The Data Reference Model

Version 2.0

November 17, 2005

CONTENTS

1. Introduction 1

2. Overview of the DRM 3

2.1. Target Audience and Stakeholders 5

2.2. DRM Implementation Framework 6

2.3. DRM Abstract Model 11

2.4. Security and Privacy 14

3. Data Description 17

3.1. Chapter Organization 18

3.2. Introduction 18

3.2.1. What is Data Description and Why is it Important 18

3.2.2. Purpose of the Data Description Section of the DRM Abstract Model: 19

3.3. Guidance 19

3.4. Data Description Section of the DRM Abstract Model 21

3.5. Data Description Attributes 24

3.6. Data Description Example 27

4. Data Context 31

4.1. Chapter Organization 32

4.2. Introduction 32

4.2.1. What is Data Context and Why is it Important: 32

4.2.2. Purpose of the Data Context Section of the DRM Abstract Model: 33

4.3. Guidance 34

4.3.1. The COI, its Participants and Processes: 34

4.3.2. The Data Context Artifact Creation Activity: 35

4.4. Data Context Section of the DRM Abstract Model 36
4.5. Data Context Attributes 41
4.6. Data Context Example 42
5. Data Sharing 44
5.1. Chapter Organization 45
5.2. Introduction 45
5.2.1. What is Data Sharing and Why is it Important 45
5.2.2. What is the Purpose of the Data Sharing Section of the DRM Abstract Model: 46
5.2.3. Structures Used for Data Sharing: 46
5.3. Guidance 53
5.4. Data Sharing Section of the DRM Abstract Model 54
5.5. Data Sharing Attributes 57
5.6. Data Sharing Example 59
6. DRM Abstract Model 61
6.1. Introduction 61
6.2. Data Description Section of the DRM Abstract Model 64
6.3. Data Context Section of the DRM Abstract Model 69
6.4. Data Sharing Section of the DRM Abstract Model 74
6.5. Data Sharing Attributes 76
APPENDIX A: Glossary of Selected Terms 78

LIST OF FIGURES

Figure 2-1 DRM Standardization Areas 7

Figure 2-2 Data Description Usage Example 8

Figure 2-3 Data Context Usage Example 9

Figure 2-4 Data Sharing Usage Example 10

Figure 2-5 DRM Abstract Model 12

Figure 3-1 DRM Data Description Abstract Model 21

Figure 3-2 Recreation One Stop Information Classes 27

Figure 3-3 DOI Three Business Focus Areas 28

Figure 3-4 COIs Identified Data Subject Areas 29

Figure 3-5 FEA BRM Logical Data Models 30

Figure 4-1 Data Context Section of the DRM Abstract Model 37

Figure 4-2 Carols Linnaeus Taxonomy 39

Figure 4-3 DOI DRM classification schemes 43

Figure 5-1 Data Supplier-to-Consumer Matrix 47

Figure 5-2 Data Sharing Section of the DRM Abstract Model 55

Figure 6-1 DRM Abstract Model 62

Figure 6-2 DRM Data Description Abstract Model 64

Figure 6-3 Data Context Section of the DRM Abstract Model 70

Figure 6-4 Carols Linnaeus Taxonomy 71

Figure 6-5 Data Sharing Section of the DRM Abstract Model 74

1. Introduction

The Data Reference Model (DRM) is one of the five reference models of the Federal Enterprise Architecture (FEA). The DRM is a framework whose primary purpose is to enable information sharing and reuse across the federal government via the standard description and discovery of common data and the promotion of uniform data management practices. The DRM describes artifacts which can be generated from the data architectures of federal government agencies. The DRM provides a flexible and standards-based approach to accomplish its purpose. The scope of the DRM is broad, as it may be applied within a single agency, within a Community of Interest (COI)[1], or cross-COI.

The DRM provides a standard means by which data may be described, categorized, and shared. These are reflected within each of the DRM's three standardization areas:

- **Data Description:** Provides a means to uniformly describe data, thereby supporting its discovery and sharing.
- **Data Context:** Facilitates discovery of data through an approach to the categorization of data according to taxonomies. Additionally, enables the definition of authoritative data assets within a COI.
- **Data Sharing:** Supports the access and exchange of data where access consists of ad-hoc requests (such as a query of a data asset), and exchange consists of fixed, re-occurring transactions between parties. Enabled by capabilities provided by both the Data Context and Data Description standardization areas.

[1] Communities of Interest are collaborative groups of user who require a shared vocabulary to exchange information to in pursuit of common goals, interests, and business objectives.

As a reference model, the DRM is presented as an abstract framework from which concrete implementations may be derived. The DRM's abstract nature will enable agencies to use multiple implementation approaches, methodologies and technologies while remaining consistent with the foundational principles of the DRM.

The following chapters and appendices are included in this specification:

- **Chapter 2 - Overview of the DRM:** Provides a brief overview of the DRM, its value to federal agencies, a summary of the DRM standardization areas, and more.
- **Chapter 3 - Data Description:** Describes the Data Description standardization area of the DRM.
- **Chapter 4 - Data Context:** Describes the Data Context standardization area of the DRM.
- **Chapter 5 - Data Sharing:** Describes the Data Sharing standardization area of the DRM.
- **Chapter 6 – Abstract Model:** Provides a consolidated view of the DRM Abstract Model.
- **Appendix A:** Glossary of Selected Terms.

2. Overview of the DRM

This document presents the DRM, one of the five reference models of the FEA. The DRM is sponsored by the Office of Management and Budget (OMB) and the Federal Chief Information Officer (CIO) Council. It is the FEA mechanism for identifying what data the federal government has and how that data can be shared in response to business/mission requirements. The DRM provides a frame of reference to:

- **Facilitate COIs (which may be aligned with the LoBs delineated in the FEA Business Reference Model) in establishing common language.**
- **Enable needed conversations to reach credible cross-agency agreements around: governance, data architecture and an information sharing architecture.**

The DRM provides guidance to enterprise architects and data architects for implementing repeatable processes to enable data sharing in accordance with federal government-wide agreements, including

agreements encompassing state, local, tribal governments, as well as other public and private non-governmental institutions. The intent is to mature, advance and sustain these data agreements in an iterative manner.

The DRM can provide value for agency data architecture initiatives by:

- **Providing a means to consistently describe data architectures:** The DRM's approach to Data Description, Data Context, and Data Sharing enables data architecture initiatives to uniformly describe their data artifacts, resulting in increased opportunities for cross-agency and cross-COI data sharing.

- **Bridging data architectures:** The DRM provides a "Rosetta Stone" to facilitate communications between enterprise and data architects about data and data architecture in their efforts to support the business/mission needs of the COIs that they support.

- **Facilitating compliance with requirements for good data architectures:** The DRM's standardization areas provide a foundation for agency data architecture initiatives to put forth requirements that can result in increased compatibility between agency data architectures.

As a reference model, the DRM is presented as an abstract framework from which concrete implementations may be derived. The DRM's abstract nature will enable agencies to use multiple implementation approaches, methodologies and technologies while remaining consistent with the foundational principles of the DRM. For example, the DRM abstract model can be implemented using different combinations of technical standards. As one example, the Exchange Package concept in the Data Sharing standardization area may be represented via different messaging standards (e.g. eXtensible Markup Language (XML) schema[2], Electronic Data Interchange (EDI) transaction set) in a concrete system architecture for purposes of information sharing. Other ways to implement DRM capabilities may be put forward by other agencies or stakeholders. By associating elements of concrete architectures with the DRM abstract model, those elements may therefore be associated with each other, which can help promote interoperability between cross-agency architectures/implementations. Thus the abstract nature of the DRM as a reference model provides tremendous implementation flexibility.

The DRM can accelerate enterprise and joint action around new opportunities afforded by standardized approaches for accomplishing goals such as the following:

---

[2] The word "schema" in this context refers to any of a number of XML-based schema languages,

- Enabling increased visibility and availability of data and data artifacts[3];
- Fostering increased information sharing;
- Facilitating harmonization within and across COIs to form common data entities that support shared missions;
- Increasing the relevance and reuse of data and data artifacts via uniform categorization techniques;

The remainder of this chapter is organized as follows:

- **Target Audience and Stakeholders:** Describes who will most benefit from reading this specification and from specific implementations of the DRM;
- **DRM Implementation Framework:** Presents the DRM guidance and rationale for the standardization areas, the purpose of each standardization area, and a brief usage example for each standardization area;
- **DRM Abstract Model:** Presents the DRM abstract model, which is described in greater detail in subsequent chapters;
- **Security and Privacy:** Discusses security and privacy considerations for the DRM;

2.1. Target Audience and Stakeholders

The target audience for DRM 2.0 is:

- Enterprise architects
- Data architects

The following additional stakeholders may make use of the DRM, depending on their individual interest and needs:

- **Senior Federal Managers:** This includes CIOs, Chief Financial Officers (CFOs), Assistant Secretaries, and other executives and managers engaged in federal information management;
- **Congressional Stakeholders:** This includes relevant Congressional committees and their staff who have legislated requirements relating to federal information and data management including subsection 207(d) of the E-Government Act;
- **External Stakeholders:** This includes:

[3] In this specification, the term “data” is often used alone to collectively mean data, data artifacts (e.g. documents, XML schemas, etc.) and data assets. At times, the term "data artifact" and/or "data asset" may be used separately, or together with "data", as appropriate for the intended meaning. The reader should consider the context of each reference.

- Citizen-centered stakeholders working in support of eGov initiatives;
- State and local government in their role as information exchangers with federal agencies;
- Industry/vendors engaged in providing Information Technology (IT) support and tools to the federal government;

2.2. DRM Implementation Framework

This section presents the DRM Implementation Framework. The DRM Implementation Framework is depicted in the table below. This framework provides a roadmap to be used by enterprise architects and data architects to guide their efforts in supporting data sharing within the COIs that they support. The roadmap is based upon the following basic assertions.

- Data Context is a standardization area within the DRM. A COI should agree on the context of the data needed to meet its shared mission business needs. A COI should be able to answer basic questions about the Data Assets that it manages. "What are the data (subject areas) that the COI needs? What organization(s) is responsible for maintaining the data? What is the linkage to the FEA Business Reference Model (BRM)? What services are available to access the data? What database(s) is used to store the data?" Data Context provides the basis for data governance with the COI.

- Data Description is a standardization area within the DRM. A COI should agree on meaning and the structure of the data that it needs in order to effectively use the data.

- Data Sharing is a standardization area within the DRM. A COI should have common capabilities to enable information to be accessed and exchanged. Hence the DRM provides guidance for the types of services that should be provisioned within a COI to enable this information sharing.

		DRM Chapters		
		Context	**Description**	**Sharing**
DRM Sections	**Introduction**	- What are the data needed to support the business/mission needs of a COI? - What core information does the COI need to make the data discoverable and establish governance?	- How will the meaning and structure of the data be conveyed?	- What is the data sharing architecture? (i.e., How will the data be made sharable)
	Guidance	- Define subject areas and entities of interest - Identify data sources and stewardship - Establish governance	- Establish semantic and syntactic standards	- Establish the Data Sharing services required to support the data sharing needs of the COI
	Abstract Model	- Document in accordance with the DRM abstract model	- Document in accordance with the DRM abstract model	- Describe services specifications in accordance with the DRM abstract model

DRM Implementation Framework

These three standardization areas are shown in Figure 2-1 below:

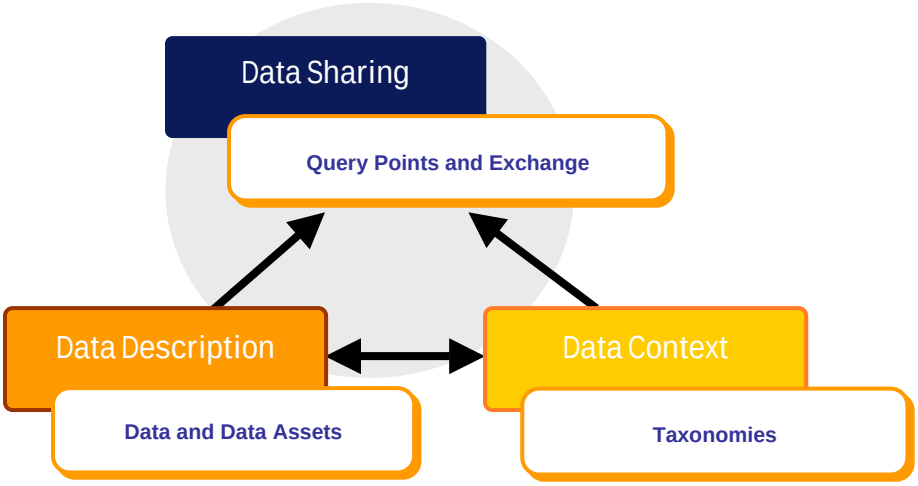


Figure 2-1 DRM Standardization Areas

The arrangement of the standardization areas in the above figure indicates how Data Sharing is supported by the capabilities provided by the Data Description and Data Context standardization areas, and how Data Description and Data Context capabilities are mutually supportive. These relationships will become clearer in the subsequent chapters in which the standardization areas are described in detail.

The following is a brief description of each standardization area, along with its purpose and a usage example.

Data Description: The Data Description standardization area provides a means to uniformly capture the semantic and syntactic structure of data. This enables comparison of metadata (data about data) for purposes of harmonization, and supports the ability to respond to questions regarding what is available in terms of Data Descriptions (metadata).

Figure 2-2 depicts a usage example for the Data Description standardization area:

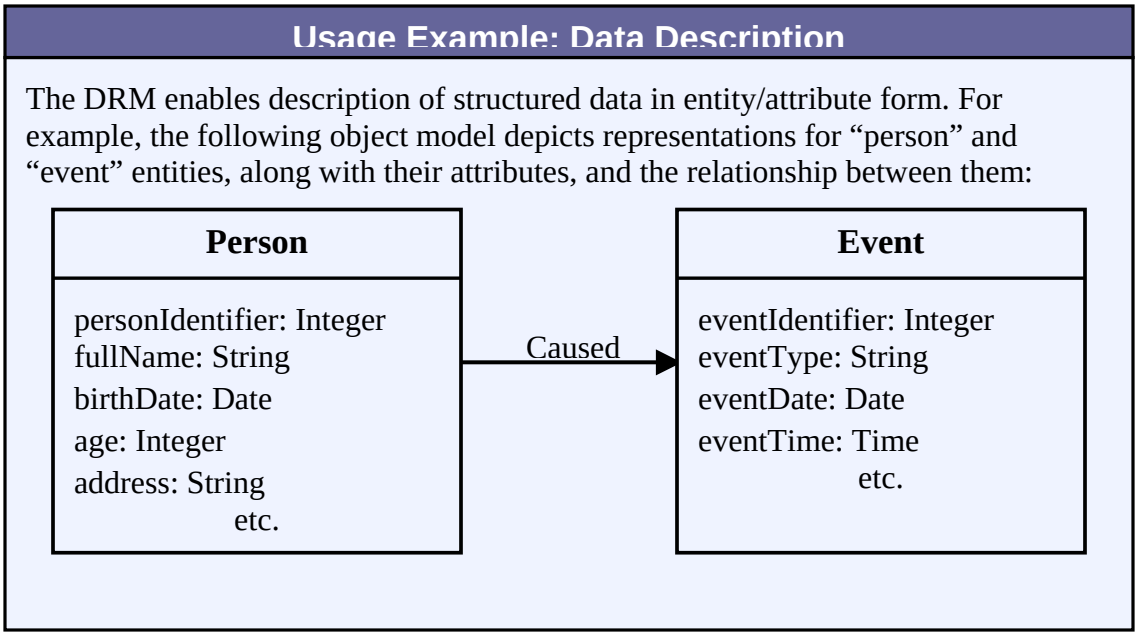


Figure 2-2 Data Description Usage Example

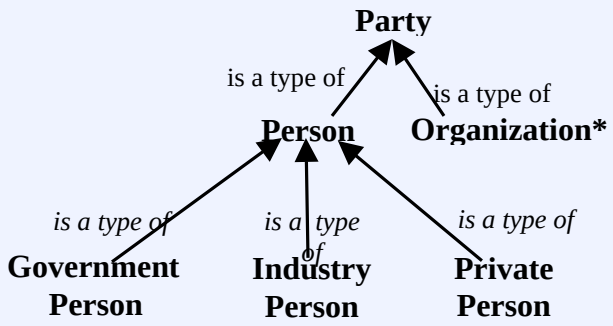
Data Context: The Data Context standardization area establishes an approach to the categorization of data assets using taxonomies and other descriptive information. In general, Data Context answers key questions about the data required within a COI and establishes the basis for data governance. Data Context also enables discovery of data, and can provides linkages to the other FEA reference models, which are themselves taxonomies.

It should be noted that context also includes business rules. However, business rules will be covered in a later version of the DRM.

The following is a usage example for the Data Context standardization area:

Usage Example: Data Context

The following example expands on the earlier "Data Description" usage example. It depicts a partial taxonomy that includes a "Person" entity, with several levels of categorization (subtopics) beneath it. Relationship types are specified near each arrow – for example, "Private Person" (e.g. a retiree) is a type of "Person". Although all relationships in this example are the same, that is not always the case.



*expansion not shown, for purposes of brevity

Figure 2-3 Data Context Usage Example

Data Sharing: The Data Sharing standardization area describes the access and exchange of data, where access consists of recurring requests (such as a query of a Data Asset), and exchange consists of fixed, recurring information exchanges between parties. Data sharing is enabled by capabilities provided by both the Data Context and Data Description standardization areas.

The Data Sharing standardization area is supported by the Data Description and Data Context standardization areas in the following ways:

- **Data Description:** Uniform definition of Exchange Packages and Query Points supports the capability to effectively share them within and between COIs;

- **Data Context:** Categorization of Exchange Packages and Query Points supports their discovery, and their subsequent use in data access and data exchange.

The following is a usage example for the Data Sharing standardization area:

Usage Example: Data Sharing

This example is based on an existing implementation of the DRM at U.S. Department of the Interior (DOI), for the Recreation One Stop initiative. It depicts the sharing of data that resides in the Recreation Information Database (RIDB). A query point exists for this data asset, and it is a CGI (Common Gateway Interface) program. A query is specified which acts on the query point to obtain information regarding a recreation area whose identifier is "1577". This query results in an XML document that conforms to the RecML standard, and the XML schema with which this XML document is associated can be considered a "payload definition" that is associated with the exchange package that describes this data exchange. Finally, the result is an HTML page that displays the contents of the XML document along with various Web page features and formatting (headings, etc.) on the Recreation.gov Web site, created by applying an XSLT style sheet to the RecML XML document.

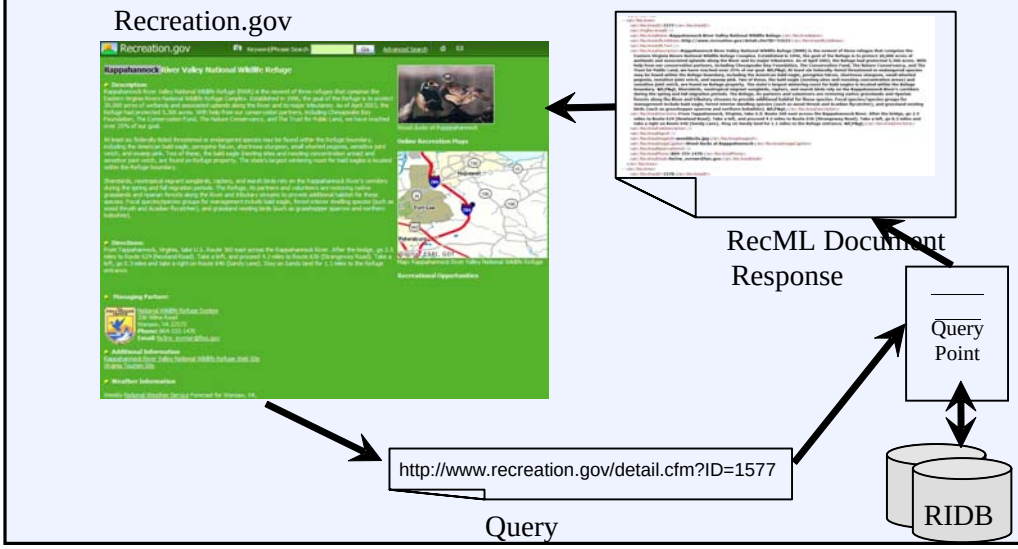


Figure 2-4 Data Sharing Usage Example

2.3. DRM Abstract Model

Figure 2-5 presents the DRM abstract model. It depicts the major concepts from each standardization area and the relationships between them. Concepts highlighted in red are described further below. Concepts are expressed as boxes, while relationships are expressed as arrows.

