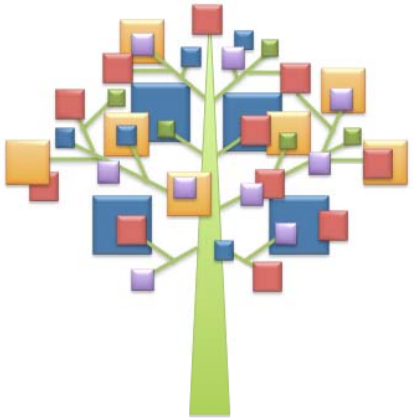


Ontology-Related Metadata Standards

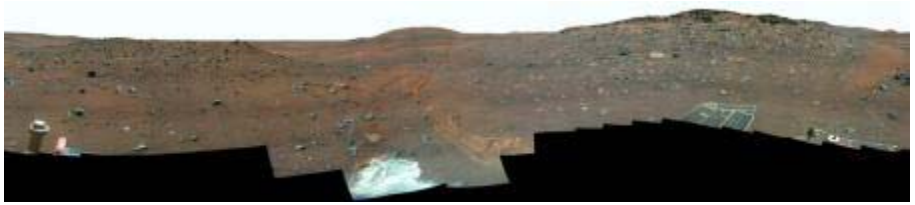
ISO Metadata & Metamodel Standards for Ontology Work & Related Activities at OMG

Elisa Kendall
Thematix Partners LLC
3 Nov 2011

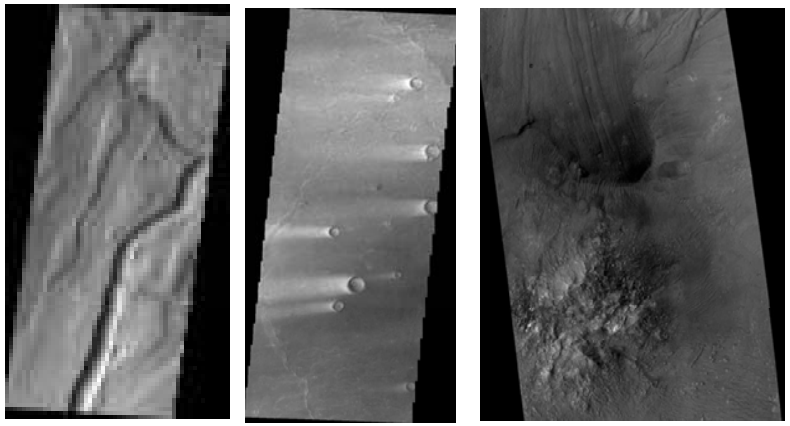
Content management

Mars was photographed by the Hubble Space Telescope in August 2003 as the planet passed closer to Earth than it had in nearly 60,000 years. Image Credit: NASA, J. Bell (Cornell U.) and M. Wolff (SSI)

A sunset on Mars creates a glow due to the presence of tiny dust particles in the atmosphere. This photo is a combination of four images taken by Mars Pathfinder, which landed on Mars in 1997. Image credit: NASA/JPL

