

Next Generation of Metadata Registries and Open Ontology Registries

☞ Workshop on ROR, On-Demand Model Selection and
Semantic Metadata Mapping Process

SKLSE (State Key Laboratory of Software Engineering), Wuhan University, Wuhan,
China, September 4-6, 2008

Bruce Bargmeyer
Lawrence Berkeley National Laboratory
University of California, Berkeley
Tel: +1 510-495-2905
bebargmeyer@lbl.gov



Notions Useful for This Workshop



- ☞ Issues for WG 2 (from Horiuchi-san)
- ☞ NCI advances linking terminology to data (metadata)
- ☞ ISO/IEC11179 Edition 3
- ☞ Open Ontology Repository
- ☞ BioPortal and XMDR

Issues for WG2

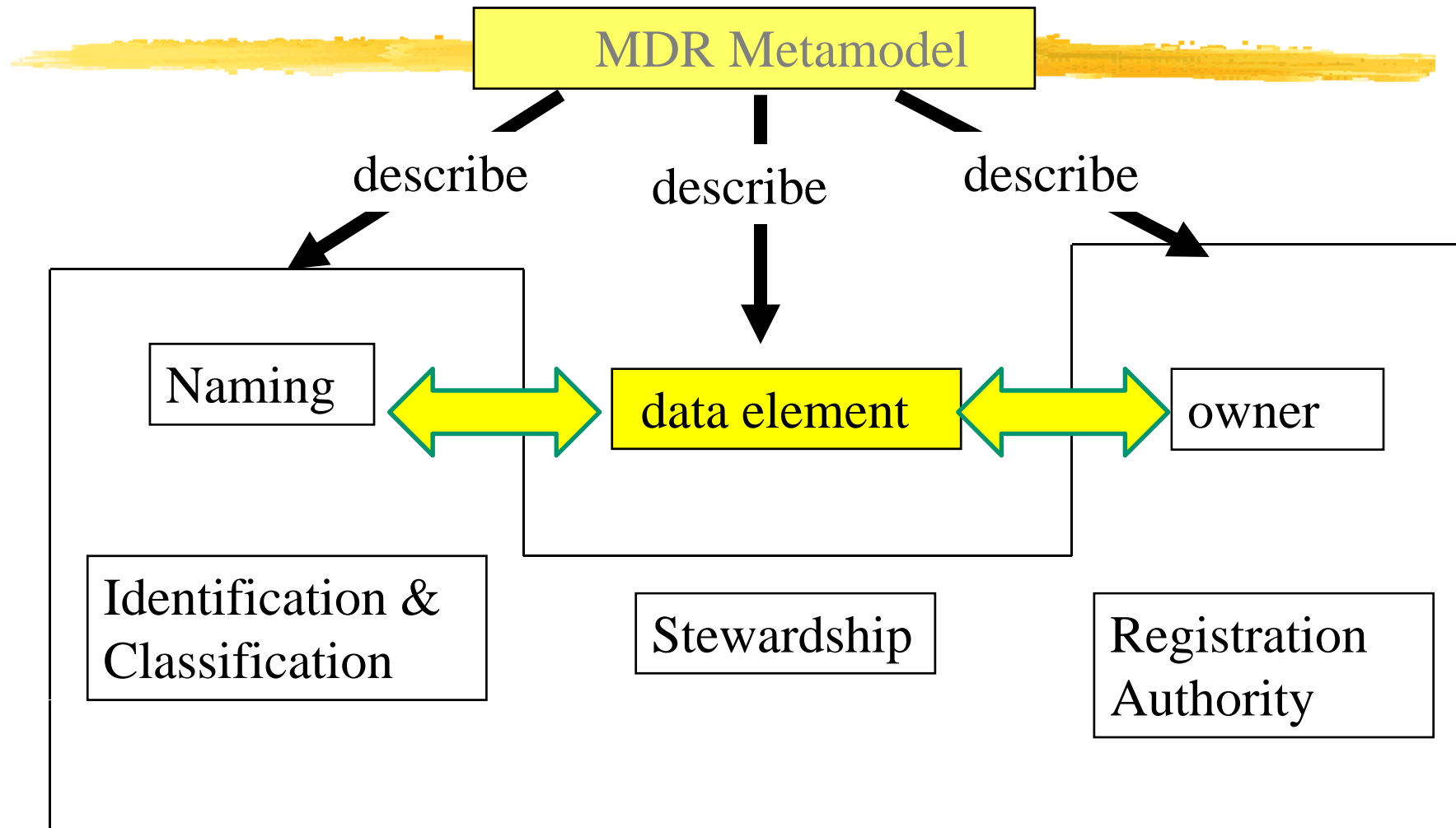


☞ Consolidation of MDR & MFI

☞ Service Interfaces

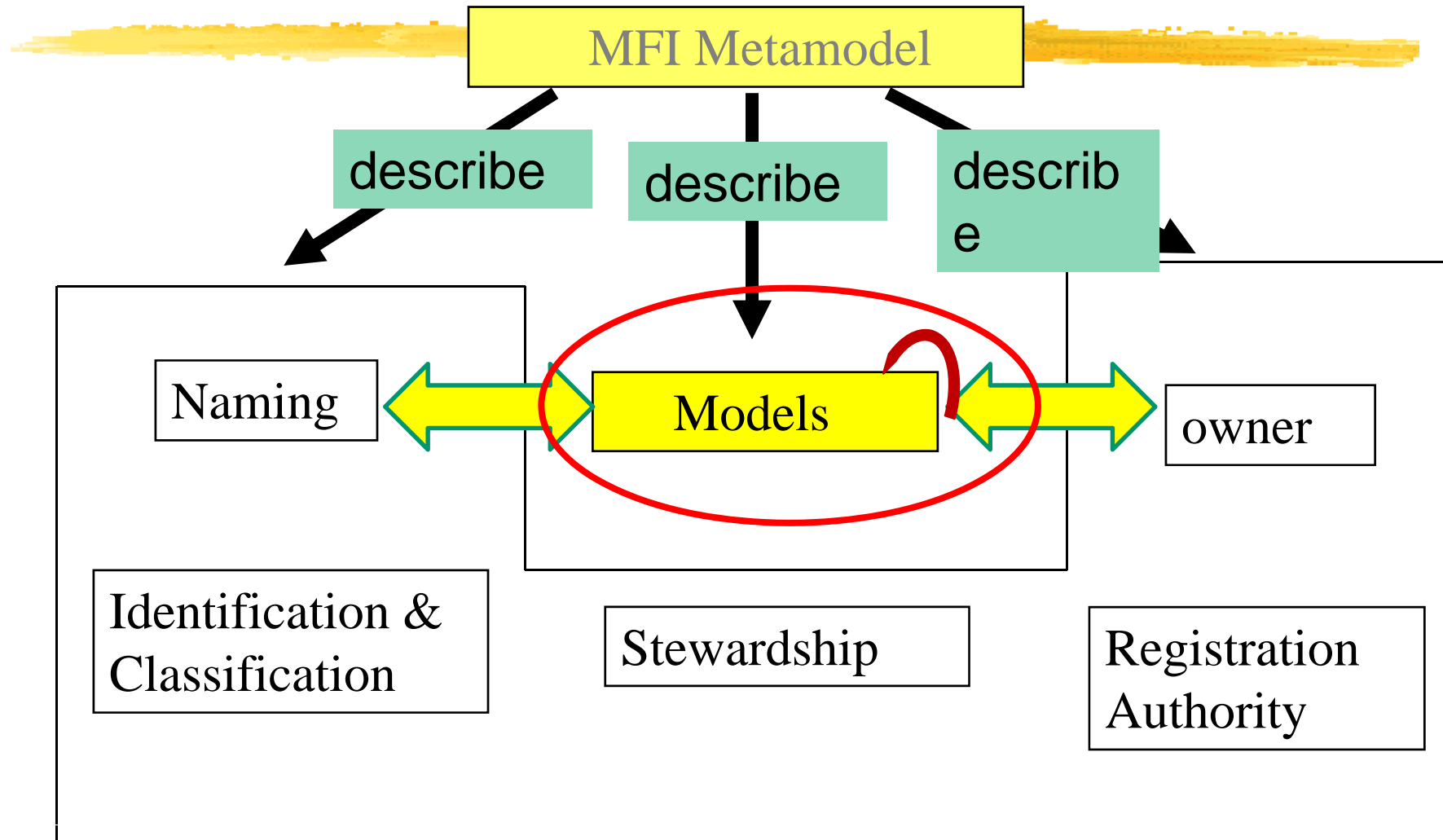
From presentation given by Prof. Hajime Horiuchi at Sydney OFMR, May 2008

