FEA DRM Data Management Strategy

DRAFT

FEAPMO | FEDERAL ENTERPRISE ARCHITECTURE PROGRAM MANAGEMENT OFFICE

December, 2003

The following were major contributors or participants in the development of this document. The FEA PMO is very grateful for the time they volunteered and their commitment to produce this document.

Major Contributors:

Elliot Christian, USGS
John Dodd, CSC/IAC, Consultant
Jim Feagans, Department of Justice
Ken Gill, Department of Justice
Terry Hargrove, Pearson/FSA, Consultant
Michael Lang, MetaMatrix, Executive VP
Brand Niemann, EPA
Davis Roberts, Unisys, Consultant
Craig Tanner, Department of Interior, Consultant

Subject Matter Experts:

Suzanne Acar, Department of Interior, Chief Data Architect
Bob Greeves, Department of Justice,
Joan Karrie, EPA, Consultant
John Sullivan, EPA, Chief Enterprise Architect

OMB Lead Architect:
Diane Reeves

FEA DRM - Data Management Strategy

Executive Summary ……

Introduction ……..

DRM – Program Management ……
Business Value Proposition
Stakeholders
Strategic Goals
Performance Objectives
Principles
Product Schedule
FEA DRM Services

DRM Business Drivers ……

"Three Pillars" of Successful Data Management ……

DRM – Data Governance ……

DRM - Data Architecture ……

DRM – Data Sharing Architecture ……
Business-Centric Data Value
Data Harmonization
Data Source Automation
Information Exchange
Security

Final Words ……..

Executive Summary

The development of the Data and Information Reference Model (DRM) has been cautious and taken longer than any of the other models. This was done intentionally to ensure that the proper focus and structure was provided for this foundational concept that could potentially have the most impact of all the FEA models. The DRM will pave the way for policies and standards needed to establish government-wide information sharing and interoperability.

The DRM Data Management Strategy was developed guided by "Business Drivers" that represented key factors or challenges to government data sharing. These issues were discussed and thoroughly vetted by the strategy team. Each business driver was assigned and addressed by team members. Their strategies were presented to the group for refinement and were then built into the document as they related to the DRM Matrix of Information

We have researched many aspects of data management and emerging technologies to identify key elements that should be addressed by the DRM. The DRM Data Management Strategy includes best practices and valuable input from industry experts, as well as government representatives. The DRM Matrix includes the three D's from Bryan Aucoin's "Three Pillars" for the development of a successful data management program.[1] This has effectively guided the development of DRM components, ensuring that these pillars of success are appropriately addressed. The following are the three pillars built into the strategy.

> ***DRM Data Governance***. Data governance is the practice of making enterprise-wide decisions regarding an organization's information holdings.
>
> ***DRM Data Architecture.*** Data architecture is an element of an enterprise architecture that comprises a data model and assigns accountability for data. The data architecture reflects business area entities with attributes and establishes accountability for this information in business process improvements.
>
> ***DRM Data Sharing Architecture.*** The Data Sharing Architecture describes technology considerations for government agencies to participate in global information sharing communities.

Based on requirements set by e-Gov initiatives or Communities of Practice, each agency will define their data and information exchange formats. The Communities of Practice will define or adopt standard data definitions and data sharing schemas. A centralized dictionary and directory will maintain this information for general use. It is the responsibility of each agency to know and understand their environment (data, applications, and infrastructure) in order to map their data to requirements for information sharing.

There are several benefits to approaching government-wide data and information from the perspective proposed by the DRM program. It supports a federated approach to data management that allows each agency to maintain their existing data programs without disruption, but provides opportunities to participate in global information sharing. The methodology for identifying and defining information exchange packages allows agencies to map these packages to their own data exchange requirements with the opportunity to re-use DRM schemas and data components designed by others. DRM data standardization will allow agencies to have a common understanding of data contents and format. DRM data quality policies ensure that data exchanged is classified and maintains a certain level of integrity.

The DRM program as defined in this strategy document brings together key ideas and concepts that are needed to improve government operations. Individual communities of practice are beginning to face the challenges identified in this document. To address them in a way that many can benefit from the actions of a few is the most advantageous approach to bring value and improved access to government information assets.

Introduction

Information is being generated all around us. The challenge is how to find the right pieces of information, how to grasp the appropriate understanding, how to effectively manage and maintain it so that data can effectively meet the needs of our Government. Regardless of where we are and who we are, we all have a need to make decisions based on the most reliable and efficient data available. Can this be our reality?

