

The Federal Enterprise Architecture Program Management Office



The Data and Information Reference Model Version 1.0:

A Foundation for Government-wide Improvement

January 2004

A MESSAGE FROM THE CHIEF ARCHITECT

The Federal Enterprise Architecture (FEA) exists to bring together "best practices" to facilitate better management of IT portfolios from a government-wide perspective. We work with agency and industry partners to identify concepts that are most applicable and effective for government use. This has been especially important in the development of the Data and Information Reference Model (DRM), our most complex undertaking yet.

We envision the DRM as enabling information efficiencies on a scale that is unprecedented in the history of government. That said, we understand that there will be many challenges and we are working hard to provide a holistic approach to managing the DRM and its components.

Our success, however, relies a great deal upon the collaborative efforts of organizations who will partner to make these concepts operational. We need the synergistic efforts of those who have formed Communities of Practice to provide their insight and guidance as we develop a path for others to follow. We need the support and financial backing of senior managers who indeed will benefit from the information we can generate to help transform their business operations. We see our role as facilitators of your success and will make every effort to achieve that end.

Bob Haycock

ONAD OL: - C E. - L - - - - - A - - - 1 - : L - - L

TABLE OF CONTENTS

TABLE OF CONTENTS 3	
EXECUTIVE SUMMARY 5	
INTRODUCTION 6	
THE FOCUS THE NEED OTHER VOLUMES RELATIONSHIP TO OTHER FEA REFERENCE MODELS VALUE Outcomes of using DRM concepts Benefits DATA AND INFORMATION REFERENCE MODEL VERSION 1.0 11 PURPOSE 11 DEFINITION 11	6 8 9
BUSINESS CONTEXT 13	
Subject AreaSuper-Type	
	13
INFORMATION FLOW14 DATA ELEMENT DESCRIPTION 14	
DATA ELEMENT DESCRIPTION 14	14
DATA ELEMENT DESCRIPTION 14 Data Object	14
DATA ELEMENT DESCRIPTION 14 Data Object	14

EXECUTIVE SUMMARY

As part of its effort to transform the Government into one that is citizen-centered, results-oriented, and market-based, the Office of Management and Budget (OMB) is developing the Federal Enterprise Architecture (FEA), a business-based framework for Government-wide improvement. As illustrated in Figure 1, the FEA is being constructed through a collection of interrelated "reference models" designed to facilitate cross-agency analysis and the identification of duplicative investments, gaps, and opportunities for collaboration within and across Agencies.

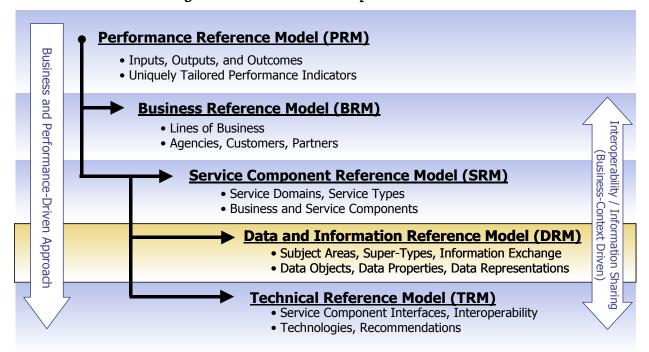
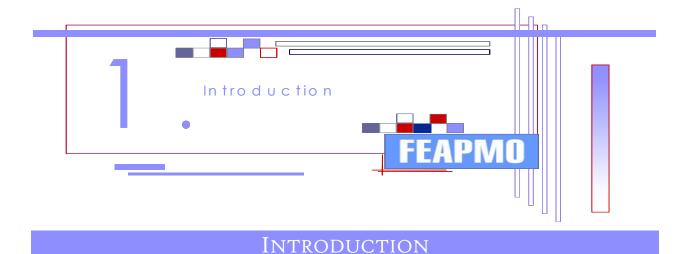


Figure 1 - The Federal Enterprise Architecture

The FEA Data and Information Reference Model (DRM) is a business-driven, functional framework that classifies Data and Information with respect to how it supports the business of the government. The model is intended to provide a common, consistent way of categorizing and describing data. An organized structure for data facilitates data sharing and data integration. It also enhances the ability to understand the types of interactions and exchanges that occur between the federal, state and local governments and their various customers, constituencies, and business partners. The DRM is structured to accommodate data and information needs that cross the vertical lines of business.