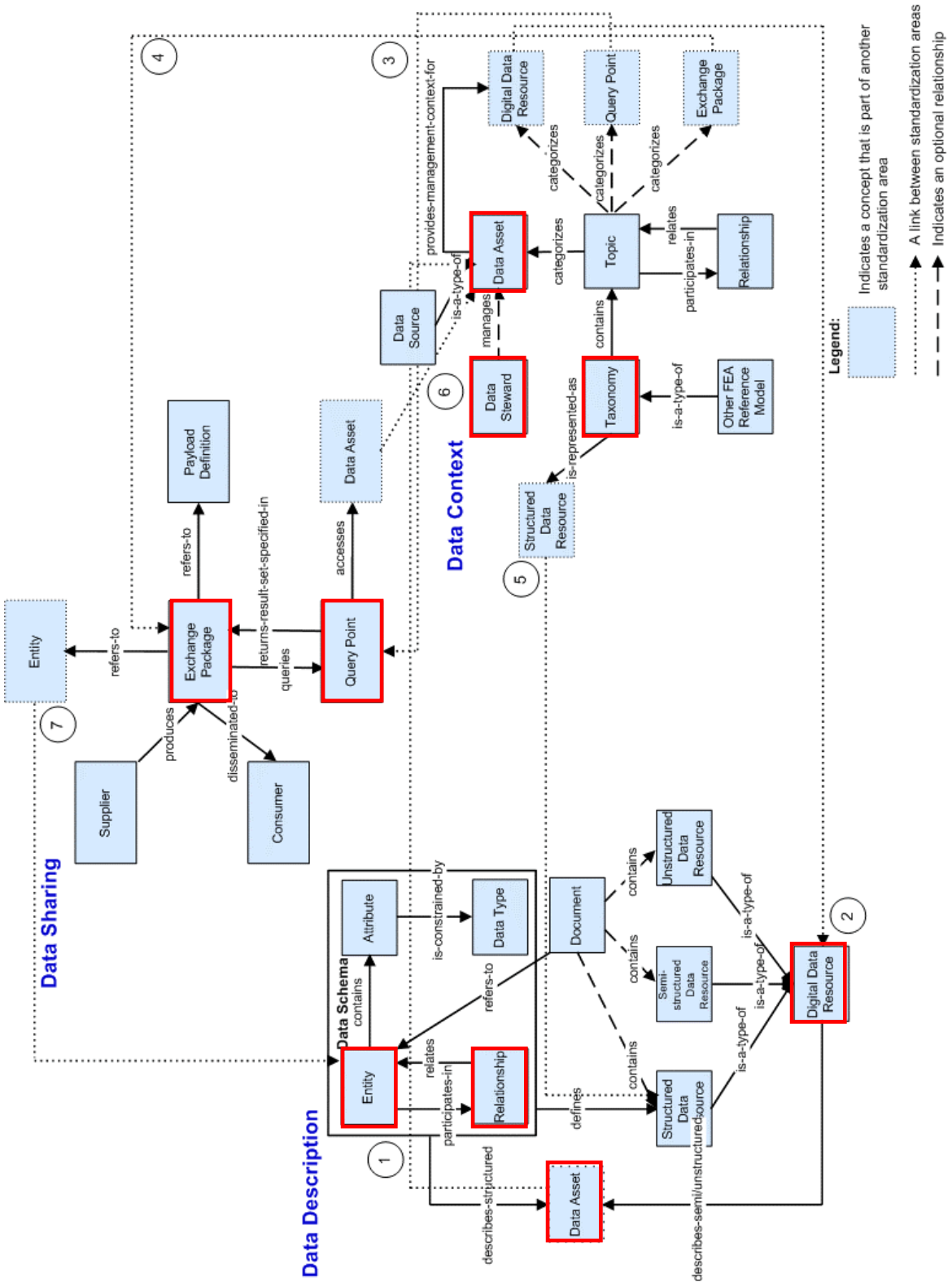


Figure 2-5 DRM Abstract Model

The DRM abstract model is an architectural pattern to optimize agency data architectures. It is abstract in that it allows multiple technical implementations; for example, the Department of Defense could use the DOD Discovery Metadata Specification (DDMS) for Digital Data Resource attributes while another agency may choose to use the Dublin Core elements, and both could demonstrate how their implementation maps to the DRM abstract model. This architectural pattern is designed to optimize an agency's data architecture for information integration, interoperability, discovery and sharing. The pattern achieves this optimization by defining, arranging and relating the standard concepts in a data architecture, specifying common attributes for each concept (presented in tables following the abstract model section figure in each chapter) and demonstrating a use case of the model in each chapter. Figure 2-5 depicts all the concepts and relationships in DRM the abstract model.

Before defining each concept, it is important to understand the highlights of the model in the three standardization areas. In the Data Description standardization area, the focus is on understanding the data at two levels of abstraction: the metadata artifacts required to understand the data and how those metadata artifacts are aggregated into a managed Data Asset. There are two basic types of metadata recommended in the Data Description section of the DRM abstract model: logical data models to describe Structured Data Resources, and Digital Data Resource metadata (such as Dublin Core elements) to describe Semi-Structured and Unstructured Data Resources. The division of data along these two axes is intended to support harmonization (via comparison of logical data models) and registration (via description of universal resource attributes). Implementation of the Data Schema concept group would take the form of Entity-Relationship diagrams, class diagrams, etc. Implementation of the Digital Data Resource could be records in a content management system or metadata catalog.

In the Data Context standardization area, the focus is on management mechanisms to capture the context of data in an organization or COI. Those mechanisms are Taxonomies (a hierarchical set of Topics connected by relationships) and a Data Asset description (captured in an inventory). A Data Asset is a collection of Digital Data Resources that is managed by an organization, categorized for discovery, and governed by a data steward. A key attribute of a Data Asset is whether it is authoritative and if so designated, authoritative on which Entity or Attribute of the logical data model (see Data Schema in the Data Description section of the DRM abstract model). Implementation of Taxonomies could take the form of extensible Markup Language (XML) Topic Maps, Web Ontology Language (OWL) hierarchies or ISO11179 Classification schemes. Implementation of a Data Asset inventory could be records in a metadata registry.

Lastly, in the Data Sharing standardization area, the focus is on how information is packaged for and/or exposed to members of a COI. The key concepts are Exchange Packages as containers for fixed messages and Query Points as descriptions of data access points. Implementation of Exchange Packages could be standard XML messages or EDI transaction sets. Implementation of Query Points could be descriptions in a

Universal Description, Discovery and Integration (UDDI) or ebXML registry of a data access Web service.

Taken as a whole, the DRM abstract model should be used by agencies to assess the current state of their data architectures and to chart a roadmap to an improved architecture. In inter-agency collaborations, this abstract model becomes a Rosetta Stone to decipher specific implementations of these common concepts and thus speed effective communication to deliver cross-organizational agility to a COI.

Subsequent chapters will "drill down" into the details of this abstract model. Chapter 6 also describes the DRM Abstract Model in its entirety. Each section of the DRM abstract model represents the core concepts and the relationship of those concepts within its respective standardization area. Each section represents the *minimal* level of detail necessary to convey the major concepts for the standardization area, with COIs extending the model as necessary for their implementations.

2.4. Security and Privacy

Security and privacy considerations apply to all three of the DRM's standardization areas. Security defines the methods of protecting information and information systems from unauthorized access, use, disclosure, disruption, modification, or destruction in order to provide integrity, confidentiality and availability, whether in storage or in transit. Privacy addresses the acceptable collection, creation, use, disclosure, transmission, and storage of information, its accuracy, and the minimum necessary use of information.

The DRM allows for the integration of existing federal information security and privacy policies within each of its standardization areas. The table below describes several sets of security/privacy policies and legislation that are applicable to the DRM.

Policy/Legislation	Description
Federal Information Security Management Act (FISMA) (Title III – Information Security)	FISMA is the premier legislation governing federal information security. It provides a comprehensive framework for ensuring the effectiveness of information security controls over information resources that support Federal operations and assets. FISMA is part of the E-Government Act.
National Institute of Standards and Technology (NIST) FIPS (NIST FIPS 199)	FIPS 199 provides standards for the security categorization of federal information and information systems.
E-Government Act of 2002 (Title III, Section 208 – Privacy Provisions)	Title III, Section 208 of the E-Government Act of 2002 requires that OMB issue guidance to agencies on implementing the privacy provisions of the E-Government Act.
OMB Circular A-11 (Section 31-8)	Section 31-8 of OMB Circular A-11 addresses management improvement initiatives and policies for agencies, to include security and privacy.
NIST 800-60 (Volume I)	NIST 800-60 provides guidance on mapping types of information and information systems to security categories. Its objective is to facilitate provision of appropriate levels of information security according to a range of levels of impact or consequences that might result from the unauthorized disclosure, modification, or loss of availability of the information or information system.

A Security and Privacy Profile (SPP) has been created for the FEA. The FEA SPP provides guidance to agencies to integrate security and privacy requirements across their enterprise architecture, and to ensure security and privacy requirements are addressed in IT programs from their inception. The FEA SPP is currently in the Validation stage. During this stage, the FEA SPP approach and methodology will be validated against Federal experience and insight.

An institutional process that includes roles and responsibilities for data stewardship for each project or program in the agency needs to be defined as part of a policy that governs data Quality, Security, Privacy and Confidentiality.

There are a number of areas that should be addressed in building a Security, Privacy and Confidentiality Policy for an agency. These include:

- Constructing a policy that is compliant with legislation, Executive Orders and Standards
- Addressing sensitivity of information that eliminates possible compromise of sources and methods of information collection and analysis

- Establishing the practices of data stewardship
- Addressing specific data access policies defined by the responsible steward; for example:
    - Data is available for open, unrestricted access
    - Data is accessible only to a group
    - Data access is a function of the person (his or her identity), data about that person (e.g., current position), and data about the environment (e.g., physical location)
    - Data is self protecting through digital rights management[4] or similar technologies

The successful categorization, describing and sharing of data are dependent on the implementation of security regarding the data being exchanged. Security requirements must be considered at each level of the DRM and, in particular, regarding the sharing of data. The DRM is designed to allow for the integration of existing federal information security and privacy policies within each of its standardization areas.

Future versions of the DRM will relate the DRM to the FEA SPP, and will apply the results of the FEA SPP validation in expanding on the security and privacy considerations for the DRM.

[4] Digital Rights Management is also abbreviated DRM. Hence, the reader should be aware of context when this abbreviation is encountered.

3. Data Description

This chapter describes the Data Description standardization area of the DRM. The purpose of this standardization area is to enable the uniform description of data in order to enable mission-critical capabilities such as data discovery, reuse, harmonization, sharing and exchange, as well as rapid coordination and communication clarity in cross-government actions. The Data Description standardization area addresses the question of "How do you understand what data is available and what it means?" Through the generation of Data Description artifacts, data within an agency can be categorized, discovered, and shared. It further enables data to be clearly tied to LoBs and specific agency missions. The chapter establishes guidance for the description of the types of data depicted in the DRM abstract model.

3.1. Chapter Organization

This chapter is organized as follows:

- **Introduction:** Provides introductory information regarding the Data Description standardization area, the nature of the related business issues and the business reasons for sound Data Description;
- **Guidance:** Provides a description of the key issues affecting Data Description;
- **Data Description Section of the DRM Abstract Model:** Presents and describes the Data Description section of the DRM abstract model;
- **Data Description Example:** Provides a usage example to further explain the Data Description standardization area;

3.2. Introduction

3.2.1. What is Data Description and Why is it Important

The purpose of the Data Description standardization area is to provide a means for a COI to agree to the structure (syntax) and meaning (semantics) of the data that it uses. Within the context of the DRM, these agreements are documented as Data Description artifacts that are captured in accordance with the DRM abstract model. Hence, Data Description artifacts are an output of the process of providing data syntax and semantics and a meaningful identification for a data resource so as to make it visible and usable by a COI.

The FEA Program Management Office (FEA PMO) recognizes that data has a significant role in the FEA. Historically, when executives, managers, operations personnel, etc. hear the terms "data" and "data management", they have equated it to a low level, "bits and bytes" technical task that is taken care of by data people on application development projects. In reality, the data in a COI are the basis for sound business decision making. If Data Description is done right it has a positive impact on mission effectiveness. If it is done wrong it impedes that effectiveness, sometimes with disastrous results when data needed for decision making cannot be found.

Comprehensive management of data, throughout its life cycle, is critical to providing high quality information to all aspects of government operations. The inclusion of the DRM in the FEA not only elevates the significance of sound data management practices, it is also a catalyst for federal government agencies to improve the quality, efficiency, and effectiveness of their data. Data Description is the foundation of those practices. It enables the following critical mission support capabilities:

- **Data Discovery:** The capability to quickly and accurately identify and find data that supports mission requirements. This is possible through the means of uniformly describing data that are presented in this chapter, as well as through the categorization, search and query capabilities described in subsequent chapters.
- **Data Reuse**: The capability to increase utilization of data in new and synergistic ways in order to innovatively and creatively support missions.

- **Data Sharing:** The identification of data for sharing and exchange within and between agencies and COIs, including international, state, local and tribal governments, as appropriate.
- **Data Entity Harmonization:** An enhanced capability to compare data artifacts across government through a common, well-defined model that supports the harmonization of those artifacts and the creation of "common entities".
- **Semantic Interoperability[5]:** Implementing information sharing infrastructures between discrete content owners (even with using service-oriented architectures or business process modeling approaches) still has to contend with problems with different contexts and their associated meanings. Semantic interoperability is a capability that enables enhanced automated discovery and usage of data due to the enhanced meaning (semantics) that are provided for data.

The Data Sharing services described in Chapter 5 describe the underlying capabilities that enable a COI to successfully perform these functions --- when the data within a COI has been adequately described.

3.2.2. Purpose of the Data Description Section of the DRM Abstract Model:

The Data Description section of the DRM abstract model exists to identify the various data types used for Data Description artifacts and their interrelationships. They are the artifacts also generated and used as a matter of course in good data management practices. The specific focus of the section is twofold, the identification of entities, and designation of the information describing them. The process of identifying entities is part of the analysis as to what data supports what aspects of a Line of Business (LoB). When the Data Description artifacts are developed with high quality standards they support an agency's or COI's data architecture and enable Data Sharing services.

3.3. Guidance

The guidance for data architects is straightforward: generate the appropriate artifacts for the data collections that will have the greatest benefit if they become shared. The artifacts used for Data Description are the ones that data architects have been using for decades. They are Data Schemas and document descriptions that provide metadata to be associated with the various databases, documents and files that are stored on the agency's or COI's computers. The DRM abstract model shows the relationships of those artifacts. Data architects should create them and make them available to provision Data Sharing services that are described in Chapter 5.

---

[5] From *Adaptive Information,* by Jeffery T. Pollock and Ralph Hodgson, John Wiley and Sons, Inc., ISBN 0-471-48854-2, 2004. p. 6.

As a first step to realize these capabilities, data needed with a COI should be architecturally tied to the LoBs that it supports. This linkage is established by processes that will be chosen by data architects, and documented within an enterprise architecture. The artifacts of the Data Description standardization area, as defined in the Data Description section of the DRM abstract model, however, were purposely defined at the most abstract level. Thus, if used by data architects, they will support data architecture at various levels (e.g. agency, COI).

Metadata developed in accordance with the DRM abstract model should be provided that is appropriate for each type of data. This activity should be guided by the data architect and have several phases, each applying an 80/20 type of rule. During the transition to EA processes that incorporate support for the DRM's Data Sharing activity, data architects should interact with the COIs that can identify and prioritize key data collections and related services within their domain of expertise; these may already exist or they may be in development. This prioritized list will provide a focus for near-term COI initiatives to create metadata, to advertise the data, and ensure that the data is available in a stakeholder-accessible space. Artifacts may need to be mapped to the Data Description section of the DRM abstract model. When creating such artifacts there is an opportunity to adopt practices that would improve maturity scores on the EA Assessment Framework.

When a COI is established, the architect should support establish a mechanism to capture common semantic and syntactic information (e.g. a data dictionary). The Data Description section of the DRM abstract model shows that there are Structured Data Resources. These are distinguished by having a description, a Data Schema, a pre-defined self-consistent form, one independent of the actual values of the data that it describes. Such data is typically managed with a tool suite that supports documentation of Data Schemas. The Data Description artifacts should be generated using those tools. They provide the syntax of the structured data and some associated data semantics. Further, best practices require that the names of the Entities and Attributes in the Data Schemas should be associated with an additional textual description of their meaning. Taken together these textual descriptions are called a data dictionary. When Data Schemas are published for databases they should be accompanied by their data dictionaries, which are also instances of a Structured Data Resource.

The Data Description section of the DRM abstract model further identifies that in addition to Structured Data Resources, there are Unstructured Data Resources and Semi-Structured Data Resources. The latter combines the former two. The latter two also have a *contains* relationship to a Document, meaning that a Document may contain unstructured and/or semi-structured data. The distinction is made at a high level of abstraction because the government's data holdings encompass textual material, fixed field databases, web page repositories, multimedia files, scientific data, geospatial data, simulation data, manufactured product data and data in other more specialized formats. Whatever the type of data, however, COIs specializing in them have developed within the government and external stakeholder organizations. These COIs have a long history of

understanding how such data should be described. The standards which these groups use should guide federal Data Description efforts.

3.4. Data Description Section of the DRM Abstract Model

The Data Description section of the DRM abstract model is shown in figure 3-1. It depicts the concepts that comprise the Data Description standardization area and the relationships between them. Concepts are expressed as boxes, while relationships are expressed as arrows. A *concept group*, an aggregation of related concepts, is also expressed in this section of the DRM abstract model as the *Data Schema* concept group.

NOTE: The "Document" concept below represents an example of one kind of data object.

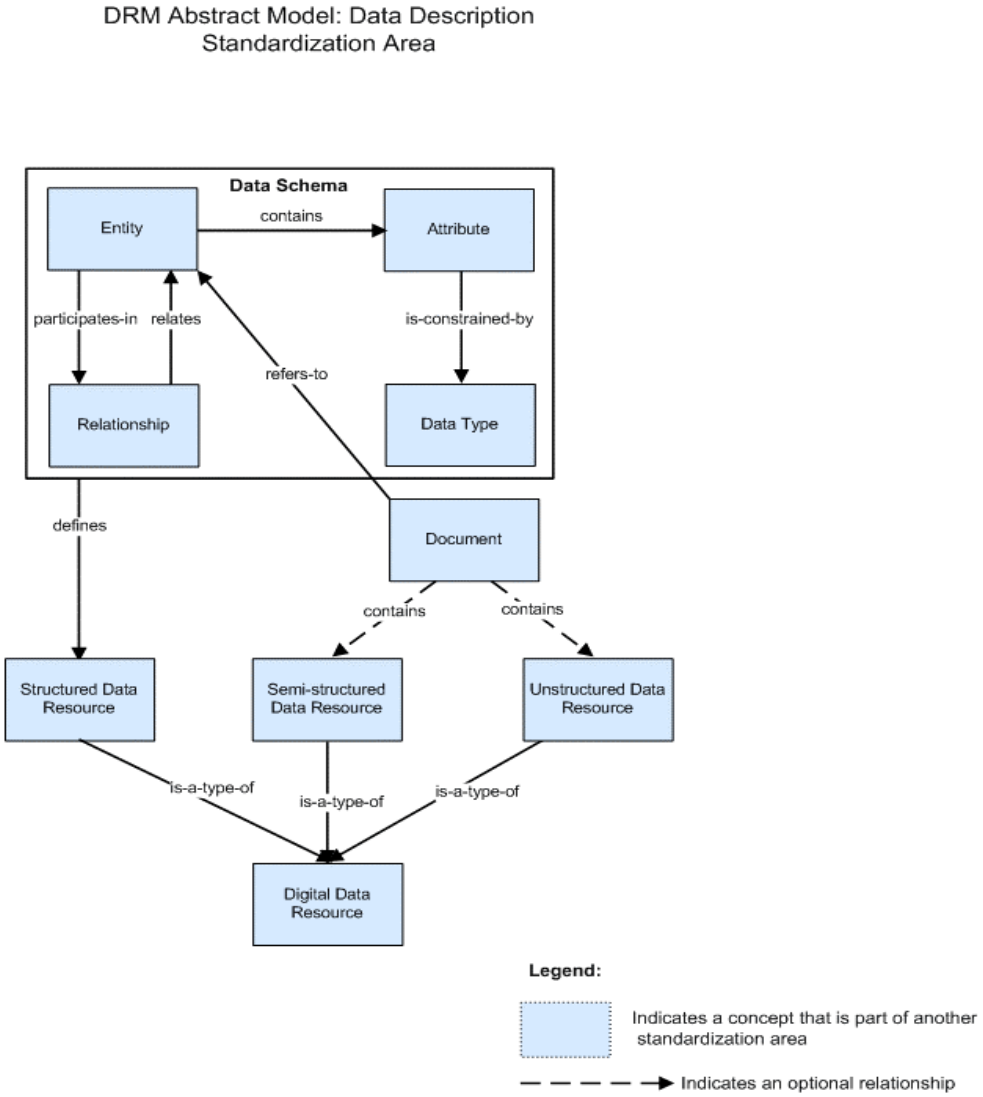


Figure 3-1 DRM Data Description Abstract Model

The following are definitions for each of the concepts and relationships within the figure shown above. Conventions used are:

- Only "outbound" relationships are listed (i.e. those that originate from the concept);
- The concepts are presented in an order that will ensure the best possible understanding, and specific examples are provided where appropriate;
- Though cardinality is not expressed in the figure, the descriptions below may include cardinality (e.g. "one or more") for purposes of clarity;
- Concept names will be capitalized as in figure itself (e.g. "Digital Data Resource"), while relationship names will be expressed in italics, and without any hyphens that may appear in the relationship name in figure (e.g. *"is constrained by"*). This is done so that the definitions below can take on as narrative a tone as possible. The reader should therefore be able to easily visually navigate through the figure as they read the definitions below.
- Each concept will be referred to in a quantity of one (e.g. "An Entity *contains* an Attribute") for purposes of simplicity as figure does not depict cardinality. However, implementations based on the DRM will introduce cardinality as needed according to their requirements.

Data Schema: A representation of metadata, often in the form of data artifacts such as logical data models or conceptual data models. The *Data Schema* concept group is comprised of those concepts pertaining to the representation of structured data. A Data Schema provides a means to provision data sharing services that is independent of the values of the data in the data resource that it describes.

- **Relationships:**
  - A Data Schema *defines* a Structured Data Resource
  - A Data Schema *describes a Structured* Data Asset

Entity: An abstraction for a person, place, object, event, or concept described (or characterized) by common Attributes. For example, "Person" and "Agency" are Entities. An *instance* of an Entity represents one particular occurrence of the Entity, such as a specific person or a specific agency.

- **Relationships:**
  - An Entity *contains* an Attribute
  - An Entity *participates in* a Relationship with another Entity

Data Type: A constraint on the type of physical representation that an instance of an Attribute may hold (e.g. "string" or "integer").

- **Relationships:**

- None

Attribute: A characteristic of an Entity whose value may be used to help distinguish one instance of an Entity from other instances of the same Entity. For example, an Attribute of a "Person" Entity may be "Social Security Number (SSN)". An SSN is used to distinguish one person (i.e. one instance of a "Person" Entity) from another.

- **Relationships:**
  - An Attribute *is constrained by* a Data Type

    Example: The "SSN" Attribute of a "Person" Entity may have a Data Type of "string" (if hyphens are included with the SSN) or "integer" (if hyphens are not included).

Relationship: Describes the relationship[6] between two Entities.

- **Relationships:**
  - A Relationship *relates* an Entity

    Example: A "Person" Entity may have a Relationship with an "Agency" Entity of "works for".

Digital Data Resource: A digital container of information, typically known as a file. A Digital Data Resource may be one of three specific types of data resources, each corresponding to one of the three types of data described earlier, and each described below (see "Structured Data Resource", "Semi-Structured Data Resource", and "Unstructured Data Resource"). It will be a container for the metadata about the data resource.

- **Relationships:**
  - A Digital Data Resource *describes a Semi-structured* Data Asset
  - A Digital Data Resource *describes an Unstructured* Data Asset

Structured Data Resource: A Digital Data Resource containing structured data. This data can be accessed in a uniform manner, independent of data values, once the Data Schema is known.

- **Relationships:**
  - A Structured Data Resource *is a type of* Digital Data Resource

Semi-Structured Data Resource: A Digital Data Resource containing semi-structured data. This will generally consist in part of structured data and in part of unstructured data.

---

[6] It should be noted that the term "relationship" is used in two ways here. The concept named "Relationship" participates in relationships with other concepts in the abstract model, and also defines the relationship between entities when it is applied to a specific scenario.

