

eXtended Metadata Registry (XMDR): Input for Open Ontology Repository

**OOR Panel - “Ontology Registry and Repository
Technology & Infrastructure Landscape”**

February 28, 2008

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



Topics



- ❖ Describe the technology/infrastructure that XMDR brings to the table for the OOR project.
- ❖ How does that contribute to the overall OOR initiative
- ❖ How does that fit in with the other things that the rest of the teams are bringing to the table

What XMDR Brings to the Table



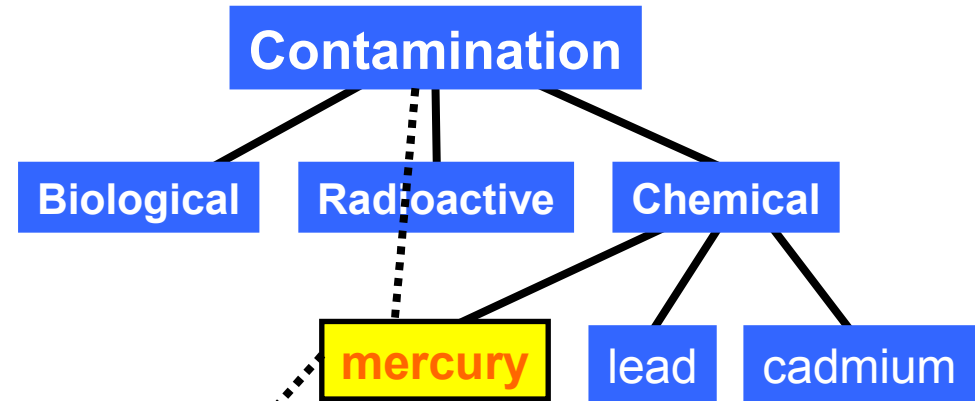
- ❖ Use cases - semantics challenges - and Requirements
- ❖ Proposed specifications for ISO/IEC 11179 Edition 3 – Model, definitions, ontology
- ❖ Modular software architecture and open source software modules
- ❖ Open Source XMDR software
- ❖ Test content

Challenge: 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”

Concept system:



