

Data Reference Model: Update on Status

Brand Niemann

Co-Chair, Semantic Interoperability Community of Practice (SICoP)
Best Practices Committee (BPC), CIO Council, and
Enterprise Architecture Team, Office of Environmental Information
U.S. Environmental Protection Agency

March 7, 2005

Overview

- 1. Data Interoperability Paradigm Shifts
- 2. Limitations of ISO/IEC 11179
- 3. DRM Objectives and Use Cases
- 4. DRM Volume Strategy
- 5. Semantic Technology Profiles

1. Data Interoperability Paradigm Shifts

- Organizational:
 - The National Infrastructure for Community Statistics (NICS) Community of Practice (CoP) wants to make its data “NICS Ready” by publishing it to the Web in such a way that others can easily reuse it! (Like buying a new TV that is HDTV Ready.)
- Technical:
 - The conceptualization of new technical systems suffers from “technological presbyopia” – the condition of being able to envisage things more clearly the farther they are from the present realization – even though the prospective users may grow weary and skeptical while waiting for the future to arrive.
- Semantic:
 - Ontology and ontology patterns are the applied use of two basic tenets of software design and architecture, indirection and abstraction (see Appendix).

Source: Adding Value While Having Fun With EPA Data! Briefing to the EPA Office of Environmental Information Board of Directors, March 2, 2005.

1. Data Interoperability Paradigm Shifts

- Data is...*
- Data can be re-purposed
- Data can be mined
- Data can be modeled
- Data can be integrated & published
- Data standards can evolve (e.g. ISO 11179)
- Data architecture can implemented in ontology-driven information systems
- Appendix on Indirection & Abstraction

*This is a semantic approach!

2. Limitations of ISO/IEC 11179

- Initial DRM Work:
 - IAC White Paper, May 28, 2003 (Mike Lang, MetaMatrix) (See next slide).
- EPA Comments on the DRM, November 15, 2004.
- Ontolog Forum, at the EIDX "Semantic Harmonization" Panel Session (Jon Bosak), December 1, 2004:
 - "Explicit Semantics for Business Ontology - an interim report from the Ontolog Forum"
 - http://ontolog.cim3.net/cgi-bin/wiki.pl/wiki.pl?ConferenceCall_2004_12_01

2. Limitations of ISO/IEC 11179

- Business Integration Driven by Business Lines: A perspective on the Data Reference Model as it relates to Cross Agency Challenges. Standards Based Architecture to Support Federated Data Management. Concept Level WHITE PAPER Developed for the Federal Enterprise Architecture Program Management Office (FEA -PMO), Federal CIO Council, NASCIO, and Public Cross Government Initiatives Industry Advisory Council (IAC) Enterprise Architecture SIG, May 28, 2003:
 - This white paper discusses the limitations of ISO 11179 on page 46 as well as limitations of ebXML on page 50.

2. Limitations of ISO/IEC 11179

- Mike Daconta, February 11, 2005:
 - Set up a meeting with the ISO/IEC 11179 editors (Larry Fitzwater, Sam Chance, Nancy Lawler) on the evolution of 11179 to OWL?
 - First understand the plan for evolving 11179 and second evolve it towards greater semantics in its metamodel (e.g. rewrite Volume 1 to specify OOP and OWL principles).
- Ontolog Forum Discussions:
 - February 24, 2005, "Ontologies and Meta-Ontologies: Practical Considerations" (11179 to OWL, Upper Ontology Conversion, etc.):
 - http://ontolog.cim3.net/cgi-bin/wiki.pl?OntologDiscussion/MetaOntologies_And_Ontologies
 - March 3, 2005, Annual Ontolog Community Strategic & Work Planning Work Session (Collaborations with Duane Nickull, etc.):
 - http://ontolog.cim3.net/cgi-bin/wiki.pl?ConferenceCall_2005_03_03

2. Limitations of ISO/IEC 11179

- DRM WG Meeting, February 23, 2005, Informal Discussions:
 - No vendor implementation (Mike Daconta).
 - Only for legacy data (structured) holdings (Larry Fitzwater).
- Introducing Semantic Technologies and the Vision of the Semantic Web (DKR Version) ("DRM of the Future")
Delivered by SICoP to the CIO Council's Best Practices Committees, February 28, 2005.
 - Machine-processable with strong semantics for all three types of data (unstructured, semi-structured, and structured).
- Adding Value While Having Fun With EPA Data! Briefing to the EPA Office of Environmental Information Board of Directors, March 2, 2005 (See next two slide).