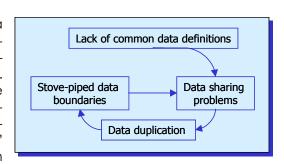
This document is the first of four volumes that define the DRM and provide guidance on its implementation. Volume 1 presents an overview of the DRM structure, its relation to the other FEA reference models, DRM value and applicability, and a high-level discussion on strategy for the DRM going forward.

THE FOCUS

The FEA-PMO focus for the DRM is to establish initial guidance and strategy for information sharing among collaborative partners. Central to the concept of information sharing is an understanding of Information Flow, the middle section of the DRM. The DRM focuses on attributes and metadata that specify processing criteria for a semantic understanding of the data being passed, in addition to the actual data that is in motion. For each package of data that is passed, the structure to fully understand the package and its contents from the point of origination to its consumption is extremely valuable for effectively re-using the data. Collaboration among agencies is key to the development of information sharing capabilities.

THE NEED

There are a number of contributing factors to data inefficiencies throughout the government. Unfortunately, data problems hinder the government's ability to perform its business and meet customer needs. The lack of a common way to describe or define data contributes to data sharing problems and results in duplicating data rather than sharing or reusing it. This leads to the creation of "stove-piped" data boundaries and finally comes full circle when



those "stove-piped" data boundaries contribute to even more data sharing and consistency problems. Some of the primary issues and barriers associated with information sharing include the following:

- No common framework or methodology to describe the data and information that supports the processes, activities, and functions of the business
- Existing systems offer diffused content that is difficult to manage, coordinate, and evolve
- Information is inconsistent and/or classified inappropriately
- Without a common reference, data are easier to duplicate than integrate
- Data and Information context is rarely defined
- No single authoritative data source
- Data sensitivity and security of data can easily be compromised
- New laws/issues result in continuous adding of databases that cannot share data

The inability to share or exchange data efficiently not only costs time and money, but it also works against the citizen-centered focus of today's government. The results of data inefficiencies include:

- Increased burden on finding and accessing the right data
- Increased delays to satisfy citizen and stakeholder requests
- Unclear knowledge of who to contact for specific data
- Increasing costs to manage and integrate data
- Increased corruption of the data
- Decreased ability to interoperate

OTHER VOLUMES

There are three additional volumes that follow Volume 1, providing greater detail and including implementation strategies for Agency consideration. They are outlined in Section 3, Phased Development of the DRM.

Volume II: Business Information Context. This document discusses the definition and application of data classifications that will categorize commonly shared Government information.

Volume III: Information Flow. This document establishes criteria for describing data created or used by common business processes and the implementation of strategies that effectively support the transfer of this information.

Volume IV: Data Element Description. This document identifies concepts and requirements for defining and using data elements with the use of registries and repositories to manage data integrity and availability.

RELATIONSHIP TO OTHER FEA REFERENCE MODELS

The DRM is the last of the five FEA reference models to be developed. It is also the most difficult to construct because data is at the core of everything we do in the government. Below is an example of how the DRM might be used in conjunction with the other models in the response to the implementation of a hypothetical piece of legislature.