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

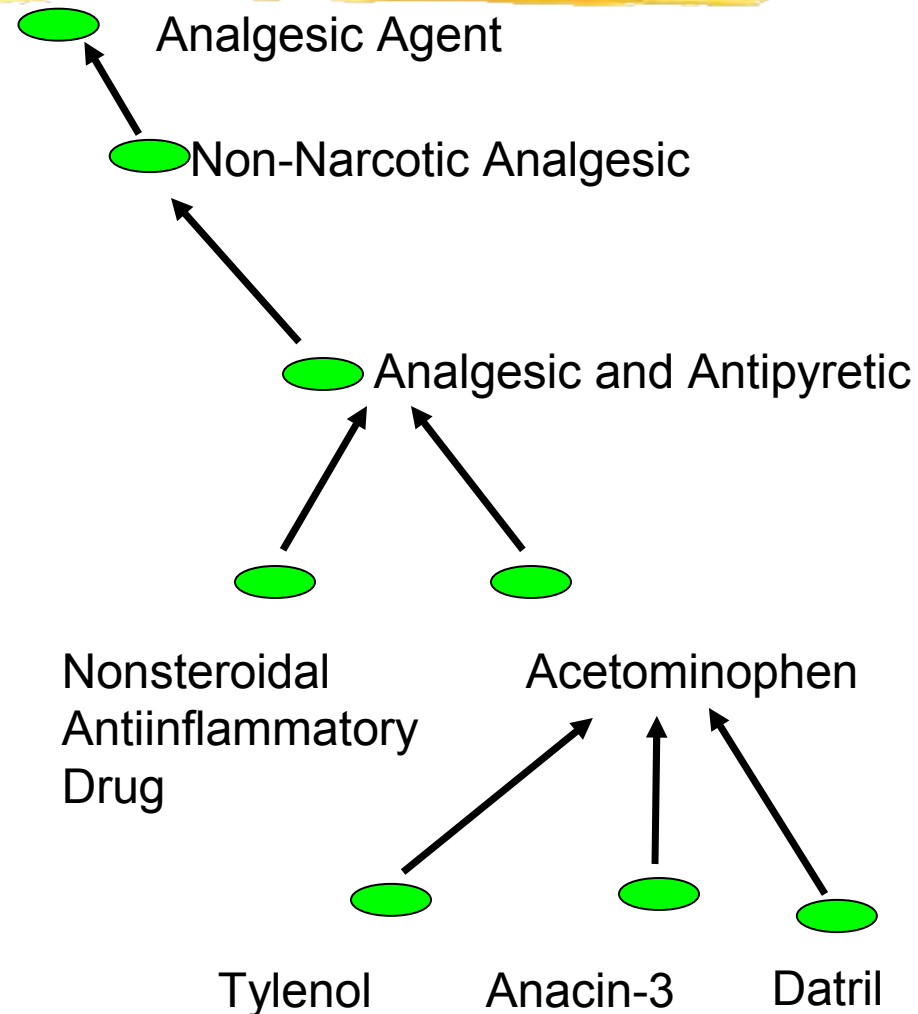


Challenge: Find and process non-explicit data

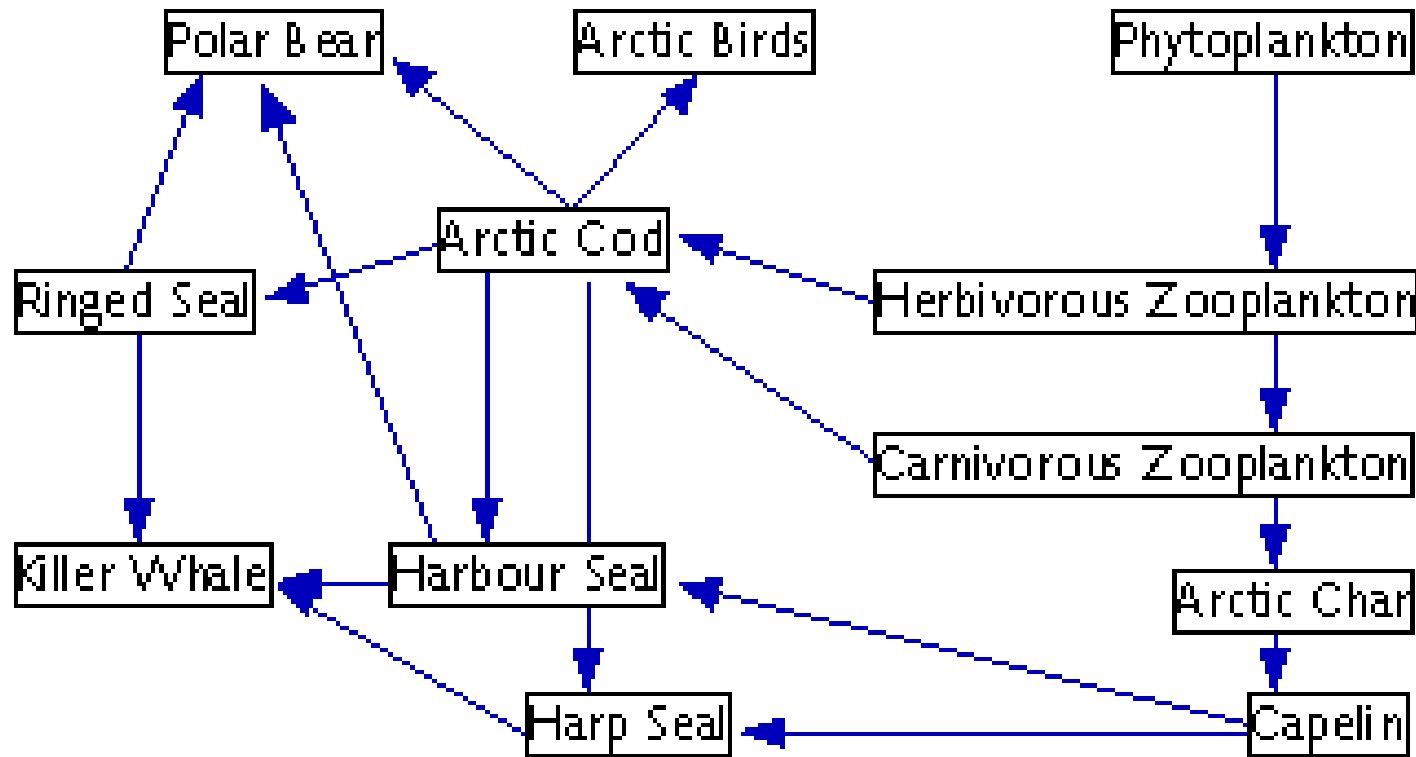
For example...

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

However, want to study patients taking analgesic agents



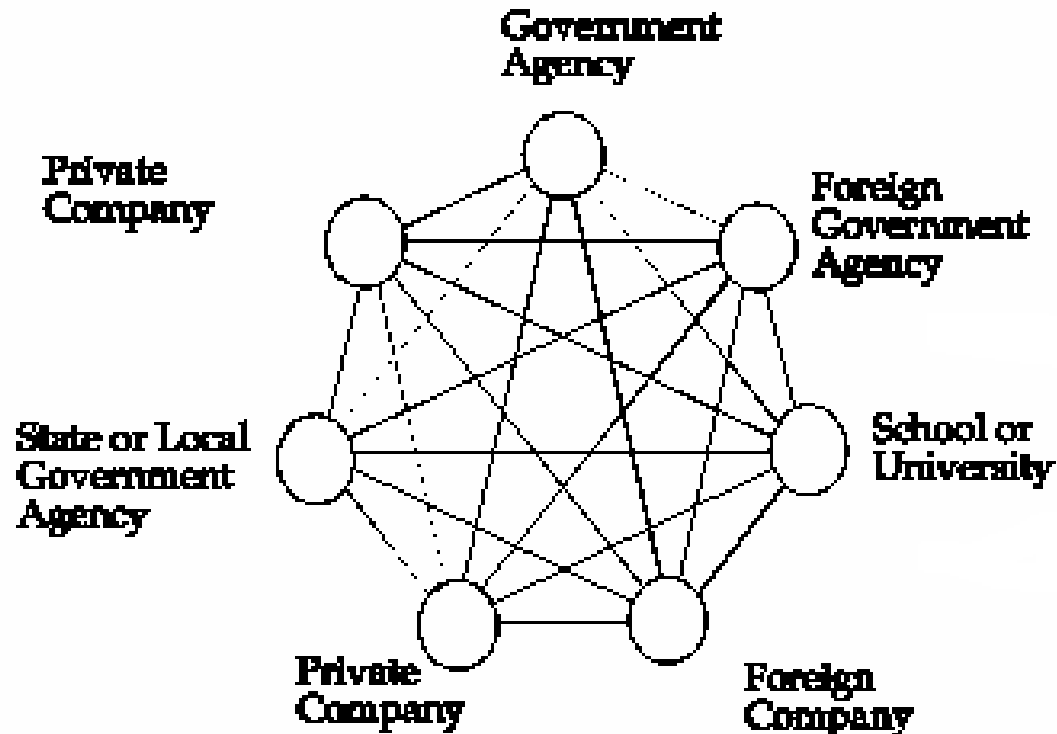
Challenge: Specify and compute across Relations, e.g., within a *food web* in an Arctic ecosystem



