

MDA Journal

December 2004



David S. Frankel

David Frankel Consulting

df@DavidFrankelConsulting.com

MDA, UML, and CORBA are Registered Trademarks of the Object Management Group. The logo at the top of the second page is a Trademark of the OMG.

www.bptrends.com

Comments on the U.S. Federal Government's Data Reference Model

<u>Introduction</u>	1
<u>Issue: The Scope of Commonality</u>	2
<u>A Basic Separation of Concerns</u>	2
<u>A Common Means for Defining Data Elements</u>	3
<u>Common Data Element Definitions</u>	3
<u>Common Data</u>	4
<u>OMG MDA® Modeling Levels</u>	4
<u>Organizational Scope</u>	5
<u>A Topology of Sharing</u>	5
<u>Value Chain Scope</u>	6
<u>Issue: Leveraging Semantic Web and MDA Technology</u>	7
<u>Leveraging the Semantic Web</u>	7
<u>Leveraging MDA7</u>	
<u>Using the Semantic Web and MDA Synergistically</u>	10
<u>Issue: Contexts and Business Collaborations</u>	11
<u>Conclusion</u>	11
<u>References</u>	12

Introduction

The U.S. Federal Government's Office of Management and Budget (OMB) has mandated that every federal agency's IT organization develop an enterprise architecture. To assist the organizations in developing such architectures, the OMB created the Federal Enterprise Architecture Program Management Office¹ (FEAPMO). (This move is closely related to the "E-Government" initiative that is pushing federal agencies toward Web Services based business transactions.)

The FEAPMO has released the Federal Enterprise Architecture (FEA).² The FEA has five parts:

- Business Reference Model (BRM)
- Performance Reference Model (PRM)
- Service Component Reference Model (SRM)
- Technical Reference Model (TRM)
- Data Reference Model (DRM)

The FEAPMO released the Data Reference Model (DRM) recently and has been soliciting feedback on it. The DRM is an architecture intended to support information sharing. It leverages the FEA's Business Reference Model and ISO's 11179 data standard³.

¹ [FEAPMO]

² [FEA]

³ [ISO 11179]



With the sponsorship of the General Services Administration (GSA), I have examined an edition of the federal DRM, dated September 2004. My evaluation follows.

(A summary of the GSA's approach to implementing the FEA was presented by George Thomas, Chief Architect of the GSA's Office of the CIO, at the OMG meeting in Washington D.C. a few weeks ago.⁴ GSA has taken ownership of five E-Government projects from inception through launch.)

Issue: The Scope of Commonality

There have been many efforts to share data within work groups, companies, government agencies, and whole industries, and some efforts to share data globally. Progress has been uneven, often two steps forward and one step backward.

One issue is the difficulty of obtaining broad agreement on data definitions. As a sharing effort expands from the small local work group, human, organizational, and technical barriers arise that make agreement more difficult.

While we need to share our efforts with the federal government, based on past experience, we would be wise to define modest and achievable objectives, while laying the groundwork for more ambitious objectives we may hope to achieve later with more solid commitment from the government.

A Basic Separation of Concerns

We need to distinguish between three forms of sharing:

- Having a *common means for defining* data elements
- Arriving at *common definitions for those data elements*
- Working on *data elements we have in common*

For example, two healthcare organizations might conceivably share the same means for defining data elements, but that does not necessarily mean that they share specific data element definitions. On the other hand, suppose they do, in fact, share the definitions of some data elements that describe an immunization schedule; that does not necessarily mean that they share actual immunization schedules that realize the data element definitions.

It is important to clearly differentiate among these three levels of commonality in order to separate concerns properly in a data architecture, and in order to allow organizations to develop an organized approach to gradually progressing from relatively easy to more difficult tasks. In any of its discussions, the DRM must make it crystal clear which of these concerns it is talking about.

A Common Means for Defining Data Elements

Even the most modest of the three objectives, establishing a common means for defining data elements would be no small achievement. Happily, ISO/IEC 11179 provides a common means for defining the structure of a data element

⁴ [GSA OCIO]

and, thus, supports this objective. A means to define the business context of a data element, as the DRM envisions, also supports this objective.