Overview of MDR Metamodel

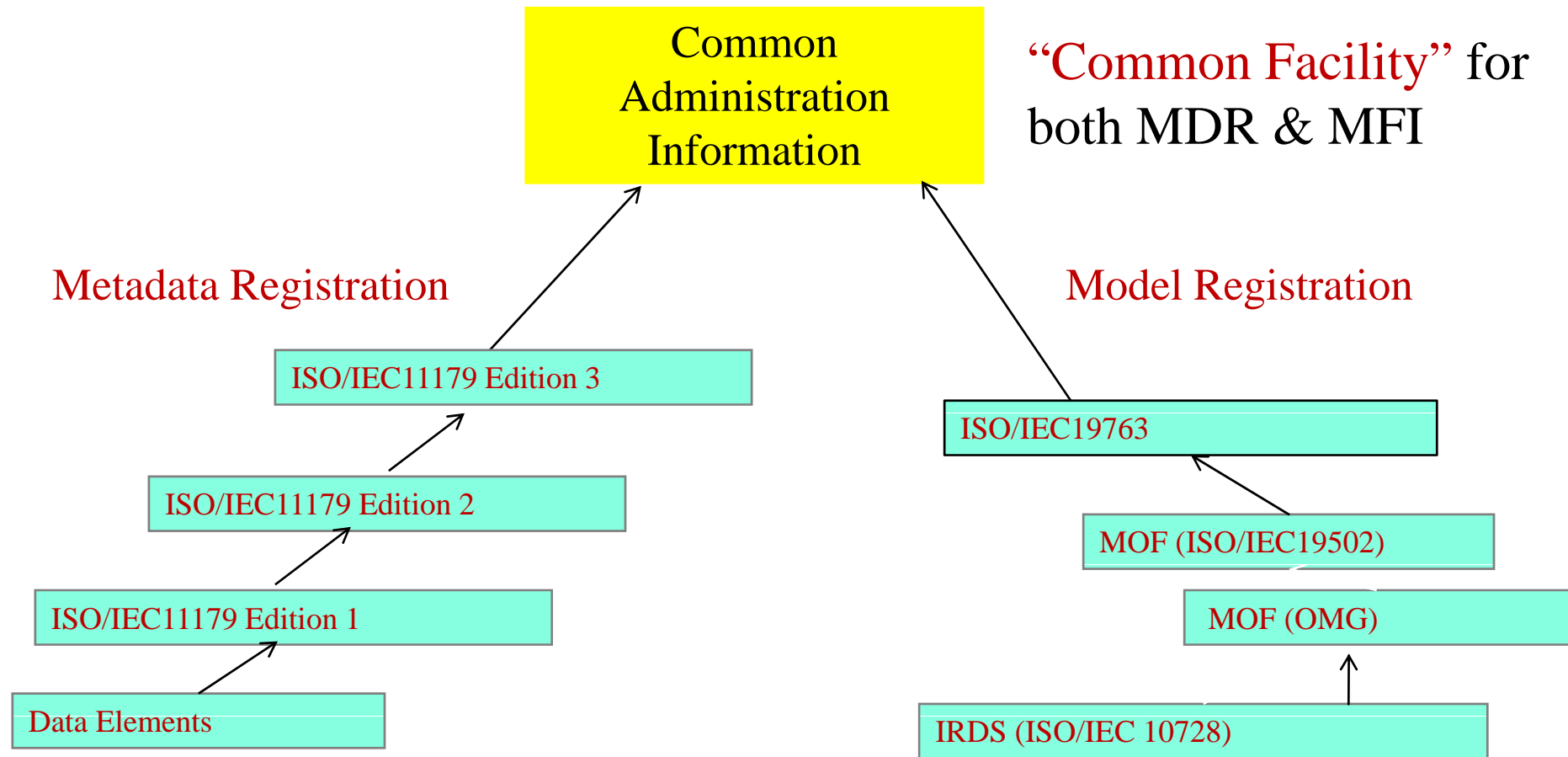


From presentation given by Prof. Hajime Horiuchi at Sydney
OFMR, May 2008

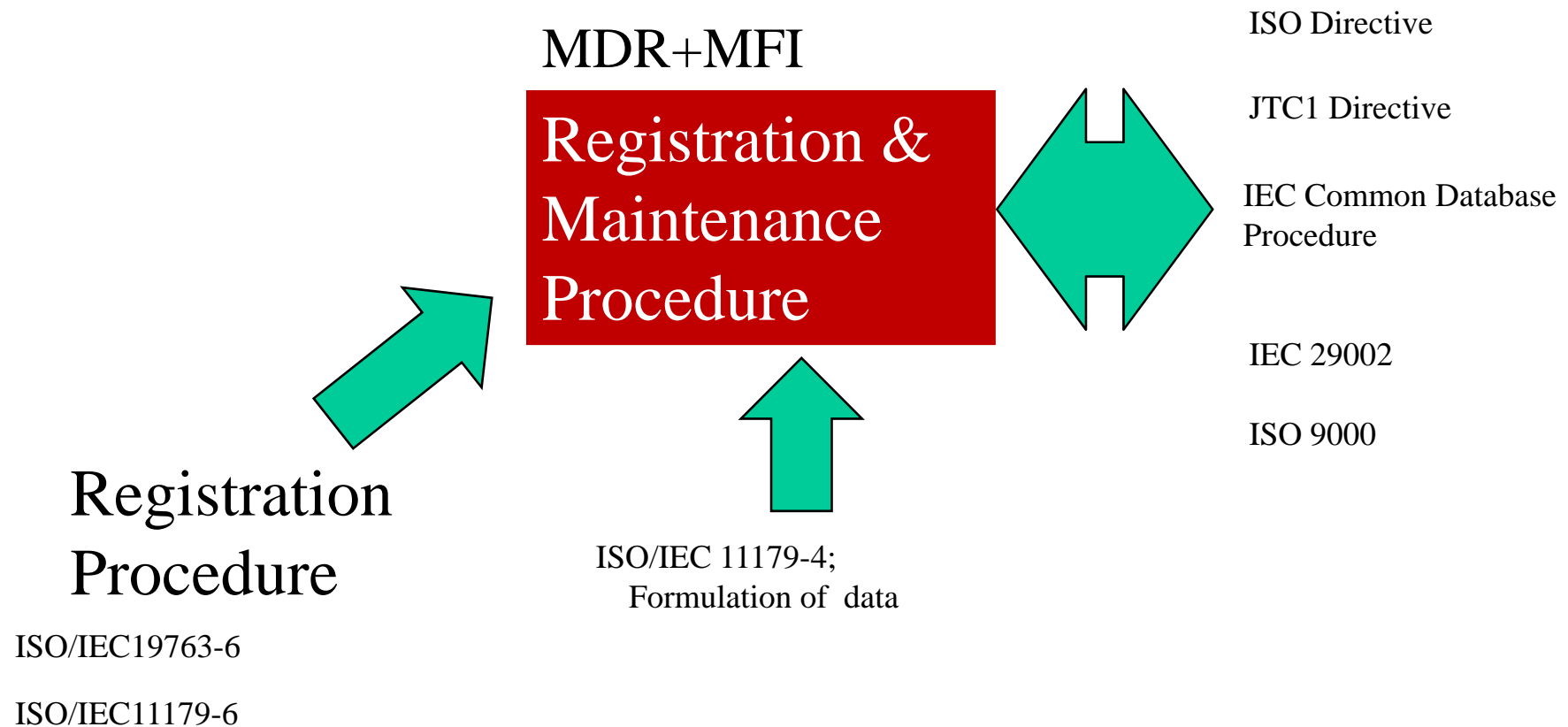
Overview of MFI Metamodel



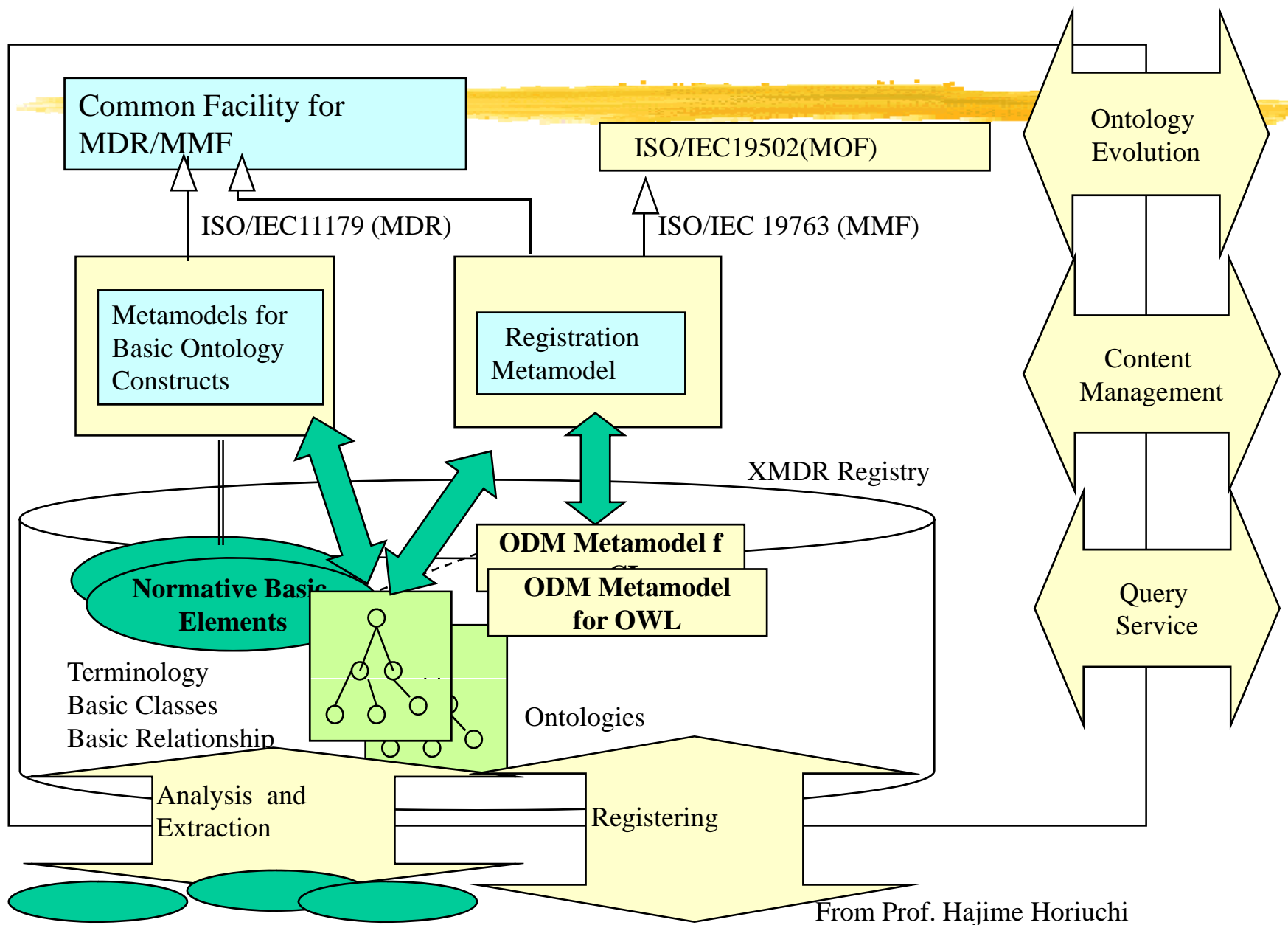
Expected Evolution of MDR & MFI (1)



