Formalizing the DRM using Standards Evaluation

Data Access Technologies, Inc.-

Oct 24th, 2005

Overview

At the request of Mike Daconta (DRM) and Richard Soley (OMG) Cory Casanave (Data Access Technologies) has reviewed the DRM 2.0 documents for the purposes of evaluating;

- Aligning the DRM with applicable standards
- Utilizing these standards to further the purpose of the DRM

The documents evaluated were;

- DRM Draft Version 2.0
- DRM Management Strategy (Oct 17th)
- DRM Implementation through iteration and testing version 1.0

The standards in question are;

- Object Management Group (OMG) Model Driven Architecture
  - Unified Modeling Language (UML)
  - Meta Object Facility (MOF)
  - XML Model Interchange (XMI)
- W3C XML Schema
- W3C RDF and OWL

Note; There were no structured data resources provided to document the DRM, resources that could be provided such as; UML Model, XML Schema, Ontology.

Summary of findings

The DRM uses a casual style of modeling that is sufficient for communicating ideas but does not take advantage of state-of-the-art techniques in modeling, model driven architecture and Ontologies. Additional work is recommended to formalize the DRM and provide a consistent "matched set" of artifacts that define the DRM in terms of leading standards. Informal diagrams in the DRM should be replaced by standards compliant diagrams to aid in the precision and understandability of the document.

The potential down-stream advantages of a more precise DRM would be greater data sharing and more widespread adoption.

Potential

The founding principle of the DRM is to support human and machine interoperability at a wide scale embracing legacy and future information technologies.

The potential exists for the DRM to be precisely specified using existing standards such that full and automated interoperability would exist between DRM instances defined using different tools by parties that have never communicated. Further, the DRM could embrace multiple data representation standards with full interoperability. This would allow tools of choice to be used without sacrificing interoperability. For example, the DRM could be specified in such a way as to allow for the specification of DRM resources in UML, XML-Schema, Ontologies, SQL or even proprietary formats.

> The recommended path to this precision and interoperability would be to produce the normative description of the DRM as a MOF-UML model with standards based mappings to XML-Schema, an OWL Ontology as well as a UML Profile. This would allow DRM information to be created in or mapped to any of the above representations without data loss.

By providing a representation of the DRM that is "above" the wars between these technology camps we can embrace a much wider scope of existing and emerging data repositories, allowing for faster integration of the DRM into the mainstream of information technology.

Further, this specification can achieve a greater degree of precision, enabling automated integration of information sources and leveraging knowledge based techniques of information integration. Formalizing the DRM based on a model may also provide a mechanism to resolve some of the open DRM issues.

Findings

The diagrams in the document are non-standard and lacking in critical detail

The diagrams are informal and lack cardinalities, types and other detail that would allow for a precise interpretation of the document. The lack of precision may, at first, appear friendly, but on closer inspection would be cause for confusion.

This critical detail is not necessarily found in other forms, such as the attribute tables.

The attribute tables added detail to the diagrams but did not clear up critical information. See the questions in the appendix.

There is no formal specification of the DRM in a data modeling representation.

The are many ways to deliver data models such as the DRM, at least one and perhaps several should be specified and those artifacts carefully documented. Interoperability tests should be conducted based on independent implementation and sharing of information.

Note; It is our understanding that there is an intent to deliver a DRM XML-Schema.

The structure of the DRM attributes is useful for human consumption but not sufficiently well specified to provide for automated interoperability

A set of DRM documents intended for machine interoperability would require precise structure, definitions for attributes, enumerated values and mechanisms for references. None of these were found in the document. For Example; "An unambiguous reference to the resource within a given context" does not specify how such a reference is to be dereferenced or how to identify the context. "Format" does not say what the formats are or where to find known formats.

Structure of attributes specified as "strings"

Most of the attributes would suggest some kind of structure that is not specified, almost all are simple strings.