With the onslaught of change in our cities and in our offices, we face the challenge of provisioning data that crosses organizational boundaries, data that is self-sufficient – carrying its own processing needs and context. In other words, data that is smart. Understanding that our data is a most critical organization asset is no longer an issue. The concern now is how agencies can better manage data to achieve their mission, to meet the needs of citizens, to ensure a better future.

In considering these aspects of data management, we have realized that there are certain barriers and hindrances. We have identified DRM Business Drivers that highlight issues from various data management perspectives. Our approach to federal data management was drafted in response to and around these issues.

Our Approach

The Data and Information Reference Model (DRM) provides a structure that facilitates the development of government data that can be effectively shared across agency boundaries to improve mission performance. The DRM is a service-oriented model that provides the pathway for “Services to Citizens” to become operational. At the same time, the DRM provides an impetus for agencies to better understand their data and how it fits in the total realm of government information.

The DRM addresses fundamental aspects necessary to enable information sharing opportunities and position agencies to operate in an environment of global information. Our approach directs the definition of data requirements and operational elements to three key aspects referred to as “The Pillars” of data management or the “Three Ds”; Data Governance, Data Architecture, and Data Sharing Architecture. These three are key elements of any successful data management program and will be critical moving into a global information arena. Each volume of the DRM will include these three aspects of data management with discussions specific to that portion of the DRM. The matrix on the following page outlines topics that will likely be addressed in the future volumes of the DRM as we continue to develop this approach.

DRM Matrix of Information

	Data Governance	Data Architecture	Data Sharing Architecture
VOLUME II BUSINESS CONTEXT	**Governance Structure, Policy & Procedures** Purpose: Define the use of data for government information sharing.	**Information Categories Data Groups, and Security Profile** Purpose: Catalogue and Index Government Information consistent with the E-Gov Act.	**Information Categories, Data Groups, and Exchange Security Requirements** Purpose: Identify and define federated data classifications to discover commonalities and opportunities for re-use.
VOLUME III INFORMATION FLOW	**Governance Structure, Policy & Procedures** Purpose: Define the use of data in the protection of information packages available in the DRM registry.	**Information Exchange Packages, Security Profile** Purpose: Define data groups (tables, records, messages, and text) and attributes that reflect business process needs common to a Community of Practice.	**Information Maps, Exchange Security Requirements** Purpose: Define data transformation patterns and key attributes that determine data exchange processing requirements.
VOLUME IV DATA ELEMENT DESCRIPTION	**Governance Structure, Policy & Procedures** Purpose: Define standards for Data Standardization.	**Data Element Descriptions, Security Profile** Purpose: Define and maintain data structures that reflect business data entity attributes and relationships.	**Object Descriptions, XML Schemas, Exchange Security Requirements** Purpose: Define and maintain metadata required to provide or support a specific service pattern.

The matrix above does not represent a hierarchical structure of information, but a matrix of information, products, and activities that will be accommodated by the DRM. Each portion of the DRM has a different set of requirements for governance that needs to be understood by IT management. Each portion has a different set of criteria that needs to be addressed as a result of data analysis activities. And still another set of criteria is required to adequately focus on the information sharing aspects of data management.

DRM Program Management

The FEA PMO has primary responsibility for the development and oversight of Federal Enterprise Architecture (FEA) models and their application. To increase government use and understanding of the DRM, the FEA PMO collaborates with government and industry partners to enhance the development and use of best practices in managing government information assets. Information contained within this document is intended to support the needs of government managers and staff responsible for the use of DRM concepts and products.

The strategy contained within assumes that a Governance structure will be established to oversee effective management of government data and information sharing activities. This document will highlight various elements of program management that should be undertaken and executed in a permanent structure to ensure that excellence is built into these activities.

FEA DRM Services

- *Briefings*
- *Management Consultation*
- *Product Review*
- *Training*

Business Value Proposition

Information sharing has evolved to the center of management focus as a critical organization capability. The globalization of corporate interests and organization outreach has placed significant stress on technology to effectively provide access to timely, quality information without geographical boundaries. In many cases, organizational performance is highly leveraged by integrated support functions enabled by technology.

The DRM will provide a structure that drives consistency in standards of data exchange. DRM standards allow government agencies to share data in common formats with common definitions. This will result in consist application of data across government.