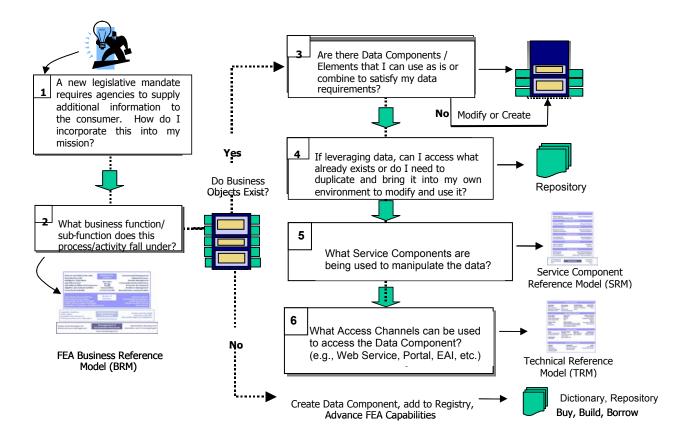


Figure 2 - FEA / DRM Use Case Example

Figure 2 describes the steps that an agency might take to locate data to satisfy new requirements. Does this new legislation necessitate the establishment of new processes to generate the needed data? Where does this activity fall within the scope of work already being performed? Further, are we able to look into our current Data Architecture to identify this data? Do we have agency partners that have done this and whose work we can possibly leverage?

The DRM provides support for all of these considerations to meet an agency's information needs. Once we understand the process that creates or uses the data, we can define the needed data. In doing this, we have specified the criteria to design mechanisms that can manage the data effectively. Breaking this process down into self-contained modules identifies components that can be used interchangeably in other environments to support similar processing requirements. These are the service components referred to in step 5 and can be defined in such granularity as messages being passed in an application service. Step 6 refers to the technology used to support the transfer of data and will be different for each agency.

The BRM provides a business perspective for the data, influencing how it is defined or used. The most direct link between the DRM and BRM exists at the business process level where data is defined within the context of how it is manipulated and exchanged. The alignment and mapping of the DRM to both the BRM and agency enterprise architectures is one of the next steps towards implementing the model across the government. A method for doing this analysis is being developed and is referred to as the "Common Process View".

Linking the DRM to the SRM begins with analysis of the business activities. Understanding business patterns allows for the discovery of service components that manipulate data and information in support of a given business process. This discovery process leads to identifying patterns that can be used to support other similar activities and furthers the building of a repository of reusable components that can be shared throughout government.

VALUE

Outcomes of using DRM concepts

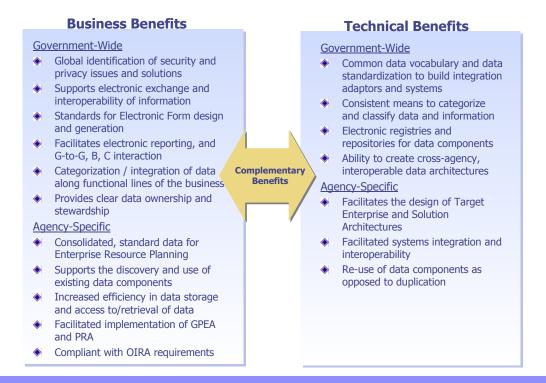
Outcomes of DRM use are beneficial to both the business manager and the technical manager. For the business manager, an important outcome of using DRM concepts will be an alignment of data elements and information packages to the processes that an agency or government body uses to conduct business operations. As a result, a greater understanding of data and information can be leveraged and shared throughout the business cycle. Another important outcome of DRM implementation will be more clearly defined authoritative sources for data. Authoritative data sources ensure the integrity and accuracy of data from its origin. These organizations play a critical role in developing data registries and dictionaries that enable consistent use of data categorizations and classifications as data components are developed, published and re-used.