Standardization across the federal government agencies on the means for defining data elements would make it easier for one governmental organization to examine and assess the data element definitions of another organization. Common tooling could be used to support the following functions:

- Defining data elements
- Storing data element definitions in repositories
- Registering data element definitions
- Querying repositories of data element definitions
- Interchanging data element definitions
- Securing access to data element definitions

Thus, even in the absence of common use of the same data element definitions across organizations, having a common means for defining data elements is useful.

For example, consider our immunization data scenario. Two organizations might have different data elements defined for immunization records, but if they share the means of definition, they can readily review each other's definitions.

Common Data Element Definitions

Sharing data element definitions across organizations is a more ambitious undertaking. Having a common means of definition is a necessary but not sufficient condition for sharing data element definitions. The additional conditions that have to be satisfied in order to share data element definitions across organizations include:

- Agreement on the business context of the data element
- Agreement on the structure (syntax) of the data element
- Agreement on the meaning (semantics) of the data element

Agreement on all of the above is necessary to share the definition of a business element. This is a considerably higher hurdle than that posed by simply sharing the means of definition. In our immunization example, agreeing on how to structure the immunization data elements and on the meaning of each of the data elements depends on organizational and interpersonal factors, not to mention the technical challenges posed by disparate ways in which different organizations maintain and use immunization records.

Commonality of data definitions, when achieved, makes it possible to share applications based on those definitions across organizations, even when the organizations do not share the actual data that the applications manipulate.

Note that the current version of the DRM does not have provisions for the semantic grounding of data element definitions. The DRM relegates the specification of

the meaning of data elements to informal descriptions in English. That Semantic Web technologies have something to offer here is discussed later in this paper.

Common Data

Sharing actual data among organizations—i.e. sharing instances of the defined data elements—is even more difficult. The sharing of data element definitions is a necessary but not sufficient precondition for the sharing of data. Additionally, the organizations must:

- Satisfy security concerns, including confidentiality, authentication, and authorization
- Resolve potential conflicts over administrative control of the data, i.e., determine who is responsible for the data's integrity under what conditions
- Resolve technical problems of sharing data across firewalls (API-based access across firewalls is usually not possible, so architects usually have to rely on HTTP-based bulk transfers, which have performance issues when transferring large amounts of data)

OMG MDA® Modeling Levels

These different levels of sharing are related to the modeling levels that the OMG's Model Driven Architecture® (MDA) defines. This section briefly establishes that linkage.

The OMG modeling levels—called *metalevels*⁵— are related to the levels of sharing as follows:

- *Metalevel M2*: The specification of a means for defining data elements is at level M2. A formal M2 specification—i.e., a formal specification of a means for definition—is called a *metamodel*. ISO 11179 is an M2 thing.
- *Metalevel M1*: A particular data element definition is at level M1. Thus, a definition of an immunization date data element—defined via the means specified at M2—is an M1 thing.
- *Metalevel M0*: A data record that contains a specific immunization date—which is an instance of the immunization date data element—is an M0 thing.

⁵ MDA did not invent the concept of *metalevels*. *Metalevels* have been a concept in the field of knowledge representation for some time.

⁶ Note that the Federal Enterprise Architecture's Business Reference Model (BRM) contains both M2 and M1 specifications. It specifies a means for defining business contexts in terms of Business Areas, Lines of Business, and Sub-Functions. This means is an M2 thing. It also contains particular business context definitions, including four Business Areas, 39 Lines of Business, and 153 Sub-Functions, which are M1 things.

Thus, we can use MDA shorthand to refer to “sharing at the M2 level,” which would mean settling on a common means of definition such as ISO 11179, or “sharing at the M1 level,” or “sharing at the M0 level.”⁶

MDA also has an M3 *metalevel*, which is the language in which M2 things are specified. This language is called MOF (Meta Object Facility). There is an M3 because there is more than one kind of definition. A data definition is one kind of definition. A service definition is another kind of definition. There are yet more kinds of definitions, including business process definitions, definitions of software deployments, and more. A full treatment of MOF is beyond the scope of this document.