The DRM will address the need for semantic understanding of government information. The DRM will facilitate common understanding of data context by including descriptive data that conveys the overall intent and use of data.

The DRM will provide an index of government information that can facilitate the exchange of electronic information. The DRM will provide classifications of government data that allows general users (citizens) to locate information with ease.

The DRM will define requirements for establishing standard information exchange formats that can be shared across government. The DRM infrastructure enables the collaborative development of information exchange, defining policies for development, storage, and use throughout government.

Strategic Goals and Objectives

I. Promote Multi-level, Government-wide Data Efficiencies.

Performance Objectives:

Reduce data inefficiencies caused by organizational boundaries.
Increase opportunities for federal, state, and local agencies to collaborate on data sharing initiatives.

DRM Volume II – Business Context

DRM Data Governance:

- Principles used to manage data consistency
- Partner with State and Local organizations

DRM Data Architecture:

- Establish a framework that facilitates development of Common Solutions and Re-use

II. Implement Re-usable Information Module Development.

Performance Objectives:

- Increase the availability of information modules.
- Increase the availability of guidance to develop information modules.

DRM Volume III – Information Exchange

DRM Data Architecture:

- Provide information Exchange Package development methodology.
- Adopt an Information Mapping approach.
- Collaborate with Communities of Practice to develop data components.

III. Enable Government-wide Information Sharing.

Performance Objectives:

- Increase access channels to federated Government Information.
- Increase data standardization among government agencies.

DRM Volume III - Information Exchange

DRM Data Governance:

- Establish data exchange format standards

DRM Data Sharing:

- Define requirements for secured data transfer

DRM Volume IV – Data Element Description

DRM Data Governance:

- Establish Meta Data Standards

DRM Data Sharing:

- Establish federal registry/repository/dictionary capabilities

DRM Principles

The following represent ideals or beliefs that reflect major characteristics of the DRM Program and will guide management decisions regarding oversight and implementation of DRM program goals, activities, and products.

Component-Based: DRM solutions emphasize development of self-contained data modules that can be re-used.

Data Consistency: Data components should be designed according to standards that support consistent definition and use of federated data.

Data Source Automation: Data is gathered at the source of its creation and transformation of any sort is mapped in order to maintain integrity.

Extensibility: Every data component should be designed so that it can be adapted to government-wide information needs.

Knowledge-Based: Each data component is developed as a result of a collaborative effort that increases organizational learning and results in quality information.

Services-Oriented: Each data component is a standardized, re-usable module that represents a discrete piece of logic that supports process-centric applications.

Solution Supportability: Each solution should be fully supportable by government operations and infrastructure or be a viable alternative for a contracted provider, either locally or globally.

Stakeholder Analysis

- *OMB* is leading the effort to transform government operations to reflect more modern approaches to improve mission performance and efficiencies in IT portfolio management. The DRM helps agencies better understand information needs and points to opportunities to streamline and improve agency IT portfolios.

- *Federal, State, and Local Agencies* will be able to use the DRM to identify commonalities in data requirements, taking advantage of information sharing opportunities and re-use of common service components.

- *Congress* will see the requirements set forth in the E-Government Act begin to generate results in the electronic dissemination of catalogues Government information, enhancing the availability and the value of this information.

- *Communities of Practice* will benefit from a structured, standardized approach to managing data that needs to be shared and understood throughout government.

DRM Business Drivers

The Business Drivers represent issues voiced by managers or identified by working groups trying to build cross-agency information solutions. We have used these issues to guide the development of strategies that will be supported or set in operation by the DRM. These strategies are also linked to strategic goals and performance objectives of the DRM program to ensure that DRM activities continue to work toward its intended end results.

The strategies identified for each business driver may be executed through policy and procedures that are described in more than one volume of the DRM. Each strategy discussed potentially has many interacting components or facets that will be developed over time, supporting more than one functional aspect of data management. For example, aspects of data quality and security will be addressed in all three volumes as each portion of the DRM structure will need to address both data quality and security aspects of data management.

DRM Business Drivers

Business Driver 1: Data-oriented barriers to eGov success need to be eliminated.

Problem Statement:

- Current methods for data sharing are inadequate and a consistent, reliable approach is needed across Communities of Practice to facilitate E-Government.

Strategy:

- Facilitate the development of Communities of Practice and common documents such as data sharing agreements
- Provide policy direction for Communities of Practice to publish templates and data access and sharing patterns in their registries

Business Driver 2: Data standardization requirements need to be implemented.