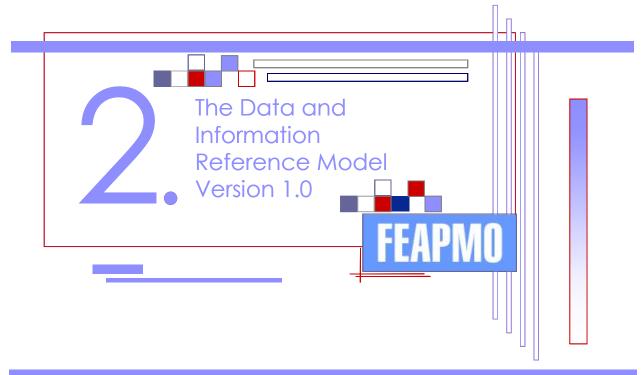
For the technical manager an important outcome will be the ongoing population of an XML registry and repositories for the purpose of providing re-usable data exchange modules. These modules enable data sharing between systems of the federal, state and local governments and their various customers, constituencies, and business partners. The DRM provides classifications of government data with a registry that directs users to appropriate representations of data within the repositories. This will save time and money as data modules can be selected from the registry and applied repeatedly without modifications.

Benefits

The outcomes of a DRM provide both business and technical benefits. Figure 3 portrays the benefits of the DRM.

Figure 3 – Benefits of a DRM





Data and Information Reference Model Version 1.0

PURPOSE

Many organizations are seeing an increased need to exchange and share information. A problem they are encountering is that each organization has developed and managed their data based on the inner workings of their business units. As such, data and systems are tightly bound to the "problem at hand", lacking a more holistic perspective. With an increasing focus on external and global operations, data must now be integrated at a level not previously required.

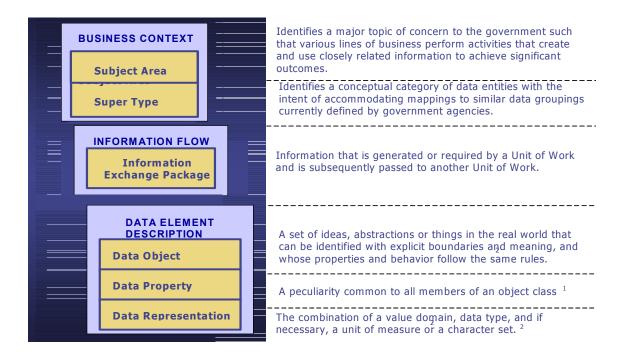
The DRM provides a common, consistent way of categorizing and describing data to facilitate data sharing, re-use and integration. DRM concepts address the need for semantic understanding, while maintaining data quality and integrity. The need to manage unstructured data will also be addressed as part of the DRM Data Management Strategy, a management document to be released shortly after Volume I.

DEFINITION

The DRM reference model provides classifications of data and information that support government program and business line operations. It describes in part the interactions and information exchanges that occur during the execution of government operations and the delivery of services and information from the government to citizens, businesses, partners and customers.

The structure of the DRM is defined in figure 4.

Figure 4 - DRM Structure



The DRM structure has three major categories: Business Context, Information Flow, and Data Element Description. Within the categories are six layers (Figure 4) that further define the business context and specific characteristics of the data elements themselves. Within the Business Context, each Subject Area utilizes one or more data Super-Types. These Super-Types represent categories of data whose elements may be specified or used within one or more Information Exchange packages. The Subject Areas and Super-Type classifications are included to help define the context in which the data was defined and applied.

Each Information Exchange package contains one or more data elements that are required or generated by a business process. Data elements, defined in accordance with ISO 111791, consist of a data object, a data property, and a data representation.

The DRM structure is designed to enable semantic understanding for those who need to integrate the data or may desire to utilize schemas listed in the registry. The structure itself is not hierarchical and is not intended to be. The DRM is not a data model and does not address data management from the perspective of data modeling and analysis techniques. We leave this level of analysis to the agencies themselves, and the Communities of Practice in which they participate.

¹ ISO/IEC 11179-3:2003 Information Technology – Metadata registries – Part 3: Registry Metamodel and basic attributes

² ISO/IEC 11179-3:2003 Information Technology – Metadata registries – Part 3: Registry Metamodel and basic attributes

BUSINESS CONTEXT

Because the DRM must accommodate data and information from all levels of government and numerous agencies and departments, the business environment in which data is created and commonly used becomes rather critical for complete understanding. Subject Areas are classifications of government-wide interests that influence or impact the development and use of data as needed to address issues in a particular business environment. The Super-Types are generic groups of data that are commonly related to a specific topic. They represent categories of data and information and may be related to more than one topic.