- **Relationships:**
  - A Semi-Structured Data Resource *is a type of* Digital Data Resource

Unstructured Data Resource: A Digital Data Resource containing unstructured data. Unstructured data is collection of data values that are likely to be processed only by specialized application programs.

- **Relationships:**
  - An Unstructured Data Resource *is a type of* Digital Data Resource

Document: A file containing Unstructured and/or Semi-Structured Data Resources.

- **Relationships:**
  - A Document *may contain* an Unstructured or Semi-Structured Data Resource
  - A Document *refers to* an Entity

    Example (relationship with Entity): A query that states "Find all Documents in which the following person is referenced".

    NOTE: While a Document can contain structured data, it normally has explanatory material included, which would cause it to therefore be considered semi-structured. It is for this reason that there is no *"contains"* relationship from Document to Structured Data Resource. It is very important to separate Documents from Structured Data Resources because they are processed very differently. The difference between a Document and a Digital Data Resource, therefore, is that a Digital Data Resource can contained structured data.

3.5. Data Description Attributes

This section will expand on the concepts presented above to include attributes[7] that are associated with each concept in the Data Description section of the DRM abstract model. A description will be provided for each attribute, along with an example where necessary for clarity. All Unstructured Data Resource attributes and their descriptions are taken from the Dublin Core Metadata Initiative (DCMI), Version 1.1, available at http://dublincore.org/documents/dcmi-terms/. All references to "resource" within descriptions of Unstructured Data Resource should therefore be interpreted as "Unstructured Data Resource". The above URL provides additional information on attribute descriptions and usage.

[7] It should be noted that the term "attribute" is used here in a different way than for the concept named "Attribute". Here, an "attribute" is used to describe characteristics of each of the concepts in the abstract model.

Concept	Attribute	Description	Example
Entity	Identifier[8]	A unique string associated with an Entity for identification purposes.	"200XCB"
	Name	The name of an Entity.	"Person"
	Description	A description of an Entity.	
Data Type	Name	The name of a Data Type.	"string"
	Description	A description of a Data Type.	
Attribute	Name	The name of an Attribute.	"Date Of Birth"
	Description	A description of an Attribute.	
Relationship	Name	The name of a Relationship.	"works-for"
	Origin	Name of the concept that is the origin (i.e. the "from" concept) of a Relationship.	
	Destination	Name of the concept that is the destination (i.e. the "to" concept) of a Relationship.	
Digital Data Resource	See "Structured Data Resource", "Semi-Structured Data Resource", and "Unstructured Data Resource"[9]		
Structured Data Resource	See all concepts within "Data Schema" group		
Semi-Structured Data Resource	See "Structured Data Resource" and "Unstructured Data Resource"		
Unstructured Data Resource[10]	Title	A name given to the resource.	"Information Exchange Report – July 2005"
	Resource Identifier	An unambiguous reference to the resource within a given context.	"200XCB"
	Date	A date of an event in the lifecycle of the resource. Will typically be associated with	

[8] The "Identifier" attribute is described at an abstract level in order to be consistent with the abstract nature of the reference model. Therefore, there are no references to aspects such as identifier uniqueness, representation format, or similar. Implementations based on the DRM will introduce such aspects as needed according to their requirements.
[9] As shown in the abstract model, a Digital Data Resource may be one of these three specific types of data resources. The same general idea applies to the entries for the "Semi-Structured Data Resource" and "Data Object" concepts above.

Concept	Attribute	Description	Example
		the creation or availability of the resource.	
	Creator	An entity[11] primarily responsible for making the content of the resource.	
	Format	The physical or digital manifestation of the resource. Typically, format may include the media-type or dimensions of the resource.	"text/plain"
	Description	An account of the content of the resource.	
	Source	A reference to a resource from which the present resource is derived. Recommended best practice is to reference the resource by means of a string or number conforming to a formal identification system.	"300YDC"
	Subject	A topic of the content of the resource.	
	Resource Type	The nature or genre of the content of the resource.	"Service"
	Publisher	An entity responsible for making the resource available.	
	Contributor	An entity responsible for making contributions to the content of the resource.	
	Language	A language of the intellectual content of the resource.	"eng"
	Relation	A reference to a related resource.	"400ZED"
	Coverage	The extent or scope of the content of the resource.	"Chicago"
	Rights Management	Information about rights held in and over the resource.	"Public domain"
Document	See "Structured Data Resource" and "Semi-Structured Data Resource"		

[11] It should be noted that the term "entity" here, and in subsequent Dublin Core attributes, does not have the same exact meaning as the "Entity" concept of the Data Description abstract model.

3.6. Data Description Example

This section provides a usage example for the Data Description standardization area. It is based on an existing implementation of the DRM at the Department of the Interior (DOI), for the Recreation One Stop initiative[12].

The DOI recreation functions deliver services that make up Recreation One Stop. DOI has created various "information classes" that describe the data required for Recreation One Step – these are shown in Figure 3-2:

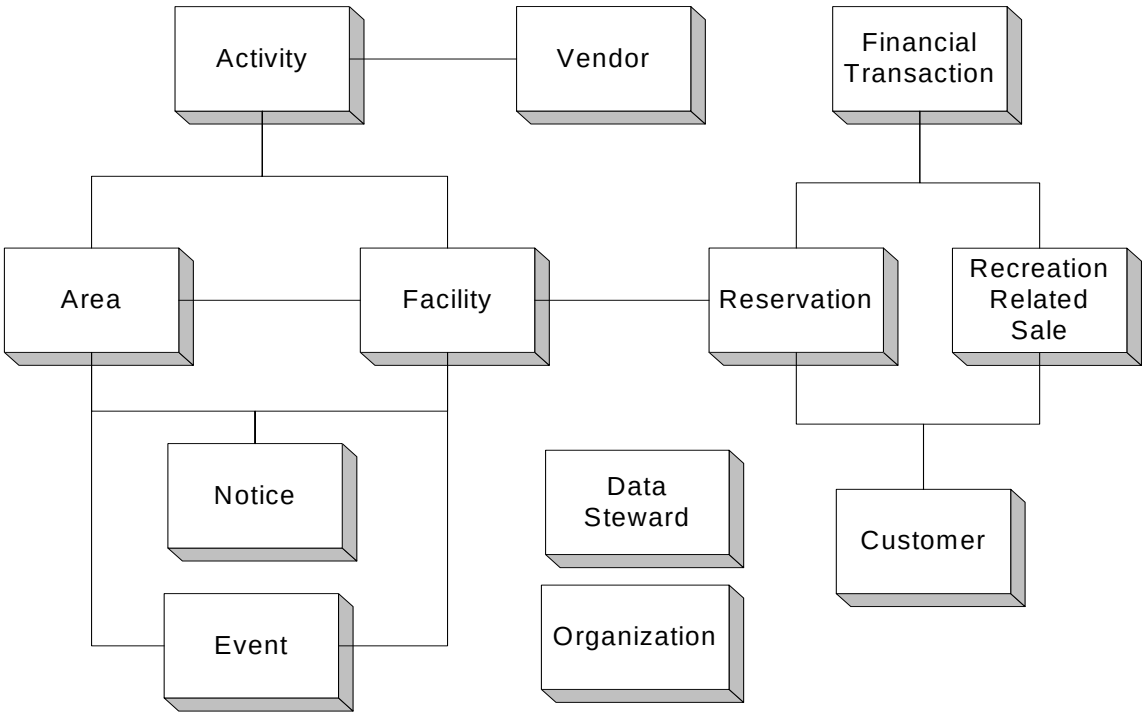


Figure 3-2 Recreation One Stop Information Classes

The above figure represents a conceptual data model, in which each information class is equivalent to Data Description's Entity. Attributes are not represented in the conceptual data model – however, they are represented in *logical data models* that are derived from the conceptual data model. Names of relationships between classes are omitted from the above figure below for purposes of simplicity; however, some are generally evident (such as Customer *makes-a* Reservation).

[12] As it is a taken from an existing operational system the terminology used in the description may differ than that described in the DRM abstract model, but it is offered to demonstrate the various ways that an agency uses a variety of logical data models to characterize the data description/sharing constructs

DOI used the International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) 11179 Metadata Registries standard for the metadata attributes that describe its data. ISO/IEC 11179 is a Metadata Registry standard that can be used by implementations based on the DRM to register and represent the metadata describing data within their data assets.

Using techniques that are standard in data architecture, DOI identified those data subject areas[13] that needed to be shared between business areas of the DOI enterprise. Figure 3-3 depicts one such example involving three "business focus areas" and the citizen. Several information classes shown earlier are evident – for example:

- Customer
- Event[14]
- Financial Transaction

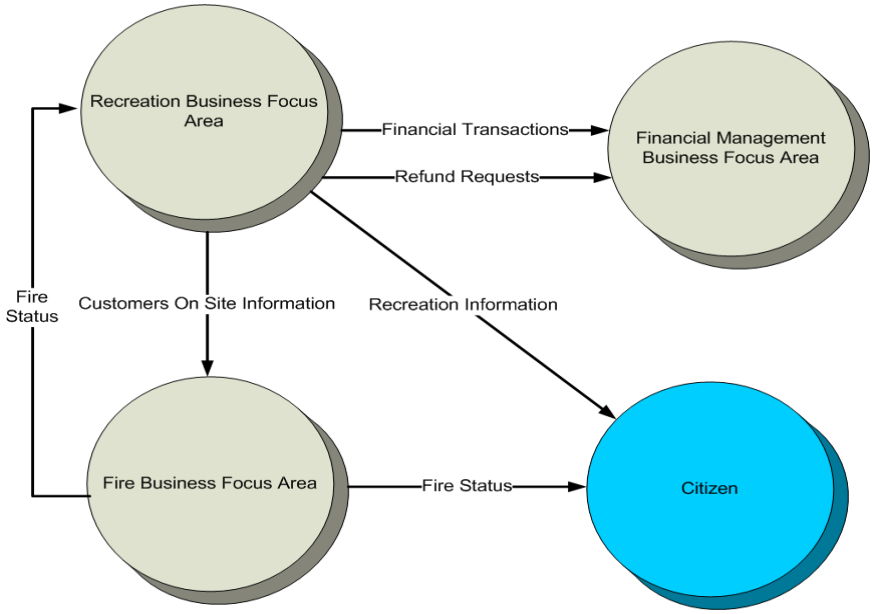


Figure 3-3 DOI Three Business Focus Areas

[13] A data subject area is comprised of one or more information classes.
[14] In this example, a specific type of event is depicted (a fire).

Common data and data sharing opportunities were also identified using identified data subject areas as a unifying mechanism across COIs, as shown in Figure 3-4:

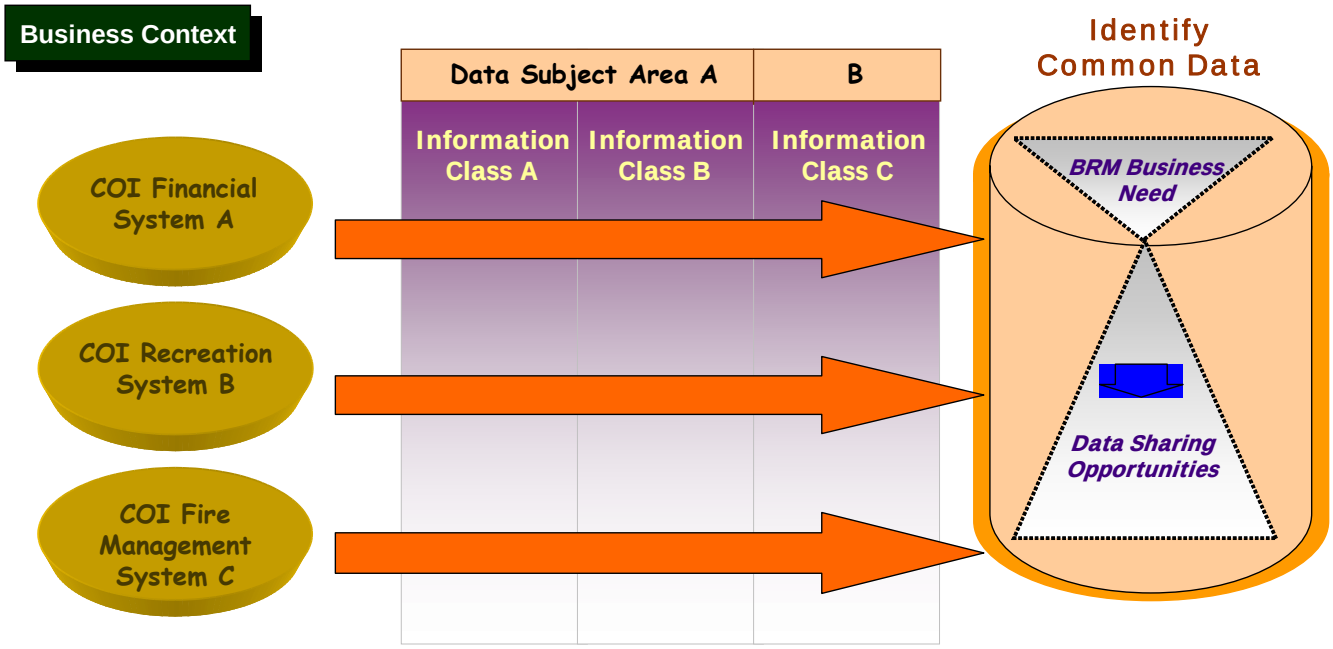


Figure 3-4 COIs Identified Data Subject Areas

Logical data models were also developed according to business context, using the FEA BRM. Figure 3-5 is an example of one such logical data model:

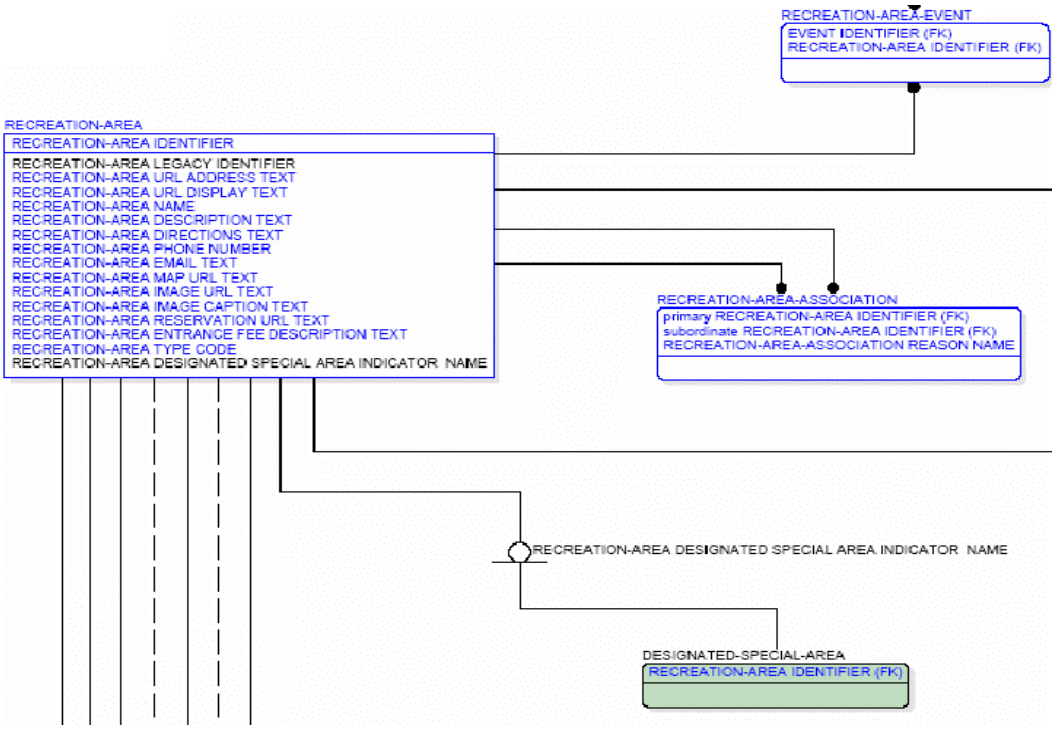


Figure 3-5 FEA BRM Logical Data Models

Figure 3-5 depicts a RECREATION-AREA entity along with various attributes (RECREATION-AREA LEGACY, RECREATION-AREA URL, etc.). Each attribute name is followed by its data type (e.g. “IDENTIFIER”, “TEXT”), and several relationships are shown. For example, the relationship between a RECREATION-AREA entity and a RECREATION-AREA-EVENT entity is depicted at the top right, with the relationship based on a mapping between a RECREATION-AREA identifier and an EVENT identifier.

4. Data Context

This chapter describes the Data Context standardization area of the DRM. The purpose of the Data Context standardization area is to enable identification and discovery of data and to provide linkages to the other FEA reference models. The content of the Data Context captured by each COI is determined by its needs and the need to support the related LoBs. Data Context is the basis for data governance. Understanding Data Context provides the means for informed government decision making with regard to its information holdings. This chapter also describes the overall structures, methods and benefits for providing data with context and gives practical examples on their use.

4.1. Chapter Organization

This chapter is organized as follows:

- **Introduction:** Provides introductory information regarding the Data Context standardization area, explaining why there is a Data Context section of the DRM abstract model and what can be done because of it;
- **Guidance:** Provides a description of how to start the process of creating Data Context artifacts and the goals of the process;
- **Data Context Section of the DRM Abstract Model:** Presents and describes the Data Context section of the DRM abstract model – documenting Data Context in compliance with this section of the DRM abstract model provides a common mechanism for communication within the COI and among COIs;
- **Data Context Example:** Provides a usage example to further explain the Data Context standardization area.

4.2. Introduction

4.2.1. What is Data Context and Why is it Important:

Data Context is any information that provides additional meaning to data to relate it to the purposes for which it was created and used. It is the information that makes it possible to provision a Context Awareness Service to support a COI or collaboration among COIs. Within a Context Awareness Service, one identifies the existence of a Data Asset and enables a user to discover whether it is potentially relevant to a given information need. The service makes Data Context artifacts, developed in accordance with the Data Context section of the DRM abstract model, available for use. These artifacts are chosen by the COI to reflect government related business needs and contain adequate information to support government related decision making. Typical examples of Data Context for a given Data Asset may include a Topic identifying a subject area, a data stewardship assignment, sources of record, etc. At a minimum, the Data Context for a given Data Asset should answer the following questions:

- What are the data (subject areas/Topics and entities of interest) contained within the Data Asset?
- What organization is responsible for maintaining the Data Asset?
- What is the linkage to the FEA BRM?
- What services are available to access the Data Asset? (See Data Sharing)

There may be more than one context for a Data Asset. Context can be considered a “lens” and one may view something through a number of different “lenses”, one for each of the different contexts in which a Data Asset may be of interest. Data Context artifacts should

be developed to reflect the understanding of the relevant Data Assets from the perspective of a COI.

To satisfy a broad, general audience, as in the case of citizen access to public information, modern search engines are an effective means of discovering and retrieving unstructured and semi-structured information. Search technology, like the popular Google™ search engine, indexes unstructured and semi-structured documents (like Web pages) and returns a result set in response to a keyword-based query. The speed of the returned results often offsets the large quantity of hits (or matches) in the result set. In summary, search effectively serves information sharing to citizens and the techniques expressed in this section effectively serve information sharing within communities of interest.

Agencies and organizations, participating in COIs, are called upon to categorize their data using taxonomies that may be defined and/or exchanged using the DRM's Data Context standardization area. Once shared in data registries, these taxonomies become vehicles for discovering data that offers value for data sharing. Additionally, data consumers can subscribe to topics published within data registries, further enhancing data discovery. Lastly, for citizen-access to semi-structured and unstructured information, enterprise search technologies should be used.

4.2.2. Purpose of the Data Context Section of the DRM Abstract Model:

The Data Context section of the DRM abstract model exists to identify the structures[15] used for Data Context artifacts.

Context often takes the form of a set of terms, i.e. words or phrases, that are themselves organized in lists, hierarchies, or trees; they may be referred to as "context items". Collectively, Data Context can be also be called "categorization" or "classification". In this case the groupings of the context items can be called "categorization schemes" or "classification schemes." More complex Data Context artifacts may also be generated, e.g. networks of classification terms or schemes.

Classification schemes can include simple lists of terms (or terms and phrases) that are arranged using some form of relationship, such as

- sets of equivalent terms,
- a hierarchy or a tree relationship structure, or
- a general network.

[15] The term "structure" is used here in the formal Computer Science sense of a data structure. Examples are networks, trees and hierarchies. The choice of a specific data structure impacts the type of relationships that can be represented.

Many classification schemes are formally created and administered by organizations or consortiums using a set of rules describing how concepts are named and designated as terms, related artifacts designed and how they can be used. In some applications of context an entity may be related to one or more terms in a classification scheme in a formal manner. In other applications, these associations are more informal, and a relationship may be only implied or exist just in the form of a co-occurrence. One example of a more complex scheme is a high-level directory that brings together important terms from multiple classification schemes; it may also show their relationships.

4.3. Guidance

4.3.1. The COI, its Participants and Processes:

Data Context artifacts are generated by members of COIs, so the first step to be taken by an agency's enterprise architect is to identify them. The Data Assets maintained or accessed by the federal government describe facts about

- the physical world,
- the world of social relationships,
- the individuals within those worlds and
- the joint interactions within those worlds over time.

Agencies, foreign governments, state and local governments, tribal governments, private sector organizations and individual persons differ as to what facts and relationships are of interest, but share some interests in common. It is therefore through the COIs that it is possible to meaningfully establish contexts in which data sharing should be supported. Therefore COIs may be organized around an LoB, a sub-part of the LoB or be cross-cutting across several LoBs. The data architect of an agency may well be aware of current COIs; over time some COIs may become less active and new ones may be created.

For the DRM the role of the members of the COI is to agree on the form and content of the context data needed to meet shared mission business needs. A COI should be able to answer basic questions about the Data Assets that it manages, such as those shown earlier in Section 4.2.1. Being able to answer these questions enables governance over all the relevant artifacts.

Having identified subject areas/Topics and the entities of interest, the COI should determine the syntax and semantics of its own data. The principle is that it should be able to explain its data to its own community first. That means obtaining agreement on relevant terms, their organization and structures before trying to explain them to others. In many areas there are external organizations that have done related work and this work should be built upon wherever possible.

The next step should be the designation of data sources within a COI. A data source is a Data Asset distinguished by the type of business rules that are used to ensure its data integrity. For any data there is a time when the data is first recorded in a government system. The government may be the original collector of the data, or it may acquire it from an external organization. As soon as the data is collected there is a quality control step associated with it. That step could be a data integrity check which might cause a data item to be rejected and therefore not retained. The step, however, might merely record the time and possibly other environmental variables surrounding the data acquisition process, or it could be more complex. The above description applies to base data, data that is recorded without further processing. Some base data is the input to other processes and the resulting data is of more interest. This data may become a separate Data Asset. Examples are

- materialized views in formatted databases,
- document versions perhaps time-stamped,
- "Level-X" datasets in a repository of scientific data.