Problem Statement:

- Inattention to standards for syntax and semantics impedes the efficiency and effectiveness of agency programs, especially missed opportunities for data sharing and cross-agency collaboration.
- A diverse set of registries will continue to be maintained by the many communities that primarily benefit from them.

Strategy:

- Deploy a standardized network interface service for the discovery and interchange of specific characteristics of data elements and value sets in all registries.
- Agree on a common meta-model for registry implementations, based on standards such as the ISO 11179 international standard and the emergent ebXML industry-led standard.
- A central agency of the U.S. Federal Government will maintain a public list of all those registries that participate in the DRM.

Business Driver 3: The Federal Government needs a common approach for addressing data quality

Problem:

- The Federal Government needs rigorous methodologies for ascertaining and reporting data quality.
- The Federal Government needs to adopt a methodology for designing stewardship for various classes of data.

Strategy:

- Provide policy direction to government organizations regarding the reporting of data quality
- Provide policy direction to government organizations regarding attention to data quality throughout the data life cycle

Business Driver 4: Data semantics issues that impede community of practice work need to be resolved.

Problem Statement:

- The availability of increasing amounts of information presents the challenge of delivering the right information, to the right person, at the right time.
- The data must be relevant, meaningful and at the appropriate level of depth. (i.e. data quality and data quantity)
- The context of business processing for data creation and manipulation is needed to ensure complete understanding of the data being exchanged among agencies.

Strategy:

- Facilitate the publishing of data dictionaries, structured vocabularies and other data documentation facilities by Communities of Practice
- Ensure the DRM Meta-model includes maturing semantic technologies.

Business Driver 5: Need to demonstrate information value by establishing ties between information assets and business performance value.

Problem Statement:

A process is needed to place business value on data exchange, connecting that to mission performance.

Strategy:

- Facilitate Discovery of Business Need for information exchange by business process modeling
- Create data and information maps to define exchange requirements
- Assign value by given scenarios/conditions of data exchange considering risk contingencies
- Use models and best practices to accomplish these

Business Driver 6: Need an approach to adopt emerging integration technologies.

Problem Statement:

- Government needs an approach to more completely understand and convey the full depth of shared information characteristics, including but not limited to provenance, lineage, quality, transformations, sources, methods, relationships, etc.

Strategy:

- Ensure that the DRM Meta-model supports capturing needed classes/categories of data relationships (mapping and transformation).
- Adopt a method of information mapping.

Business Driver 7: Need an approach to implement data transfer and interoperability.

Problem Statement:

- Incompatible tools, technology, data formats, and processes hinder data and information exchange.

Strategy:

- Facilitate the discovery of government organization and the services by which they exchange data and information
- Provide guidance for government organizations to publish their services

Business Driver 8: Implement DRM Governance.

Problem Statement:

- Federal departments must begin to implement measures of governance in order to properly coordinate and support the Federal ~~DRM program~~ FEA.
- The DRM must be developed through a governance framework that considers the broad scope of the Federal Program.
- The individual departments need guidance and documented best practices.

Strategy:

- Implement policy to ensure Data Quality
- Oversee DRM Stewardship activities
- Establish Data Ownership
- Maintain Data and Information Inventories

"Three Pillars" of Successful Data Management

The DRM has evolved for three basic reasons: control, information exchange and efficiency of data operations. The "Three Pillars" of data management provide a framework for the operational focus of the DRM, addressing the previously mentioned reasons within the following categories.

A. Data Governance
B. Data Architecture
C. Data Sharing Architecture

The following discussion of the "Three Pillars" is taken from a Position Statement presented by Bryan Aucoin at the Eighth International Conference on Information Quality (ICIQ-03). He and his staff continue to work with Dr. Wang of MIT to address data quality issues within the focus of this framework. The "Three Pillars" of data management emphasizes major focal points for the success of the DRM and are integrated into the DRM Matrix of Information on page 4.

"Data architecture is an element of an enterprise architecture that comprises a data model and assigns accountability for data integrity (this is also referred to as Information Architecture). The data architecture reflects business area entities with attributes and establishes accountability for this information in business process improvements. The data model that describes the data architecture illustrates the interrelationships among real-world objects and events integral to an enterprise's business. This data architecture is independent of hardware, software, or machine performance considerations.