Data standards can evolve

- ISO 11179:
 - EPA Date:
 - The Date Data Standard provides for a standard representation of calendar date in data files for data interchange.
- Suggested Upper Merged Ontology (SUMO):
 - Date:
 - According to WordNet, the noun "date" has 8 sense (s) (see next slide).
 - SUMO is written in the SUO-KIF language (declarative semantics and machine processible) which has been translated to OWL – Web Ontology Language.
 - See <http://www.ontologyportal.org/>

Data standards can evolve

- Date:
 - 1. The specified day of the month; "what is the date today?".
 - 2. A particular day specified as the time something will happen; "the date of the election is set by law".
 - 3. A meeting arranged in advance; "she asked how to avoid kissing at the end of a date".
 - 4. A particular but unspecified point in time; "they hoped to get together at an early date".
 - The present; "they are up to date"; "we haven't heard from them to date".
 - 5. The present; "they are up to date"; "we haven't heard from them to date".
 - 6. A participant in a date; "his date never stopped talking".
 - 7. The particular year (usually according to the Gregorian calendar) that an event occurred; "he tried to memorizes all the dates for his history class".
 - 8. Sweet edible fruit of the date palm with a single long woody seed.

A Bit of Semantic Humor

- Enterprise Architecture:
 - Enterprise: A Star Trek Spaceship
 - Architecture: Blueprints
- So, Blueprints of the Spaceship Enterprise!

3. DRM Objectives and Use Cases

- Draft V.1 – February 17, 2005:
 - Objectives: “...continually analyze and optimize the cost of structuring data against the benefits of knowledge discovery, reuse and sharing.”
 - Comment: Maps exactly to recent workshop purpose and presentations (e.g. SAA’s PolicyNet):
 - See Semantic Conflict, Mapping, and Enablement: Making Commitments Together:
 - » http://colab.cim3.net/cgi-bin/wiki.pl?ExpeditionWorkshop/SemanticConflictMappingandEnablement_MakingCommitmentsTogether_2005_02_22
 - 1. What the DRM is.. “Processes, methods, and techniques for using metadata to enable the interoperability and integration of live information systems: This is what the DRM implementation profiles will accomplish.”
 - Comment: Maps exactly to SICoP SCOPE emphasis, namely that data architecture can implemented in ontology-driven information systems (See next slide).

Data architecture can implemented in ontology-driven information systems

- **Ontology-Driven Information Systems:**
 - **Methodology Side** – the adoption of a highly interdisciplinary approach:
 - Analyze the structure at a high level of generality.
 - Formulate a clear and rigorous vocabulary.
 - **Architectural Side** – the central role in the main components of an information system:
 - Information resources.
 - User interfaces.
 - Application programs.

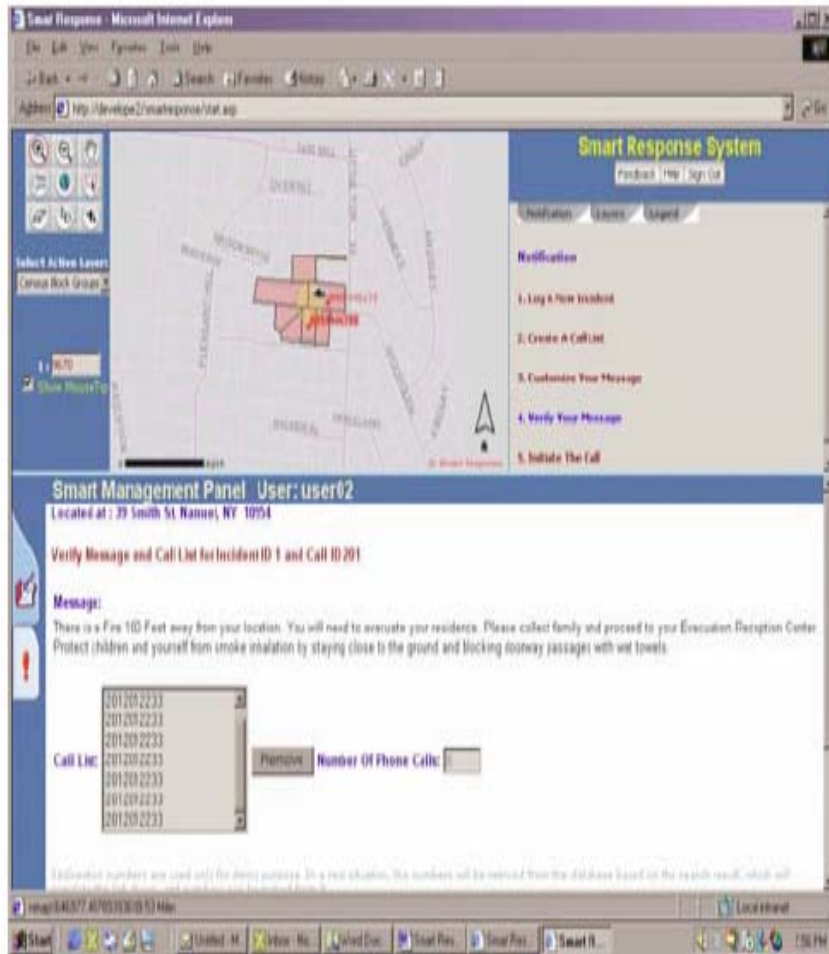
See for example: Nicola Guarino, Formal Ontology and Information Systems, Proceedings of FOIS '98, Trento, Italy, 6-8 June 1998.