The structure of the DRM was designed to categorize and classify data in a manner that provides both the business context of the information being exchanged, as well as specific characteristics of the supporting data elements within the information exchange.

Subject Area

The Subject Areas provide a collection of data classifications that represent broad categories of information. This layer will contain the major areas of information and data subjects that support the business of an organization. Typically, this layer is aligned to the data areas or highest layer of data within an agency's enterprise data architecture. The Subject Area layer of the DRM will facilitate discovery of data and information common to various lines of business. Additionally, these classifications improve the ability to discover specific data and metadata for semantic understanding.

Examples of this layer might include Public Health, Global Justice, Environment, and Recreation.

Super-Type

The second business context layer, Super-Type, serves to identify data classifications by providing generic groupings of data related to a specific topic (Subject Area), and should be considered conceptual entities. These high-level groupings provide sufficient context for Communities of Practice to discover data commonality for business process information needs. Using information in this layer, agencies can map their data descriptions to the DRM, while maintaining their existing data architectures and descriptions. This approach leaves data stewardship responsibilities with the agencies and does not require any modifications to existing data models.

Examples of this layer might include *Immunization* within the *Public Health* Subject Area, *Warrant* within the *Justice* Subject Area, and *Wildlife* within the *Environment* Subject Area.

Categorization of Information is a requirement of the E-Gov Act, subsection 207(d). DRM Volume II will provide specifics on how this will be done and the taxonomies that will support this requirement. Mapping agency data architectures to DRM super-types enables the categorization of an Agency's data, as well as other information assets to enable data sharing. By mapping agency data to DRM Subject Areas and Super-Types, the top two layers of the DRM, we begin the process of harmonizing and normalizing government data. These two layers also provide the classification of Government information to be used as standard taxonomies.

INFORMATION FLOW

The DRM Information Exchange Package defines data elements that are communicated or passed between processes. This concept aligns with the ISO/IEC 11179 concept of Information Interchange: "The process of sending and receiving data in such a manner that the information content or meaning assigned to the data is not altered during the transmission."

This layer of the model is based upon the definition of information exchanges as defined by agencies, Communities of Practice, and authoritative data sources. A single flow of information from or to a business process is referred to as an "Information Exchange Package". Data elements included in each "package" of information are defined individually using standard definitions and are available for re-use in other information flows and XML schemas. Process types include all levels of processing that carry groupings of data needed for activity or task execution.

An example of an information exchange package and its respective data elements is shown in Figure 5. The name of the Information Exchange package is Vaccination Delivery and the needed data elements for this package includes Vaccination Name, Vaccination Arrival Date, Carrier Name, and Facility Address. This is representative of a typical information exchange package at the business process level. An application process would be more transaction oriented and might include the name of the application module. Each information exchange package should be created by one process and then consumed by at least one process.

Figure 5 - Data and Information Exchange Package Layer



Other examples of Information Exchange packages include Grant Application, Camping Permit Request, and Budget Submission.

DATA ELEMENT DESCRIPTION

At this most detailed level of the reference model, we focus on the physical description of the data. The ISO11179 defines a data element as follows: "A unit of data for which the definition, identification, representation, and permissible values are specified by means of a set of attributes."

Data Object

The Data Object layer is the first layer that defines a data element. The Data Object layer is a set of ideas, abstractions or things that can be identified with explicit boundaries. Typically, this

layer is a breakdown of the nouns that make up the data elements within an Information Exchange package.

Examples could include Vaccination, Carrier or Facility as the Data Objects.

Data Property

The Data Property layer is a peculiarity common to all members of a Data Object. This layer will include descriptors or attributes of the Data Object. A property of the Data Object is a characteristic of that object that is of importance to the enterprise. For example, the data object Vaccination has attributes such as the Name of the vaccination – Vaccination Name, where Name is a descriptor. Vaccination resides within the Data Object layer and Name would reside with the Data Property layer as its descriptor.