Expected Evolution of MDR & MFI (2)

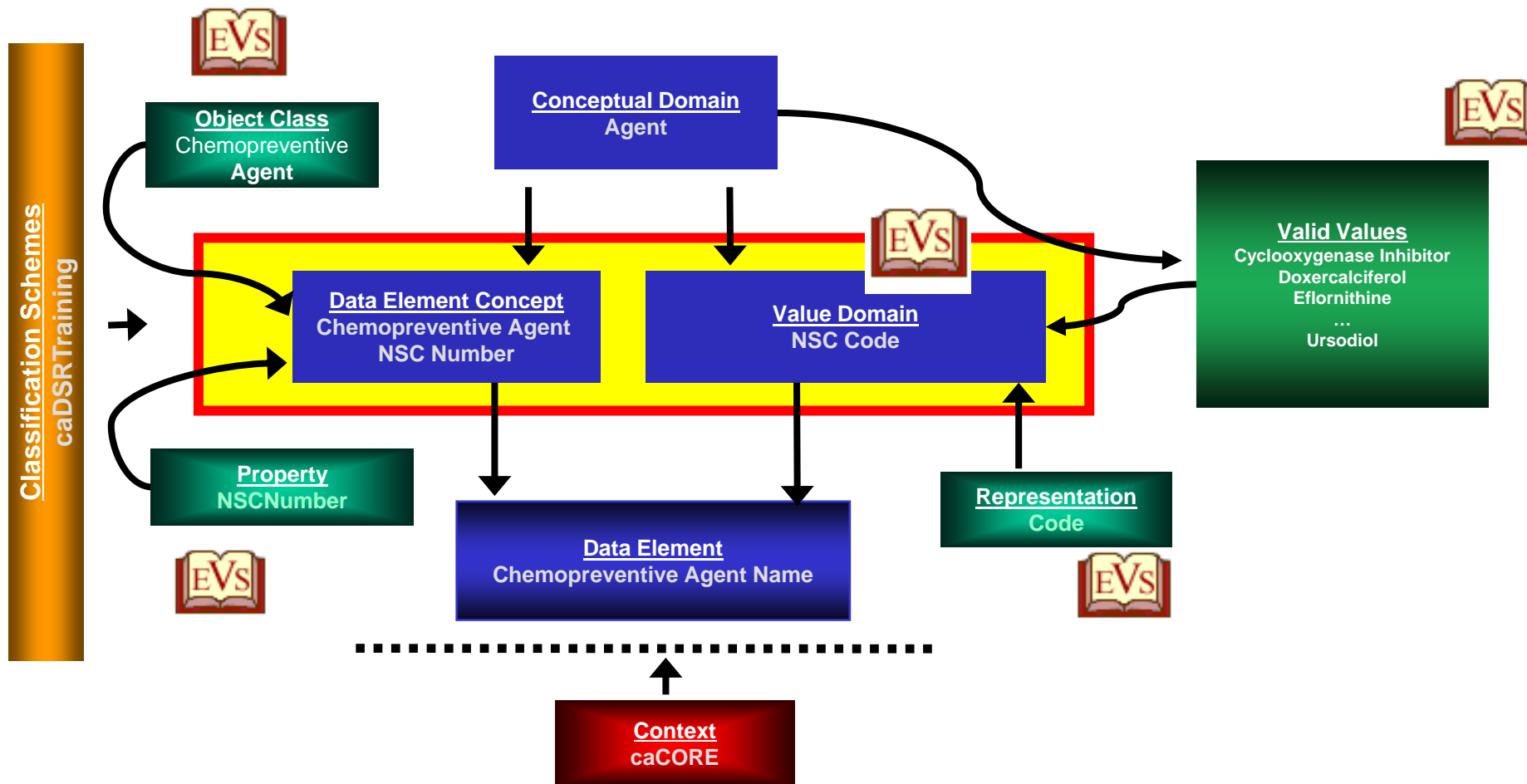


ISO/IEC 11179 & 19763 and OOR

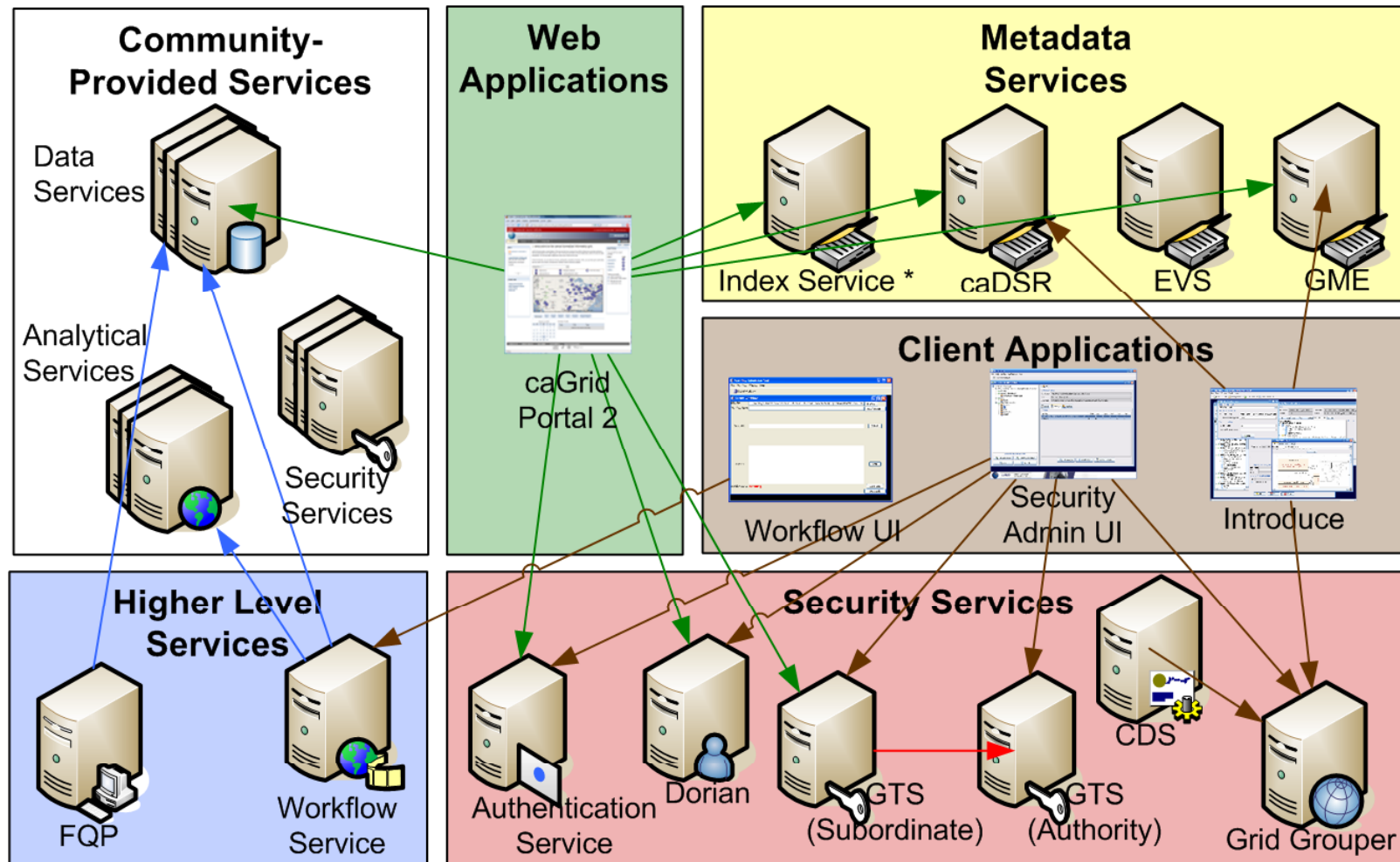


From Prof. Hajime Horiuchi

NCI enhancements: 11179 with Enterprise Vocabulary Services (EVS) Concepts Unite NCI MDR



NCI caGrid Production Environment



*All Services Register with the Index Service

Overview



- ☞ A framework which enables fast and easy creation of Globus based grid services
- ☞ Provide easy to use graphical service authoring tool
- ☞ Hide all “grid-ness” from the developer
- ☞ Integration with other core grid services and architecture components
 - ◆ GAARDS Security Infrastructure (Dorian, GridGrouper, CSM)
 - ◆ Globus Index Service
 - ◆ Global Model Exchange (GME)
 - ◆ Cancer Data Standards Repository (caDSR)
- ☞ Extension framework for integrating with other architecture components

Metadata Services



☞ Cancer Data Standards Repository (caDSR)

- ◆ caBIG projects register their **data models** as Common Data Elements (CDEs) which are semantically harmonized and then centrally stored and managed the caDSR
- ◆ The caDSR grid service provides:
 - Model discovery and traversal
 - caGrid standard metadata generation capabilities

☞ Enterprise Vocabulary Services (EVS)

- ◆ EVS is set of services and resources that address the need for **controlled vocabulary**
- ◆ The EVS grid service provides:
 - Query access to the data semantics and controlled vocabulary managed by the EVS

Metadata Services cont.