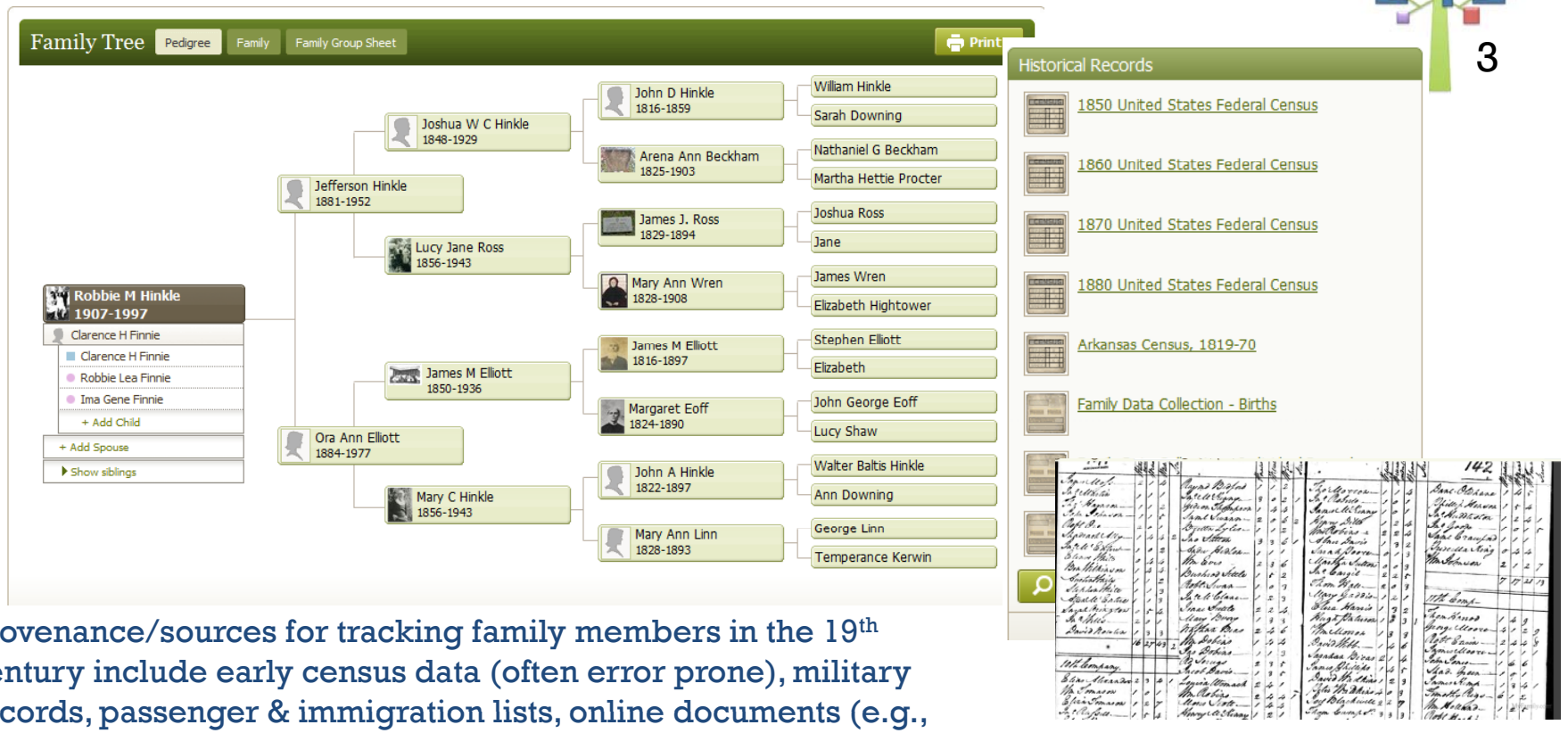


Recent images from instruments on board the Mars Reconnaissance Orbiter take much more detailed, narrower views of specific features of the Martian surface. Image credit: NASA/JPL



The Planetary Data Store (PDS) is a distributed repository of 40+ years' imagery & data taken by a range of instruments on many diverse missions, available for scientific research.

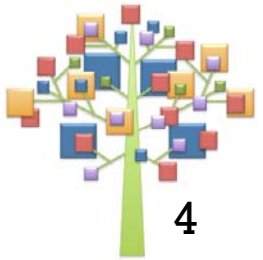
Smart search



Provenance/sources for tracking family members in the 19th century include early census data (often error prone), military records, passenger & immigration lists, online documents (e.g., county histories, church histories, etc.)

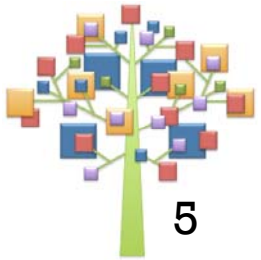
- ***Historical/forensic research requires cross-domain search of a wide variety of resources within a given geo-spatial/temporal context***
- ***Similar capabilities are essential for business intelligence, law enforcement, government applications – all require terminology reconciliation***

Terminology & metadata standards landscape



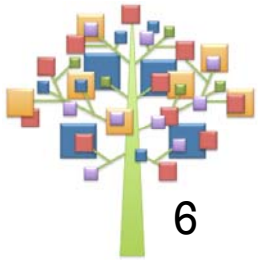
- There are many –
  - domain independent – such as ISO terminology standards maintained by TC 37, and metadata standards maintained by SC32
  - domain specific – such as ISO 20022, a widely used financial industry messaging scheme
  - community-driven standards that have broad appeal & uptake, such as the DCMI Dublin Core Metadata Terms & the W3C Simple Knowledge Organization System (SKOS)
  - many more developed by vertical industry organizations, such as the Open Travel Alliance, who develop the metadata and messaging schemes for travel data interchange