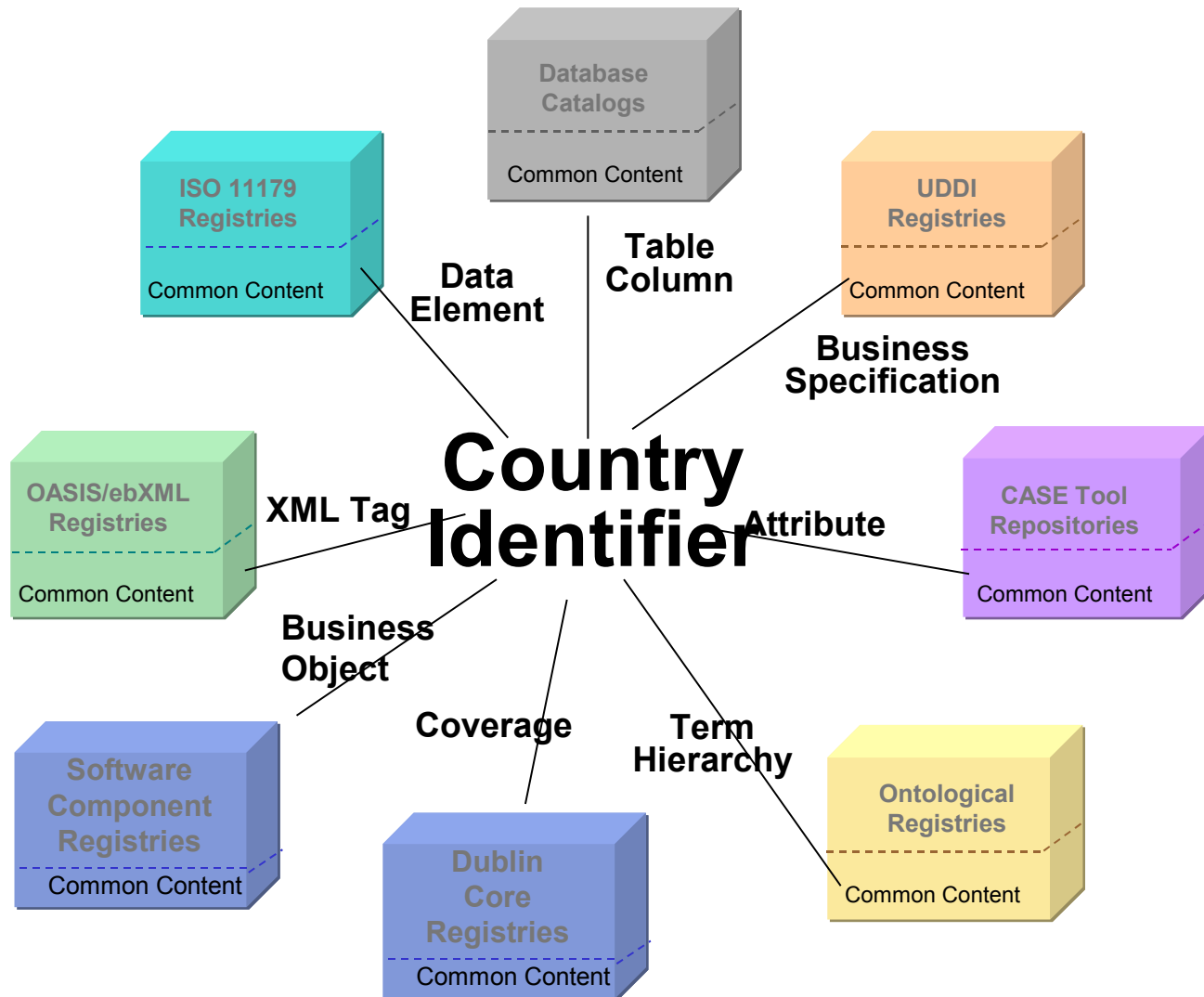
An organism is connected to another organism for which it is a source of food energy and material by an arrow representing the direction of biomass transfer.

Challenge: Use data from systems that record the same facts with different terms

- ❖ Reduce the human toil of drawing information together and performing analysis -> shift to computer processing.



Challenge: Use data from systems that record the same facts with different terms



Same Fact, Different Terms

Data Element Concept

Name: Country Identifiers
Context:
Definition:
Unique ID: 5769
Conceptual Domain: →
Maintenance Org.:
Steward:
Classification:
Registration Authority:
Others

Algeria
 Belgium
 China
 Denmark
 Egypt
 France
 ...
 Zimbabwe

Data Elements

Name:
Context:
Definition:
Unique ID: 4572
Value Domain:
Maintenance Org.:
Steward:
Classification:
Registration Authority:
Others