3. DRM Objectives and Use Cases

- Draft V.1 – February 17, 2005 (continued):
 - 3.1 Inter-agency information sharing: As agencies coordinate and document information models over time, a bottom-up “government ontology” evolves over time.
 - Comment: Maps exactly to upcoming Semantic Web Applications for National Security Conference (SWANS) where Trade Show vendors demonstrate support for RDF/OWL.
 - At least one of those vendors is showing the DHS/DOJ National Information Exchange Model (NIEM) Information Sharing Use Case (1)* (See next slide).

*Global JXDM Executive Briefing, February 28, 2005.

GeoResponse - Voice + GIS = Multimodal Notification



- Features:
 - Report an Event
 - Geocode the Event
 - Define the Call List
 - Customize your Message
 - Make the Call
 - Track and Map Responses
 - Trigger another Process
- Award-winning VoiceXML Web Service from Broadstrokes at GeoResponse.com to be featured at the SWANS Conference Trade Show, April 7-8, 2005.
- Recently integrated with WSRP/CAP in cooperation with Starbourne/Oracle Team.
- Similar to DHS/DOJ Information Sharing Use Case (February 28, 2005)!

Implementing the Norfolk Southern Graniteville Derailment Scenario for the new Emergency Response Architecture!

4. DRM Volume Strategy

- Option A (Mike Daconta, Presenter):
 - Volume 1: DRM Overview
 - Volume II: DRM Management Strategy
 - Volume III: Data Description
 - Volume IV: Data Sharing
 - Volume V: Data Context

Note: I have asked the SCoP membership to review and vote by the March 7, 2005 COB Deadline. I think that most will prefer Option A.

4. DRM Volume Strategy

- Option B (Terry Hardgrave, Presenter):
 - Volume 1: DRM Overview
 - Volume II: Database Taxonomy and Exhibit 300 DRM Guidance
 - Volume III: DRM Metadata Repository Guidance for Federal Agencies
 - Volume IV: Catalog of Federal, Experimental and Commercial Metadata Repositories

5. Semantic Technology Profiles

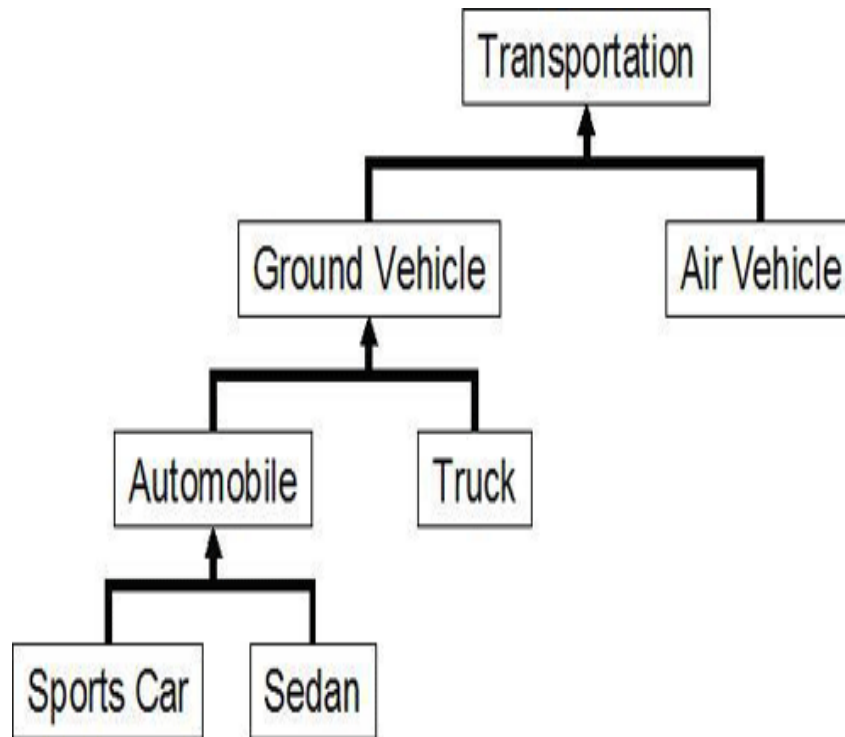
- Mike Daconta's proposal to XML CoP and SICoP, September 17, 2005:
 - If you think of the three areas (context, exchange and data element description) – an XML profile would look something like this:
 - 1. Several XML Topic Maps (business, security and service) with links to #2.
 - 2. an XML Schema (like the Watchlist schema) that can be exchange via #3.
 - 3. Web Services.
 - Of course in the SICOP, the above 3 would be:
 - 1. OWL Ontologies for each context with links to #2.
 - 2. XML Schema (or possibly an RDF Schema) that can be exchanged via...
 - 3. Semantic Web Services (that could be composed into larger services).
 - Note: The SICOP version has less current tool support but potentially better inference and rule integration.