Our core premise in developing data architecture is that it must be defined incrementally to meet specific business requirements. While one should start with a high-level framework, specific business requirements for shared information should drive the definition of standards. The road to success in developing architecture is to bind it to the business. If the architecture is used to resolve specific business needs, it will be accepted.

Data governance is the practice of making enterprise-wide decisions regarding an organization's information holdings. Data governance includes the determination of data sources, responsibilities for integrity, defining requirements for business process development and change, and mechanisms of arbitrating differences among stakeholders. In a nutshell, there has to be way to make hard decisions within the enterprise.

Data sharing is the practice of provisioning data from an information source to an information consumer in response to a business requirement. A data sharing architecture is a standard, repeatable technical pattern for sharing data. If an enterprise can enforce its architecture through a governance process as data is shared to support real business needs, then enterprise has a good chance of creating quality data. In our *v*iew, once data is viewed as a shared corporate asset, it opens the possibility of aggregating data into corporate repositories used by many applications.

Data Governance

Success for the DRM is dependent on the identification, acceptance and proliferation of data standards, across several communities of practice and at various levels of government. The complexities of information exchange and semantic understanding of data will require agreement to accept standard definitions and procedures. Governance provides the policy and structure that allows this to take place.

With the DRM about to be launched, it is important that all departments and agencies begin to implement some measure of governance in order to properly coordinate and support the use of information exchange packages.

Cross-community coordination and feedback administration will necessitate a well-designed set of information sharing governance processes. For each proposed data standard, the sheer volume of department, agency and individual person interactions requires that a governance framework be developed to handle all aspects of proper DRM development.

At the second level, the individual departments and agencies need guidance and documented best-practices. These will serve as a Rosetta stone, representing the "translation" of the "silent symbols" of the FEA DRM into a living language for implementing department- or agency-level governance. A key precept of this guidance is the necessity for their programs to be supportive and aligned with the FEA DRM.

FEA DRM Standards Adoption Process

The DRM Standards Adoption Process will be a well-defined, repeatable process. At the heart of the process is a suite of guiding procedures. Their appropriate application will contribute to the adoption of overall best practice for identifying, labeling/documenting, categorizing, registration, and dissemination of DRM Standards.

We are basing the Standards Adoption Process on the current FGDC Standards Development process as shown on the following page. Modification and refinement of this process is likely as we continue to evolve DRM processes.

Standards Adoption Process:

PROPOSAL STAGE

Step 1.
Develop Proposed Data Element Standard - A new data element standards project proposal is submitted.
Step 2.
Review Proposal - The Data Standards Committee and the appropriate Subject Area Focus Team reviews and evaluates the standard proposal. The proposal goes out for public comment.

PROJECT STAGE

Step 3.
Set Up Project – The Data Standards Committee establishes the project and activates standards development.

DRAFT STAGE

Step 4.
Produce Working Draft - The Data Standards Committee proceeds with standard development.
Step 5.
Review Working Draft - The Data Standards Committee submits a working draft for pre-public review and then prepares a committee draft for public review.

REVIEW STAGE

Step 6.
Review and Evaluate Committee Draft - The Data Standards Committee evaluates the Committee Draft of the standard and makes a recommendation for public review to the Coordination Group.

Step 7.
Approve Standard for Public Review - The Coordination Group reviews the recommendation of the Data Standards Committee and approves standard for public review.
Step 8.
Coordinate Public Review - The Coordination Group announces and coordinates a public review of the proposed standard. Testing and validation of the standard take place at this time.
Step 9.
Respond to Public Comments - The Data Standards Committee reviews all comments and produces a revised standard and a comment response document. Results from testing and validation of the standard are documented.
Step 10.
Evaluate Responsiveness to Public Comments - The proposed standard and a public response document are reviewed by the Data Standards Committee.
Step 11.
Approve Standard for Endorsement - The Coordination Group reviews the recommendation of the DSC and approves the proposed standard for DRM endorsement.

FINAL STAGE

Step 12.
Endorsement - The Data Standards committee reviews the recommendation of the Coordination Group and endorses the standard.

Data Architecture

The need for organizations to better understand and manage information could not be more critical than it is today. Yet, we continue to see a lack of consistency in process and analytical depth in approaches to manage data across the government. Data and information needs are often ignored in the more strategic aspects of Enterprise Portfolio Management. However, advancement in performance improvement and mission outcomes will only result from calibrated efforts to strategically direct the use of information assets.