Algeria
 Belgium
 China
 Denmark
 Egypt
 France
 ...
 Zimbabwe

ISO 3166
English Name

L` Algérie
 Belgique
 Chine
 Danemark
 Egypte
 La France
 ...
 Zimbabwe

ISO 3166
French Name

DZ
 BE
 CN
 DK
 EG
 FR
 ...
 ZW

ISO 3166
2-Alpha Code

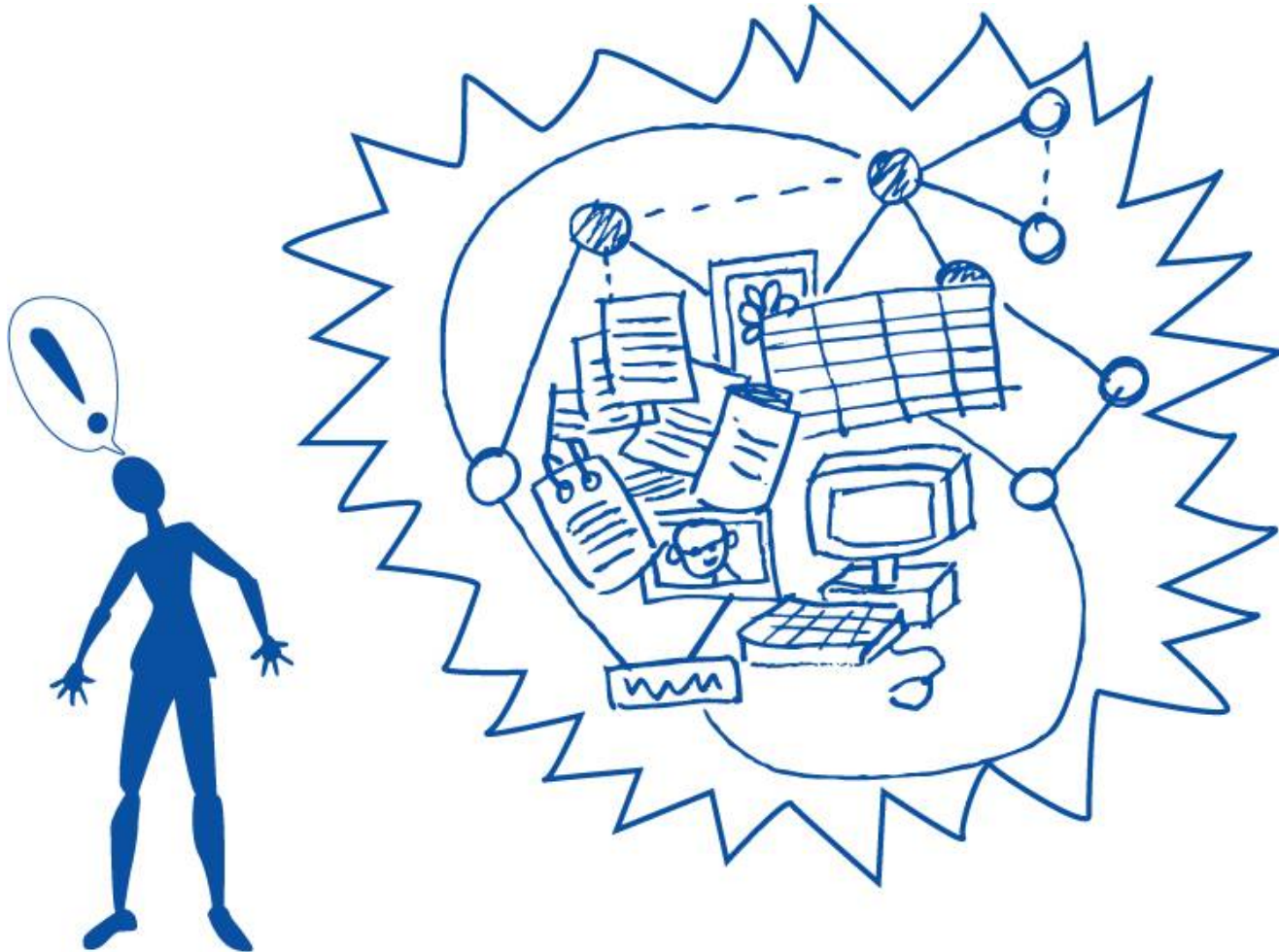
DZA
 BEL
 CHN
 DNK
 EGY
 FRA
 ...
 ZWE

ISO 3166
3-Alpha Code

012
 056
 156
 208
 818
 250
 ...
 716

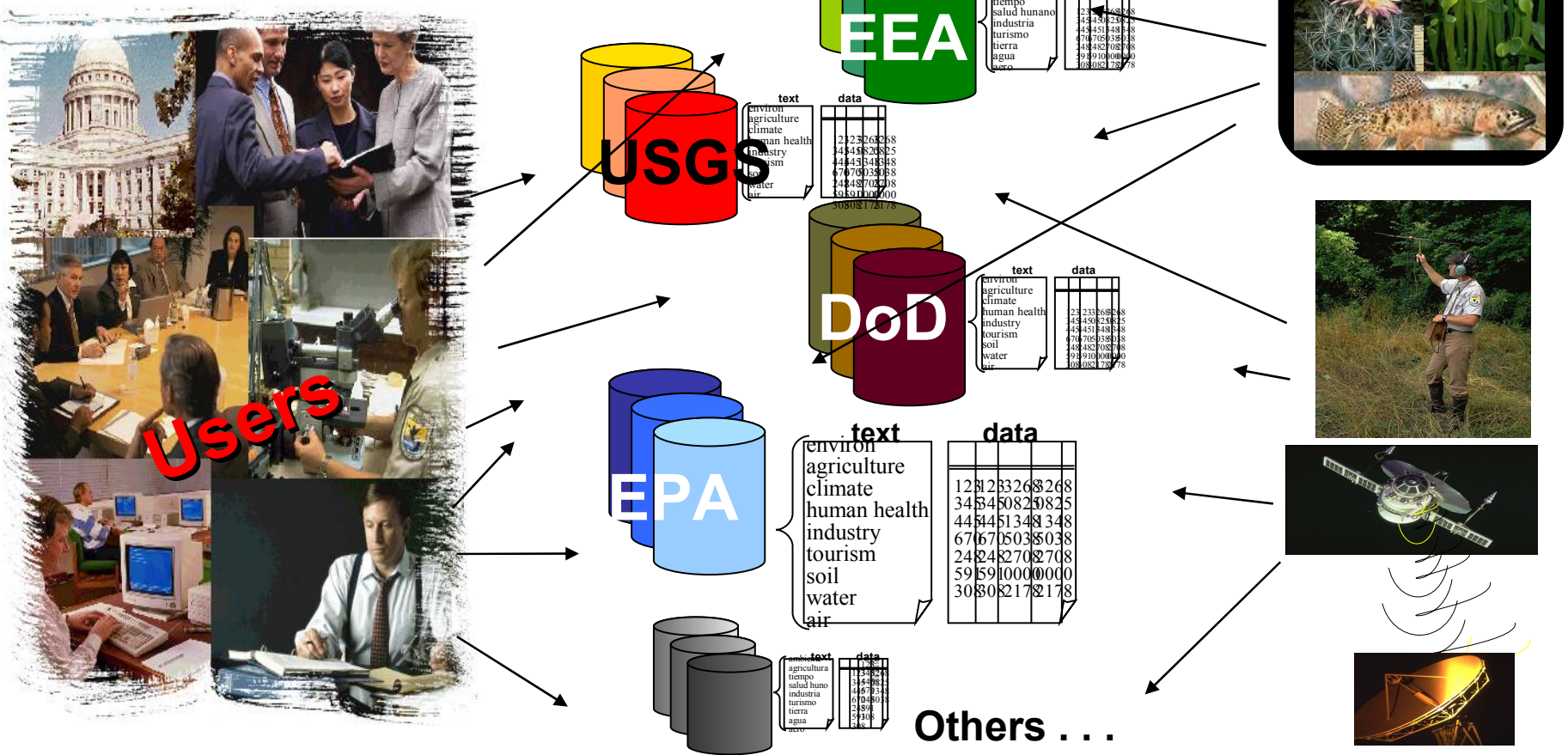
ISO 3166
3-Numeric Code

Challenge: Draw information together
from a broad range of studies,
databases, reports, etc.



Challenge: Gain Common Understanding of meaning between Data Creators and Data Users