Organizational Scope

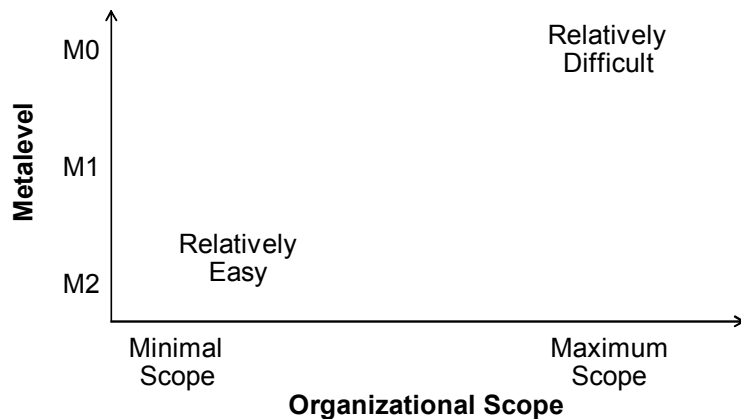
We have seen that we can categorize sharing according to metalevels. We also can categorize it according to the organizational scope of the sharing.

As described earlier, we might intend to share only within a work group, or among work groups within one department (department in the corporate sense), or among departments within one division, or among divisions within a company, or among companies within an industry. We can also describe a continuum of organizational scope using governmental rather than corporate terms, such as agency, department, and so on.

A Topology of Sharing

Thus, the topology of sharing has at least two dimensions—metalevel and organizational scope. Figure 1 illustrates these two dimensions combining to form a sharing topology.

Figure 1: Topology of Sharing



Value Chain Scope

A particular way of scoping a sharing effort organizationally is by value chain. Continuing our example scenario, suppose that two health care organizations that maintain immunization records have a legacy whereby they have no commonality at any metalevel, yet they need to exchange data in the context of one or more e-government value chains.

The value chain scenario requires sharing at the M0 level, the most challenging metalevel for sharing, which, of course, requires sharing at the M1 level which, in turn, requires sharing at the M2 level. However, by restricting the organizational scope of sharing to a particular value chain, sharing at all three metalevels becomes more tractable. In practice this means that the organizations would not share common operational data stores, but would exchange M0 data in the context of the value chain.

Thus, within the context of the value chain, the organizations must establish commonality at all three metalevels. However, value chain scope does not require commonality within the *internal* scope of the individual organizations that play the roles of partners in the value chains. When one of the organizations

receives the M0 data via a link in the value chain, the organization's systems transform the data into the internal form as the data passes out of the context of the value chain and into the internal context of the organization. The transformation may result in changes at the M0 level only, or at the M0 and M1 levels only, or at all three metalevels.

It is more tractable and achievable to agree on a means (M2) of defining data elements and on a common set of definitions (M1) for the scope of a value chain than for the entire scope of an organization. In this approach, organizations leverage the common means of definition to publish data element definitions that they use to support various value chains, without attempting to publish all of their internal data element definitions. Over time, to varying degrees, organizations may find utility in publishing internal definitions, but they would not be used directly by e-government value chains.

Issue: Leveraging Semantic Web and MDA Technology

An impressive amount of effort is going into building out infrastructure that supports the Semantic Web; infrastructure based on MDA's MOF is spreading as well. Reasoners based on RDF and OWL, the two primary Semantic Web specifications, are proliferating and becoming more sophisticated. At the same time, MOF technology is deeply wired into Eclipse, an open-source integrated development environment (IDE) that is gaining enormous traction in tooling for enterprise systems development.

Leveraging the Semantic Web

As indicated earlier, grounding the DRM semantically using Semantic Web technology would be the key to leveraging Semantic Web reasoners that can help detect inconsistency in definitions and aid architects in finding definitions that meet certain criteria. Space does not permit going into this matter in detail, but suffice it to say that generating an OWL-based definition of the M2 data element definition mechanisms that DRM proposes would be a critical step along this path. This would entail producing an OWL-based ontology of the ISO 11179 definition mechanisms, and probably of 11179's registration mechanisms as well.

