meeting these properties will be considered simple ontologies.

- *Finite controlled (extensible) vocabulary*
- *Unambiguous interpretation of classes and term relationships*
- *Strict hierarchical subclass relationships between classes*

Deborah McGuinness. Ontologies come of age. In Dieter Fensel et al. Spinning the Semantic Web, MIT, 2002. [http://www.ksl.stanford.edu/people/dlm/papers/ontologies-come-of-age-mit-press-\(with-citation\).htm](http://www.ksl.stanford.edu/people/dlm/papers/ontologies-come-of-age-mit-press-(with-citation).htm)



Ontology servers

- DAML Ontology Library
 - 194 ontologies, browsing, query by name of class or property, statistics on usage of DAML property features <http://www.daml.org/ontologies/>



Agents	http://www.cyc.com/2003/04/01/cyc
Agents	http://www.cyc.com/cyc-2-1/cyc-vocab.daml
Agreements	http://opencyc.sourceforge.net/daml/cyc.daml
Agreements	http://www.cyc.com/2002/04/08/cyc.daml
Agreements	http://www.cyc.com/2003/04/01/cyc
Agreements	http://www.cyc.com/cyc-2-1/cyc-vocab.daml
aircraft	http://opencyc.sourceforge.net/daml/cyc-transportation.daml
airport	http://www.daml.org/2001/10/html/airport-ont
airport	http://www.daml.ricmu.edu/ont/AirportCodes.daml
AirportCode	http://www.daml.ricmu.edu/ont/AirportCodes.daml
alias	http://orlando.drc.com/daml/ontology/UniversalProperty/current/
annotation	http://www.w3.org/2000/10/annotation-ns#
annotation	http://www.w3.org/2000/10/annotationType#
annotation	http://www.w3.org/2001/03/thread
Army	http://orlando.drc.com/daml/Ontology/Thesaurus/CALL/current/
Art	http://mr.teknowledge.com/DAML/ArtOntology.daml
assembly	http://edge.mcs.drexel.edu/MUG/2001/05/16/sbf.daml
association	http://orlando.drc.com/daml/Ontology/POC/current/
Assorted	http://purl.org/net/swn
aviation	http://mnemosyne.umd.edu/~aelkiss/weather-ont.daml
baseball	http://www.daml.org/2001/08/baseball/baseball-ont
Beer	http://www.cs.umd.edu/projects/plus/DAML/onts/beer1.0.daml
Bibliographic	http://orlando.drc.com/daml/ontology/Bibliographic/current/
Bibliographies	http://www.daml.ricmu.edu/ont/homework/atlas-publications.daml
Bibliography	http://www.cs.yale.edu/~dvm/daml/bib-ont.daml
bibliography	http://www.isi.edu/webscripiter/bibtex.o.daml
BibTex	http://www.isi.edu/webscripiter/bibtex.o.daml
hibtex entries	http://www.daml.ricmu.edu/ont/homework/cmu-ri-nublications-ont.daml

Ontology servers (2)

- Ontosaurus
 - Browser providing query and browse interface for knowledge bases
<http://www.isi.edu/isd/ontosaurus.html>



? Theory: AIRCRAFT Show View... Hold Window Options...

any Find Match Substring

Concept ENTITY

Theory: [AIRCRAFT](#) Package: TEST

[Find Matching Instances](#) [Bookmark](#)

Definition

(defconcept **ENTITY**)

Child Concepts (depth 12345)

ENTITY

[PHYSICAL-ENTITY](#)

[ENGINE](#)

[DIESEL-ENGINE](#)

[GAS-TURBINE-ENGINE](#)

[GASOLINE-ENGINE](#)

[JET-ENGINE](#)

...

[PROPELLOR-ENGINE](#)

...

[RUBBER-BAND-ENGINE](#)

[STEAM-ENGINE](#)

[FLYING-OBJECT](#)

[AIRCRAFT](#)

Ontology servers (3)

- SHOE repository
 - provides descriptions and links to existing SHOE ontologies
<http://www.cs.umd.edu/projects/plus/SHOE/onts/>
- WebODE
 - links to a number of ontologies
<http://babage.dia.fi.upm.es/webode/>
- WebONTO
 - supports the collaborative browsing, creation and editing of ontologies
<http://kmi.open.ac.uk/projects/webonto/>
- Stanford Knowledge Systems Laboratory Server
 - ontology repository
<http://www-ksl-svc.stanford.edu:5915>



Commonalities between metadata and ontology initiatives

- Similar interest in developing ‘core ontologies’ and ‘core ‘ element sets, which are
 - Flexible
 - Extensible
 - Have shared data model
- Similar interest in integration between multiple schema/ontologies
- Similar interest in developing language to express semantics
- Possibility of shared tools



Differences...

- Scale
- Complexity
- Focus on semantics
 - Metadata schemas focus on definitions
 - Ontologies focus on relationships between terms
- Instantiation
 - Metadata Schemas focus on ‘metadata instances’ describing resources
 - Ontologies on delineation of knowledge space



Metadata concepts

Layer 3	(a) Attribute Space e.g LOM, Dublin Core MES, indecs	(b) Value Space e.g. classification schemes, controlled vocabularies, taxonomies
Layer 2	Representation e.g. XML, RDF, DAML-OIL	
Layer 1	Transport and Exchange e.g HTTP Get, OAI Protocol for Metadata Harvesting	

From: Thomas Baker et al., Principles of Metadata Registries, DELOS 2003



MEG case study.....