In the end some organization or perhaps a federated set of organizations, within the COI maintains the system(s) with the data. This organization or federation becomes the data source in one of two senses. A "data source of record" is a Data Asset that satisfies the following business rule: the data contained within it is designated by the owning organization as having been generated by policy compliant business processes that ensures its integrity. A "data source of reference" is a Data Asset containing data that may replicate the data in a data source of record.

On an ongoing basis the data within an organization undergoes transformations, such as changes and deletions or versioning; the same is true for Data Context artifacts. There should be a person who can determine if this activity as well as the initial creation is being performed in accordance to policy after properly following the correct procedures. That person is a Data Steward; there is a Data Steward for both data sources of record and of reference.

For each COI once a data source and a Data Steward are identified a governance process can be put in place to ensure that there are Data Context artifacts to provision the services described in Chapter 5.

4.3.2. The Data Context Artifact Creation Activity:

Within a COI it is recognized that data categorization, or better a more strict classification through a controlled vocabulary, aids in the process of data discovery, comprehension, and data sharing. The vocabulary terms provide a perspective, significance, connotation, of the contents of the Asset and an understanding of the environment in which a Data Asset is defined and used. This consistent method of defining, using, and sharing information about a Data Asset's context improves the likelihood of data sharing and reuse across diverse and large organizations, including the government as an enterprise and all of its stakeholders.

The members of the COI recognize that a hierarchy is one of the most effective techniques for organizing the *terms* – single words and phrases – that are commonly used to identify concepts of interest. These hierarchies in general go from more general concepts to more specific ones. The terminology most often associated with these hierarchies is “taxonomy”. This term can be understood loosely in the sense above or in a stricter, more mathematical sense. In this document the looser sense is used. Similarly the word “term” will be used synonymously below with less precise word “topic”.

The COIs may be totally within the government, in which case the selection of terms is at the discretion of the program managers. Many COIs, however, interact with other non-federal government organizations or with non-governmental bodies. In such cases there are often externally developed taxonomies. Common terms that reflect such taxonomies should be used to establish a Data Context to increasing data sharing potential.

4.4. Data Context Section of the DRM Abstract Model

The Data Context section of the DRM abstract model is shown in Figure 4-1. It depicts the concepts that comprise the Data Context standardization area and the relationships between them. Concepts are expressed as boxes, while relationships are expressed as arrows.

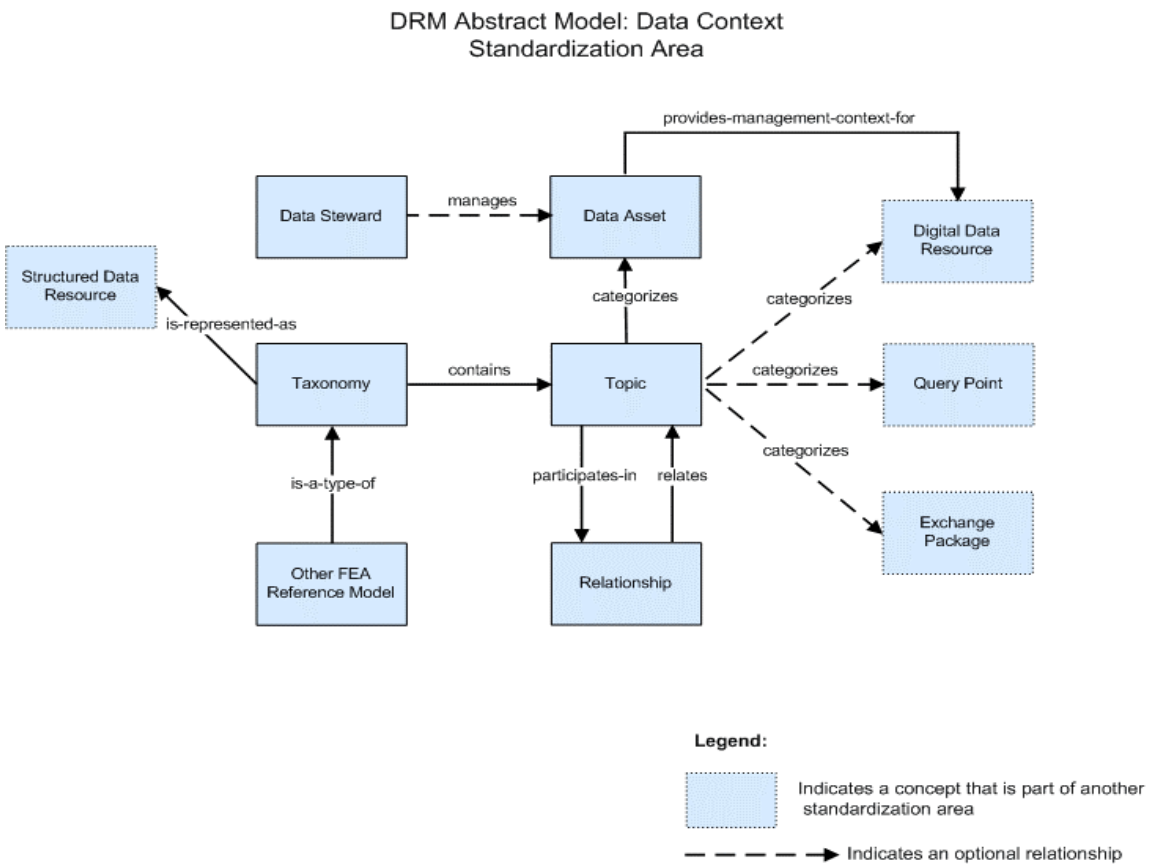


Figure 4-1 Data Context Section of the DRM Abstract Model

The following are definitions for each of the concepts and relationships within the figure shown above. Conventions used are:

- Only "outbound" relationships are listed (i.e. those that originate from the concept);
- The concepts are presented in an order that will ensure the best possible understanding, and specific examples are provided where appropriate;
- Though cardinality is not expressed in the figure, the descriptions below may include cardinality (e.g. "one or more") for purposes of clarity;
- Concept names will be capitalized as in the figure itself (e.g. "Data Asset"), while relationship names will be expressed in italics, and without any hyphens that may appear in the relationship name in the figure (e.g. *"provides management context for"*). This is done so that the definitions below can take on as narrative a tone as possible. The reader should therefore be able to easily visually navigate through the figure as they read the definitions below.
- Each concept will be referred to in a quantity of one (e.g. "A Topic *categorizes* a Data Asset") for purposes of simplicity as the figure does not depict cardinality. However, implementations based on the DRM will introduce cardinality as needed according to their requirements.
- In some cases, concepts that are part of another standardization area are included in definitions and examples below. These concepts will not be described further in this chapter; the reader should reference the pertinent chapter for definitions and examples for those concepts.

Taxonomy: A collection of controlled vocabulary terms organized into a hierarchical structure. Taxonomies provide a means for categorizing or classifying information within a reasonably well-defined associative structure, in which each term in a Taxonomy is in one or more parent/child (broader/narrower) relationships to other terms in the Taxonomy. A common example of a Taxonomy is the hierarchical structure used to classify living things within the biological sciences from Carols Linnaeus, as shown in Figure 4-2:

Category	Value for Humans
Kingdom	Animalia
Phylum	Chordata
Class	Mammalia
Order	Primates
Family	Hominidae
Genus	Homo
Species	Sapiens

Figure 4-2 Carols Linnaeus Taxonomy

- **Relationships:**
  - A Taxonomy *contains* a Topic
  - A Taxonomy *is represented as a* Structured Data Resource[16]

  Example: A taxonomy expressed in W3C Web Ontology Language (OWL) format.

Structured Data Resource: See the Data Description chapter.

Topic: A category within a Taxonomy. A Topic is the central concept for applying context to data. For example, an agency may have a Taxonomy that represents their organizational structure. In such a Taxonomy, each role in the organizational structure (e.g. CIO) represents a Topic. Topic is often synonymous with "node".

- **Relationships:**
  - A Topic *categorizes* a Data Asset
  - A Topic *may categorize* a Digital Data Resource
  - A Topic *may categorize* a Query Point
  - A Topic *may categorize* an Exchange Package
  - A Topic *participates in* a Relationship with another Topic

Digital Data Resource: See the Data Description chapter.

Query Point: See the Data Sharing chapter.

Exchange Package: See the Data Sharing chapter.

Relationship: Describes the relationship[17] between two Topics.

[16] Because a Taxonomy is represented as a Structured Data Resource, and a Data Asset provides management context for a Digital Data Resource, it follows that a Taxonomy may be stored and managed within a Data Asset.

- **Relationships:**
  - A Relationship *relates* a Topic

    Example: A "Person" Entity may be represented in one Data Asset in a "Customer" context because it is part of a CUSTOMER_INFO table. However, the same Entity may be represented in a "Suspect" context on law enforcement Web site. The metadata that is associated with the "Person" Entity would be different in each context – for example, the "Suspect" context would likely include physical characteristic metadata (height, hair color, etc.), while the "Customer" context would not.

Data Asset: A managed container for data. In many cases, this will be a relational database; however, a Data Asset may also be a Web site, a document repository, directory or data service.

- **Relationships:**
  - A Data Asset *provides management context for* a Digital Data Resource

    Example: A document that is stored and managed within a data asset (such as a document repository) has management context provided for it through the metadata that is associated with that document within the document repository. Such metadata may include the Dublin Core attributes that are described in the Data Description chapter.

Data Steward: A person responsible for managing a Data Asset.

- **Relationships:**
  - A Data Asset *may be managed by* a Data Steward

Other FEA Reference Model: This concept represents the four other FEA reference models – the Business Reference Model (BRM), the Service Component Reference Model (SRM), the Technical Reference Model (TRM), and the Performance Reference Model (PRM). Its purpose is to provide a linkage to these other reference models, which are themselves Taxonomies. These are depicted as a special kind of Taxonomy due to their importance in overall classification of information.

- **Relationships:**
  - The Other FEA Reference Models *are types of* Taxonomies

---

[17] It should be noted that the term "relationship" is used in two ways here. The concept named "Relationship" participates in relationships with other concepts in the abstract model, and also defines the relationship between topics when it is applied to a specific scenario.

4.5. Data Context Attributes

This section will expand on the concepts presented above to include attributes that are associated with each concept in the Data Context section of the DRM abstract model. A description will be provided for each attribute, along with an example where necessary for clarity.

Concept	Attribute	Description	Example
Taxonomy	Identifier[18]	A unique string associated with a Taxonomy for identification purposes.	"200XCB"
	Name	The name of a Taxonomy.	"Geographic Areas"
	Description	A description of a Taxonomy.	
Topic	Name	The name of a Topic.	"Country"
	Description	A description of a Topic.	
Relationship	Name	The name of a Relationship.	"part-of"
	Origin	Name of the concept that is the origin (i.e. the "from" concept) of a Relationship.	
	Destination	Name of the concept that is the destination (i.e. the "to" concept) of a Relationship.	
Data Asset	Identifier	A unique string associated with a Data Asset for identification purposes.	"333XBD"
	Type	Type of Data Asset – e.g. database, Web site, registry, directory, data service, etc.	"database"
	Geospatial Enabled	Designates whether or not the Data Asset supports or provides Geospatial data.	"yes"
Data Steward	Employee ID	Data Steward's employee ID.	
	Department	Department for which Data Steward works.	
	Initial Date	The date that the Data Steward became associated with the Data Asset.	
Other FEA Reference Model	Acronym	Reference model acronym.	"BRM"
	Name	Reference model name.	"Business Reference Model"

[18] The "Identifier" attribute is described at an abstract level in order to be consistent with the abstract nature of the reference model. Therefore, there are no references to aspects such as identifier uniqueness, representation format, or similar. Implementations based on the DRM will introduce such aspects as needed according to their requirements.

4.6. Data Context Example

This section provides a usage example for the Data Context standardization area. It is based on an existing implementation of the DRM at DOI, for the Recreation One Stop initiative.

One or more contexts for an entity may be conveyed by creating an association between the entity and a context item that is part of a classification scheme. For example, an exam may be given at a university for different purposes. One purpose may be to evaluate the student's ability to meet the requirements of a course, as with a midterm or final exam for a given semester. Another purpose may be that of a comprehensive exam for a graduate program, in which the exam is intended to evaluate the student's capabilities as an expert in their primary field of graduate study. In each of these cases, the "exam" entity has a different context because it is associated with a different context item – one context item relating to a semester, another relating to a graduate program. Each of these context items can be considered to be part of a classification scheme involving types of exams.

Figure 4-3 depicts examples of five different classification schemes as applied to a single entity within the DOI DRM implementation:

The entity in this example is a data entity called RECREATION-AREA. Classification scheme (1), which provides subject area and information class context, represents part of a high-level data architecture listing subject areas and information classes. Two topics (more precisely, a topic and a subtopic) from this classification scheme are shown, and a "subclass-of" relationship exists between the parent[19] topic RECREATION and the child topic RECREATION INVENTORY. This conveys that the RECREATION-AREA is part of the RECREATION INVENTORY.

Classification scheme (2), which provides organization context, represents part of an organization hierarchy for a federal department. One topic from this classification scheme is shown, and relating the RECREATION-AREA entity this topic ("National Park Service") indicates that a recreation area is used or processed by the organization known as National Park Service. This categorization capability also provides a mechanism to identity common data across organizations.

[19] In a Taxonomy given any topic a second topic is a parent topic if it is higher in the hierarchy or a child topic if it is directly lower in the hierarchy.

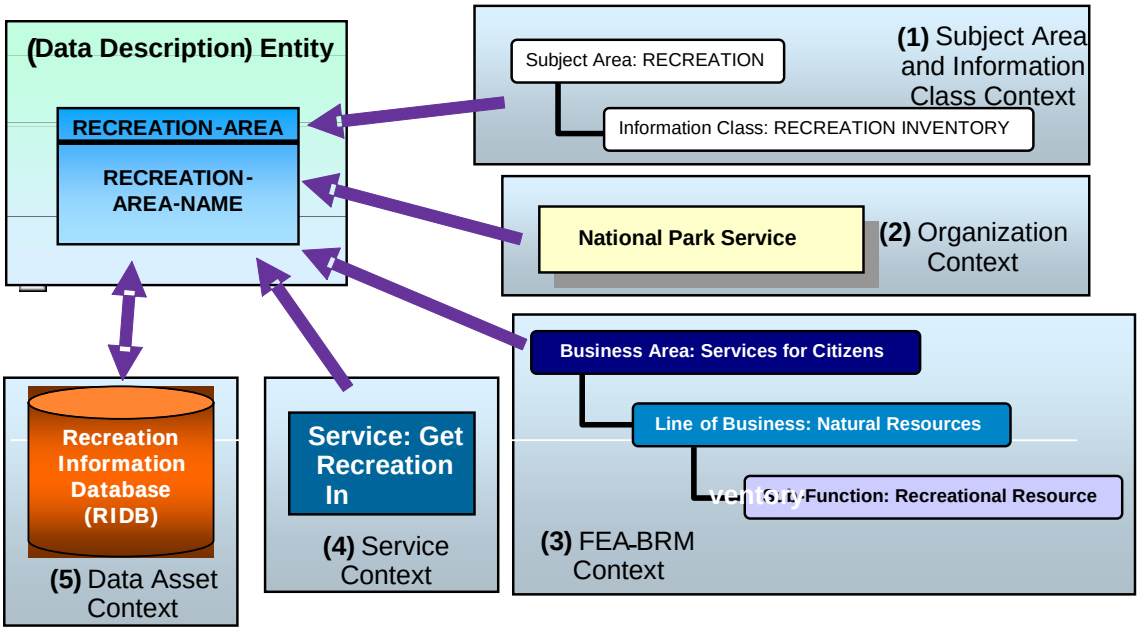


Figure 4-3 DOI DRM classification schemes

Classification scheme (3), which provided business context using the FEA BRM, represents part of the FEA BRM taxonomy. One particular sub-function topic ("Recreational Resource") is shown, along with the its parent hierarchy topics for LoB ("Natural Resources") and Business Area ("Service for Citizens"). The RECREATION-AREA entity is related to the FEA BRM sub-function of "Recreational Resource", which establishes the business context for this entity. This indicates that data about a RECREATION-AREA is typically created, updated, processed or deleted by systems that support the Recreational Resource sub-function.

Classification scheme (4), which provides service context, indicates specific services related to the processing of RECREATION-AREA data. One topic from this classification scheme is shown, which represents the specific purpose of a given service. Relating the RECREATION-AREA entity this topic ("Service: Get Recreation Inventory") indicates that the entity RECREATION-AREA is part of the information model associated with this service – that is, it is a key piece of data that is provided when this service is invoked, and indicates the exact recreation area for which an inventory of recreation assets should be obtained.

Classification scheme (5), which provides data asset context, indicates specific systems, applications, or physical data stores that process data related to RECREATION-AREAs. One topic from this classification scheme is shown, and relating the RECREATION-AREA entity to this topic ("Recreation Information Database (RIDB)") indicates that instances of RECREATION-AREA data exist as records in the Recreation Information Database (RIDB). This type of context may also describe the process method that a particular system may apply to an entity, such as creating instances of the entity, updating instances, deleting instances, or simply referencing instances.

5. Data Sharing

This chapter describes the Data Sharing standardization area of the DRM. This chapter conveys an architectural pattern for the sharing and exchange of data, with examples for its use. To guide architects in its use, a Data Supplier-To-Consumer Matrix is provided for planning the services required for data access and exchange within and between COIs to support their business/mission needs. These COIs may include international, federal, state, local and tribal governments, as well as other public and private non-governmental institutions. The Chapter also provides guidance on how architects can document these services using the DRM Abstract Model as a guide. The concepts and relationships of Data Sharing to the Data Description and Data Context standardization areas were introduced in the Overview.

5.1. Chapter Organization

This chapter is organized as follows:

- **Introduction:** Provides introductory information regarding the Data Sharing standardization area, explaining why there is a Data Sharing element to the DRM and what can be done because of it.
- **Guidance:** Provides a description of how to start the process of creating Data Services and the goals of the process.
- **Data Sharing Section of the DRM Abstract Model:** Presents and describes the Data Sharing section of the DRM abstract model. Documenting data sharing in compliance with the Data Context section of the DRM abstract model provides a common mechanism for communication within the COI and among COIs;
- **Data Sharing Example:** Provides a usage example to further explain the Data Sharing standardization area;

5.2. Introduction

5.2.1. What is Data Sharing and Why is it Important

Data sharing is the use of information by one or more consumers that is produced by another source other than the consumer. The need for data sharing often manifests itself in ways that are difficult to predict in advance. This is illustrated by a July 2005 Washington Post article entitled "Pilots Claimed Disability but Kept Flight Status". In this article, the Washington Post reported a curious correspondence between records from Social Security Administration (SSA) and Federal Aviation Administration (FAA). Forty pilots who claimed to FAA they were fit to fly were arrested in Northern California, because they had reported debilitating illnesses to SSA that should have grounded them. The data sharing between FAA and SSA that led to the discovery of criminal wrongdoing was somewhat ad hoc in this case; however, it demonstrates how the approaches to data sharing that are described in this chapter could facilitate uncovering many other correlations of interest.

Such data sharing is of importance on the local to federal level as well. On August 17, 2005, in the article entitled "L.A. Holdups Linked to Islamic Group, Possible Terrorist Plot," the Washington Post reported that a police probe of gas station holdups in Los Angeles grew into an investigation of a possible terrorist plot with connections to a radical Islamic group. The local investigation into the holdups was taken over by the FBI's Joint Terrorism Task Force when L.A. police discovered jihadist literature, bulletproof vests and a list of addresses for local synagogues, the Israeli consulate, National Guard centers and more, in the home of one of the suspects. An anonymous

U.S. official was quoted as saying there was reason to believe that terrorist attacks were planned with some of these locations as targets.

While it may have been physical evidence that led the local authorities to contact the FBI in this case, it is easy to imagine how the FBI might have decided to become involved by examining the data collected (reported) by L.A. police.

5.2.2. What is the Purpose of the Data Sharing Section of the DRM Abstract Model:

The Data Sharing section of the DRM abstract model exists to provide a reference for describing the services offered by a COI to enable access to and exchange of data.

5.2.3. Structures Used for Data Sharing:

The constructs used for data sharing are depicted in the **Data Supplier-to-Consumer Matrix.** An architect can use this matrix to ascertain which services should be provisioned to support an agency's or COI's information sharing requirements. This section also defines the principles for identifying a capability or service for sharing data. The section also identifies standards or best practices and technologies that support repeatable consistent exchange or discoverable and presented content.

- **Data Supplier-to-Consumer Matrix Overview**

Data are managed and stored in ways to optimize their use. The Data Supplier-to-Consumer Matrix is organized by the typical optimization patterns, and can be used for identifying the use of a data repository (from the perspective of a COI), the information exchange methods appropriate for these uses, and the services that should be provisioned for each use. Note that these repositories are Data Assets within the DRM abstract model. The Matrix is comprised of four quadrants, each related to the primary use of an underlying data repository. Figure 5.1 below depicts the FEA DRM Supplier-to-Consumer Matrix:

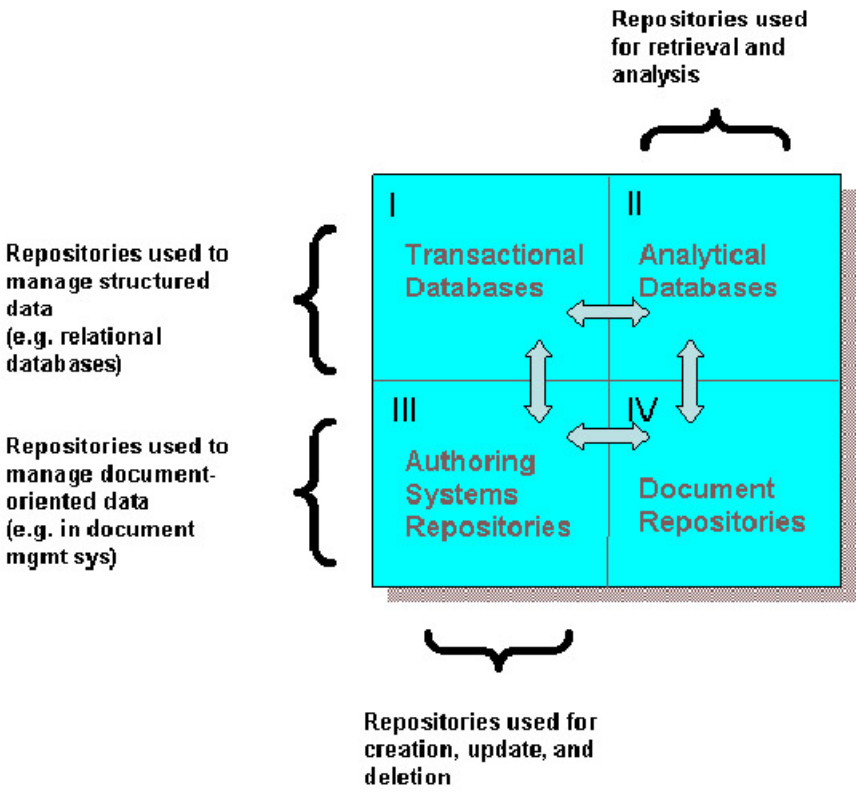


Figure 5-1 Data Supplier-to-Consumer Matrix

The four quadrants contained with the Matrix are as follows:

Quadrant I - Transactional Databases: These databases contain structured data objects that support business process and workflow. These structured databases, when well designed, tend to be highly normalized and optimized for transactional performance. Quadrant I repositories include the databases supporting On-Line Transaction Processing (OLTP) Systems, Enterprise Resource Management Systems (ERPs), and other "back-office" systems that implement core business processes and workflows. The data within these repositories tend not to be directly accessible to create, read, update, and delete (CRUD) operations, except through services usually in the form of application program interfaces (APIs) because of the need to enforce business logic and referential integrity within the database.