Leveraging MDA

The build-out of MOF-based infrastructure is evolving to support enterprise-class repositories that manage UML models, entity-relationship models, relational database models, OLAP star schema models, workflow and business collaboration models, software deployment models, and additional models that are part of a model-driven enterprise architecture. Until now, these different kinds of models have been isolated from each other in silos within the enterprise, but MDA's MOF-based common approach to managing disparate kinds of models in an integrated fashion is beginning to knit these islands together.

There are several approaches to leveraging MDA's MOF technology and thereby positioning the DRM to take advantage of the growing Eclipse ecosystem and other MOF tooling. One approach involves creating MOF metamodels of the M2 data element definition mechanisms that DRM proposes. MOF metamodels



are very much like UML class models, and MOF architects use UML tools to create them.

For example, Figure 2, Figure 3, and Figure 4 are parts of a prototype MOF metamodel of ISO 11179 Part 5 developed and released into the public domain by Visa International's Visa Data Authority. It is not a complete metamodel of 11179. The blue boxes represent classes for which constraints expressed in UML's Object Constraint Language (OCL) have been written. For example, the Equivalence class in Figure 4 has a formal constraint that reads:

```
context Equivalence
inv:
  isDirectional = false
```

Figure 2: Fragment of a MOF Metamodel of ISO 11179-5

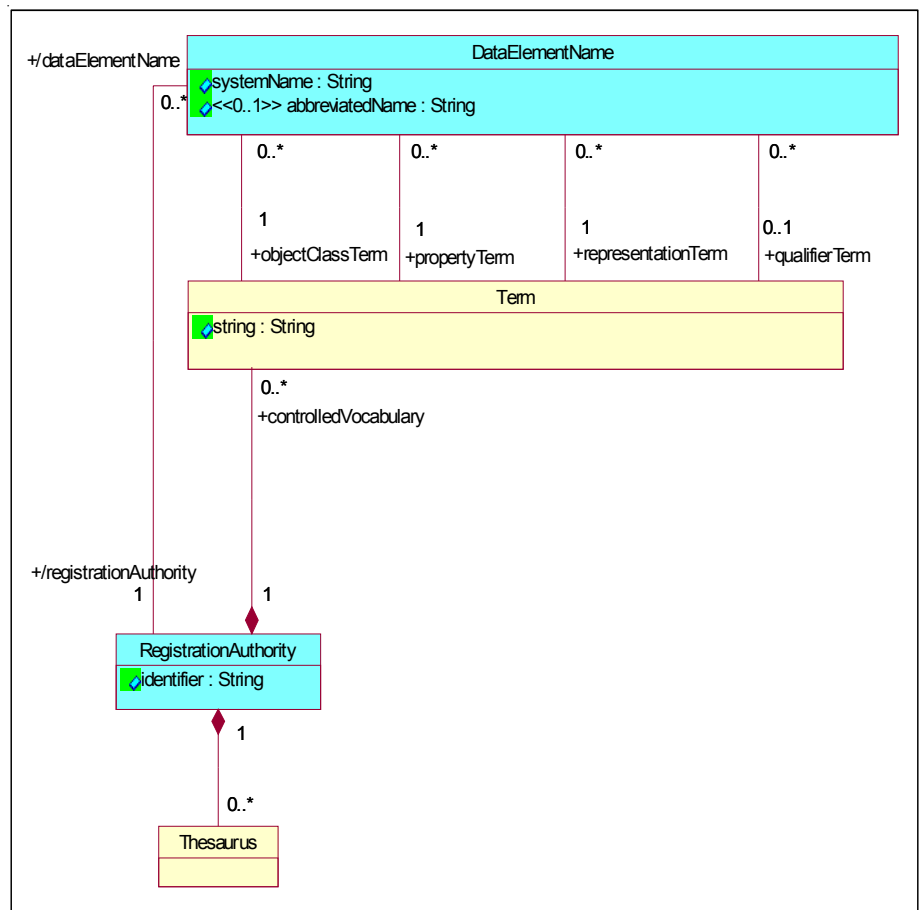


Figure 3: Another Fragment of a MOF Metamodel of ISO 11179-5

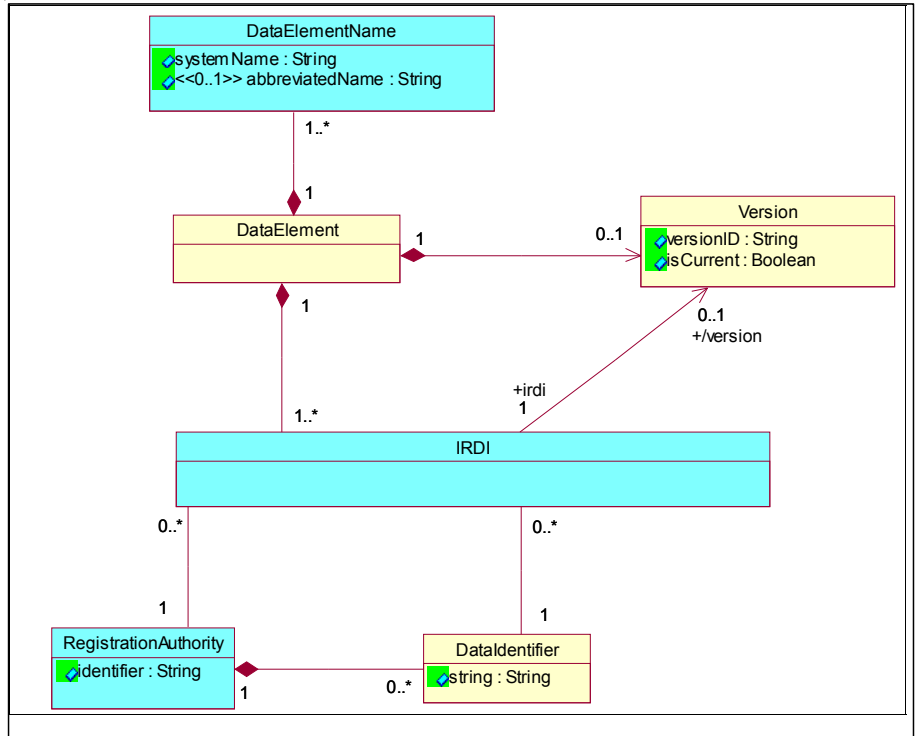
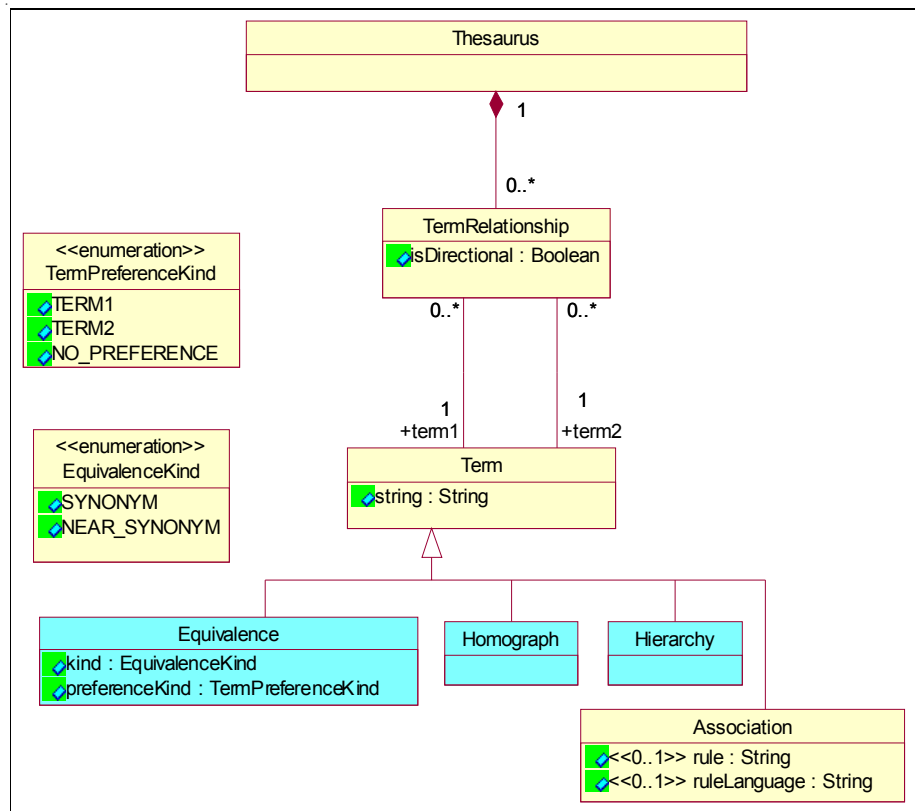