☞ Global Model Exchange (GME)

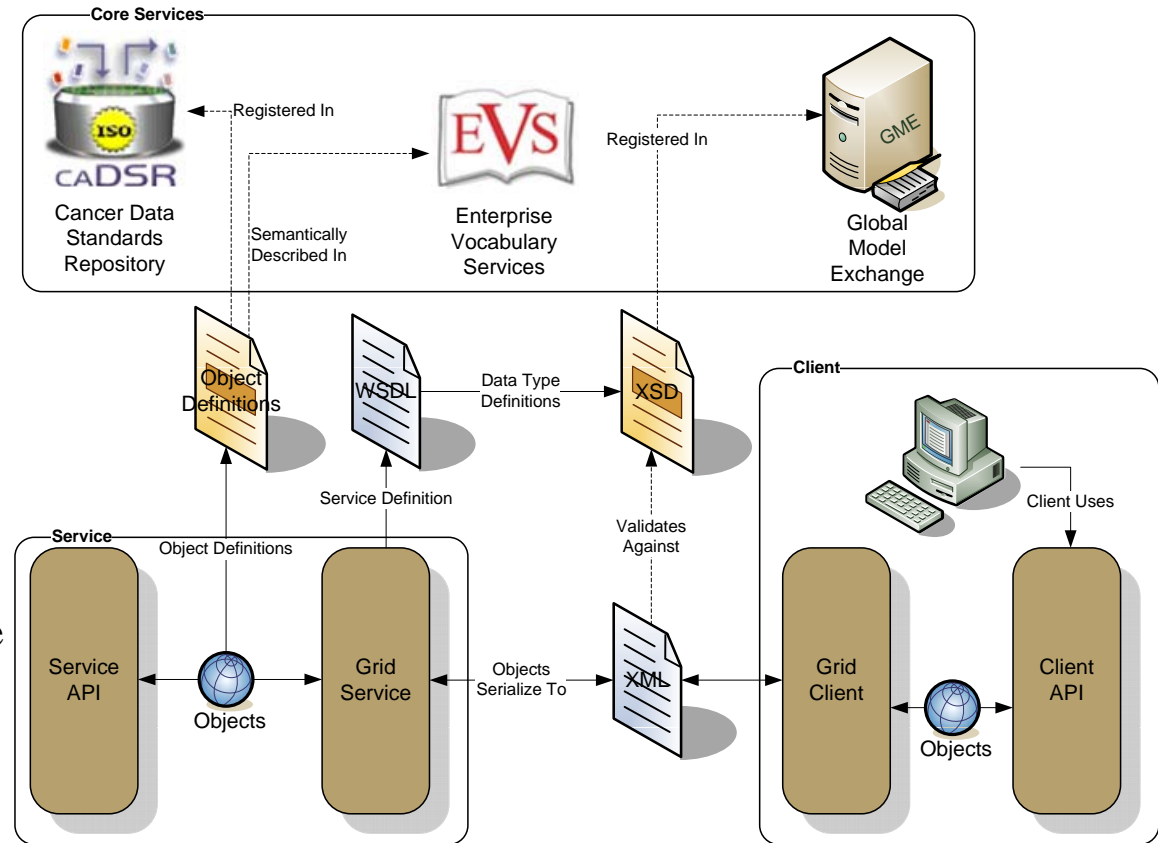
- ◆ GME is a DNS-like data definition registry and exchange service that is responsible for storing and linking together **structural data models** in the form of XML schema.
- ◆ The GME grid service provides:
 - Access to the authoritative structural representation of data types on the grid

☞ Globus Information Services: Index Service

- ◆ The Globus Information Services infrastructure provides a generic framework for **aggregation of service metadata**, a registry of running Grid services, and a dynamic data-generating and indexing node, suitable for use in a hierarchy or federation of services
- ◆ The Index grid service provides:
 - Yellow and white pages for the grid

caGrid Data Description Infrastructure

- Client and service APIs are object oriented, and operate over well-defined and curated data types
- Objects are defined in UML and converted into ISO/IEC 11179 Administered Components, which are in turn registered in the Cancer Data Standards Repository (caDSR)
- Object definitions draw from controlled terminology and vocabulary registered in the Enterprise Vocabulary Services (EVS), and their relationships are thus semantically described
- XML serialization of objects adhere to XML schemas registered in the Global Model Exchange (GME)



Ontology Summit



- NIST, NSF, MITRE, National Center for Ontological Research (NCOR), and others organized an Ontology Summit meeting in April 2008.
- The theme of the 2008 Ontology Summit was the vision for an Open Ontology Repository. This vision forms the basis of the international Open Ontology Repository community.
- The summit participants supported the idea to produce such a repository (or set of repositories). An OOR-Forum will serve as a venue - both virtual and face-to-face - in which many of the issues relating to the design, implementation, and ongoing use of an ontology repository can be discussed and ultimately resolved.
- This was the third in a series of Ontology Summit meetings
- There are bi-weekly OOR telecons – all are invited

Two Thrusts

- ☞ The theoretical aspects discussed in Ontology Summit Forum
 - ◆ See: <http://ontolog.cim3.net/cgi-bin/wiki.pl?OntologySummit2008>
 - ◆ See: <http://ontolog.cim3.net/cgi-bin/wiki.pl?OntologySummit2008>
- ☞ The practical aspects of building an Open Ontology Repository. This is pursued in the OOR-Forum.
 - ◆ See: <http://ontolog.cim3.net/cgi-bin/wiki.pl?OpenOntologyRepository>

OOR Charter



- Promote the global use and sharing of ontologies by: 1. establishing a hosted registry-repository; 2. enabling and facilitating open, federated, collaborative ontology repositories; 3. establishing best practices for expressing interoperable ontology and taxonomy work in registry-repositories.

Key Issues

- OpenOntologyRepository_Scope - documentation related to this OOR initiative's mission, charter, objectives, goals, terms of reference, definitions, scope, ... etc.
- OpenOntologyRepository_Requirement - documentation related to the requirements of the "open ontology repository" we are planning to implement through this OOR initiative.
- OpenOntologyRepository_UseCases
- OpenOntologyRepository_Approach - documentation related to this OOR initiative's approach, and the process we will be using.
- OpenOntologyRepository_Architecture - documentation related to the architecture of the "open ontology repository" we are planning to implement through this OOR initiative.
- OpenOntologyRepository_Plan - documentation related to this OOR initiative's deliverables, timeline, milestones & deadlines, project plans, etc.

Scope



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

Message from Leo Obrst (MITRE)



- “I'd advise folks to look over xmdr.org (eXtended Metadata Registry) before they re-invent 95% of the wheels, including definitions, architecture, requirements, use cases, etc. Although XMDR, like the ISO
- “11179 Metadata Registry standard, is concerned with more than an ontology registry/repository, their notions are extremely salient to our effort. And no, I am not involved in XMDR, and am only interested in seeing that good work is recognized, and possible duplication of effort is avoided.”

BioPortal and XMDR



☞ Demo BioPortal here

☞ Demo XMDR here

What XMDR Brings to the Table



- ☞ Use cases - semantics challenges - and Requirements
- ☞ Proposed specifications for ISO/IEC 11179 Edition 3 – Model, definitions, ontology
- ☞ Modular software architecture and open source software modules
- ☞ Open Source XMDR software
- ☞ Test content

XMDR Use Cases



- OOR - what do we want to construct?
- XMDR contributed use cases to the OOR web site. It would be good to update and flesh out the use cases. The XMDR use cases are based on proposals for ISO/IEC 11179 Edition 3. The intent is to use semantic artifacts, such as ontologies, to help manage data.
- The core OOR functionality may be registering and disseminating ontologies. Do we want to extend our effort to the linkage of concept systems (e.g., ontologies) to metadata that describes data?
- Do we want to be able to reason across the content of an OOR? XMDR enables this.