A common interpretation of what the data represents



Users

Information systems

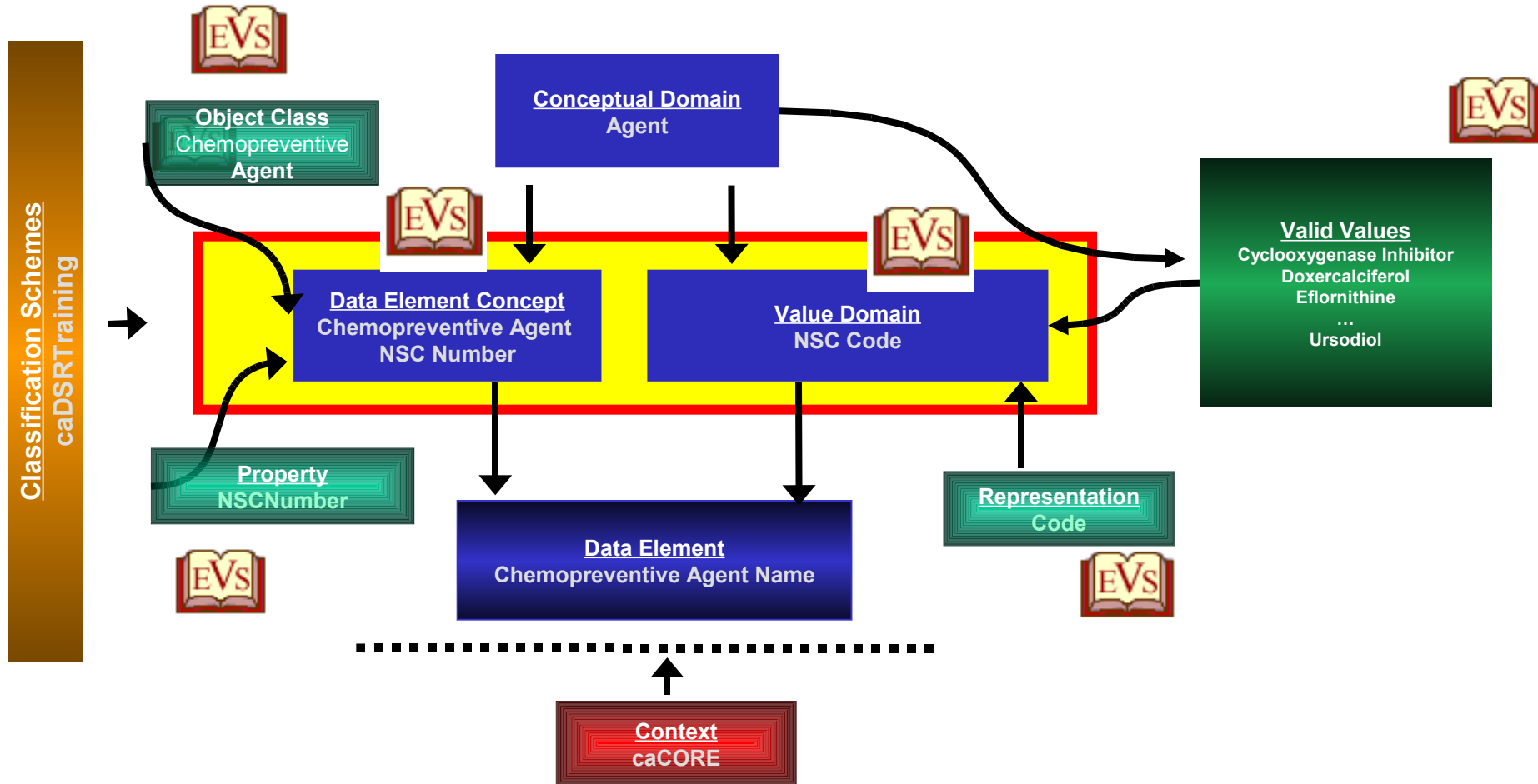
Data Creation

Semantics Challenges



- ❖ Managing, harmonizing and vetting semantics is important for traditional data management.
 - ◆ In the past we just covered the basics
- ❖ Managing, harmonizing, and vetting semantics is essential to enable enterprise semantic computing

Enterprise Vocabulary Services (EVS) Concepts Unite NCI MDR



XMDR Prototype



Demonstrate capabilities:

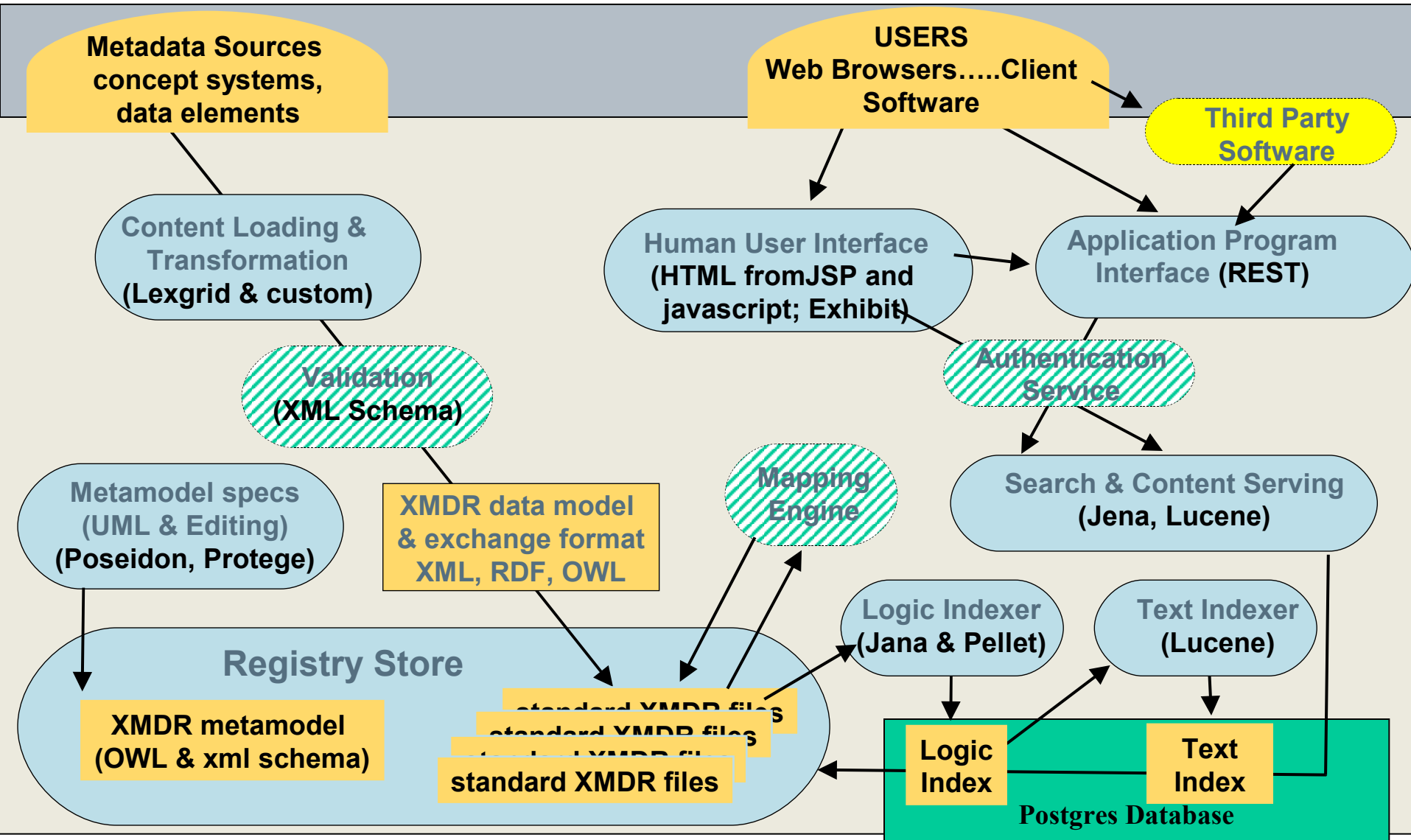
- ❖ Register existing concept systems, based on their underlying structures, such as graphs of varying complexity.
- ❖ Interrelate concepts systems with each other.
 - E.g., register mappings between multiple vocabularies
- ❖ Support harmonization and vetting of concept systems for a community of interest.
 - E.g., Register, harmonize, validate, and vet definitions and relations
- ❖ Interrelate concepts in concept systems with concepts in metadata and concepts in databases, knowledgebases, and text.
- ❖ Provide semantic services needed to support traditional computing as well as semantic computing.
 - ◆ E.g., dereferencing the URIs used in creating RDF statements, by providing relevant information describing the referenced concept and its authoritative standing within some community of interest.
- ❖ Register and manage the provenance of data
- ❖ XMDR is part of the infrastructure for semantics and data management.