Figure 4: Yet Another Part of the MOF Metamodel of ISO 11179-5



Eclipse-based and other MOF-based tooling uses such M2 metamodels to generate XML schemas and Java APIs for representing M1 instances of the metamodel. In this case, the generated XML schema and Java APIs for ISO 11179-5 would be used to render specific data element definitions in the form of XML documents and Java objects, respectively. Such tooling also generates code that parses conforming XML documents into Java objects—that is, it converts XML based representations of the definitions into Java objects, where the XML documents conform to the generated schema and the Java objects expose the generated APIs. It also generates code that serializes definitions from Java form into XML document form. All of this generation follows patterns that have been codified in the MOF standards, and the consistent patterns facilitate the integration of different kinds of models.

Using the Semantic Web and MDA Synergistically

There are a number of ways to use the Semantic Web and MDA synergistically. A project is underway in the OMG to define a UML Profile for OWL and a more general UML-to-OWL mapping, which would include a MOF-to-OWL mapping. The project is also defining MOF metamodels of RDF and OWL.⁷

The basic value proposition for the DRM to leverage both the Semantic Web and MOF is to gain the reasoning capabilities of the Semantic Web, and to gain the model and metadata management facilities of MOF-based tools and thereby to

- Ease the integration of the Semantic Web ontologies into enterprises where MOF-based tooling is being applied to the management of UML models, entity-relationship models, and more
- Enable MDA to take advantage of the Semantic Web reasoners

Furthermore, a UML Profile for OWL also makes it possible for UML-savvy architects to use UML tools to model OWL ontologies.

Issue: Contexts and Business Collaborations

The DRM proposes to use the BRM to establish the business context of a data element definition. Others have posted feedback indicating that the business context requires more than simply specifying Business Areas, Lines of Business, and Sub-Functions, which is what the BRM currently supports.

In particular, it may be advisable to also establish a business collaboration context. E-government value chains are collaborations in which a number of services are invoked among parties to the value chain. These collaborations constitute protocols that are an important part of the context of data element definitions, when such definitions are used by value chains.

There are a number of ways to model collaborations. UML interaction models, the UML Profile for EDOC's component collaboration models, and other means exist. Business collaboration models are closely related to business process models, which also can help to define the business context of a data element. A number of means exist to define business processes, including BPMN, UML, and so on.

⁷ See [MDA SW], a paper which explains the potential for synergy between the Semantic Web and MDA in more detail, written cooperatively about this project by Semantic Web and MDA experts.

Conclusion

The main points of this review of the DRM are:

- The DRM must separate concerns among
 - Sharing the means for defining data elements (M2 commonality)
 - Sharing specific data element definitions (M1 commonality)
 - Sharing instance data (M0 commonality).
- The DRM must guide users to carefully define their sharing topology along the dimensions of metalevel and organizational scope.
- The DRM must leverage the progress in building out infrastructure for the Semantic Web and the OMG's MDA.
- Business collaborations and business processes are essential parts of the business context of a data element.

References

[FEA] The Federal Enterprise Architecture, <http://www.feapmo.gov/fea.asp>

[FEAPMO] The Federal Enterprise Architecture Program Management Office
<http://www.feapmo.gov>

[GSA OCIO] "Executable EA at GSA," presentation by George Thomas, Chief Architect, GSA Office of the CIO, November, 2004, <http://www.omg.org/cgi-bin/doc?finance/2004-11-02>

[ISO 11179] ISO/IEC 11179, "Information technology — Specification and standardization of data elements," consisting of six parts.

[MDA SW] Frankel, Kendall, Hayes, and McGuinness, "A Model-Driven Semantic Web," in the book *The MDA Journal: Model Driven Architecture Straight from the Masters*, Meghan-Kiffer Press, 2004, edited by David S. Frankel and John Parodi.