Use Case: Combine Data, Metadata & Concept Systems

Inference Search Query:

“find water bodies downstream from Fletcher Creek where chemical contamination was over 10 micrograms per liter between December 2001 and March 2003”

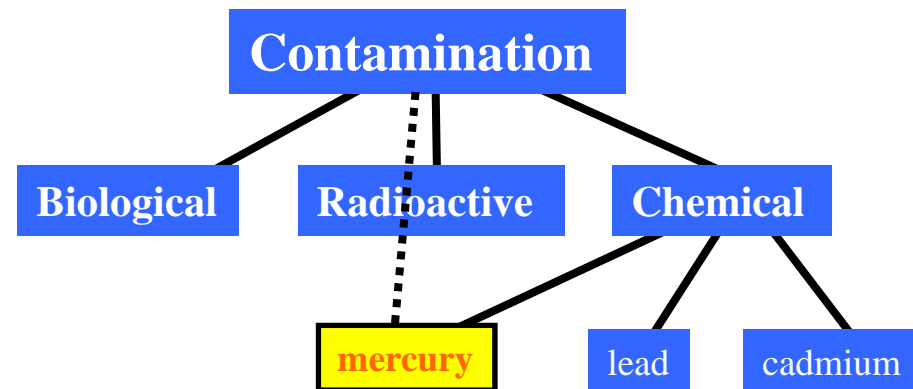
Data:

<u>ID</u>	<u>Date</u>	<u>Temp</u>	<u>Hg</u>
A	06-09-13	4.4	4
B	06-09-13	9.3	2
X	06-09-13	6.7	78

Metadata:

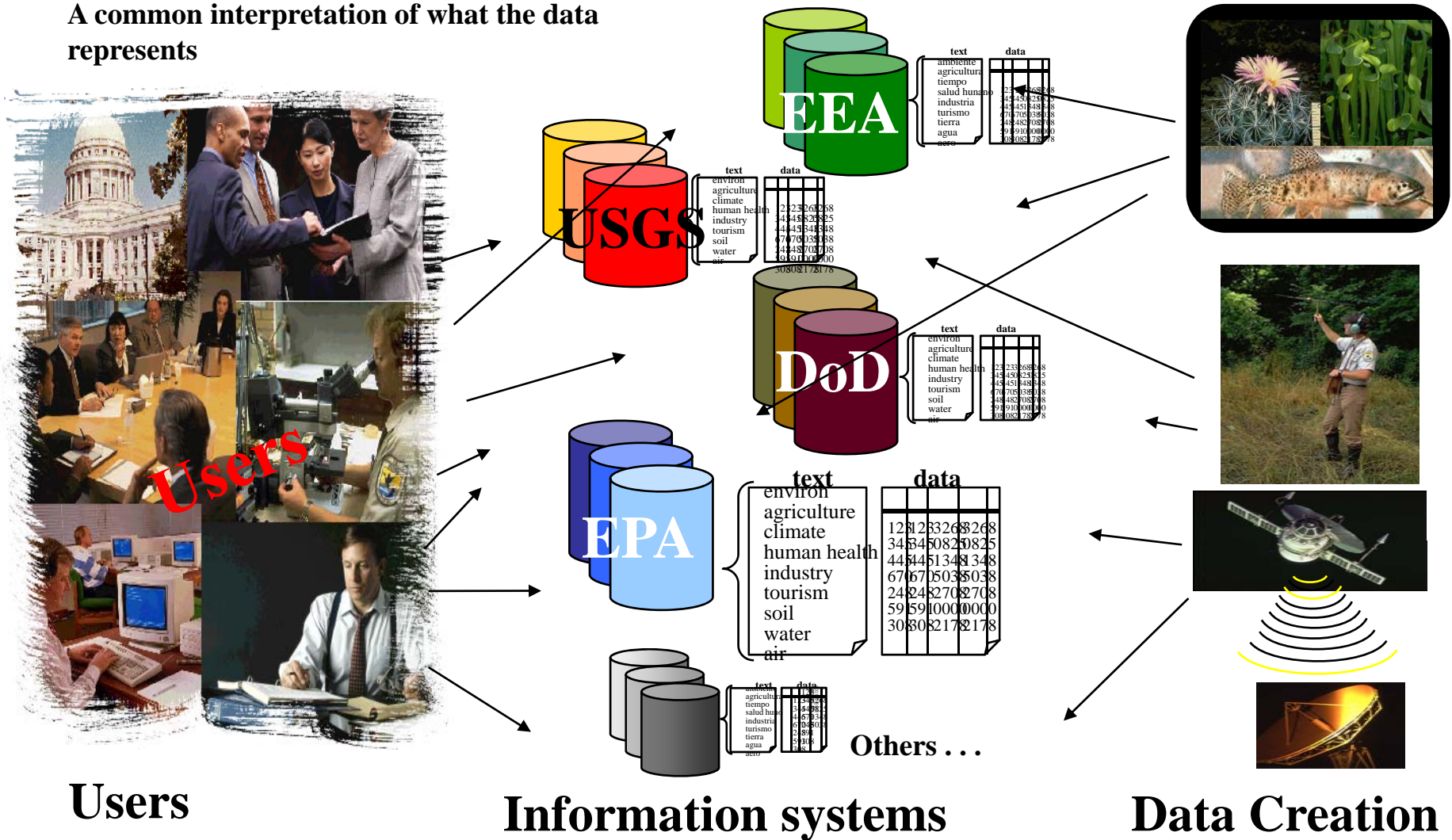
<u>Name</u>	<u>Datatype</u>	<u>Definition</u>	<u>Units</u>
ID	text	Monitoring Station Identifier	not applicable
Date	date	Date	yy-mm-dd
Temp	number	Temperature (to 0.1 degree C)	degrees Celcius
Hg	number	Mercury contamination	micrograms per liter

Concept system:



Use Case: Gain Common Understanding of meaning between Data Creators and Data Users