5. Semantic Technology Profiles

- Current List:
 - EPA Region 4.
 - Enterprise Architecture – FEA-Reference Model Ontology (FEA-RMO).
 - Taxonomies – Formal Ontologies (Michael Daconta Recently Published Paper – See Next Slide).
 - Indicators – See previous.
 - Community Statistics – In process with NICS.
 - FHA/NHIN – In process with Ontolog Forum.
 - ISO 11179 – In process with DHS, xmdr.org, Ontolog Forum, etc.
 - More to be announced as part of the SICoP Module 3 White Paper Development – Implementing the Semantic Web.

Formal Taxonomies for the U.S. Government

Transportation Class Hierarchy



OWL Listing:

```
<?xml version="1.0"?> <rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:daml="http://www.daml.org/2001/03/daml+oil#" xmlns="http://www.owl-ontologies.com/unnamed.owl#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xml:base="http://www.owl-ontologies.com/unnamed.owl"> <owl:Ontology
  rdf:about=""/> <owl:Class
  rdf:ID="Transportation"/> <owl:Class
  rdf:ID="AirVehicle"> <rdfs:subClassOf
  rdf:resource="#Transportation"/> </owl:Class>
  <owl:Class rdf:about="#GroundVehicle">
  <rdfs:subClassOf
  rdf:resource="#Transportation"/> </owl:Class>
  <owl:Class rdf:about="#Automobile">
  <rdfs:subClassOf> <owl:Class
  rdf:ID="GroundVehicle"/> </rdfs:subClassOf>
  Etc.
```

Source: Formal Taxonomies for the U.S. Government, Michael Daconta, Metadata Program Manager, US Department of Homeland Security, XML.Com, <http://www.xml.com/pub/a/2005/01/26/formtax.html>

5. Semantic Technology Profiles

- Toward a National Unified Geospatial Enterprise Architecture: Seeing the Way Forward Together:
 - The Geospatial Semantic Web Interoperability Experiment – Joshua Lieberman, Traverse Technologies, OGC, DARPA/DAML, etc.:
 - http://colab.cim3.net/cgi-bin/wiki.pl?ExpeditionWorkshop/TowardaNationalUnifiedGeospatialEA_SeeingtheWayForwardTogether_2005_03_15

Appendix on Indirection & Abstraction

- Ontology and ontology patterns are the applied use of long-time, fundamental engineering patterns of indirection and abstraction.
 - Chapter 7 in *Adaptive Information: Improving Business Through Semantic Interoperability, Grid Computing, and Enterprise Integration*, Pollock and Hodgson, Wiley Inter-science, 2004.

Appendix on Indirection & Abstraction

- Selected tidbits:
 - Ontology is simply the enabler for software engineers and architects to apply core problem solving patterns in new and innovative ways.
 - Indirection is a concept that is use to plan for future uncertainty.
 - Simply put, indirection is when two things need to be coupled, but instead of coupling them directly, a third thing is used to mediate direct, brittle connections between them.
 - By leveraging indirection in the fundamental aspects of the technology, semantic interoperability is built for change, and this built-in flexibility differentiates semantic technologies from other information-driven approaches.

Appendix on Indirection & Abstraction

- Architects of both software and physical structures routinely use the principle of abstraction to isolate complex components and reduce the scope of a problem to be solved (“see the forest for the trees”). By definition, ontology is abstraction and is the ultimate abstraction tool for information.
- Example: Imagine a scenario of using a pivot data model without abstraction – it would require the aggregation of all of the data elements in a particular community – the result could be the a community of 500+ applications, each application with approximately 100 data elements, requiring a pivot model with about 50,000 data elements – an abstracted model could conceivably be capable of representing this information in far fewer than about 100 date elements!
 - See Demonstrations of SICoP Pilot Projects for EPA Managers, August 16, 2004, Semantic Information Management (Unicorn): Integrating Health and Environmental Information to Protect American Children”, at <http://web-services.gov>