XMDR Use



- ◆ Upside
 - ◆ Collaborative
 - Supports interaction with community of interest
 - Shared evolution and dissemination
 - Enables Review Cycle
 - ◆ Standards-based – don't lock semantics into proprietary technology
 - ◆ Foundation for strategic data centric applications
 - ◆ Lays the foundation for
Ontology-based Information Management
 - ◆ Content is reusable for many purposes
- ◆ Downside
 - ◆ Managing semantics is **HARD WORK**
 - No matter how friendly the tools
 - ◆ Needs integration with other components

Modular XMDR Architecture



Initial XMDR REST-style Application Programming Interface (API)

- ❖ Search Methods (GET)
 - ◆ Text Search
 - ◆ SPARQL Search
 - ◆ XMDR Search (not documented yet)
- ❖ Registry Information Methods
 - ◆ Summary information
 - ◆ registered models
 - ◆ Identified Items
- ❖ Method Parameters
 - ◆ can be included as part of any method
 - ◆ as part of URL
 - ◆ Accept_type (what xml components to expect)
 - ◆ Stylesheet (how to display results)

REST API (Search Methods)

Resource	URI (relative to application root)	Method	Representation	Accept Request	Description
Text Search	search/text?query={queryText}	GET	application/xml (searchResult)	Any (ignores)	Start a text search.
Text Search Results	search/text/{queryID}?offset={offset}&maxResults={maxResults}	GET	application/xml (textResultSet)	application/xml, application/*, or */*	Retrieve the results of a text search.
			application/exhibi*	application/exhibit	
SPARQL Search	search/sparql?query={queryText}&model={modelNameN}	GET	application/xml (searchResult)	Any (ignores)	Start a SPARQL search.
SPARQL Search Results	search/sparql/{queryID}?offset={offset}&maxResults={maxResults}	GET	application/xml (sparqlResultSet)	application/xml, application/*, or */*	Retrieve the results of a SPARQL search.
			application/sparql-results+xml**	application/sparql-results+xml	
			application/sparql-results+json***	application/sparql-results+json, application/json	
			application/exhibit *	application/exhibit	

XMDR

- ◆ Content (selected portions of):
 - ◆ ISO/IEC 11179
 - ◆ ISO/IEC 3166 – Country codes
 - ◆ ISO 4217 – Currency codes
 - ◆ EPA Environmental Data Registry content (ISO/IEC 11179 based registry)
 - ◆ Standard Industrial Codes
 - ◆ North American Industrial Classification System
 - ◆ Mapping NAICS 02 to SIC 87
 - ◆ Adult Mouse Anatomical Dictionary
 - ◆ Defense Technology Info. Center Thesaurus
 - ◆ NBII Biocomplexity Thesaurus
 - ◆ GEneral Multilingual Environmental Thesaurus
 - ◆ NCI_Thesaurus
 - ◆ Cancer Data Standards Repository (NCI registry based on ISO/IEC 11179)
- ◆ Loading new content (ongoing)
 - ◆ OMEGA linguistic ontology
 - ◆ OpenCyc ontology
 - ◆ SIC – NAICS codes
 - ◆ Mapping of NAICS to SIC codes

Contribution



How does that contribute to the overall OOR initiative?

- ❖ It is free for the taking
- ❖ Save time on development of use cases, specifications, architectures, software, etc.

Fitting In

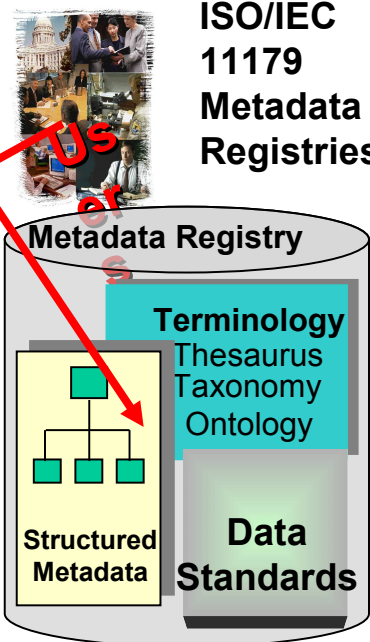


How does that fit in with the other things that the rest of the teams are bringing to the table?

- ❖ Collaboration on standards development
- ❖ Collaboration on prototype development and demonstration
- ❖ Collaboration on proposals?

Align, Coordinate, Integrate Standards/Recommendations/Specifications for Semantic Computing