A common interpretation of what the data represents

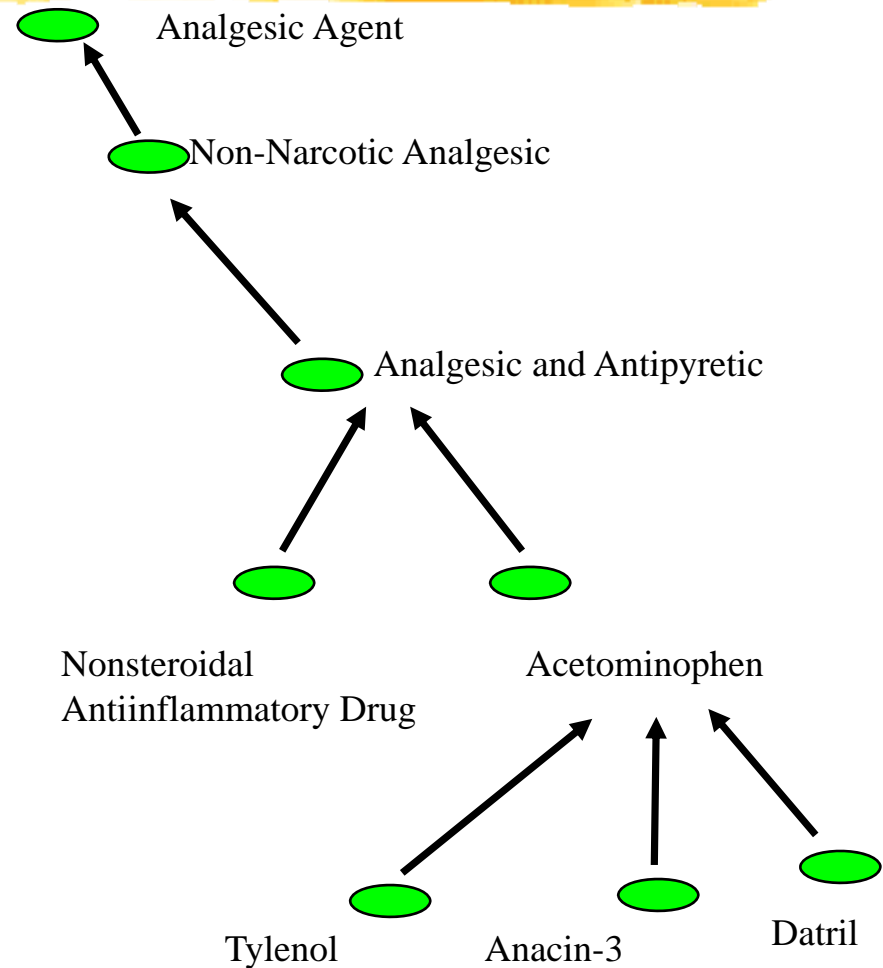


Use Case: Find and process non-explicit data

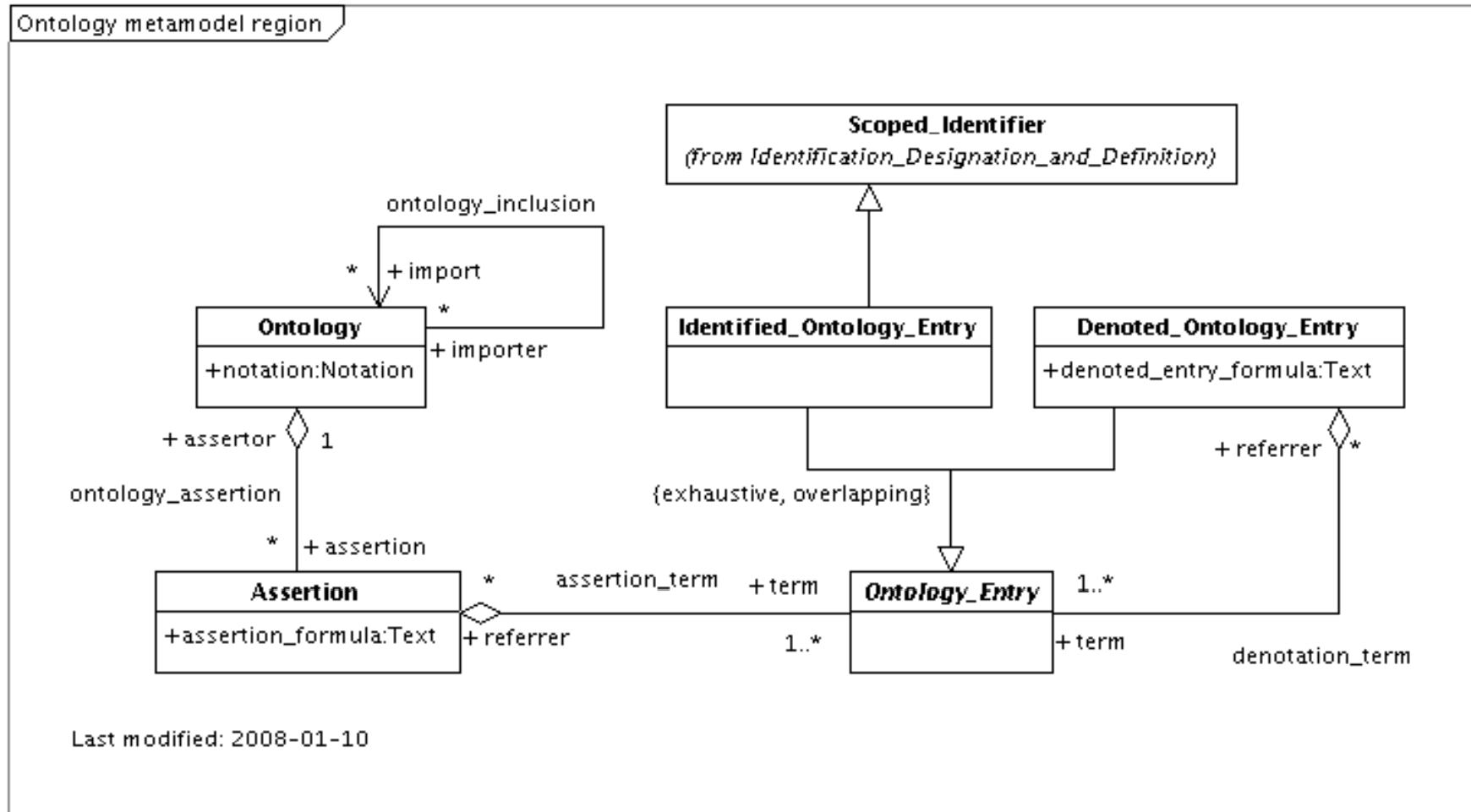
For example...

Patient data on drugs contains brand names (e.g. Tylenol, Anacin-3, Datril,...);

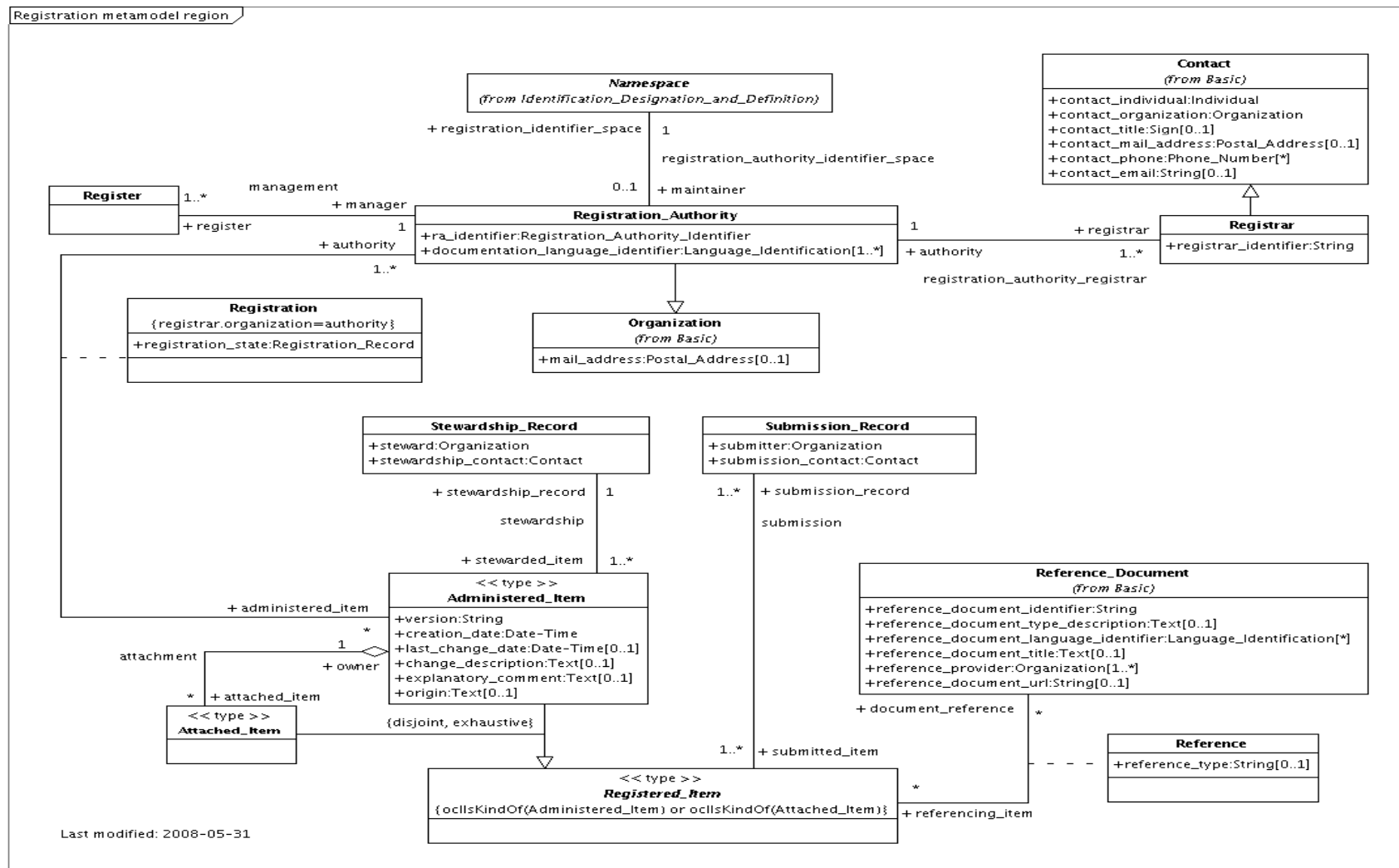
However, want to study patients taking analgesic agents