Terminology & metadata standards landscape



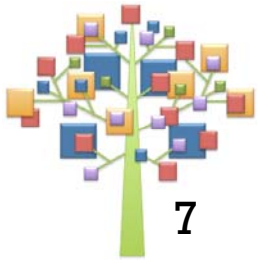
- Ontology alone isn't enough to address broad-based, cross-domain, cross-organizational interoperability requirements, but neither is metadata alone
  - emergency services, intelligence, law enforcement, security, cancer research, other government infrastructure & distributed repositories such as the planetary data store – well understood domains where the use of rigorous metadata standards is expected, and where ontology work has been promoted
  - industry-wide information interchange networks for financial services, healthcare, travel have been using and promoting metadata standards for many years, but are just beginning to adopt knowledge-based technologies
  - cloud computing – which will ultimately need information and systems interoperability on a scale that is just beginning to be understood

ISO Joint Task Force (JTC) 1 Structure



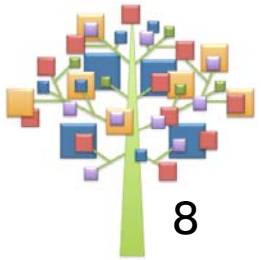
- JTC1 – responsible for all information technology standards
- SC32 – responsible for data management and interchange standards, http://www.jtc1sc32.org/
- SC32 WG1 – e-Business
- SC32 WG2 – Metadata
- SC32 WG3 – Database Languages
- SC32 WG4 SQL Multimedia and Applications Packages
- A number of the SC32 standards are freely available – the list and links to the documents, which includes ISO 11179 & Common Logic, is on the sub-commitee home page

ISO Metadata Standards Working Group – JTC1 SC32 WG2



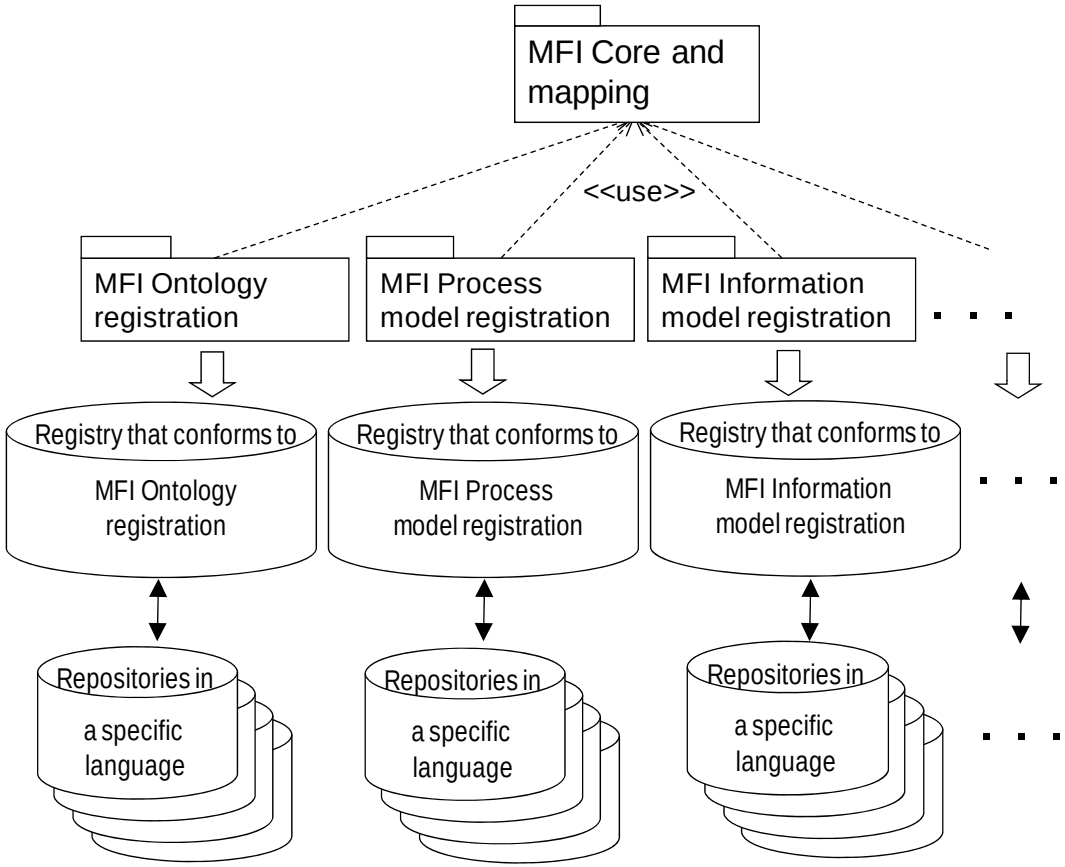
- SC32 WG2 is the ISO working group responsible for metadata and related standards – http://www.metadata-standards.org/
- Best known for development & maintenance of ISO 11179 Metadata Registration, which is now in FDIS for Edition 3
- Responsible for ISO PAS support for several well known OMG standards, including MOF (the Meta Object Facility) and XMI (MOF XML Interchange)
- Manage ISO Common Logic, ISO 24707, and will publish the Ontology Definition Metamodel (ODM) Version 1.1 (due to be published spring 2012) through the ISO PAS process
- Several metadata registration standards that are related to / depend on ISO 11179 are also maintained by SC32 WG2, including ISO 19763