Quadrant II – Analytical Databases: These databases contain structured data objects that support query and analysis. These structured databases tend to be purposefully de-normalized and optimized for query ease and performance. The data in these repositories are typically obtained from one or more Quadrant I databases and structured to support answering of specific questions of business and/or mission interest. Quadrant II repositories include On-Line Analytical Processing (OLAP) systems, data warehouses, and data marts. Quadrant II also includes directories (e.g., repositories that support the Light Weight Directory Access Protocol (LDAP) or X.500). Data in these repositories tend to be directly

accessible for query and read. Create, update and delete operations are typically performed more indirectly than in transactional databases through an extract, transform, and load (ETL) process.

Quadrant III – Authoring Systems Repositories: The term "document" within the DRM context is broadly defined to encompass a wide range of information objects. These objects may be in any of a variety of formats: multimedia, text documents with embedded graphics, XML Schema or DTD instances. Generically, in this context, the term "authoring system" is equally broad in scope. At one extreme, an "authoring system" may be a digital camera. At the other, an authoring system may implement a complex workflow used for the production of a formal publication. In either extreme, the products of an authoring system are documents. The underlying repositories used by authoring systems may also be of any of a variety of constructs to store data objects, file systems and relational databases being the most common. In general, as in Quadrant I repositories, direct data-level access to the repositories underlying enterprise-level authoring systems is not prudent. Bypassing the business logic within the authoring system may affect the integrity of the data (e.g., version control of documents).

Quadrant IV – Document Repositories: Like Quadrant II repositories, document repositories store data objects so as to optimize discovery, search and retrieval. These repositories include the file systems underlying websites, relational databases underlying content management systems, XML registries and repositories. In general, as in Quadrant II repositories, data tend to be directly accessible to query. Create, update and delete operations are not generally available to end users, but are provided through a publication function performed through an authoring system.

- **Sharing Data Through Data Exchange Services**

Using this Supplier-to-Consumer Matrix, the DRM Team analyzed the types of data interchanges between repositories (data assets, database to database information sharing) focusing on the Exchange Package payload using tangible examples of such payloads. These information exchanges vary in their structure based upon the data objects being exchanged.

Based upon this analysis, the architect may specify several types of services to support the sharing of information between databases within a collection used by a COI. These services address the *data exchange* element of the DRM abstract model. These services fall within the following categories:

- **Extract, Transform, Load (Structured Data to Structured Data):** Extract, Transform, Load (ETL) is the process of reading structured data objects from a data source (the extract), changing the format of the data objects to match the structure required by a target database (transform), and updating the target

database with the transferred data objects (load). Services that perform ETL processes range from extremely simple to extremely complex. They may also be a component of other services. The payloads for all of these exchanges are structured data.

This service applies to exchanges between:

Supplier	Consumer
Transactional (I)	Transactional (I)
Transactional (I)	Analytical (II)
Transactional (I)	Authoring (III)
Analytical (II)	Transactional (I)
Analytical (II)	Analytical (II)
Analytical (II)	Authoring (III)
Authoring (III)	Transactional (I)
Authoring (III)	Analytical (II)

- **Publication: (Structured data or documents to aggregate documents):** Publication is the process of assembling a document from its component pieces, putting into a desired format and disseminating it to target databases. The payload of this type of service is a document.

  This service applies to exchanges between:

Supplier	Consumer
Transactional (I)	Document Repository (IV)
Analytical (II)	Document Repository (IV)
Authoring (III)	Authoring (III)
Authoring (III)	Document Repository (IV)

- **Entity/Relationship Extraction (Unstructured documents to structured documents or structured data objects):** Entity/Relationship Extraction is the process of identifying and pulling out specified facts from documents. Entities are nouns that designate a specific person, place or thing. Relationships are the association or affiliation of one entity to another. Typically, the entities identified during an entity/relationship extraction process may be incorporated into the source document as metadata, inserted into a separated document (such as a metadata record used to support discovery), or incorporated into a structured database. The payloads for all of these exchanges are structured data.

  This service applies to exchanges between:

Supplier	Consumer
Document Repository (IV)	Transactional (I)
Document Repository (IV)	Analytical (II)

- **Document Translation (Document to document):** Document translation is the process of transforming a document from its original format to a format required to support a target application. The transformations may be structural (e.g., transforming MS Word to PDF format), language-oriented (e.g., changing English to French), or special purpose (e.g., the development of abstracts from longer documents.) The payload of this type of service is a document.

This service applies to exchanges between:

Supplier	Consumer
Document Repository (IV)	Authoring (II)
Document Repository (IV)	Document Repository (IV)

- **Sharing Data through Data Access Services**

The discussion above focused on the transfer of data between repositories. Additional services are required to make data accessible to other services, to the applications that used them, and ultimately to the consumers of the data. The DRM Team performed a similar analysis to determine the services required to implement data access. The architect should ascertain the services that are required to support the COI in the use of its collection. These services address the *data exchange* element of the DRM abstract model.

The services that the architect may be required to provision to support a COI's information sharing requirements are delineated below.

- **Context Awareness Services:** A context awareness service allows the users of a collection to rapidly identify the context (as defined above) of the data assets managed by the COI. Context information may be captured in a formalized data architecture, a metadata registry or a separate database.

  The architect should plan for this service for ***all*** quadrants.

- **Structural Awareness Services**: A structural awareness services allows data architects and database administrators to rapidly identify the structure of data within a data asset (i.e., a structural awareness services makes the Data Description as defined within the DRM available for use). Data Description information may be captured in a formalized data architecture, a metadata registry, or a separate database. Also, a number of commercial products are available to analyze and report data structures.

  Again, the architect should plan for this service for ***all*** quadrants.

- **Transactional Services:** A transactional services enables a transactional create, update or delete operations to an underlying data store while maintaining business and referential integrity rules. These services allow external services or end users to execute data related functions as a part of a

workflow or business process. Most commercial products provide application programming interfaces that implement this type of service.

The architect should plan to provision these services for the transactional and document authoring quadrants.

- **Data Query Services:** A data query services enables a user, service or application to directly query a repository within a collection.

  The architect should plan to provision these services for the transactional and analytical quadrants.

- **Content Search and Discovery Services**: A search and discovery service enables free text search or search of metadata contained within the documents in a repository. The searchable metadata should include the Data Context as defined within the DRM abstract model.

  The architect should plan to provision these service for the authoring and document repository quadrants.

- **Retrieval Services:** A retrieval services enables an application to request return of a specific document from a repository based upon a unique identifier, such as a URL.

  The architect should plan to provision these services for the authoring and document repository quadrants.

- **Subscription Services:** A subscription service enables another service or an end user to nominate themselves to automatically receive new documents added to a repository in accordance with a predetermined policy or profile.

  The architect should plan to provision these services for the authoring and document repository quadrants.

- **Notification Services:** A notification service automatically alerts another service or an end user of changes of the content of a repository in accordance with a predetermined policy or profile.

  The architect should plan to provision these services for the transactional, authoring and document repository quadrants.

5.3. Guidance

As stated in the Overview, once the Data Context and the Data Description standardization areas for a COI have been defined, the COI should then plan and implement common capabilities to enable information to be accessed and exchanged.

Enterprise architects and data architects supporting COIs *may* use the following table as a guide to plan implementation of data exchange and access services for each data asset under management within the COI.

Supplier	Service Requirement	Consumer
Data Asset Identifier (see abstract model below)	Data Exchange Service Type (e.g. Extract, Transform, Load)	Data Asset Identifier
… (populate as needed)	…	…
Data Asset Identifier	Access Service Type (e.g., Context Awareness)	*Access Services typically support many consumers. Generally, there is no need to populate these cells.*
… (populate as needed)	…	…

Table 5.1 – Data Sharing Service Requirements Matrix

Once the architect has populated this matrix, he or she has a clear understanding of the types of data access and exchange services that they should provision to support a COIs information sharing requirements. Once these requirements are captured, the architect may use the Data Sharing section of the DRM abstract model to fully document the data access and the exchange services required to support the COI.

5.4. Data Sharing Section of the DRM Abstract Model

The Data Sharing section of the DRM abstract model covers two primary aspects of data sharing:

- **Data Exchange:** Fixed, recurring transactions between parties, such as the regular exchange of environmental testing data among federal, state, local, and tribal entities. These exchanges, as describe above, are implemented with data exchange services;
- **Data Access:** Requests for data services, such as a query of a Data Asset[20]. These requests, as described above, are supported by Data Access Services;

The Data Sharing standardization area is supported by the Data Description and Data Context standardization areas in the following ways:

- **Data Description:** Uniform definition of Exchange Packages and Query Points supports the capability to effectively share them within and between COIs;
- **Data Context:** Categorization of Exchange Packages and Query Points supports their discovery, and their subsequent use in data access and data exchange.

Detailed information about these aspects are defined within the DRM. As described above, the architect may use the Data Sharing section of the DRM abstract model as a means to organize and share information about the information sharing within the agency/COI that he or she supports.

The Data Sharing section of the DRM abstract model is shown in Figure 5-2. As in all previous chapters, it depicts the concepts that comprise relevant to the chapter. In this instance the model depicts the Data Sharing concepts for the DRM and the relationships between them. Concepts are expressed as boxes, while relationships are expressed as arrows.

[20] The term “data asset” is synonymous with “data source”. It is described within the Data Context chapter.

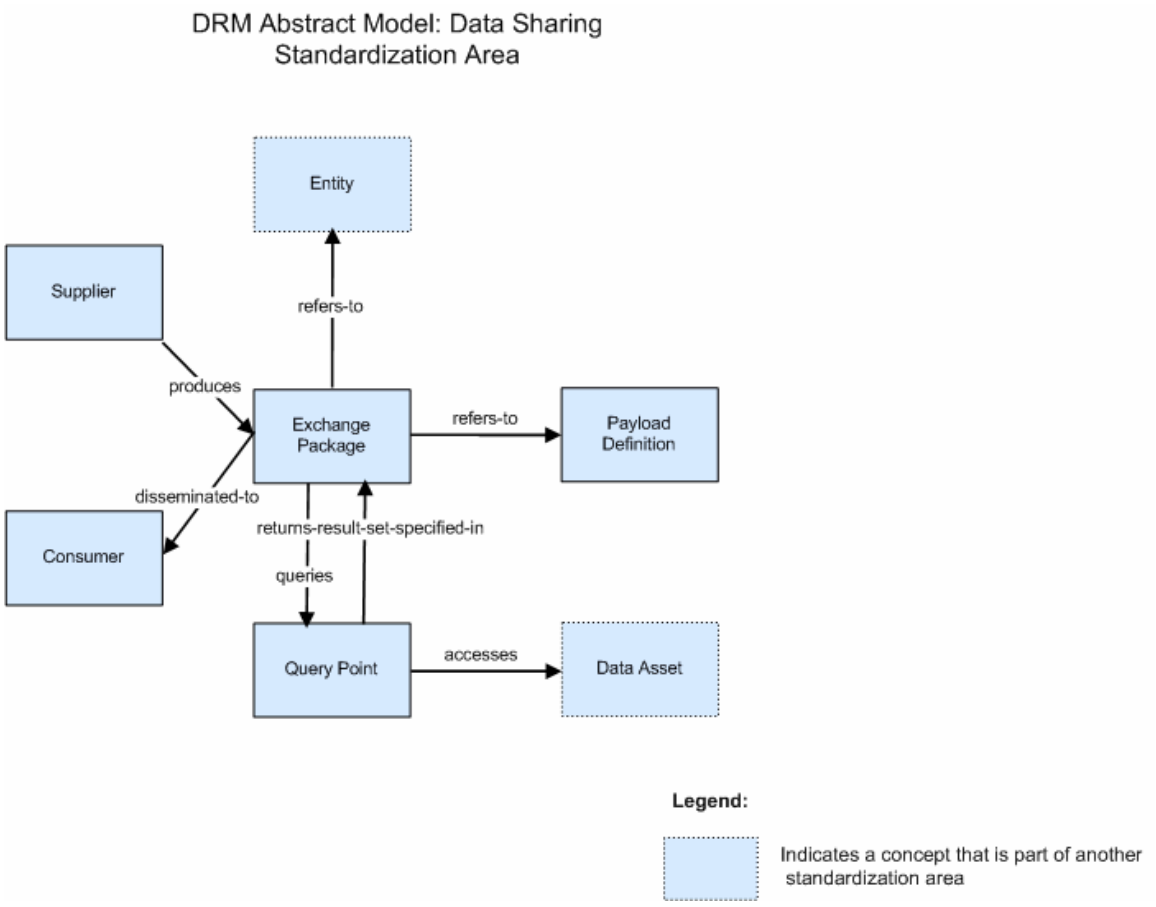


Figure 5-2 Data Sharing Section of the DRM Abstract Model

The following are definitions for each of the concepts and relationships within the figure shown above. Conventions used are:

- Only "outbound" relationships are listed (i.e. those that originate from the concept);
- The concepts are presented in an order that will ensure the best possible understanding, and specific examples are provided where appropriate;
- Though cardinality is not expressed in the figure, the descriptions below may include cardinality (e.g. "one or more") for purposes of clarity;
- Concept names will be capitalized as in the figure itself (e.g. "Exchange Package"), while relationship names will be expressed in italics, and without any hyphens that may appear in the relationship name in the figure (e.g. "*refers to*"). This is done so that the definitions below can take on as narrative a tone as possible. The reader should therefore be able to easily visually navigate through the figure as they read the definitions below.
- Each concept will be referred to in a quantity of one (e.g. "An Exchange Package *refers to* an Entity") for purposes of simplicity as the figure does not depict

cardinality. However, implementations based on the DRM will introduce cardinality as needed according to their requirements.

- In some cases, concepts that are part of another standardization area are included in definitions and examples below. These concepts will not be described further in this chapter; the reader should reference the pertinent chapter for definitions and examples for those concepts.

Exchange Package: A description of a specific recurring data exchange between a Supplier and a Consumer. An Exchange Package contains information (metadata) relating to the exchange (such as Supplier ID, Consumer ID, validity period for data, etc.), as well as a reference to the Payload (message content) for the exchange. An Exchange Package can also be used to define the result format for a query that is accepted and processed by a Query Point in a data sharing scenario.

Relationships:

- An Exchange Package *refers to* an Entity
- An Exchange Package *is disseminated to* a Consumer
- An Exchange Package *queries* a Query Point
- An Exchange Package *refers to* a Payload Definition

Example: An Exchange Package describes a specific recurring data exchange involving shipment information.

Entity: See the Data Description chapter.

Supplier: An entity (person or organization) that supplies data to a Consumer.

- **Relationships:**
  - A Supplier *produces* an Exchange Package

  Example: A federal agency that supplies data to one or more other federal agencies.

Consumer: An entity (person or organization) that consumes data that is supplied by a Supplier.

- **Relationships:**
  - None

  Example: A federal agency that consumes data from one or more other federal agencies.

Payload Definition: An electronic definition that defines the requirements for the Payload (data) that is exchanged between a Supplier and a Consumer.

- **Relationships:**
  - None

Example: A specific message set expressed as an XML schema or an EDI transaction set that contains information about a "Person" entity.

Query Point: An endpoint that provides an interface for accessing and querying a Data Asset. A concrete representation of a Query Point may be a specific URL at which a query Web Service may be invoked.

- **Relationships:**
  - A Query Point *accesses* a Data Asset

    Example: A specific URL at which a data service may be invoked.
  - A Query Point *returns a result set specified in* an Exchange Package

Data Asset: See the Data Context chapter.

5.5. Data Sharing Attributes

This section will expand on the concepts presented above to include attributes that are associated with each concept. A description will be provided for each attribute, along with an example where necessary for clarity.

Concept	Attribute	Description	Example
Exchange Package	Identifier[21]	A unique string associated with an Exchange Package for identification purposes.	"200XCB"
	Name	The name of an Exchange Package.	"Bill of Lading Message Set"
	Description	A description of an Exchange Package.	
	Classification	The security classification for an Exchange Package.	"U" (Unclassified)
	Frequency	The frequency at which the exchange occurs.	"Daily"
Supplier	Identifier	A unique string associated with a Supplier for identification purposes.	"04091967J"
	Name	The name of a Supplier.	
	Primary	The name and contact	

[21] The "Identifier" attribute is described at an abstract level in order to be consistent with the abstract nature of the reference model. Therefore, there are no references to aspects such as identifier uniqueness, representation format, or similar. Implementations based on the DRM will introduce such aspects as needed according to their requirements.

Concept	Attribute	Description	Example
	Contact	information for the Supplier's primary contact for this particular exchange.	
Consumer	Identifier	A unique string associated with a Consumer for identification purposes.	"03081956K"
	Name	The name of a Consumer.	
	Primary Contact	The name and contact information for the Consumer's primary contact for this particular exchange.	
Payload Definition	Identifier	A unique string associated with a Payload Definition for identification purposes.	"B5102078L"
	Name	The name of a Payload Definition.	"Bill of Lading XML Schema"
Query Point	Identifier[22]	A unique string associated with a Query Point for identification purposes.	http://www.example.com/querypoint3
	Name	The name of a Query Point.	"Latest Monthly Report Information"
	Description	A description of a Query Point.	
	Query Languages	A stipulation of the query languages that are supported by a Query Point (e.g. SQL-92, CQL (Z39.50), XQuery, HTTP GET, etc.).	"SQL-92"

[22] For a Query Point, an identifier represents the electronic address at which the Query Point may be accessed.

5.6. Data Sharing Example

This section provides a usage example for the Data Sharing standardization area using a data sharing scenario. It is based on an existing implementation of the DRM at DOI, for the Recreation One Stop initiative. This example is based upon the use of XML, but the architect should not construe this as guidance. This example represents one way to successfully implement the DRM abstract model to support a COI. There are many.

This example references the DOI Recreation Information Database (RIDB), which is available at http://RecData.gov/RIDBWeb/index.jsp. At this URL, there is a menu option titled "RIDB Data Sharing" which is the RIDB online interface for data sharing. Selection of this menu option results in the presentation to the user of what is commonly known as a "picklist", or set of choices, that enable the user to select one or more organizations[23].

Selection of the organization "Fish and Wildlife Service" results in the following three choices:

View **All** RecElements: http://64.241.25.115/RIDBWebService/RIDBService.jws/**getAllRecElements**ForOrgID?anOrgID=127
View **RecArea-related** RecElements: http://64.241.25.115/RIDBWebService/RIDBService.jws/**getAllRecAreaElements**ForOrgID?anOrgID=127
View **Facility-related** RecElements: http://64.241.25.115/RIDBWebService/RIDBService.jws/**getAllFacilityElements**ForOrgID?anOrgID=127

Each of the above URLs represents a different query on the RIDB data asset, with each containing a different "get" operation (i.e. "getAllRecElementsForOrgID", "getAllRecAreaElementsForOrgID", "getAllFacilityElementsForOrgID"). Each of these URLs represents a query point. Each query point has in common a single Java Web Service ("RIDBService.jws") that implements each of the above operations; this Web Service itself may also be considered a "composite" query point (i.e. one that contains several query points). Each query point returns recreation data about Fish and Wildlife

[23] Although this menu option does not present what would normally be considered an "ultimate result set" of data (i.e. the user is still in the process of formulating the query via the interface), the URL that displays this menu option (http://www.recdata.gov/RIDBWeb/goToWebServiceView.do) may still be considered an Query Point (more specifically, an "intermediate" Query Point). This is because the URL that displays this menu option may indeed query a data asset to obtain the list of organizations it displays, depending on how the processing was designed.

Service, but the quantity and structure of the data varies depending upon which query point was selected. In each case, the data returned (the result set) is an XML document that conforms to an Exchange Package defining the result format. The Exchange Package payload is expressed as a RecML (Recreation Markup Language) XML schema.

6. DRM Abstract Model

6.1. Introduction

This chapter presents the DRM abstract model. The DRM abstract model depicts the major concepts from each standardization area and the relationships between them. It represents an architectural pattern that represents the minimal level of detail necessary to convey the major concepts for the standardization area, with COIs extending the architectural pattern as necessary for their implementations.

The DRM abstract model is shown in figure 6-1. The 3 DRM standardization areas are arranged in the same general configuration as in figure 6-1 (i.e. with the Data Sharing standardization area on top, supported by the Data Description and Data Context standardization areas). The following conventions are used:

- Concepts are expressed as boxes, while relationships are expressed as arrows;
- Only "outbound" relationships are listed (i.e. those that originate from the concept);
- Links between standardization areas are marked as indicated in the legend, and each is provided a circled number intended for reference in the explanation of that link in the table below;
- Cardinality (e.g. 1-to-many, many-to-many) is not depicted in the abstract model. Implementations based on the DRM will introduce cardinality as needed according to their requirements.

Subsequent sections in this chapter will "drill down" into the details of this abstract model and describe the concepts that comprise each standardization area's section of the DRM abstract model.