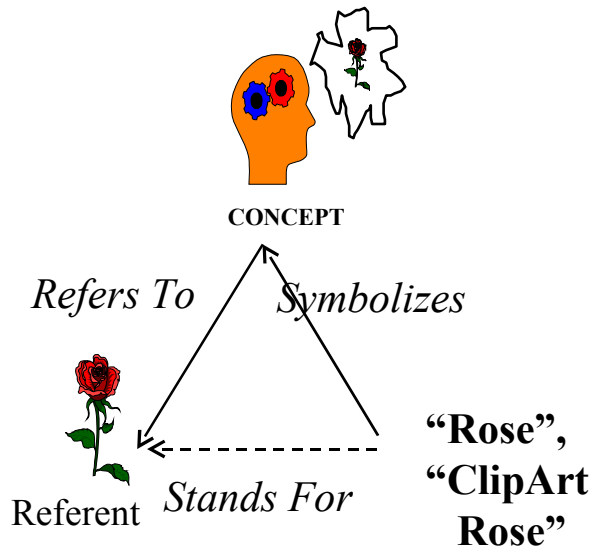
ISO/IEC
11179
Metadata
Registries



11179 Metadata Registry

ISO/IEC JTC 1/SC 32

Terminology



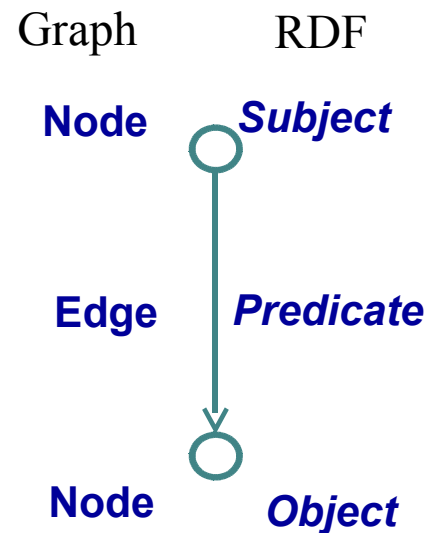
ISO TC 37

Object Management

MOF
ODM
CWM
IMM

OMG

Semantic Web

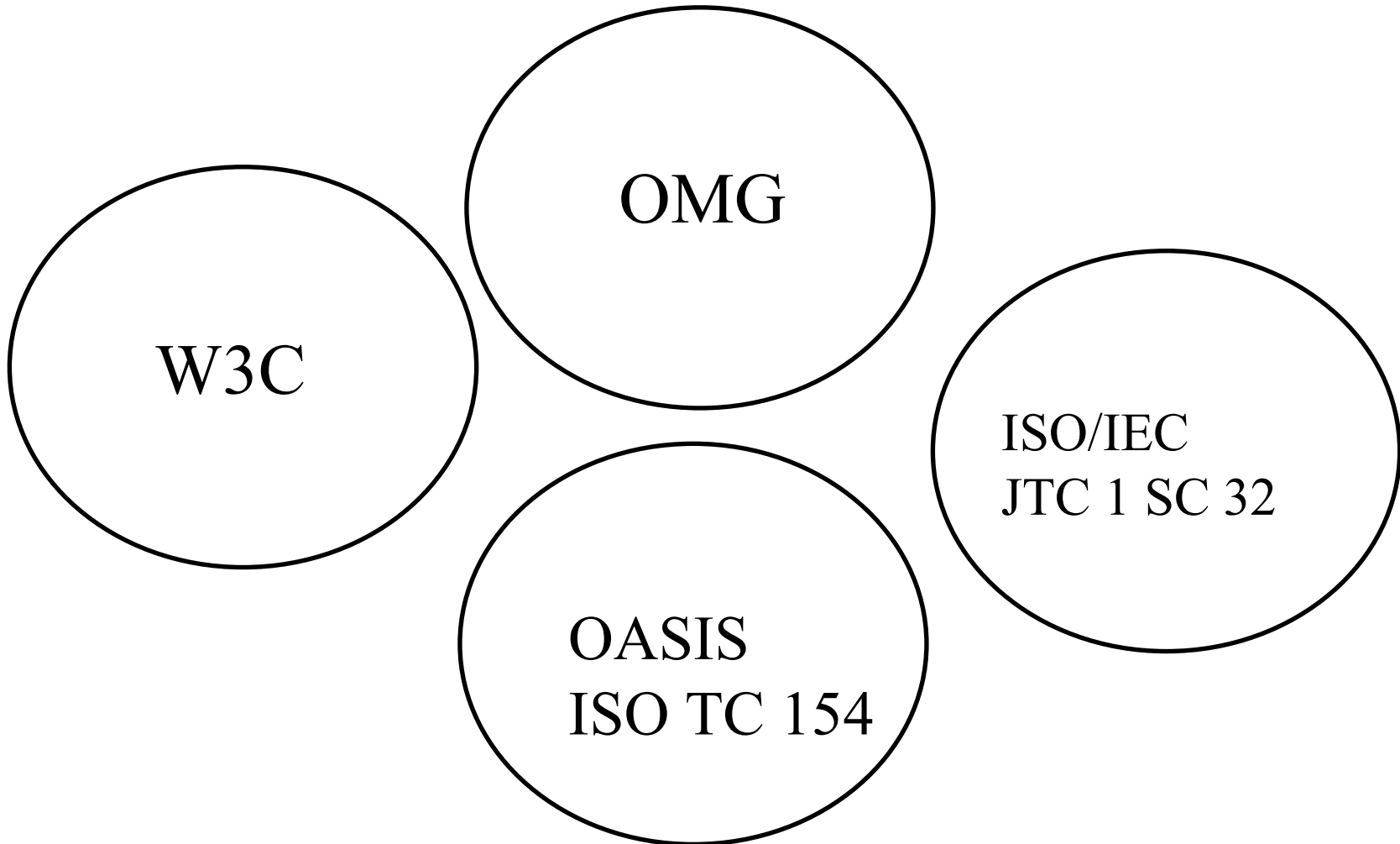


W3C

Standards Development

Semantics Management and Semantics Services – Semantic Computing

Align, Co-develop, Fast Track, PAS Submission ...