Other examples of Data Properties common to the Data Object "Vaccination" could include RecommeDosage or *Primary Side Effect*.

Data Representation

The last layer, the Data Representation layer, describes how the data is represented (value domain + data type), and is selected from a controlled word list. The controlled word list is generated based on common formats of data within the government. Value domains can be included when a data element is defined at time of data exchange, but this is optional.

Examples of data representation can include Name, Text, and Date. An example of a data representation for "Date" could have a controlled format of YYYY/MM/DD; where Y = year, M = month, D = day, and O should precede where needed.

THE "THREE PILLARS" 3

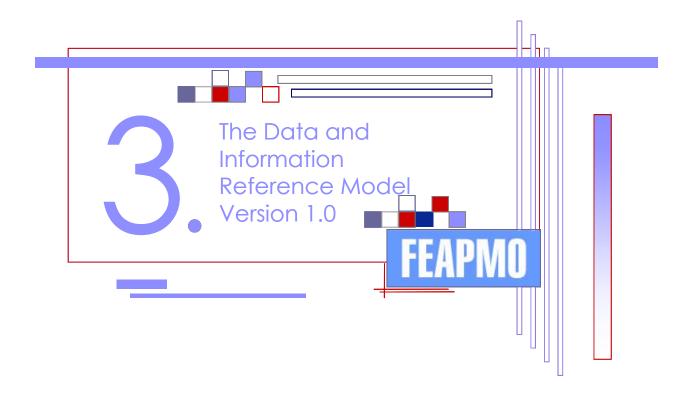
To ensure that we present the most complete and most value-added implementation of this model, we are including perspectives of each layer of the model that address the most fundamental aspects of a successful data management program. Each aspect must be addressed concurrently in order to bring an enterprise forward, using standards, policy, process, and technology. They will be referred to as the "3 Ds" for simplicity.

Data Architecture: reflecting data categories and establishing accountability for this information in business process improvements. Defining data objects and events, using specific business requirements for shared information to drive the definitions of standards is a responsibility shared by agencies and the Community of Practice.

Data Governance: establishing a structure to make decisions regarding our information assets, defining requirements for business process development and change, and mechanisms of arbitrating differences among stakeholders.

Data Sharing Architecture: providing standard, repeatable processes and technical patterns for sharing data; provisioning data from an information source to an information consumer in response to a business requirement.

³ The Three Pillars, an adaptation of Information Management and Data Quality, by Bryan Aucoin, Panel 1, Proceedings of the Eighth International Conference on Information Quality, (ICIQ-03)



DRM STRATEGY

The DRM scope requires an approach that represents a departure from the focus commonly taken in managing an organization's data. The DRM must address the needs and challenges of managing "federated data". Federated data is a subset of all agency data that has been determined to have significance beyond agency boundaries. Effective management and use of this data will require government-wide acceptance of federated policies and standards to exchange this data in a meaningful way.

DRM components establish the structure and environment that allows this meaningful exchange of data to occur. Several plausible scenarios are being considered as to how the DRM approach might be supported, including "best practices" of several federal agencies and industry organizations. Currently, there are a number of initiatives working in parallel to DRM development that are developing guidance, procedures, and software that will support the implementation and use of the DRM. We plan to pilot these and begin to use these first with the E-Gov and Line of Business (LOB) initiatives. GSA is working with OMB to provide supporting registry and repository functions.