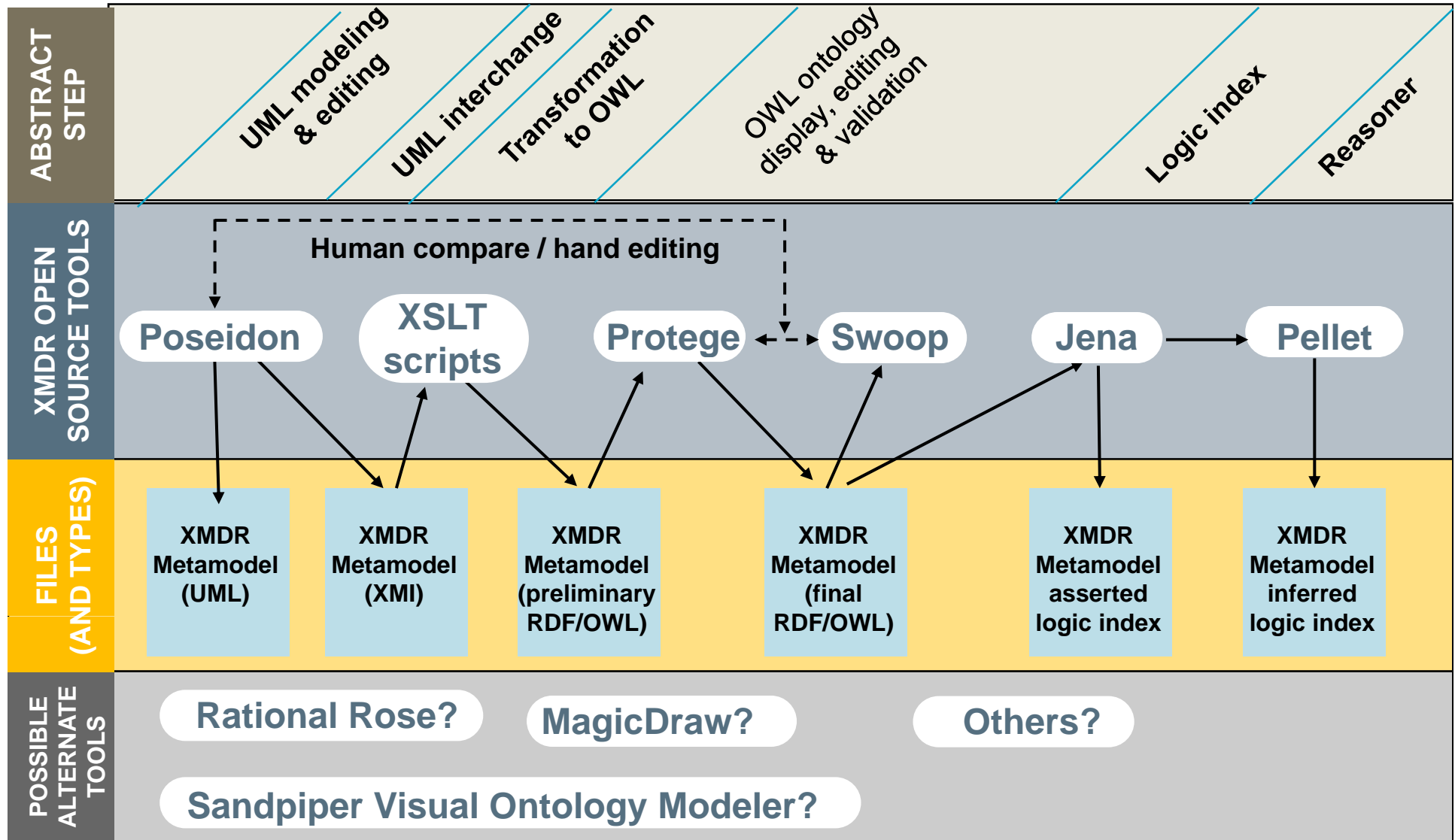
Model – Ontology Region



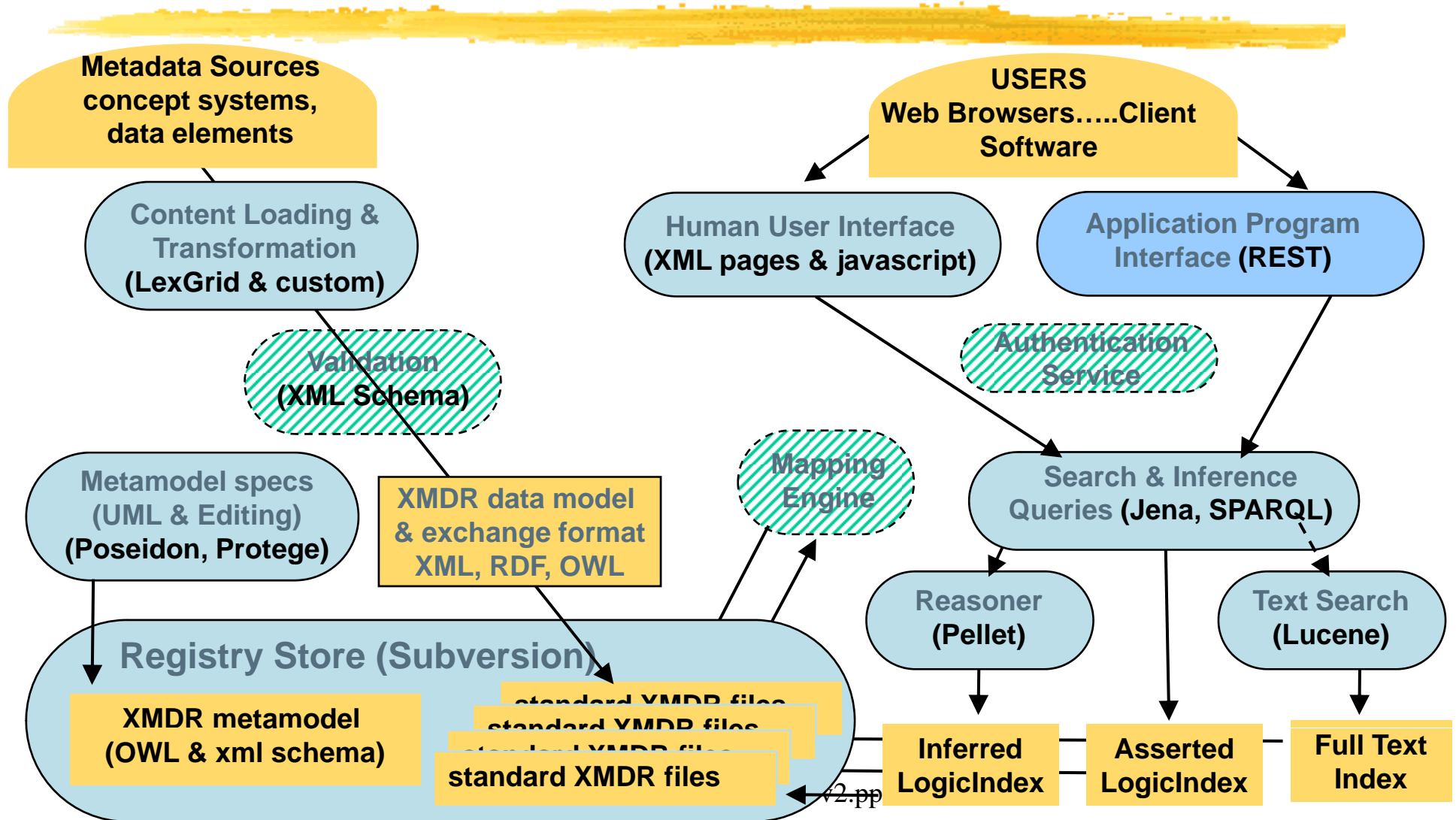
Model – Registration Region



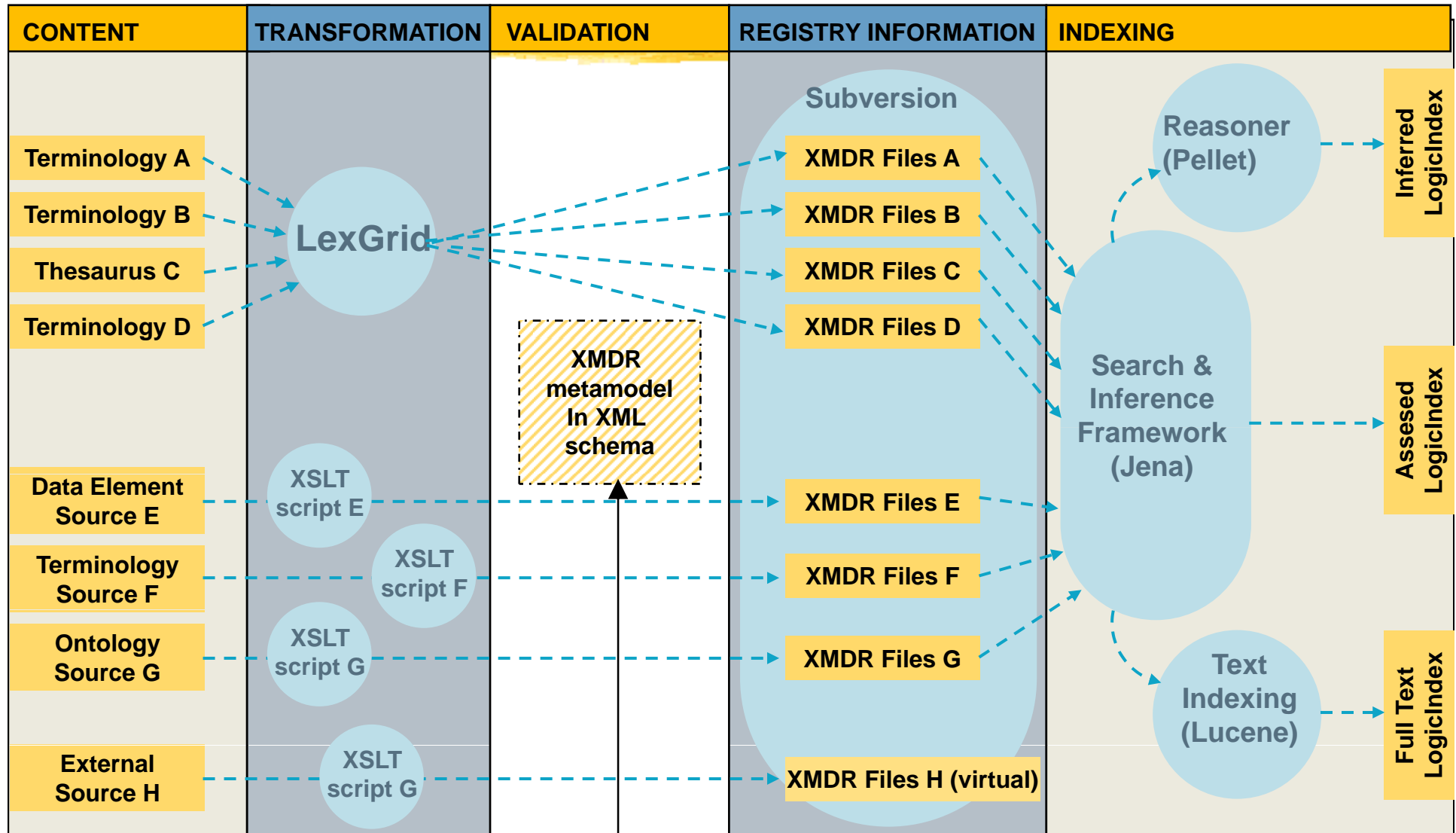
Model: XMI from 11179-ed3 UML is transformed to RDF/OWL for XML metamodel specification