ISO 19763 – Metamodel Interoperability



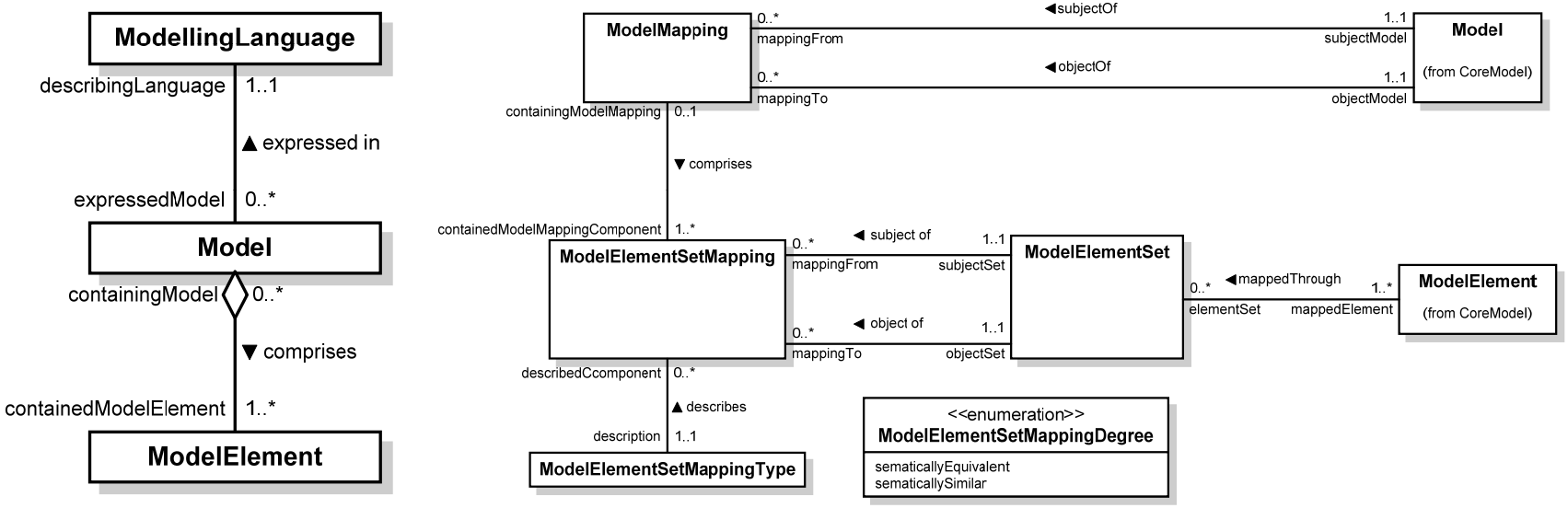
- Extends ISO 11179 for metamodel interoperability
- Provides mechanisms and additional metadata for model registration
- Includes a core set of metadata that are common across other parts of the standard, designed for general model to model mapping (primarily syntactic)
- Has a number of parts designed to support mapping of ontologies, process models, services, information models, "role and goal" models
- Parts 1 and 2, the reference and core model sections are an ISO standard
  - other sections are in process, most relatively early drafts
  - the original core model (formerly Part 2, merging with Part 4 to become Part 10) and ontology sections are further along than most, latest version is dated Oct 14, 2011, discussed last week in Crete
  - Available on the SC32 WG2 site, with target publication dates and more detail on where each section is in the process

General structure



Core and Mapping Model (new Part 10) relationship to other parts of the standard

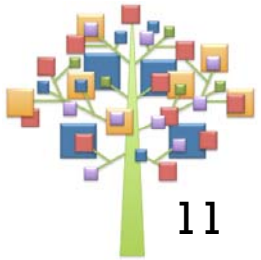
Core & Basic Mapping Model



All other parts of the standard extend this basic model, which is intended to enable model registration, alignment & mapping