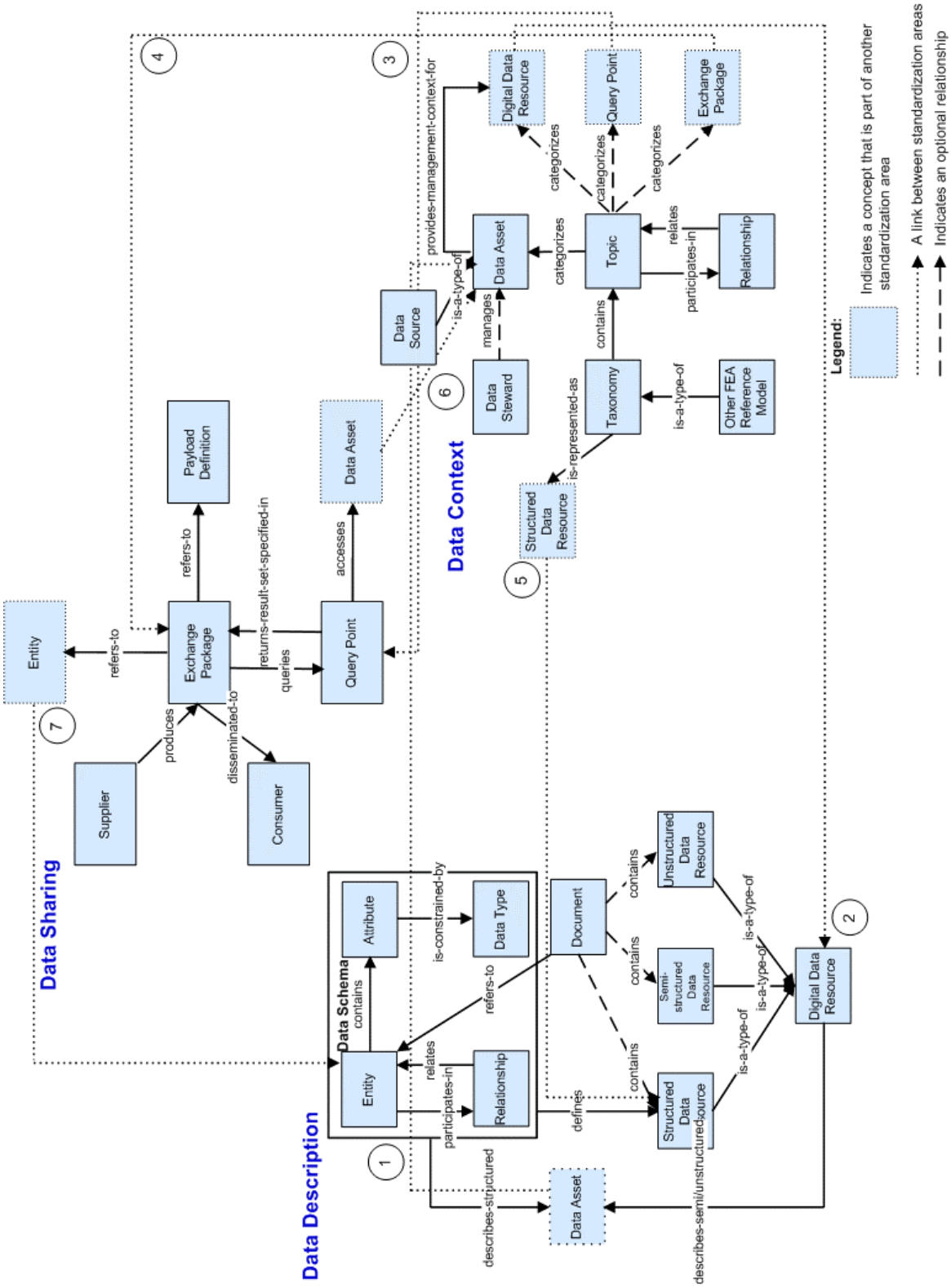


Figure 6-1 DRM Abstract Model

The table below summarizes each of the links between each standardization area's section of the DRM abstract model. It is intended to help the reader navigate through the DRM abstract model. The following information is provided for each link:

- **Link Number:** The circled number from the DRM abstract model that represents the link;
- **From Standardization Area:** The standardization area that is the source of the link;
- **To Standardization Area:** The standardization area that is the destination of the link;
- **Concept:** The concept that is common (i.e. linked) between the two standardization areas;

Link Number	From Standardization Area	To Standardization Area	Concept
1	Data Description	Data Context	Data Asset
2	Data Context	Data Description	Digital Data Resource
3	Data Context	Data Description	Query Point
4	Data Context	Data Sharing	Exchange Package
5	Data Context	Data Description	Structured Data Resource
6	Data Sharing	Data Context	Data Asset
7	Data Sharing	Data Description	Entity

The concepts that comprise each standardization area will now be described in detail.

6.2. Data Description Section of the DRM Abstract Model

The Data Description section of the DRM abstract model is shown in Figure 6-2. Links between this standardization and other standardization areas are indicated by the link number that corresponds to the link number in the DRM abstract model.

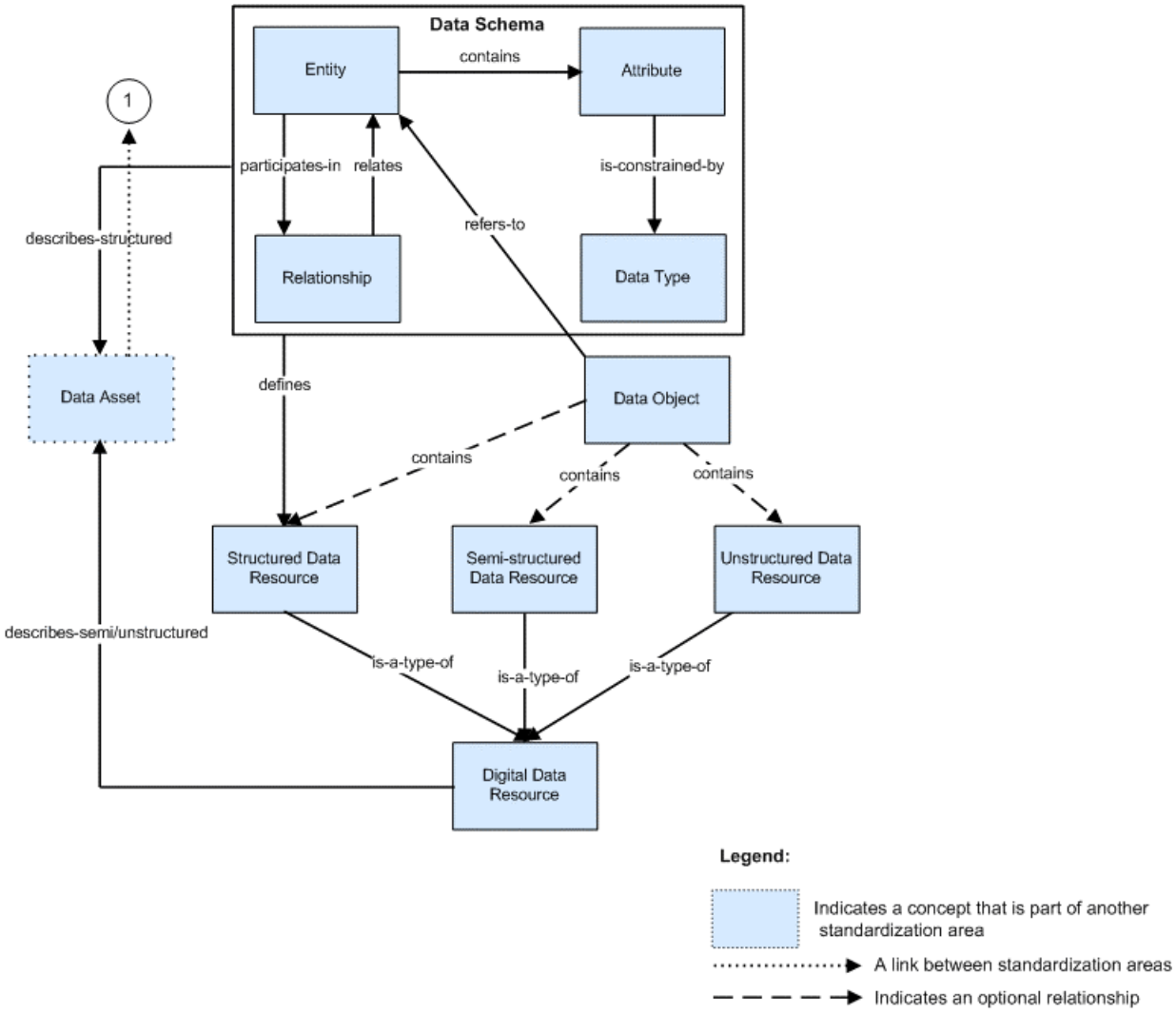


Figure 6-2 DRM Data Description Abstract Model

The following are definitions for each of the concepts and relationships shown above. The following conventions are used:

- The concepts are presented in an order that will ensure the best possible understanding, and specific examples are provided where appropriate;
- Though cardinality is not expressed in the figure, the descriptions below may include cardinality (e.g. “one or more”) for purposes of clarity. Each concept will be referred to in a quantity of one (e.g. “An Entity *contains* an Attribute”) for purposes of simplicity.
- Concept names will be capitalized as in the figure itself (e.g. “Digital Data Resource”), while relationship names will be expressed in italics, and without any hyphens that may appear in the relationship name in the figure (e.g. “*is constrained by”*). This is done so that the definitions below can take on as narrative a tone as possible. The reader should therefore be able to easily visually navigate through the figure as they read the definitions below.

Data Schema: A representation of metadata, often in the form of data artifacts such as logical data models or conceptual data models. The Data Schema concept is actually a *concept group,* which is an aggregation of related concepts. The *Data Schema* concept group is comprised of those concepts pertaining to the representation of structured data.

- **Relationships:**
  - A Data Schema *defines* a Structured Data Resource
  - A Data Schema *describes a Structured* Data Asset

Entity: An abstraction for a person, place, object, event, or concept described (or characterized) by common Attributes. For example, “Person” and “Agency” are Entities. An *instance* of an Entity represents one particular occurrence of the Entity, such as a specific person or a specific agency.

- **Relationships:**
  - An Entity *contains* an Attribute
  - An Entity *participates in* a Relationship with another Entity

Data Type: A constraint on the type of data that an instance of an Attribute may hold (e.g. "string" or "integer").

- **Relationships:**
  - None

Attribute: A characteristic of an Entity whose value may be used to help distinguish one instance of an Entity from other instances of the same Entity. For example, an Attribute of a “Person” Entity may be “Social Security Number (SSN)”. An SSN is used to distinguish one person (i.e. one instance of a “Person” Entity) from another.

- **Relationships:**
  - An Attribute *is constrained by* a Data Type

    Example: The "SSN" Attribute of a "Person" Entity may have a Data Type of "string" (if hyphens are included with the SSN) or "integer" (if hyphens are not included).

Relationship: Describes the relationship[24] between two Entities.

- **Relationships:**
  - A Relationship *relates* an Entity

    Example: A "Person" Entity may have a Relationship with an "Agency" Entity of "works for".

Digital Data Resource: A digital container of information, typically known as a file. A Digital Data Resource may be one of three specific types of data resources, each corresponding to one of the three types of data described earlier, and each described below (see "Structured Data Resource", "Semi-Structured Data Resource", and "Unstructured Data Resource").

- **Relationships:**
  - A Digital Data Resource *describes a Semi-structured* Data Asset
  - A Digital Data Resource *describes an Unstructured* Data Asset

Structured Data Resource: A Digital Data Resource containing structured data.

- **Relationships:**
  - A Structured Data Resource *is a type of* Digital Data Resource

Semi-Structured Data Resource: A Digital Data Resource containing semi-structured data.

- **Relationships:**
  - A Semi-Structured Data Resource *is a type of* Digital Data Resource

Unstructured Data Resource: A Digital Data Resource containing unstructured data.

- **Relationships:**
  - An Unstructured Data Resource *is a type of* Digital Data Resource

Document: A file containing Unstructured and/or Semi-Structured Data Resources.

---

[24] It should be noted that the term "relationship" is used in two ways here. The concept named "Relationship" participates in relationships with other concepts in the abstract model, and also defines the relationship between entities when it is applied to a specific scenario.

- **Relationships:**
  - A Document *may contain* an Unstructured or Semi-Structured Data Resource
  - A Document *refers to* an Entity

  Example (relationship with Entity): A query that states "Find all Documents in which the following person is referenced".

  NOTE: While a Document can contain structured data, it normally has explanatory material included, which would cause it to therefore be considered semi-structured. It is for this reason that there is no *"contains"* relationship from Document to Structured Data Resource. It is very important to separate Documents from Structured Data Resources because they are processed very differently. The difference between a Document and a Digital Data Resource, therefore, is that a Digital Data Resource can contained structured data.

The table below provides attributes[25] that are associated with each concept in the Data Description section of the DRM abstract model. A description will be given for each attribute, along with an example where necessary for clarity. All Unstructured Data Resource attributes and their descriptions are taken from the Dublin Core Metadata Initiative (DCMI), Version 1.1, available at http://dublincore.org/documents/dcmi-terms/. All references to "resource" within descriptions of Unstructured Data Resource should therefore be interpreted as "Unstructured Data Resource". The above URL provides additional information on attribute descriptions and usage.

Concept	Attribute	Description	Example
Entity	Identifier[26]	A unique string associated with an Entity for identification purposes.	"200XCB"
	Name	The name of an Entity.	"Person"
	Description	A description of an Entity.	
Data Type	Name	The name of a Data Type.	"string"
	Description	A description of a Data Type.	
Attribute	Name	The name of an Attribute.	"Date Of Birth"
	Description	A description of an Attribute.	
Relationship	Name	The name of a Relationship.	"works-for"

[25] It should be noted that the term "attribute" is used here in a different way than for the concept named "Attribute". Here, an "attribute" is used to describe characteristics of each of the concepts in the abstract model.
[26] The "Identifier" attribute is described at an abstract level in order to be consistent with the abstract nature of the reference model. Therefore, there are no references to aspects such as identifier uniqueness, representation format, or similar. Implementations based on the DRM will introduce such aspects as needed according to their requirements.

Concept	Attribute	Description	Example
	Origin	Name of the concept that is the origin (i.e. the "from" concept) of a Relationship.	
	Destination	Name of the concept that is the destination (i.e. the "to" concept) of a Relationship.	
Digital Data Resource	See "Structured Data Resource", "Semi-Structured Data Resource", and "Unstructured Data Resource"[27]		
Structured Data Resource	See all concepts within "Data Schema" group		
Semi-Structured Data Resource	See "Structured Data Resource" and "Unstructured Data Resource"		
Unstructured Data Resource[28]	Title	A name given to the resource.	"Information Exchange Report – July 2005"
	Resource Identifier	An unambiguous reference to the resource within a given context.	"200XCB"
	Date	A date of an event in the lifecycle of the resource. Will typically be associated with the creation or availability of the resource.	
	Creator	An entity[29] primarily responsible for making the content of the resource.	
	Format	The physical or digital manifestation of the resource. Typically, format may include the media-type or dimensions of the resource.	"text/plain"
	Description	An account of the content of the resource.	
	Source	A reference to a resource from which the present resource is derived. Recommended best practice is to reference the resource by means of a string or number conforming to a formal identification system.	"300YDC"
	Subject	A topic of the content of the resource.	
	Resource Type	The nature or genre of the content of the resource.	"Service"
	Publisher	An entity responsible for making the	

[27] As shown in the abstract model, a Digital Data Resource may be one of these three specific types of data resources. The same general idea applies to the entries for the "Semi-Structured Data Resource" and "Data Object" concepts above.

[29] It should be noted that the term "entity" here, and in subsequent Dublin Core attributes, does not have the same exact meaning as the "Entity" concept of the Data Description abstract model.

Concept	Attribute	Description	Example
		resource available.	
	Contributor	An entity responsible for making contributions to the content of the resource.	
	Language	A language of the intellectual content of the resource.	“eng”
	Relation	A reference to a related resource.	“400ZED”
	Coverage	The extent or scope of the content of the resource.	“Chicago”
	Rights Management	Information about rights held in and over the resource.	“Public domain”
Document	See “Structured Data Resource” and “Semi-Structured Data Resource”		

6.3. Data Context Section of the DRM Abstract Model

The Data Context section of the DRM abstract model is shown in Figure 6-3 below.

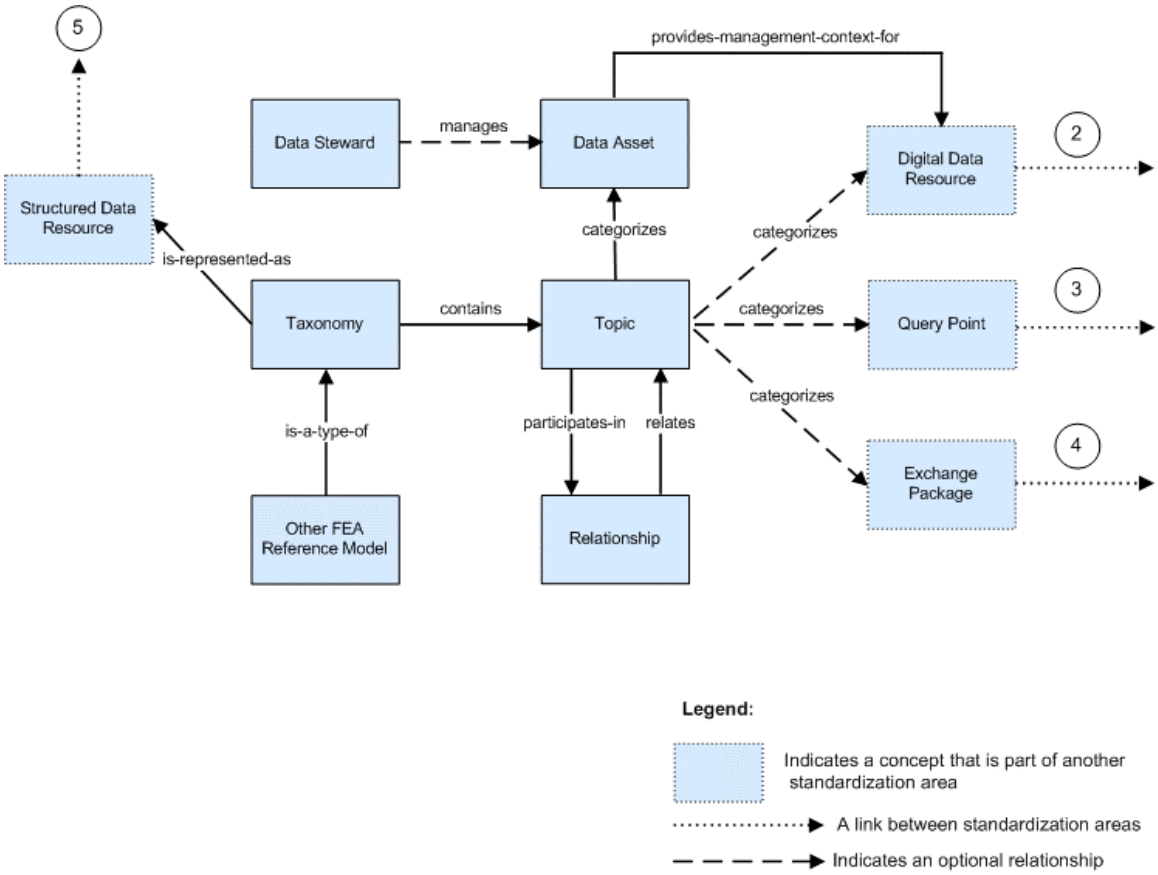


Figure 6-3 Data Context Section of the DRM Abstract Model

The following are definitions for each of the concepts and relationships shown above.

Taxonomy: A collection of controlled vocabulary terms organized into a hierarchical structure. Taxonomies provide a means for categorizing or classifying information within a reasonably well-defined associative structure, in which each term in a Taxonomy is in one or more parent/child (broader/narrower) relationships to other terms in the Taxonomy. A common example of a Taxonomy is the hierarchical structure used to classify living things within the biological sciences from Carols Linnaeus, as shown the table below.

Category	Value for Humans
Kingdom	Animalia
Phylum	Chordata
Class	Mammalia
Order	Primates
Family	Hominidae
Genus	Homo
Species	Sapiens

Figure 6-4 Carols Linnaeus Taxonomy

- **Relationships:**
  - A Taxonomy *contains* a Topic
  - A Taxonomy *is represented as a* Structured Data Resource[30]

  Example: A taxonomy expressed in W3C Web Ontology Language (OWL) format.

Structured Data Resource: See the Data Description chapter.

Topic: A category within a Taxonomy. A Topic is the central concept for applying context to data. For example, an agency may have a Taxonomy that represents their organizational structure. In such a Taxonomy, each role in the organizational structure (e.g. CIO) represents a Topic. Topic is often synonymous with "node".

- **Relationships:**
  - A Topic *categorizes* a Data Asset
  - A Topic *may categorize* a Digital Data Resource

[30] Because a Taxonomy is represented as a Structured Data Resource, and a Data Asset provides management context for a Digital Data Resource, it follows that a Taxonomy may be stored and managed within a Data Asset.

- A Topic *may categorize* a Query Point
- A Topic *may categorize* an Exchange Package
- A Topic *participates in* a Relationship with another Topic

Digital Data Resource: See the Data Description chapter.

Query Point: See the Data Sharing chapter.

Exchange Package: See the Data Sharing chapter.

Relationship: Describes the relationship[31] between two Topics.

- **Relationships:**
  - A Relationship *relates* a Topic

    Example: A "Person" Entity may be represented in one Data Asset in a "Customer" context because it is part of a CUSTOMER_INFO table. However, the same Entity may be represented in a "Suspect" context on law enforcement Web site. The metadata that is associated with the "Person" Entity would be different in each context – for example, the "Suspect" context would likely include physical characteristic metadata (height, hair color, etc.), while the "Customer" context would not.

Data Asset: A managed container for data. In many cases, this will be a relational database; however, a Data Asset may also be a Web site, a document repository, directory or data service.

- **Relationships:**
  - A Data Asset *provides management context for* a Digital Data Resource

    Example: A document that is stored and managed within a data asset (such as a document repository) has management context provided for it through the metadata that is associated with that document within the document repository. Such metadata may include the Dublin Core attributes that are described in the Data Description chapter.

Data Steward: A person responsible for managing a Data Asset.

- **Relationships:**
  - A Data Asset *may be managed by* a Data Steward

[31] It should be noted that the term "relationship" is used in two ways here. The concept named "Relationship" participates in relationships with other concepts in the abstract model, and also defines the relationship between topics when it is applied to a specific scenario.

Other FEA Reference Model: This concept represents the four other FEA reference models – the Business Reference Model (BRM), the Service Component Reference Model (SRM), the Technical Reference Model (TRM), and the Performance Reference Model (PRM). Its purpose is to provide a linkage to these other reference models, which are themselves Taxonomies. These are depicted as a special kind of Taxonomy due to their importance in overall classification of information.