The end result is that a DRM document would not be useful for automated information extraction or integration. While a human "catalog" of information is useful, this is not the only design intent of the DRM.

Units

The content of the DRM lacks common concepts that would allow for interoperability, such as a concept of "units". Note that failure to specify units caused the loss of a $125 Million mars explorer.

URI For entities

Many "entities" are referenced in the Dublin core information based on text. In an environment such as the government, entities should be more carefully identified. There are existing efforts to provide for this capability that could be leveraged or allowed for.

Suggested Path

The following is the recommended path to formalizing the DRM. Note that one or all of these items could be done, or some could be made part of a later work. Only the first two work items would be a pre-requisite to the others.

Interpret the current document as a UML Meta Model.

A "meta model" is a model of the DRM "Language". The current DRM diagrams are essentially informal meta models. The meta model would be produced based on the OMG-MOF standard, a way to use UML to specify meta models. Tool vendors will be able to use the meta model directly to support tooling, such as the open source "Eclipse" environment.

Estimated time; One Week

Validate this model with the DRM team, resolve any questions.

As the DRM is modeled questions will invariably come up. If there is general agreement on the intent of the DRM these should be relatively easy to resolve. These are questions that should be resolve regardless of the other parts of this initiative.

Estimate time; 2 Weeks

Document the model

Most of the basic concepts are already defined in the DRM, but some additional detail will be required as the questions are resolved and the model structured. The model would be documented based on the existing work and the diagrams made visually appealing and part of the DRM.

Estimate time; 2 weeks

Produce a XML-Schema from the model based on existing standards

Based on the MOF-UML model a standard "XMI" specification can be produced along with its XML-Schema. This would guarantee that the DRM Schema would agree completely with the model specified and take advantage of the model based documentation. Review with DRM team

Estimated time; One week (Based on use of available automated tools),

Produce a corresponding OWL Ontology

It is relatively easy to produce an ontology directly from a meta model based on the "Ontology Definition Meta Model" (ODM) Standard. This can be 100% compatible with the meta model such that full interoperability can be achieved between the schema, meta model, ontology and other forms. Providing an ontology allows inference tools to be applied to DRM information and to allow Ontology environments to define DRM instances directly. Review with ONTAC.

Estimated time; 2 weeks

Produce a UML Profile

The UML profile for the DRM will define how to use standard UML tools to define DRM instances. This open the DRM up to the vast number of software engineers familiar with UML-2.

Estimated time; 2 weeks

Option; Produce reference transforms

Since these are standards-based forms (mostly XML), transforms could be produced between all of the forms, allowing easier DRM interoperability. These transforms would be the "reference implementation" of the DRM and could be provided in open source.