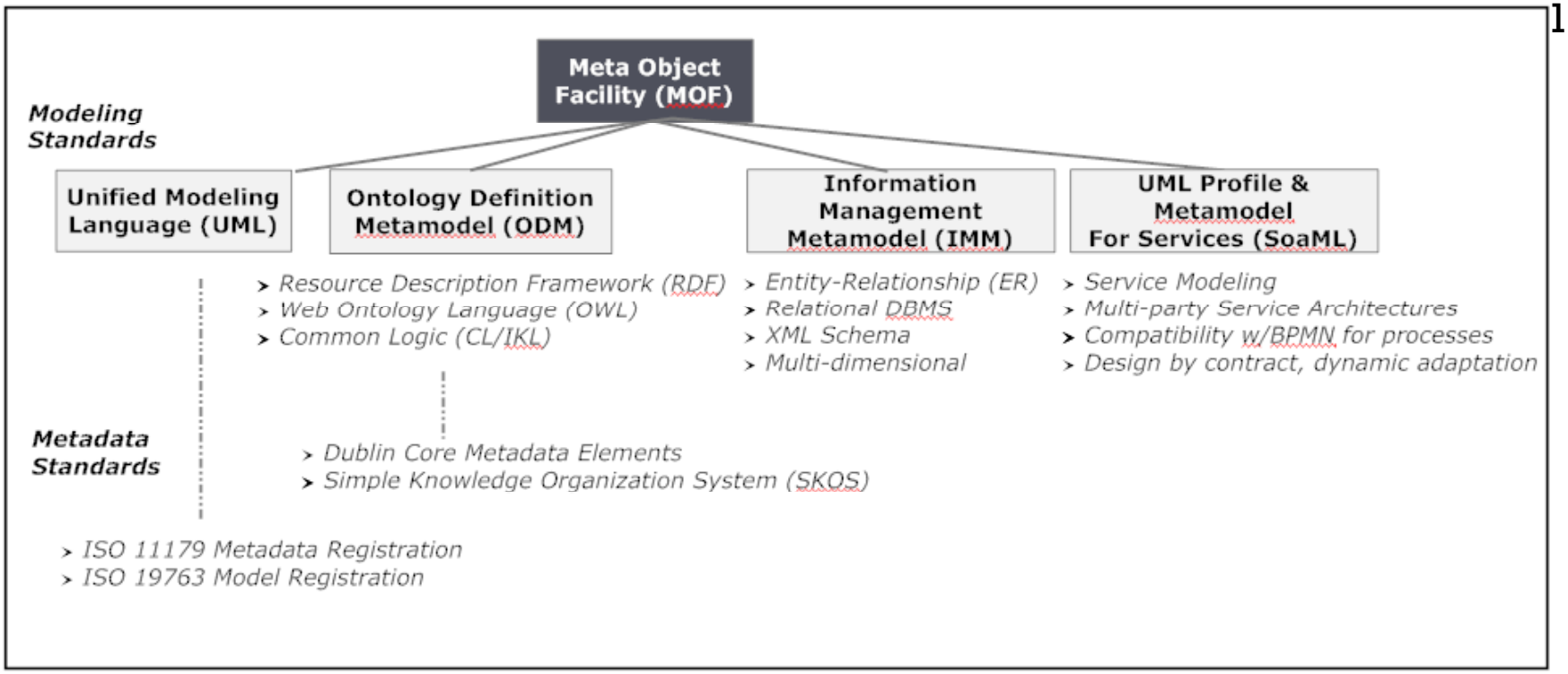
The metadata that supports this is defined in ISO 11179, in Part 3 Edition 3 in particular (currently in FDIS status)

Relationships between SC32 WG2 & the Object Management Group (OMG)



- Formal liaison relationship
- Good working discussions / intent is to limit overlap, encourage collaboration to the degree possible
- Some overlap in membership, primarily among the members from Japan and the US
- ISO PAS relationship is quite good, with use of ISO format by OMG task forces, collaboration with ISO editors on standards development
- Current work at OMG with the target of ISO PAS late next year includes the ODM and emerging Information Management Metamodel (IMM)