- **Relationships:**
  - The Other FEA Reference Models *are types of* Taxonomies

The table below provides attributes that are associated with each concept in the Data Context section of the DRM abstract model.

Concept	Attribute	Description	Example
Taxonomy	Identifier[32]	A unique string associated with a Taxonomy for identification purposes.	"200XCB"
	Name	The name of a Taxonomy.	"Geographic Areas"
	Description	A description of a Taxonomy.	
Topic	Name	The name of a Topic.	"Country"
	Description	A description of a Topic.	
Relationship	Name	The name of a Relationship.	"part-of"
	Origin	Name of the concept that is the origin (i.e. the "from" concept) of a Relationship.	
	Destination	Name of the concept that is the destination (i.e. the "to" concept) of a Relationship.	
Data Asset	Identifier	A unique string associated with a Data Asset for identification purposes.	"333XBD"
	Type	Type of Data Asset – e.g. database, Web site, registry, directory, data service, etc.	"database"
	Geospatial Enabled	Designates whether or not the Data Asset supports or provides Geospatial data.	"yes"
Data Steward	Employee ID	Data Steward's employee ID.	
	Department	Department for which Data Steward works.	
	Initial Date	The date that the Data Steward became associated with the Data Asset.	
Other FEA	Acronym	Reference model acronym.	"BRM"

[32] The "Identifier" attribute is described at an abstract level in order to be consistent with the abstract nature of the reference model. Therefore, there are no references to aspects such as identifier uniqueness, representation format, or similar. Implementations based on the DRM will introduce such aspects as needed according to their requirements.

Concept	Attribute	Description	Example
Reference Model			
	Name	Reference model name.	“Business Reference Model”

6.4. Data Sharing Section of the DRM Abstract Model

The Data Sharing section of the DRM abstract model is shown in Figure 6-5.

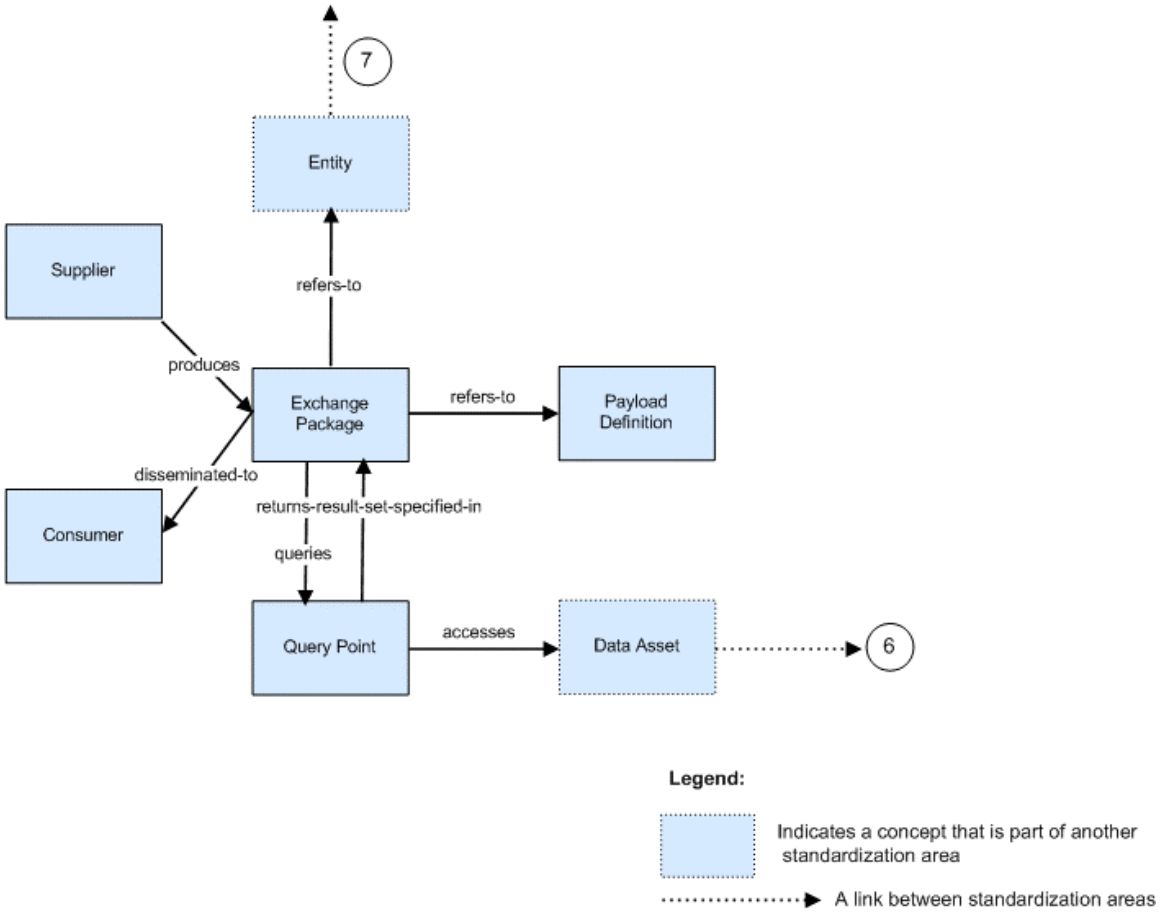


Figure 6-5 Data Sharing Section of the DRM Abstract Model

The following are definitions for each of the concepts and relationships shown above.

Exchange Package: A description of a specific recurring data exchange between a Supplier and a Consumer. An Exchange Package contains information (metadata) relating to the exchange (such as Supplier ID, Consumer ID, validity period for data, etc.), as well as a reference to the Payload (message content) for the exchange. An

Exchange Package can also be used to define the result format for a query that is accepted and processed by a Query Point in a data sharing scenario.

Relationships:

- An Exchange Package *refers to* an Entity
- An Exchange Package *is disseminated to* a Consumer
- An Exchange Package *queries* a Query Point
- An Exchange Package *refers to* a Payload Definition

Example: An Exchange Package the describes a specific recurring data exchange involving shipment information.

Entity: See the Data Description chapter.

Supplier: An entity (person or organization) that supplies data to a Consumer.

- **Relationships:**
  - A Supplier *produces* an Exchange Package

  Example: A federal agency that supplies data to one or more other federal agencies.

Consumer: An entity (person or organization) that consumes data that is supplied by a Supplier.

- **Relationships:**
  - None

  Example: A federal agency that consumes data from one or more other federal agencies.

Payload Definition: An electronic definition that defines the requirements for the Payload (data) that is exchanged between a Supplier and a Consumer.

- **Relationships:**
  - None

  Example: A specific message set expressed as an XML schema or an EDI transaction set that contains information about a “Person” entity.

Query Point: An endpoint that provides an interface for accessing and querying a Data Asset. A concrete representation of a Query Point may be a specific URL at which a query Web Service may be invoked.

- **Relationships:**
  - A Query Point *accesses* a Data Asset

  Example: A specific URL at which a data service may be invoked.

Data Asset: See the Data Context chapter.

6.5. Data Sharing Attributes

This section will expand on the concepts presented above to include attributes that are associated with each concept. A description will be provided for each attribute, along with an example where necessary for clarity.

Concept	Attribute	Description	Example
Exchange Package	Identifier[33]	A unique string associated with an Exchange Package for identification purposes.	"200XCB"
	Name	The name of an Exchange Package.	"Bill of Lading Message Set"
	Description	A description of an Exchange Package.	
	Classification	The security classification for an Exchange Package.	"U" (Unclassified)
	Frequency	The frequency at which the exchange occurs.	"Daily"
Supplier	Identifier	A unique string associated with a Supplier for identification purposes.	"04091967J"
	Name	The name of a Supplier.	
	Primary Contact	The name and contact information for the Supplier's primary contact for this particular exchange.	
Consumer	Identifier	A unique string associated with a Consumer for identification purposes.	"03081956K"
	Name	The name of a Consumer.	
	Primary	The name and contact information for the	

[33] The "Identifier" attribute is described at an abstract level in order to be consistent with the abstract nature of the reference model. Therefore, there are no references to aspects such as identifier uniqueness, representation format, or similar. Implementations based on the DRM will introduce such aspects as needed according to their requirements.

Concept	Attribute	Description	Example
	Contact	Consumer's primary contact for this particular exchange.	
Payload Definition	Identifier	A unique string associated with a Payload Definition for identification purposes.	"B5102078L"
	Name	The name of a Payload Definition.	"Bill of Lading XML Schema"
Query Point	Identifier[34]	A unique string associated with a Query Point for identification purposes.	http://www.example.com/querypoint3
	Name	The name of a Query Point.	"Latest Monthly Report Information"
	Description	A description of a Query Point.	
	Query Languages	A stipulation of the query languages that are supported by a Query Point (e.g. SQL-92, CQL (Z39.50), XQuery, HTTP GET, etc.).	"SQL-92"

[34] For a Query Point, an identifier represents the electronic address at which the Query Point may be accessed.

APPENDIX A: Glossary of Selected Terms

Notes:

1. Sources are indicated in parentheses. The phrase "(DRM usage)" denotes either a term that is unique to the DRM or that has a slightly different connotation when used in the context of the DRM.
2. Many of the context-related definitions are taken from the Z39.19-200x document, Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies. The glossary starts on page 172 of the PDF version of the document.

Abstract Model

An architectural pattern that optimizes a data architecture for Data Description, Data Context, and Data Sharing; (DRM usage). A theoretical construct that represents physical, biological or social processes, with a set of variables and a set of logical and quantitative relationships between them; (Answers.com). An abstract model is one way to establish a consistent set of concepts. An abstract model is a tool for the description of complex behavior — it is not a template for an implementation, although it should not stray so far away from reality that it is impossible to recognize how the required behaviors would be implemented. (W3C XML Protocol Abstract Model).

Ad Hoc Query

A query (i.e., search question) formed or used for specific or immediate problems or needs.

Analytical Database

A database that contains structured data objects that support query and analysis, and that tends to be purposefully de-normalized and optimized for query ease and performance.

API

An application programming interface, which is a set of definitions of the ways one piece of computer software communicates with another. It is a method of achieving abstraction, usually (but not necessarily) between higher-level and lower-level software.

Architectural Pattern

A description of an archetypal solution to a recurrent design problem that reflects well-proven design experience; (American Science Institute of Technology).

Attribute

A characteristic of an Entity whose value may be used to help distinguish one instance of an Entity from other instances of the same Entity; (DRM usage). A characteristic or property of an object, such as weight, size, or color. A construct whereby objects or individuals can be distinguished; (WordNet).

Authoring System Repository

A broad term related to document storage. At one extreme, an "authoring system" may be a digital camera. At the other, an authoring system may implement a complex workflow used for the production of a formal publication. In both cases, the products of an authoring system are documents; (DRM usage).

BRM

See "Business Reference Model".

BRM Business Area

The top tier of the BRM. Business Areas separate government operations into high-level categories relating to the purpose of government (Services for Citizens), the mechanisms the government uses to achieve its purpose (Mode of Delivery), the support functions necessary to conduct government operations (Support Delivery of Services),and the resource management functions that support all areas of the government's business(Management of Government Resources); (FEA Consolidated Reference Model).

BRM Line of Business

The middle tier of the BRM. LoBs represent the internal operations of the federal government and its services for citizens, independent of the agencies, bureaus and offices that perform them; (FEA Consolidated Reference Model).

Broader Term

A term to which another term or multiple terms are subordinate in a hierarchy. In thesauri, the relationship indicator for this type of term is BT. (ANSI/NISO Z39.19-200x)

Business

The people or organizations that are described by the BRM. In the Universal Description, Discovery, and Integration standard businesses are defined by a businessEntity. While quite often these are, in fact, businesses in the usual sense of the word, they need not be. For example, the "businesses" in a registry internal to a business might well be internal organizations; (UDDI).

Business Context

The formal description of a specific business circumstance as identified by the values of a set of Context Categories, allowing different business circumstances to be uniquely distinguished; (UN/CEFACT Core Components Technical Specification, Version 2.01). For example, in the FEA SRM, Service Types provide an additional layer of categorization that defines the business context of a specific component within a given Service Domain.

Business Reference Model (BRM)

One of the five FEA reference models. The BRM provides a framework that facilitates a functional (rather than organizational) view of the federal government's LoBs?,LoBs, including its internal operations and its services for citizens, independent of the agencies, bureaus and offices that perform them; (FEA Consolidated Reference Model).

Business Rule

Policies and other restrictions, guidelines, and procedures governing the administration and operation of a service; (Data Dictionary for Preservation Metadata: Final Report of the PREMIS Working Group, May 2005).

Categorization

The process of associating something with a category within a categorization scheme. (DRM Usage)

Categorization Scheme

A group of categories that are related in some manner, and that may be used for purposes of categorization. Categorization schemes may be less formal than classification schemes. (DRM Usage)

Category

A grouping of terms that are semantically or statistically associated, but which do not constitute a strict hierarchy based on genus/species, parent/child, or part/whole relationships. (ANSI/NISO Z39.19-200x)

CGI

Common Gateway Interface, a standard, language-neutral web technology that enables a client web browser to request data from a program executed on the Web server. CGI, invented in 1993 by NCSA, specifies a standard for passing data between the client and the program.

Class

A description of a set of objects that share the same attributes, operations, methods, relationships, and semantics; (ISO 11179-3). Classes are the "blueprints" for objects. A class wraps attributes (data) and behaviors (methods or functions) into a single distinct entity. Objects are instances of classes; (Practical UML: A Hands-On Introduction for Developers).

Class Diagram

A UML diagram that shows a collection of declarative (static) UML model elements such as classes and types, with their contents and relationships: (OMG Terms and Acronyms). A Class diagram gives an overview of a system by showing its classes and the relationships among them. Class diagrams are static -- they display what interacts but not what happens when they do interact; (Practical UML: A Hands-On Introduction for Developers).

Class Model

The process of associating something with a category within a classification scheme. (DRM Usage)

Classification

The process of associating something with a category within a classification scheme. (DRM Usage)

Classification Scheme

A method of organization according to a set of pre-established principles, usually characterized by a notation system and a hierarchical structure of relationships among the nodes. (ANSI/NISO Z39.19-200x)

Collection

An aggregation of information resources used to a support a major business function. In each of these collections data is created, retrieved, updated and deleted; (DRM usage).

Common Entities

A desired product of data entity harmonization in which the semantics and characteristics of data artifacts (elements) are compared across a Community of Interest (COI) or LoB, or government-wide to reach some level of consensus; (DRM usage). Examples might include Person, Organization, Location, and Terrorist. This concept is sometimes called "entities of interest" is certain communities.

Communities of Practice (COPs) or Communities of Interest (COIs)

Communities of Interest are collaborative groups of user who require a shared vocabulary to exchange information to in pursuit of common goals, interests, and business objectives. With the context on the DRM they may include LoBs within the government and external organizations that are dedicated to the support of business functions.

Concept

A unit of thought, formed by mentally combining some or all of the characteristics of a concrete or abstract, real or imaginary object. Concepts exist in the mind as abstract entities independent of terms used to express them. (ANSI/NISO Z39.19-200x)

Conceptual Data Model

A data model that represents an abstract view of the real world; (ISO 11179-3). A higher-level data artifact that is often used to explore domain concepts with project stakeholders. Logical data models are often derived from conceptual data models. At this level, the data modeler attempts to identify the highest-level relationships among the different entities. (More: Conceptual, Logical, and Physical Data Models).

Confidentiality

International Organization for Standardization (ISO) defines the term as "ensuring that information is accessible only to those authorized to have access" and is one of the cornerstones of information security. Confidentiality is one of the design goals for many cryptosystems, made possible in practice by the techniques of modern cryptography. Security defines the methods of protecting information and information systems from unauthorized access, use, disclosure, disruption, modification, or destruction in order to provide confidentiality, integrity, and availability [the so-called “CIA triad”], whether in storage or in transit; (DRM usage).

Consumer

An entity (person or organization) that consumes data that is supplied by a Supplier (DRM usage).

Context

As related to data, context can describe the perspective, significance, connotation, and/or environment of data assets. Context is the relationship of data assets to other concepts that aid in their discovery, use, and comprehension. See Data Context (DRM Usage). Enables the intended meaning of data to be more clearly known. This is often done through categorization of data. Such categorization also facilitates the discovery of data. (Context also includes business rules which will be covered in a later version of the DRM.)

Context Artifact

An example is a Taxonomy.

Context Item

A set of terms or phrases that are organized in lists, tree structures, or networked relationships.

Controlled Vocabulary

A list of terms that have been enumerated explicitly. This list is controlled by and is available from a controlled vocabulary registration authority. All terms in a controlled vocabulary must have an unambiguous, non-redundant definition.

NOTE: This is a design goal that may not be true in practice; it depends on how strict the controlled vocabulary registration authority is regarding registration of terms into a controlled vocabulary. At a minimum, the following two rules should be enforced: 1. If the same term is commonly used to mean different concepts in different contexts, then its name is explicitly qualified to resolve this ambiguity.

2. If multiple terms are used to mean the same thing, one of the terms is identified as the preferred term in the controlled vocabulary and the other terms are listed as synonyms or aliases. (ANSI/NISO Z39.19-200x)

Controlled Vocabulary Registration Authority

An entity that controls and makes available the set of terms within a controlled vocabulary.

CQL (Common Query Language)

A formal language for representing queries to information retrieval systems such as web indexes, bibliographic catalogs and museum collection information; (CQL home page).

CRUD

Database operations Create, Read, Update, and Delete.

Data

A value, or set of values, representing a specific concept or concepts. Data becomes "information" when analyzed and possibly combined with other data in order to extract meaning, and to provide context. The meaning of data can vary according to its context; (DRM usage). Information in a specific physical representation, usually a sequence of symbols that have meaning; especially a representation of information that can be processed or produced by a computer; (RFC2828, Internet Security Glossary). A re-interpretable representation of information in a formalized manner suitable for communication, interpretation or processing; (ISO 11179-3). A representation of facts, concepts, or instructions in a formalized manner, suitable for communication, interpretation, or processing by humans or by automatic means; (ISO 2382-4 as per ISO 11179-1).

Data Access

Requests for data services, such as a query of a Data Asset; (DRM usage). See "Query" and "Query Point".

Data Architecture

Defines how data is stored, managed, and used in a system. It describes how data is persistently stored, how components and processes reference and manipulate this data, how external/legacy systems access the data, interfaces to data managed by external/legacy systems, implementation of common data operations. Data architecture establishes common guidelines for data operations that make it possible to predict, model, gauge, and control the flow of data in the system; (Carnegie Mellon Software Engineering Institute)

Data Artifact

A collective term for electronic artifacts related to the presentation, description, representation, or storage of data. Examples are documents and XML Schemas.

Data Asset

A managed container for data; examples include a relational database, Web site, document repository, directory or data service; (DRM usage).

Data Context

Any information that provides additional meaning to data. Data Context typically specifies a designation or description of the application environment or discipline in which data is applied or from which it originates. It provides perspective, significance, and connotation to data, and is vital to the discovery, use, and comprehension of data. See Context. (DRM usage).

Data Context Standardization Area

One of the three main parts of the DRM Abstract Model. The Data Context standardization area facilitates discovery of data through an approach to the categorization of data according to taxonomies, and provide linkages to the other FEA reference models; (DRM usage).

Data Description Standardization Area

One of the three main parts of the DRM Abstract Model. The Data Description standardization area provides a means to richly describe data, thereby supporting its discovery and sharing; (DRM usage).

Data Dictionary

A database used for data that refers to the use and structure of other data; that is, a database for the storage of metadata [ANSI X3.172-1990]. See also data element dictionary; (ISO 11179-1).

Data Discovery

The process of discovering data that exists within a data asset; (DRM usage). Locating a resource on the Enterprise, using a process (such as a search engine) to obtain knowledge of information content or services that exploit metadata descriptions of enterprise IT resources stored in Directories, Registries, and Catalogs; (DDMS).

Data Element Dictionary

An information resource that lists and defines all relevant data elements; (ISO 11179-1).

Data Element Definition

A textual phrase or sentence associated with a data element within a data dictionary that describes the data element, give the data element a specific meaning and differentiates the data element from other data elements. A good definition is precise, concise, non-circular, and unbamiguous. Definitions should not refer to terms or concepts that might be misinterpreted by others or that have different meanings based on the context of a situation. Definitions should not contain acronyms that are not clearly defined or linked to other precise definitions. Standards such as the ISO/IEC 11179 Metadata Registry specification also give guidelines for creating precise data element definitions.

Data Element Registry

An information resource kept by a registration authority that describes the meaning and representational form of data elements, including registration identifiers, definitions, names, value domains, metadata and administrative attributes, etc. See also register; (ISO 11179-1).

Data Entity

An entity that describes data.

Data Exchange

Fixed, re-occurring transactions between parties, such as the regular exchange of environment testing data among federal, state, local, and tribal entities; (DRM usage).

Data Harmonization

The process of comparing two or more data entity definitions and identifying commonalities among them that warrant their being combined (harmonized) into a single data entity.

Data Integrity

The property that data has not been changed, destroyed, or lost in an unauthorized or accidental manner; (RFC2828, Internet Security Glossary).

Data Management

Principles, processes, and systems for the sharing and management of data. (CMMI V1.1)

Data Model

Representation of the information required to support the operation of any set of business processes and/or the systems used to automate them; (DRM usage). A description of the organization of data in a manner that reflects an information structure; (ISO 11179-1). A graphical and/or lexical representation of data, specifying their properties, structure and inter-relationships; (ISO 11179-3). A model that describes in an abstract way how data is represented in a business organization, an information system or a database management system.

Data Object

An aggregation of data that represents discrete information about a subject area. (DRM usage).

Data Reference Model (DRM)

One of the five reference models of the Federal Enterprise Architecture (FEA). The DRM is a framework whose primary purpose is to enable information sharing and reuse across the federal government via the standard description and discovery of common data and the promotion of uniform data management practices.

Data Registry

An information system that manages and maintains metadata about data and data-related items, such as digital data resources and data assets. A data registry is often paired with a repository; (DRM usage).

Data Repository

A repository is a central place where data is stored and maintained. A repository can be a place where multiple databases or files are located for distribution over a network, or a repository can be a location that is directly accessible to the user without having to travel across a network.

Data Representation

Describes how data is described within the property and object layers; (DRM usage).

Data Schema

A representation of metadata, often in the form of data artifacts such as logical data models or conceptual data models. The Data Schema concept group is comprised of those concepts pertaining to the representation of structured data.; (DRM usage).

Data Service

An automated process that provides a related and well described set of data related functions to other applications, systems and processes or to the end user. Data services are invoked through query points, which identify the services and its location in a Web environment; platform-neutral service (such as a Web Service) that provides access to data assets; (DRM usage).