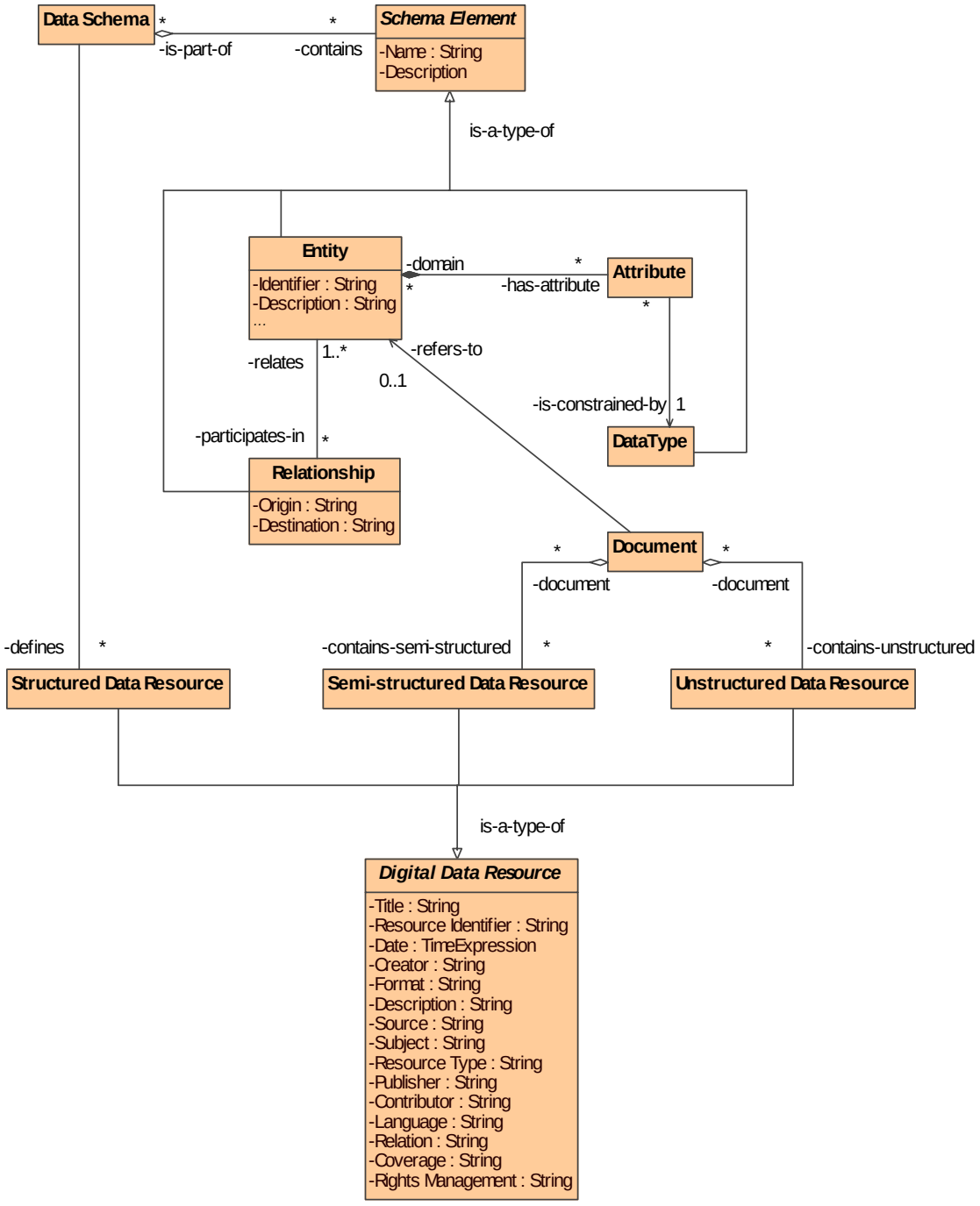
Estimated time; 4-6 Weeks

Inclusion in DRM

The standards based artifacts from the above processes could then become part of the DRM "package" and the diagrams from the meta model used to make the DRM more precise. These models could be included in the main document or in an appendix.

Appendix 1 – UML representation of the DRM Data Description

The following model was produced based on the DRM specification as an aid in evaluating the complexity and precision of the current model. A formalized DRM would look similar but have additional detail and precision.



Appendix 2 – Sample of questions encountered while modeling the DRM Data Description

The following question came up in a relatively quick review of the DRM and the attempt to model a portion of it. These are intended as examples and do not represent all the questions that would be uncovered in a full evaluation.

Can an entity/relationship/data type be part of more than one data schema?

Assume yes.

Can an attribute be part of more than one entity

Assume yes

Are data types intended to be "units" or integer/string, etc?

Assume strings & Integers but note requirement for units.

How many entities can participate in a relationship?

Assume one or more – but confusion with "Origin" and "Destination" suggest it may be a binary relationship.

Not clear how "relates" is different from attributes "Origin" & "Destination"

The "See" references in the attribute lists are so circular it is difficult to tell what the intent is. Clarify the attributes of each element type.

e.g. Relationship doesn't have a description?