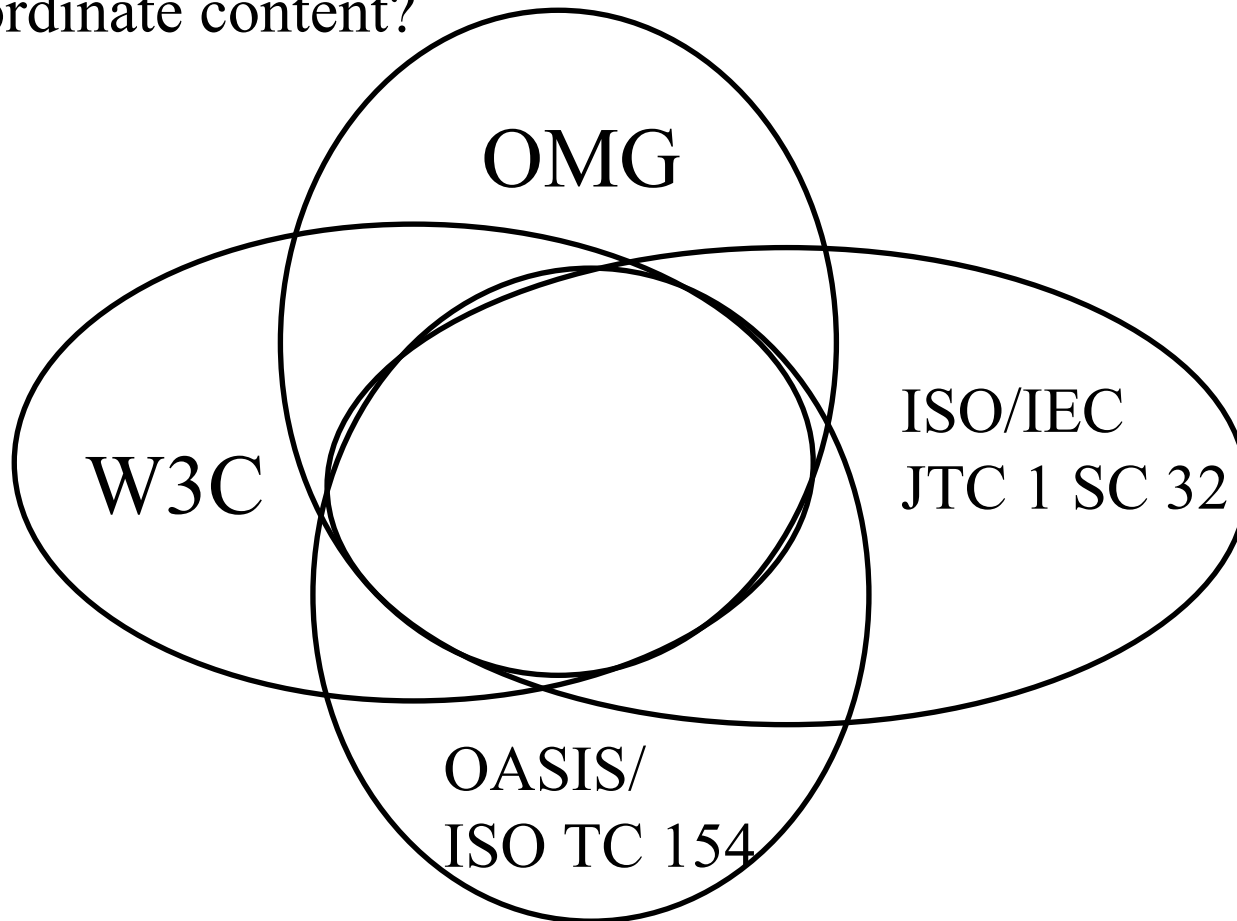


Standards Development

Semantics Management and Semantics Services – Semantic Computing

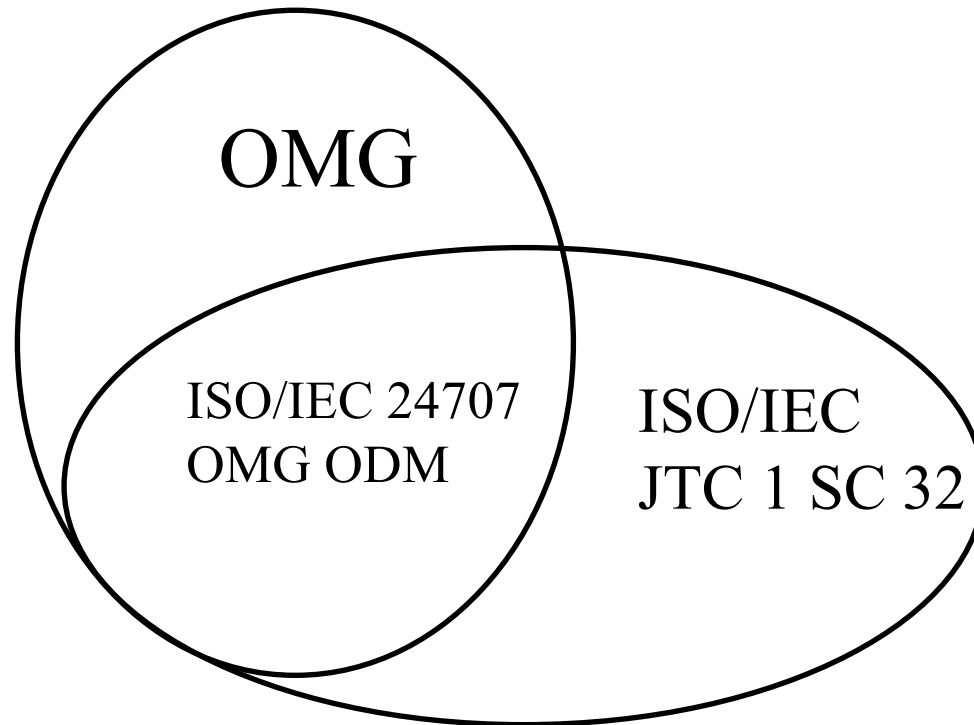
Align, integrate, co-develop, Fast Track, PAS Submission ...

Can we coordinate content?



A Success

Some text and figures are identical in the two standards.



ISO/IEC 20944 – Common Logic
OMG Ontology Definition Metamodel

Standards Development

Semantics Management and Semantics Services – Semantic Computing



Ongoing effort

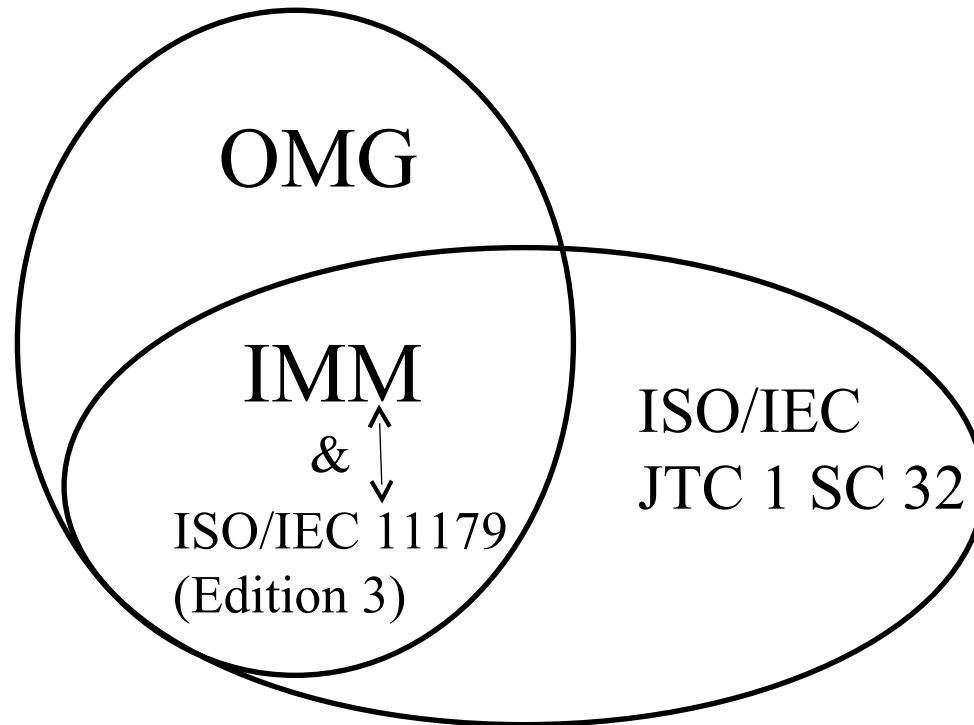
ISO/IEC 11179
(Edition 3)

ISO/IEC
JTC 1 SC 32

Standards Development

Semantics Management and Semantics Services – Semantic Computing

Hopeful?



Other Possibilities



- ❖ OASIS ebXML Registry
- ❖ W3C Semantic Web Deployment WG
- ❖ TC 37

Acknowledgements



- ❖ John McCarthy, LBNL
- ❖ Kevin Keck, LBNL
- ❖ Harold Solbrig, Apelon

- ❖ This material is based upon work supported by the National Science Foundation under Grant No. 0637122, USEPA and USDOD. 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.