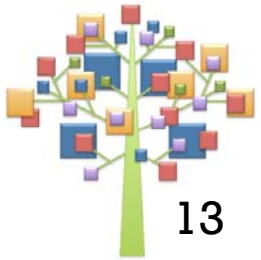
Basic structure of OMG standards

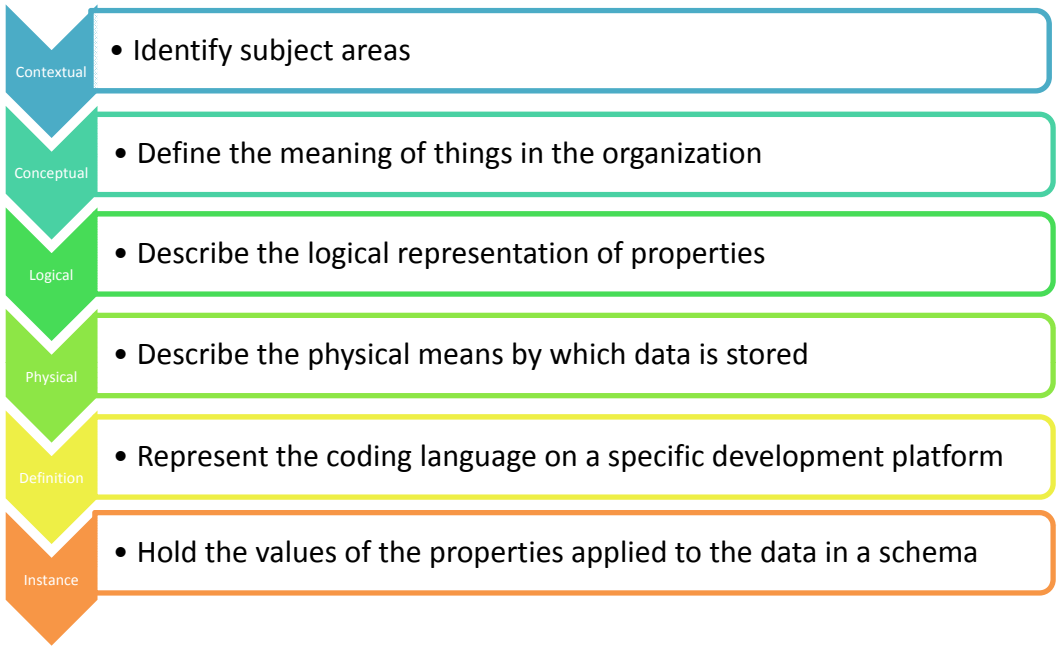


Grounding in MOF/UML facilitates

- *Model interoperability*
- *Reuse of common vocabulary, logical models across modeling approaches & asset types*

How they tend to be used





Ontology

Ontology, Conceptual ER, Conceptual Business Process …

ER, Relational, XML Schema

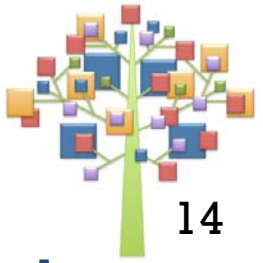
XML, source code, scripting languages, stored procedures…

Physical KBs, databases, asset repositories…

**Layering diagram courtesy Kenn Hussey*

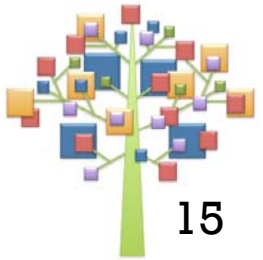
- Knowledge Representation / Management for Large Scale Applications
  - Provide broad metadata, process, service & asset management facilities (including feedback/lessons learned…)
  - Enable rich cross-domain, cross-process, cross organizational modeling supported by mapping & transformation services to provide maximum flexibility, interoperability
  - Leverage standards and best practices in information architecture, metadata modeling, management, registration, and governance, and asset management & registration
  - Provide incremental reasoning capabilities for model validation, transformation services

Status of OMG Standards



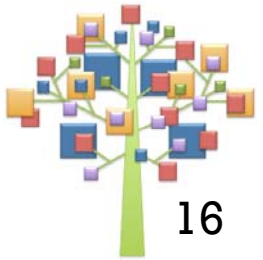
- Update to support OWL 2 is underway – ODM 1.1 revision planned for December (draft convenience document) / March (spec revision)
- CL Metamodel is identical to the UML diagrams in ISO 24707
- High degree of synergy between ODM and Topic Maps ISO 13250 working group
- Current work in ISO JTC1 SC32 to update ISO 11179 (Metadata Registration) references ODM; also addressing alignment with SKOS (Simple Knowledge Organization System) and Dublin Core
- All ODM metamodels are referenced and used in ISO CD 19763 (MMF – Metamodel Framework, Model Registry specification)
- Mappings from multiple components of IMM (*e.g.*, ER, XML Schema, etc.) are planned