Treating information as a corporate asset is a first step to improve management of the IT portfolio. Optimizing the availability of data and information without sacrificing data quality will require adoption of best practices and compliance with data quality classification policies and diagnostic procedures. Balancing accessibility without sacrificing data integrity and security necessitates the establishment and enforcement of standard data management procedures. Improving information supply chain performance and integration will require an understanding of where value is currently being provided, as well as where the potential exists for creating value.

The DRM will help to promote best practices and provide recommended procedures and direction to efforts carried out by Line of Business (LoB) and eGovernment initiatives. Specific guidelines will be provided regarding data documentation, data standards development and use applicable to both structured and unstructured data, and management of metadata of all types. DRM guidelines ensure that data entities and attributes, data models, and relationships are sufficiently defined to convey the overall meaning and use of government data and information.

Assumptions.
Each LoB or Community of Practice:

a. – will define their data and information exchange formats as needed to support their business operations
b. - will define data and metadata to accommodate related data life cycle characteristics (create, review, update, delete).
c. – will model business and data relationships that identify dependencies and behavior limitations of the data within their specific business community
d. - will have a managing agency responsible for maintaining information exchange definitions, receive comments, convene periodic creation and update meetings and register data sharing agreements with related registry and repositories.

Information Catalogues
The DRM will include and promote the use of information categories defined by the Interagency Committee for Government Information. A sub-committee group has been charged to “catalogue and index government information” as required by the E-Government Act of 2000. The resulting work of this group will form the initial components of the DRM Business Context: Information Category and Data Group.

These classifications may also be used for OMB 300 reporting to identify an investment's data categories. Further analysis of these groupings and investment proposals will identify data groups that are commonly shared and candidates for federated data management.

To enhance progress in this area, departments and agencies should create their own context diagrams and information asset value matrix with descriptions of strategic business data and information exchanges. A common set of subjects and topics as a classification mechanism for each agency and should be defined and mapped to the FEA DRM Business Context. Emphasis should be given to defining information value from various perspectives within the organization, such as executive decision-support, operational support, and process-enabling information. FEA templates can be designed to make this effort less confusing for agencies.

DRM Information Flow

Development of an information exchange requires domain-specific analysis that identifies data specifications associated with the business information needs of a LoB or Community of Practice. Following analysis of the business process, analysis of the related information flows specifies data that will be defined as exchange requirements (data records, tables, or text) in support of the process or activity. This includes developing models that identify data relationships and associated security profiles.

A business-data exchange model can be done with an extension to the use case diagrams or the business process analysis notation to link to an information container. These high level diagrams can reflect the conceptual business flows. It can be supported by annotations or related text information on the type of information container that is linked to the business process, its structure (relational, object-oriented, blobs, text, semi-structured, etc) along with a set of abstract methods of usage and value measures.

DRM Data Element Descriptions

DRM Data Element Descriptions reflect the data standards adopted by or agreed to by the LoB or Community of Practice. The format and structure of the documentation made available in the DRM dictionary, registry or repository is guided by DRM policies, including a common meta-model for registry implementations, based on standards such as the ISO 11179 international standard and the emergent ebXML industry-led standard. This ensures that consistent levels of detail are made available for re-use and general understanding throughout the government. An assigned government agency will maintain a public list of all registries that participate in DRM information exchange operations. ~~All rights and privileges regarding management and oversight of the structure and contents of DRM facilities will remain with the DRM Program Manager and governing sub-committees~~.

Data Sharing Architecture

Information sharing is the major impetus that is driving significant changes in the management of information assets. Global economic and security issues have placed unprecedented demand on data availability and optimization. Leading international governments are emphasizing consolidation and effective utilization of information assets to achieve greater stability and to gain strategic leadership.

Government information sharing must be recognized as an important program that will require business transformation and cultural change. It will not happen overnight. It must be carefully planned, including strategy development at senior management levels, balancing the need to satisfy Congressional requirements, program costs, and security and privacy implications. Cross-agency initiatives will require significant guidance and support, initially. Executive leadership and governance structures must be in place to provide policy, guidance and decision-support.

Information technology and business practices have evolved from centralized data management systems to decentralized and distributed computing information exchanges. Increasingly mature and robust infrastructures for distributing information are helping to realize the idea that information can be available to anyone, anytime, anywhere.

Data Semantics
This presents the problem of delivering the right information to the right person, at the right time. The data must be relevant, meaningful and at the appropriate level of detail. Providing the business context of shared information increases the likelihood of an information consumer selecting the right data for a particular business situation. Data Semantics is the discipline that facilitates the delivery of the right information based on the requestor's requirements. Semantic agreement within a Community of Practice is essential to facilitating meaningful and effective data exchange.