XMDR Architecture & Open Source Software



Content loading, with XMDR metamodel used for inferred indexing and validation

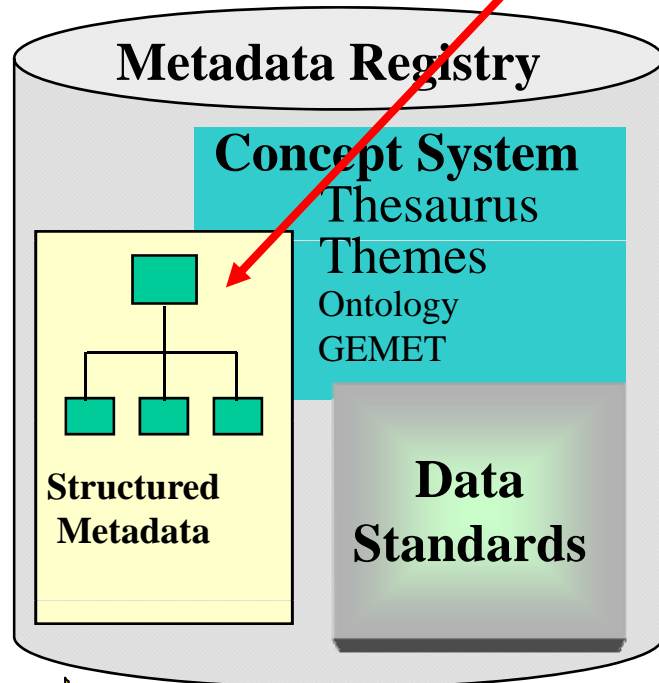


An Approach



- ➡ Use BioPortal
- ➡ Use LexGrid
- ➡ Extend with XMDR

Concept System Store



ISO/IEC 11179 Metadata Registry

Concept systems:

Keywords

Controlled Vocabularies

Thesauri

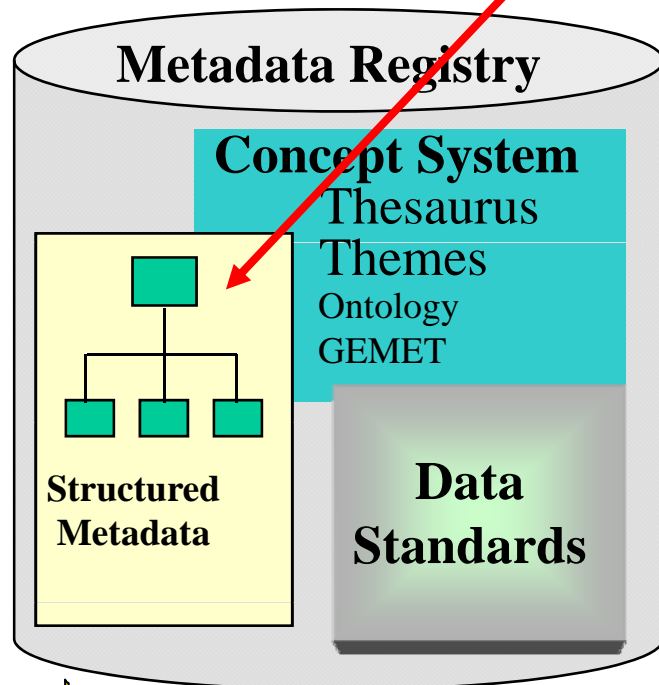
Taxonomies

Ontologies

Axiomatized Ontologies

(Essentially graphs: node-
relation-node + axioms)

Management of Concept Systems

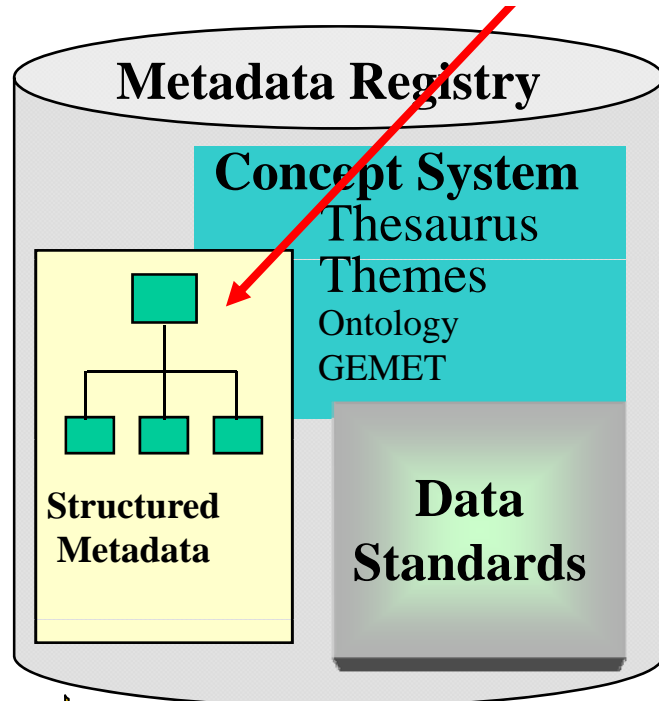


ISO/IEC 11179 Metadata Registry

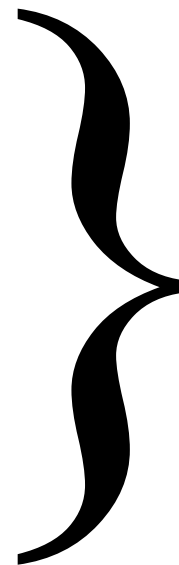


Concept system:
Registration
Harmonization
Standardization
Acceptance (vetting)
Mapping
(correspondences)

Life Cycle Management



ISO/IEC 11179 Metadata Registry



**Life cycle
management:
Data and
Concept systems
(ontologies)**

Grounding Semantics



**Metadata
Registries**

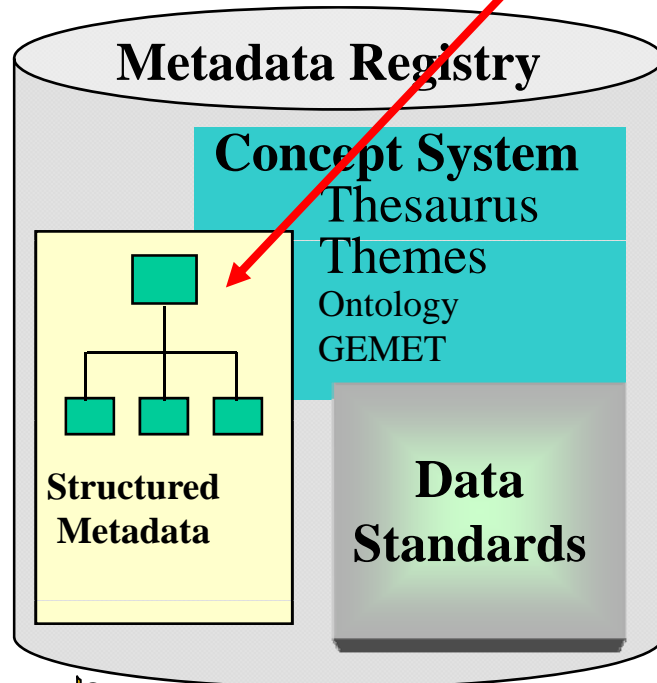
**Semantic Web
RDF Triples**

Subject (node URI)

Verb (relation URI)

Object (node URI)

Ontologies



ISO/IEC 11179 Metadata Registry

Another Application: GEO/GEOSS/GMES

☞ The work addresses information challenges identified by major initiatives including the Intergovernmental Group on Earth Observations (GEO), Global Earth Observation System of Systems (GEOSS), and Global Monitoring for Environment and Security (GMES). The techniques and technologies developed will be useful for GEO/GEOSS/GMES. In particular, this work addresses priority challenges identified by the GEOSS Architecture & Data Management (ADM) committee to “enable increased interoperability across existing data management systems:

- ◆ Identify & address integration gaps in data management systems
- ◆ Utilize community standards for data & metadata
- ◆ Enable integrated measurements, data, products & predictive models
- ◆ Examine the need for future data management requirements.”
- ◆ Architecture & Data Management (ADM) Working Group Report February 22, 2007.

See: <http://usgeo.gov/docs/USGEO%20Progress%20Report%202007-0321.pdf>

Thanks & Acknowledgements

- ☞ John McCarthy, LBNL
- ☞ Kevin Keck, LBNL
- ☞ Fred Gey, UCB
- ☞ Hajime Horiuchi, Tokyo International University
- ☞ Leo Obrst, MITRE
- ☞ Frank Olken, LBNL
- ☞ Anirban Sen, UCB
- ☞ Harold Solbrig, Mayo Clinic
- ☞ Denise Warzel, NCI
- ☞ INCITS L8 and ISO/IEC JTC1/SC 32/WG 2 Standards Committees

- ☞ Major XMDR Project Sponsors and Collaborators
 - ◆ National Science Foundation (Grant #0637122)
 - ◆ National Cancer Institute
 - ◆ U.S. Environmental Protection Agency
 - ◆ Department of Defense
 - ◆ U.S. Geological Survey
 - ◆ And others!

Acknowledgement



- ☞ This material is based upon work supported by the National Science Foundation under Grant No. 0637122, and USEPA. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation, USEPA or USDOD.
- ☞ For more information see XMDR.org