Current activities



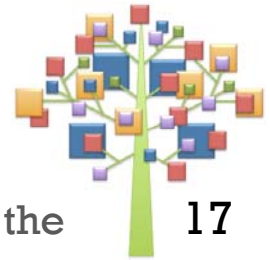
- Revision of ODM to support OWL 2, fix bugs is in work
- RFP to support APIs for knowledge base access (API4KBs) – submissions will reuse ODM RDF and OWL metamodels, initial submission effort is underway (meeting in Santa Clara in December)
- Extensions under consideration include mappings to
  - IMM Metamodels (ER, XML Schema …)
  - SysML
  - Production Rule Representation (PRR) specification
  - BPMN (Business Process Modeling Notation) and the Business Motivation Model (BMM)
- Date Time Vocabulary (in SBVR primarily, recently adopted), mapping to ODM/OWL is under development (part of the finalization task force effort)

Metadata related OMG activities



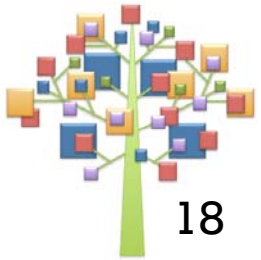
- Planned standardization of ODM models for
  - Dublin Core Metadata Terms (DCMI, http://www.dublincore.org/)
  - Simple Knowledge Organization System (W3C, http://www.w3.org/TR/skos-reference/
- OMG Common Terminology Services 2 (CTS2) uses a combination of Dublin Core, SKOS, and ISO 11179 as a basis for term registration (adopted June 2011)
- Because CTS2 uses ISO 11179-3 Metadata Registry and may require the ISO 19763 Core and Ontology models, development of a standard ODM/OWL representation is under consideration
- ISO 11179 and 19763 reuse ISO 704/ISO 1087 Vocabulary for Terminology Work – an ODM/OWL model is under development, planned for joint OMG / ISO TC37 standardization

Current Model Submission Process at OMG



- Electronic files are submitted, in xmi, to the task forces and ultimately to the OMG AB for review
- Conventions for URI specification (for the target URI) are specified in the SMSC recommendations, but are tough to find unless you know where to look
  - Document naming and versioning is described in http://www.omg.org/cgi-bin/doc?smsc/2007-09-11 and in the Hitchhiker's Guide, available at http://www.omg.org/cgi-bin/doc?hh
  - The OMG Hitchhiker's Guide specifies that URIs for the machine readable artifacts should be http://www.omg.org/spec/<spec acronym>/<date, in YYYYMMDD form>/filename.extension, but provides little guidance on naming conventions for the actual machine readable files, nor on the date (presumably the 4 week rule date and/or date of publication of the final version of the adopted specification)
  - Some starting point requirements for the contents of machine readable files (e.g., xmi) are captured in documents, see http://www.omgwiki.org/smsc/doku.php?id=document_archive, for example, and in SMSC telecon minutes, available at http://www.omgwiki.org/smsc/doku.php?id=teleconference_notes
- Because the process is not well understood, the OMG Architecture Board is now investigating best practices for our own process, including for metadata

Current Revision Task Force (RTF) Process

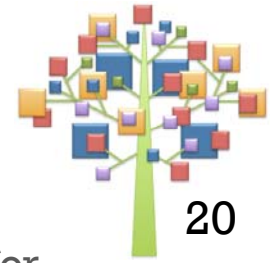


- Changes to specifications can only be made in response to issues; with an issue cut-off deadline
- Issues are (or should be) documented as a problem
- The resolution must be documented as detailed editing changes/replacements to text and/or diagram(s)
- Resolutions must be approved by a quorate majority of the RTF, and documents provided for AB review must be marked with change bars annotated with the issue number
- This approach has traditionally been document centric rather than electronic file (model) centric
  - presents challenges for providing clear change management functionality for all electronic files
  - particularly for content models
  - There is no guideline or best practice that all aspects of a model be covered within the document.
- Even for document changes there is not widespread use of markup to map document change bars to issue numbers.

Naming conventions & change management gaps

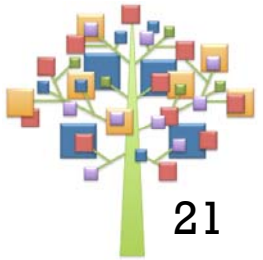
- Naming conventions and versioning policies are needed
  - for namespaces for "rapidly changing content models", if the current policy is not adequate – at a minimum, a recommendation for the potential of an additional level of hierarchy for large specifications with many modules may be needed
  - for namespace prefixes (abbreviations) to be used in individual modules, especially where a specification has multiple modules
  - for model elements – there are some rules of thumb for UML models, but these should be readily available to submitters in the context of model evaluation criteria
    - These may vary by "content community" – data modelers often use underscores at word boundaries, spaces in names – which semantic web tools may not handle well, semantic web practitioners use camel case; use of unique identifiers may not be UML-tool friendly
    - Guidelines for appropriate naming of class or class-like elements (SBVR concepts), properties or property-like elements (SBVR roles, for example), individuals (objects in UML) such that conceptual models developed in UML or a UML profile can be exported consistently to various electronic file formats are also needed
  - for managing versions of electronic files in a manner that provides traceability at the element level – especially in light of ontology changes such as the addition or deletion of owl:disjointWith relationships, as one example
  - Methodology for addressing these issues may be needed on a file type by file type basis, but some aspects should be common