Domain Data Harmonization Strategy
The vast majority of existing information systems have evolved over time with diverse requirements and different data models. As a result, the data stored in these systems have varying levels of meaning, consistency, and quality. Data harmonization is the process by which a Community of Practice agrees to the meaning and format of the data residing in its information systems by applying common definitions, attributes, and values. (See Appendix A, Best Practices)

Communities of Practice
The FEA DRM program facilitates the establishment of Communities of Practice and supports the development of Data Sharing Agreements as a collaboration tool. FEA DRM Guidance provides direction for these partnering agencies to publish templates and data access and sharing patterns in their registries, as well as publish their data dictionaries, structured vocabularies and other data documentation facilities.

Data Sharing Architecture

Data sharing involves providing the mechanisms needed to take data from an information source to an information consumer to satisfy a business need. Technology allows us to leverage this activity by developing standard, repeatable patterns for sharing data that can be re-used as platform-independent processing modules. By defining these modules to provide a specific service (business, application, other), code generated to provide a commonly used service could then be made available to be implemented by many participants.

Developing a Data Sharing Architecture that uses protected Internet protocols as the physical integration mechanism and uses models, provides for business-centric coordination of identified and managed data and information. This is an approach to discovering information needs and defining information needs in terms that business users can understand. The Data Sharing Architecture also supports development with minimal programming in a declarative model-based approach that builds in coordination, change management, exception handling, and conflict resolution.

The following are specific strategies that may be implemented in developing the Data Sharing Architecture.

1. Define a series of data access and sharing patterns and templates that can be used during the design process. Patterns for consideration include: decoupling, data access, active domain object, selection factory, domain object factory, update factory domain object assembler.
2. Provide a set of common utility services at the client and at the data and information servers and any use of directories or hubs available for use within a specific community of Line of Business.
3. Data elements will be integrated within a Metadata repository with separate areas along with federated models and information maps.
4. Define Line of Business information categories and a series of common secure data transfer patterns and related utility service components that can be tailored to provide secured data services.
5. While ISO 11179 is emphasized because of considerable legacy work and MOF is considered more useful currently, the Semantic Web technologies of RDF and OWL have much more mature and capable data models for semantic integration and interoperability and provides a convergence of the four data communities (document, Web, database, and programming).
6. Adopt a method of information mapping. Provide guidance to develop Information Production Maps (IP-MAP) based on the work by Dr Wang and MIT/TDQM and recent modeling approaches based on data quality improvement such as IP-UML, adapted to the FEA Reference Models; particularly, performance elements and include it as part of the business case process.
7. Deploy a standardized network interface service for the discovery and interchange of specific characteristics of data elements and value sets in all registries.

Shared Data Ownership Agreement
A Shared Data Ownership Agreement policy will define the practices used for share ownership and how the LoB would define their agreements and the "governance" policies that manage the commitment of participating parties on what can be shared, conditions for sharing, conditions for revocation of agreements and any exception clauses.

There are, however, a number of issues surrounding data sharing that require serious consideration. Addressing security and privacy issues has been a major hindrance to data sharing in some government areas. Methods for addressing security and privacy requirements will need to be identified with policy guidance. Additionally, the quality of data being shared can suffer greatly as data moves from one environment to another.

Data Quality Management
Designating stewardship for various classes of data is an alternative that can be developed. Requiring agencies to designate data sensitivity and quality is also an option that could help to reduce risks in this area. A sensitivity coefficient could be developed for each class of data. Given the sensitivity coefficient, each agency would need to document the classes of data quality problems that exist, particularly for sensitive data. Data Quality Diagnostics could be used to identify the data quality class. For each diagnostic execution, the agency could document and make available to partners the diagnostic information.

Appendix A.

Best Practices

A. Data Harmonization Guiding Principles

1. Data harmonization is a process not a project and should begin as early as possible.
2. Identify key internal and external stakeholders.
3. Engage existing and potential partners.
4. Understand and agree on scope of initiative.
5. Define requirements.
6. Review best practices.
7. Select a methodology and appropriate tools.
8. Identify relevant information exchanges and the data systems that support them.
9. Concepts and definitions must be universally accepted within the Community of Practice.
10. Publish the work product, so it can be consumed by practitioners and technologists.