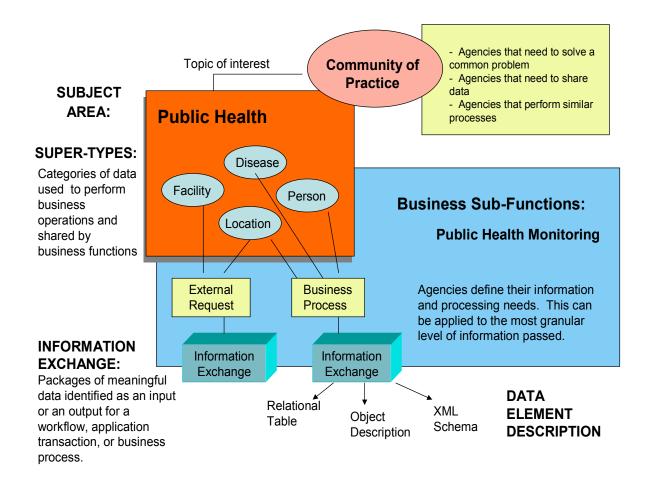
A separate DRM Data Management Strategy document is in development and will go into much greater detail on management concepts, procedures, and technology components that will be included in the implementation of the DRM. Additionally, this document defines the Business Drivers that guide strategy development, identifying issues or problems that the DRM strategy should address.

APPLICABILITY

The DRM will be applied where the potential for sharing and exchanging common data and information within and between agencies exists. Initially, it will be applied to E-Gov and LoB initiatives by agencies that share common interests and are pursuing similar performance goals. Shared interests and performance goals form the basis for identifying topics of interest government-wide. These topics are **Subject Areas**. Agencies form a "Community of Practice" when they come together to solve a common problem that carries a government-wide focus of significance such as the President's E-Gov initiatives. These "communities" define how they will share information using agreed upon data definitions and exchange formats. This approach results in the greatest return on investments when "Services-to-Citizen" processes and technology interfaces are shared among multiple agencies.

The DRM will be utilized within such targeted business areas as mentioned above to foster data and information sharing and exchange where it is needed the most. For example, **Public Health Monitoring** is a sub-function that is performed by many different agencies; some focusing on different aspects of this sub-function. To fully understand where commonalities exist, an analysis of their business processes supporting this sub-function must be done. The FEA is currently developing the methodology and software requirements to enable this process analysis. An outcome of this analysis is the identification of packages of information generated and passed to other processes, some external to the agency. That data is then reflected as an **information exchange package**.

Documenting these processes and data establish requirements for developing an information system with shared access for all participating agencies. Based on services needed to successfully perform business processes involved, **re-usable components** are designed that can be brought into agency environments for other related uses. Each component is designed to be a self-contained module that can be used in a "**plug and play**" fashion. These components will be made available to all agencies through a central registry and repositories that store approved components.



Population of DRM registries and repositories will be based upon coordinated efforts between authoritative sources such as communities of practice and agencies, our industry partners, and the FEA-PMO. The FEA-PMO will establish DRM components in a "top down" fashion by defining and populating the Subject Area and Super-Type layers of the model. Agencies, communities of practice and governing bodies will continue to "build out" the model by defining information exchange packages and standard data elements that will be represented in XML schemas to transfer data.

Figure 6 lists the stakeholders and governing bodies for defining, owning and managing the various layers of the DRM.

Figure 6 - DRM Definition, Ownership & Stewardship

		Definition (defines)	Ownership (owns)	Stewardship (manages)
Business Context	Subject Areas	FEA-PMO/Agencies	FEA-PMO	FEA-PMO/AIC
	Super-Types	FEA-PMO/Agencies	FEA-PMO	FEA-PMO/AIC
Information Flow	Information Exchange	Communities of Practice / Agencies/ Industry / ISO*	Communities of Practice/Agencies	Communities of Practice/Agencies
Data Element Description	Data Object	Communities of Practice/Agencies/ Industry/ISO*	Communities of Practice/Agencies	Communities of Practice/Agencies
(physical representation)	Data Property	Communities of Practice/Agencies/ Industry/ISO*	Communities of Practice/Agencies	Communities P rá ctice/Agencies
	Data Representation	Communities of Practice /Agencies/ Industry / ISO*	Communities of Practice/Agencies	Communities of Practice/Agencies

^{*} Thousands of data elements have already been defined using ISO 11179 conventions that the Federal Government can adopt or modify for re-use.