More on the change management gap



- Change management and traceability is not well defined at the element level for content models (and possibly electronic artifacts produced by OMG in general)
  - the approach to element-level versioning to support reasoning is to track two parallel streams: one for additions to the ontology, one for retractions; these are managed as two separate files typically
  - MOF versioning suggests that users link to a workspace, and use of the default workspace would get the latest version, and tools provide an indication of the differences
  - MOF versioning does not support reasoning about the differences so that users can determine the impact of applying the changes, which is frequently required in the semantic web community – the addition or deletion of certain axioms can (and often does) change downstream reasoning results, especially where complex dependencies exist
  - A mechanism that preserves versioning detail in artifacts that are generated from such models, such as RDF/XML serialized OWL, and is compatible with both of the above approaches is needed, balanced to support the usability/performance of those artifacts
- A methodology and best practice white paper, and instructions to submitters indicating how to support it for submissions, finalization and revision task forces is in work

Modularity gap



- Guidance on modularity for large content models is lacking
  - When is a model "too big"?
  - Can individual parts evolve independently, and if so, shouldn't that dictate module boundaries?
  - Can individual parts be used independently, and if so, shouldn't that also dictate module boundaries?
  - For ontologies, particularly OWL ontologies, recommended approach is to manage any "rarified" individuals separately from the "schema" part of the model to facilitate reasoning during development (*i.e.*, if you add an individual that creates a logical inconsistency, most reasoners won't load the ontology at all)
  - What criteria should be used to determine whether or not a particular ontology has been modularized appropriately?

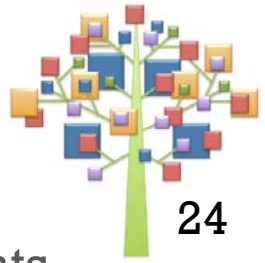
More on modularity

- Current practices in the semantic web community address modularity either statically or dynamically
  - Static approaches include adherence to "DL-Safe" rules and suggestions in papers by Alan Rector
  - Dynamic approaches include use of tools that can assist in determining whether or not an ontology is appropriately modularized, using reasoning to perform rewriting with respect to soundness and completeness for a given ontology
- OMG is still working through some of these issues to give guidance to submitters and revision task forces
  - For example, given that we modularize FIBO, (financial industry business ontology) and use constructs such as owl:imports to link ontologies together, recommend use of GRDDL to point to the latest version of an electronic file, if appropriate, rather than the explicit URI reflecting the date it was published
  - Better support for content negotiation on the OMG site may be needed

Metadata gap

- There is currently no requirement with regard to metadata that should be included in electronic files at OMG
- This is critical for content models in particular, but some elements should be required in every xmi or other electronic representation of a standard at OMG
- In September, at the OMG meeting in Kissimmee, FL, we proposed metadata requirements at two levels, the model level and the element level within *any* model
  - Most elements will be optional at the element level
  - Consistent use of the same elements (properties, tags) improves readability, facilitates automated specification generation (including change bars in documents), and enables better search
  - Model level metadata could include identification of the relevant technology areas, domain, etc., which could use taxonomies that are reused on the OMG site to enable better search (e.g., through RDFa)
- The presence of common metadata in xmi files would make it much easier for task forces and the AB to review electronic files and provide the same level of rigor in standards evolution that we do at the document level

Next steps



- Review of the potential issues, and identification of requirements for change management
- Consider the impact on work in progress and/or planned
  - SE DSIG SysML Quantities Units Dimensions and Values (QUDV – currently an informative annex to SysML)
  - BMI Date Time Vocabulary (proposed for adoption on Monday, approved by the AB)
  - FDTF Property & Casualty conceptual model (proposed for adoption today)
  - MARS Information Exchange Policy Vocabulary (IEPV – in work)
  - FDTF Financial Industry Business Ontology (FIBO – RFC anticipate in December / March)
  - Any content model in particular, but more importantly for all electronic files in general
- Provide recommendations and policies for metadata management for all OMG specifications going forward