The authoritative data source will have the legal right/responsibility to create and modify data definitions, as well as register XML schemas within an XML repository that supports the data and information exchange packages.

The strategy for providing data sharing opportunities within and between agencies will be based upon the alignment of an agency's data architecture to the DRM. For that reason, the implementation of the DRM will not affect agency data and information architectures. The DRM, like the other FEA reference models, will provide a framework to classify and categorize existing agency data, as that data relates to agency architectures and IT investments. This allows agencies to maintain their current data architectures, while also allowing the discovery of other similar data residing at external agencies through an alignment to the DRM.

PHASED DEVELOPMENT OF THE DRM

With each DRM Volume released, guidelines will be provided to assist agencies in the development of their data architectures, consistent with the objectives of the DRM; data sharing, re-use and integration. These objectives are supported by the analysis and alignment of agency data to the DRM. Over time, the discovery of opportunities to integrate data stores will surface – based upon the business needs of organizations that use common data.

The DRM will be presented in four (4) volumes. This document is the first volume. Successive volumes will contain the following information.

Volume 2 – This volume will focus on the business context of the DRM to identify subject areas of government-wide interest and super-types of the data and information, including guidance for their implementation. The FEA-PMO is engaged in conversation with various agencies to begin identifying the business context as "subject areas" and "super-types" based on their efforts at categorizing data at a high level. Information resources to construct this layer of the DRM will include data from the OMB Exhibit 300s, as well as departmental listings of Subject Areas or data areas from agency enterprise architectures. This document will also describe how Communities of Practice will register and share data based on commonly used taxonomies.

Volume 3 – This volume will provide considerable detail regarding requirements for documenting "packages" of information passed between business processes that need to share data. Information provided in this volume links the data exchange with process characteristics that are needed for effective understanding of data contents. Additionally, this volume will address policy and governance to implement needed to manage the data and supporting infrastructure.

Volume 4 – This volume will address the more granular topics to establish standard definitions and implementation of the actual data elements. It will also discuss the registries and repositories that contain DRM components, as well as supporting dictionary and Thesaurus components.

WHAT THE DRM IS NOT

The DRM is not a government-wide data model; it is a reference model. It is intended to support categorization and classification of information.

The DRM is not an all-encompassing set of XML schemas that describe all government data. XML schemas will be defined and registered by Communities of Practice as needed to effectively share information. These schemas will be made available to the government for re-use.

The DRM is not meant to replace existing data structures within the agencies, rather to provide the means to compare data among agencies in order to exchange, re-use or integrate data.

VISION FOR THE FUTURE

The goal of the DRM is to promote and facilitate the exchange and re-use of data and information, which in turn, helps to improve business performance and decrease the cost of data. To help achieve this goal, our strategy moving forward will include defining the governance structure needed to oversee the on-going development and maintenance of DRM structures and contents. Based upon the volume and breadth of effort needed to do this and to fully define and populate the layers of the model, the need for agencies and communities of practice to be involved becomes even more significant.

As with most EA efforts, the development and ongoing maintenance of the DRM is an iterative and continuous improvement process. A DRM program management plan and communication plan are being established to ensure that the DRM has all the necessary elements for success. This includes planning for agency training and continued outreach activities. Since many agencies are still in the process of developing their EAs, special consideration will be given to developing documents that are balanced between elementary guidance for those defining data architectures and more advanced concepts for those agencies who are leaders in the area of data management. Any changes to the DRM will continue to be validated through Agencies and published to the FEA-PMO Website.

In summary, OMB envisions a collaborative and mutually beneficial management plan for the DRM that will result in positive outcomes for all stakeholders. OMB will work over the coming months to obtain feedback and collaboration on the definition and usage of the DRM that is most beneficial to the government. The information contained herein is intended to provide general concepts of current thinking in this area, and is subject to modification.