Data Sharing Standardization Area

One of the three main parts of the DRM Abstract Model. Describes the sharing and exchange of data, where sharing may consist of ad-hoc requests (such as a one-time query of a particular data asset), scheduled queries, and/or exchanges characterized by fixed, re-occurring transactions between parties. Data sharing is enabled by capabilities provided by both the Data Context and Data Description standardization areas. Data sharing involves exchanges within and between agencies and COIs to support mission-critical capabilities. These COIs may include international, state, local and tribal governments. Data sharing eliminates duplication and/or replication of data, thereby increasing data quality and integrity; (DRM usage).

Data Steward

A person or organization delegated the responsibility for managing a specific set of data resources; (ISO 11179-1).

Data Stewardship

Identifying, defining, specifying, sourcing, and standardizing data assets across all business areas within a specific business subject area consisting of some set of entity types, e.g., person.

Data Supplier-to-Consumer Matrix

Presents a planning matrix to describe Data Sharing services that should be considered in meeting an agency's or COI's information sharing requirements; comprised of four quadrants: transactional databases, analytical databases, authoring systems repositories, and document repositories; (DRM usage).

Data Type (or Datatype)

A constraint on the type of data that an instance of an Attribute may hold (e.g. "date", "string", "float" or "integer"); defines the kind of data that can be stored in a variable or data element; (DRM usage). The format used for the collection of letters, digits, and/or symbols, to depict values of a data element, determined by the operations that may be performed on the data element; (ISO 11179-1).

Digital Data Resource

A digital container of information, typically known as a file; may be a structured, semi-structured, or unstructured data resource; (DRM usage). The difference between a Document and a Digital Data Resource, is that a Digital Data Resource can contained structured data, unlike a Document. See also "Document".

Digital Rights Management

An umbrella term referring to any of several technical methods used to control or restrict the use of digital media content on electronic devices with such technologies installed. The media most often restricted by DRM techniques include music, visual artwork, computer and video games, and movies. Because the "rights" — actually technical capabilities — that a content owner grants are not the same as the legal rights of a content consumer, DRM critics argue that the phrase "digital rights management" is a misnomer and the term “digital restrictions management” is a more accurate characterization of the functionality of DRM systems.

Directory

An entity in a file system which contains a group of files and other directories. In Microsoft Windows, a directory is called a “folder”.

Document

A file containing Unstructured and/or Semi-Structured Data Resources. A discrete and unique electronic aggregation of data produced with the intent of conveying information. All data within a document may be in the same format (e.g., text), or a document may be a composite that consists of sets of data in a variety of formats (e.g., MS Word files containing embedded graphics). The term “discrete” implies that a document requires no linkage to other data to convey its meaning. The term “unique” implies that each instance or version of a document can be distinguished from all others (i.e., it can be assigned a unique identifying number). Documents may be unstructured, meaning that the document follows no rigid, machine interpretable structural convention or it may contain self describing metadata that is machine interpretable. For example, an ASCII document is unstructured. Alternatively, documents may be semi-structured, meaning that they conform to a

machine interpretable structural convention or contain embedded self-describing metadata that is machine interpretable. A Microsoft Word document with headings and sub-headings is considered semi-structured, as is a XHTML document; (DRM usage). See also "Digital Data Resource".

Document Metadata

Describes an electronic document as well as the data required to file and retrieve it. It includes information fields such as To, From, Date, Subject, Document Type, Format, Location, Record Number, Version Number, File Tag, and Originating Organization. XML is the preferred format for storing document metadata. Examples of document metadata include MS Office document “Properties”, or “meta” tags in HTML/XHTML. MS Office Properties include: Title, Subject, Author, Date Modified, etc. For comparison, the Dublin Core metadata elements are Contributor, Coverage, Creator, Date, Description, Format, Identifier, Language, Publisher, Relation, Right, Source, Subject, Title, and Type; (DRM usage).

Document Repository

A data asset whose primary role is the storage and maintenance of documents.

Document Type Definition (DTD)

A set of declarations that conform to a particular markup syntax and that describe a class, or "type", of SGML, HTML, or XML documents, in terms of constraints on the structure of those documents. In a DTD, the structure of a class of documents is described via element and attribute-list declarations.

DRM

See “Data Reference Model”.

Dublin Core Metadata Initiative (DCMI)

An open forum engaged in the development of interoperable online metadata standards that support a broad range of purposes and business models; (DCMI).

E-Government Act of 2002

HR 2458 is an act of congress “to enhance the management and promotion of electronic Government services and processes by establishing a Federal Chief Information Officer within the Office of Management and Budget, and by establishing a broad framework of measures that require using Internet-based information technology to enhance citizen access to Government information and services, and for other purposes.” It addresses information security; privacy; accessibility, usability, and preservation of government information; share-in-

savings initiatives around common LoBs, and much more. (More: Library of Congress, H.R.2458 or CIO Council archives or Complete Text as PDF or selected portions on XML.gov).

E-Government Act of 2002, Section 207(d)

A section of the E-Government Act of 2002 that pertains to the categorization of information. (More: Complete Text as PDF or selected portions on XML.gov).

Electronic Data Interchange (EDI)

A standard format for exchanging business data. The North American standard for EDI is called ANSI (American National Standards Institute) X12; (TechTarget.com). Computer-to-computer exchange of structured information, by agreed message standards, from one computer application to another by electronic means and with a minimum of human intervention. EDI is still the data format used by the vast majority of electronic commerce transactions in the world.

Enterprise Architecture (EA)

The explicit description and documentation of the current and desired relationships among business and management processes and information technology. An EA describes the "current architecture" and "target architecture" to include the rules and standards and systems life cycle information to optimize and maintain the environment which the agency wishes to create and maintain by managing its IT portfolio. The EA must also provide a strategy that will enable the agency to support its current state and also act as the roadmap for transition to its target environment. These transition processes will include an agency's capital planning and investment control processes, agency EA planning processes, and agency systems life cycle methodologies. The EA will define principles and goals and set direction on such issues as the promotion of interoperability, open systems, public access, compliance with GPEA, end user satisfaction, and IT security. The agency must support the EA with a complete inventory of agency information resources, including personnel, equipment, and funds devoted to information resources management and information technology, at an appropriate level of detail. Agencies must implement the EA consistent with following principles: (i) Develop information systems that facilitate interoperability, application portability, and scalability of electronic applications across networks of heterogeneous hardware, software, and telecommunications platforms; (ii) Meet information technology needs through cost effective intra-agency and interagency sharing, before acquiring new information technology resources; and (iii) Establish a level of security for all information systems that is commensurate to the risk and magnitude of the harm resulting from the loss, misuse, unauthorized access to, or modification of the information stored or flowing through these systems;
(http://www.whitehouse.gov/omb/circulars/a130/a130trans4.html)

Entity

An abstraction for a person, place, object, event, or concept described (or characterized) by common Attributes; (DRM usage). Any concrete or abstract thing that exists, did exist, or might exist, including associations among these things; (ISO 11179-3).

Entity of Interest

An abstraction for a person, place, object, event, or concept described (or characterized) by common Attributes that is central to the information sharing requirements of a COI; (DRM Usage).

E-R (Entity-Relationship) Diagram (ERD)

A data modeling technique that creates a graphical representation of the entities, and the relationships between entities, within an information system; also includes cardinality; (More: WhatIs.com).

E-R (Entity-Relationship) Model

A way of graphically representing the logical relationships of entities (or objects) in order to create a database; (More: WhatIs.com).

Exchange Package

A description of a specific recurring data exchange between a Supplier and a Consumer. An Exchange Package contains information (metadata) relating to the exchange (such as Supplier ID, Consumer ID, validity period for data, etc.), as well as a reference to the Payload (message content) for the exchange. An Exchange Package can also be used to define the result format for a query that is accepted and processed by a Query Point in a data sharing scenario; (DRM usage).

Extract, Transform, Load (ETL)

The process of reading structured data objects from a data source (the extract), changing the format of the data objects to match the structure required by a target database (transform), and updating the target database with the transferred data objects (load).

FEA Reference Model

A series of interrelated taxonomies that comprise the FEA, and that are designed to facilitate cross-agency analysis and the identification of duplicative investments, gaps, and opportunities for collaboration within and across Federal Agencies. (http://www.whitehouse.gov/omb/egov/a-2-EAModelsNEW2.html)

FEA Security and Privacy Profile (SPP)

Provides guidance to agencies to integrate security and privacy requirements across their enterprise architecture, and to ensure security and privacy requirements are addressed in IT programs from their inception.

Federal Enterprise Architecture (FEA)

A business-based framework for government-wide improvement developed by the Office of Management and Budget (OMB). (http://www.whitehouse.gov/omb/egov/a-1-fea.html)

Federated Registries

Registries may be federated in order to enable their contents to be shared amongst other registries, causing them to appear to a user and to automated processes (such as queries) as a single registry.

FIPS 199

Federal Information Processing Standard 199 (2004 February) defines "Standards for Security Categorization of Federal Information and Information Systems". It addresses one of the requirements specified in FISMA. FIPS 199 provides security categorization standards for information and information systems. Security categorization standards make available a common framework and method for expressing security. This promotes the effective management and oversight of information security programs, including the coordination of information security efforts throughout the civilian, national security, emergency preparedness, homeland security, and law enforcement communities. Such standards also enable consistent reporting to OMB and Congress on the adequacy and effectiveness of information security policies, procedures, and practices; (FIPS By Number page). (More: FIPS home page; FIPS 199 document.)

FISMA

Federal Information Security Management Act (FISMA) of 2002, requires all federal agencies to develop, document, and implement agency-wide information security programs for the information and information systems that support the operations and the assets of the agency, including those provided or managed by another agency, contractor, or other source; (FIPS By Number page). FISMA compliance is a matter of national security, and is therefore scrutinized at the highest level of government. FISMA is also known as TITLE III—INFORMATION SECURITY of H. R. 2458, the EGovernment Act of 2002. (More: FISMA Implementation Project at NIST; FISMA document at NIST; FISMA Compliance at Watchfire.com.)

Formal Classification

Classification that involves formal relationships between topics, and includes specific rules or constraints for those relationships.

Geospatial

Pertaining to the geographic location and characteristics of natural or constructed features and boundaries on, above, or below the earth's surface; especially referring to data that is geographic and spatial in nature; (Dictionary.com). (More: Open Geospatial Consortium; National Geospatial-Intelligence Agency; Geography Markup Language.)

Harmonization

Act of bringing or coming to agreement or harmony (Dictionary.com).

Hierarchy

Broader (generic) to narrower (specific) or whole-part relationships, which are generally indicated in a controlled vocabulary through codes or indentation. (ANSI/NISO Z39.19-200x)

HTML (HyperText Markup Language)

A markup language designed for the creation of web pages and other information viewable in a browser. HTML is used to structure information – denoting certain text as headings, paragraphs, lists and so on – and can be used to define the semantics of a document. Originally defined by Tim Berners-Lee and further developed by the IETF with a simplified SGML syntax, HTML is now an international standard (ISO/IEC 15445:2000). Later HTML specifications are maintained by the World Wide Web Consortium (W3C).

HTTP (HyperText Transfer Protocol)

The primary method used to convey information on the World Wide Web. HTTP is a request/response protocol between clients and servers.

HTTP GET

The most common method used to request a specified URL. When you click on most web links (other than web forms), you are causing your browser to issue an HTTP GET request for a particular page or resource from a web server.

Informal Classification

Classification in which there may or may not be specific types of topics, and the topics that are defined may or may not have formally defined relationships. Many Web sites and search utilities offer a basic classification that may be considered informal classification.

Information Class

In the DOI Recreation.gov example, an information class is equivalent to the Entity concept of the Data Description standardization area.

ISO/IEC 11179

A standard for representing Metadata for an organization in a Metadata Registry. The specification is formally known as the ISO/IEC 11179 Metadata Registry Standard and consists of six sections: Part 1 - Framework, Part 2 - Conceptual Schema, Part 3 - Registry Metamodel and Basic Attributes, Part 4 - Formulation of Data Definitions, Part 5 - Naming and Identification Principals, and Part 6 - Registration. The specification defines how data elements are classified, specified, defined, named, and registered. Use of ISO-11179 is strongly recommended by state and federal agencies.

LDAP

In computer networking, the Lightweight Directory Access Protocol is a standardized networking protocol designed for querying and modifying directory services. LDAP defines a relatively simple protocol for updating and searching directories running over TCP/IP. No specific type of directory is an "LDAP directory". One could reasonably use the term to describe any directory accessible using LDAP and which can identify objects in the directory with X.500 identifiers. LDAP directory entries feature a hierarchical structure that reflects political, geographic, and/or organizational boundaries, usually with DNS names at the top level.

Lines of Businesses (LoBs)

Major government business areas identified in the Business Reference Model (BRM). Each LoB is comprised of a collection of Sub-Functions. Approximately 39 LoBs are identified in the BRM. About half are external; they are found in the Services for Citizens layer and describe the purpose of government in functional terms. The remaining half are internal LoBs that describe the support functions the government should conduct in order to effectively deliver services for citizens; (FEA BRM 2.0, June 2003).

List

A limited set of terms arranged as a simple alphabetical list or in some other logically evident way; the simplest type of controlled vocabularies.

Logical Data Model

A model describes the same data as a conceptual data model, but as structured in an information system. It is often referred to as a Model of the Information System. A logical data model can be directly used for database design; (ISO 11179-3). A graphical representation of the information requirements of a business area, it is not a database; (More: "Why Build a Logical Data Model" by Embarcadero). At this level, the data modeler attempts to describe the data in as much detail as possible, without regard to how they will be physically implemented in the database; (Conceptual, Logical, and Physical Data Models).

Management Context

A data artifact that represents the concepts (entities) that are specific to a domain, their attributes, and the relationships between the concepts. Logical data models may also contain data types for attributes.

Metadata

To facilitate common understanding, a number of characteristics, or attributes, of data are defined. These characteristics of data are known as “metadata”, that is, “data that describes data”. (ISO 11179-3). Information about data. For any particular datum, the metadata may describe how the datum is represented, ranges of acceptable values, its relationship to other data, and how it should be labeled. Metadata also may provide other relevant information, such as the responsible steward, associated laws and regulations, and access management policy. Each of the types of data described above has a corresponding set of metadata. Two of the many metadata standards are the Dublin Core Metadata Initiative (DCMI) and Department of Defense Discovery Metadata Standard (DDMS). The metadata for structured data objects describes the structure, data elements, interrelationships, and other characteristics of information, including its creation, disposition, access and handling controls, formats, content, and context, as well as related audit trails. Metadata includes data element names (such as Organization Name, Address, etc.), their definition, and their format (numeric, date, text, etc.). In contrast, data is the actual data values such as the “US Patent and Trade Office” or the “Social Security Administration” for the metadata called “Organization Name”. Metadata may include metrics about an organization’s data including its data quality (accuracy, completeness, etc.); (DRM usage).

Metadata Registry

An information system for registering metadata (ISO/IEC 11179). A metadata registry provides a shared understanding about the metadata that describes a data object (DRM usage).

Metamodel

A structure used to create models. For example, an XML Schema defines how to create XML vocabularies and structure XML data. In relational terms, data definition language (DDL) is used to generate (one or more) database schema (made up of related database tables) from which data can be entered. A data model that specifies one or more other data models; (ISO 11179-3).

Narrower Term

A term that is subordinate to another term or to multiple terms in a hierarchy. In thesauri, the relationship indicator for this type of term is NT. (ANSI/NISO Z39.19-200x)

Node

A specific concept or term in a taxonomy, thesaurus, classification scheme or categorization scheme. (DRM Usage)

Node Relationship

A semantic relationship (e.g. narrower-term) between nodes. (DRM Usage)

Object

Anything perceivable or conceivable. NOTE Objects may also be material (e.g. an engine, a sheet of paper, a diamond), immaterial (e.g. a conversion ratio, a project plan) or imagined (e.g. a unicorn); (ISO 11179-3).

OLAP

On-Line Analytical Processing, an approach to quickly provide the answer to complex analytical queries, providing the ability to analyze metrics in different dimensions such as time, geography, gender, product, etc. The OLAP Report has proposed the FASMI test, Fast Analysis of Shared Multidimensional Information.

OLTP

On-Line Transaction Processing, a form of transaction processing conducted via computer network. Some applications of OLTP include electronic banking, order processing, employee time clock systems, e-commerce, and eTrading. In large applications, efficient OLTP may depend on sophisticated transaction management software and/or database optimization tactics to facilitate the processing of large numbers of concurrent updates. OLTP is often integrated into service-oriented architecture and Web services.

OMB

The United States Office of Management and Budget is a body within the Executive Office of the President of the United States (EOP) which is tasked with coordinating United States Federal agencies. A "stop-and-think shop," it is a senior management team of the White House. The OMB performs this coordination by gathering and filtering budget requests, by issuing circulars dictating agency management practices, and by reviewing agency regulations.; (More: Whitehouse.gov.)

OMB 300

Also known as “Exhibit 300” and “E-300”. The Exhibit 300 business case is a high level summary of the investment’s current justification and management plans including a project plan, benefit-cost analysis, alternatives analysis, acquisition plan, risk management plan, human resources management plan, enterprise architecture and IT Security plan. In the case of proposed new IT investments, this information is used by the operating unit, the Department’s Capital Investment Technology Review Board (CITRB), and OMB to determine if the investment should be recommended for funding. For on-going investments, the Exhibit 300 is used to review the investment’s current status and, subsequently, to assess how well the investment accomplished its goals; (Department of Commerce). (More: FEA in the Budget; Current year E-3000; and Exhibit 300 XML Schema, Version 2.97 [for FY07].)

OMG

The Object Management is a consortium that sets standards in object-oriented programming as well as system modeling. The OMG created the Common Object Request Broker Architecture (CORBA) standard in 1991 and more recently, the standard for Unified Modeling Language (UML) and related technologies Meta-Object Facility (MOF) and XML Metadata Interchange (XMI). It has further expanded into Model Driven Architecture (MDA). (More: OMG web site and OMG Terms and Acronyms.)

Ontology

A controlled vocabulary expressed in a representation language that has a grammar for using vocabulary terms to express something meaningful within a specified domain of interest. The grammar contains formal constraints (e.g., specifies what it means to be a well-formed statement, assertion, query, etc.) on how terms in the ontology's controlled vocabulary can be used together. (ANSI/NISO Z39.19-200x)

OWL (Web Ontology Language)

A markup language for publishing and sharing data using ontologies on the Internet and is a vocabulary extension of RDF. Together with RDF and other components, these tools make up the Semantic Web project. (More: W3C).

Payload

The set of data objects a data service exchanges during a transaction; the message content; (DRM usage).

Payload Definition

An electronic definition that defines the requirements for the Payload (data) that is exchanged between a Supplier and a Consumer. Examples include XML Schema and EDI transactions.

Performance Reference Model (PRM)

One of the five FEA reference models. The PRM is a framework for performance measurement providing common output measurements throughout the federal government.

Physical Data Model

A representation of a data design which takes into account the facilities and constraints of a given database management system. It is typically derived from the Logical Data Model and may include all the database artifacts required to create relationships between tables or achieve performance goals, such as indexes, constraint definitions, linking tables, partitioned tables or clusters. At this level, the data modeler specifies how the logical data model will be realized in the database schema; (Conceptual, Logical, and Physical Data Models).

Polyhierarchy

Networked relationships, where each item may be related to one or more other items without the direct notion of a parent-child pair.

Preferred Term

One of two or more synonyms or lexical variants selected as a term for inclusion in a controlled vocabulary. (ANSI/NISO Z39.19-200x)

Privacy

Addresses the acceptable collection, creation, use, disclosure, transmitting, and storage of information, its accuracy, and the minimum necessary use of information. Section 208 of Title II of the E-Government Act of 2002 addresses privacy. See “FISMA”, “FIPS 199”, and “E-Government Act”.

Query

An instruction given to access a Data Asset; a request issued to receive data. A Query may be ad hoc when it is issued as an isolated access to a Data Asset (e.g., a one-time database query), or a Query may be part of a pre-planned, methodical operation, in which case it is recurring and often scheduled; (DRM usage).

Query Point

An endpoint that provides an interface for accessing and querying a Data Asset. A concrete representation of a Query Point may be a specific URL at which a query Web Service may be invoked; (DRM usage). See "Exchange Package".

RecML

Recreation Markup Language is an XML vocabulary that defines terms for recreation areas (parks), facilities (trails, campgrounds, etc.), activities (hiking, wildlife viewing, etc.), alerts (temporary closures), events, and similar recreation elements. RecML is a voluntary data sharing specification for recreation information developed by a COI, namely the recreation community, including the private sector and government organizations at the Federal, tribal, state, and local levels; (Recreation.gov site).

Recreation One-Stop

A citizen-focused E-Government Initiative managed by the Department of the Interior with two main goals: (1) customer-friendly recreation portal (Recreation.gov) with information for planning visits to Federal recreation sites and making campground/tour reservations; and (2) consistent information about Federal recreation areas via different channels (databases, websites, and publications), by standardizing data and interfacing recreation-related computer systems; (Whitehouse.gov egov pages).

Reference Models

A structure which allows the modules and interfaces of a system to be described in a consistent manner; An abstract framework for understanding significant relationships among the entities of some environment, and for the development of consistent standards or specifications supporting that environment. A reference model is based on a small number of unifying concepts and may be used as a basis for education and explaining standards to a non-specialist. A reference model is not directly tied to any standards, technologies or other concrete implementation details, but it does seek to provide a common semantics that can be used unambiguously across and between different implementations. (The Federal Enterprise Architecture Framework is defined in terms of reference models).

Register

A set of files (paper, electronic, or a combination) containing the assigned data elements and the associated information. See also Data Element Registry; (ISO 11179-1).

Registration

The assignment of an unambiguous identifier to a data element in a way that makes the metadata about those data elements available to interested parties; (ISO 11179-1).

Related Term

A term that is associatively but not hierarchically linked to another term in a controlled vocabulary. In thesauri, the relationship indicator for this type of term is RT. (ANSI/NISO Z39.19-200x)

Relationship

Association between two entities in an ERD. Each end of the relationship shows the degree of how the entities are related and the optionality; (Oracle FAQ).

RDF (Resource Description Framework)

A family of specifications for a metadata model. The RDF family of specifications is maintained by the World Wide Web Consortium (W3C). The RDF metadata model is based upon the idea of making statements about resources in the form of a subject-predicate-object expression…and is a major component in what is proposed by the W3C's Semantic Web activitiy: an evolutionary stage of the World Wide Web in which automated software can store, exchange, and utilize metadata about the vast resources of the Web, in turn enabling users to deal

with those resources with greater efficiency and certainty. RDF's simple data model and ability to model disparate, abstract concepts has also led to its increasing use in knowledge management applications unrelated to Semantic Web activity.

Result Set

A Query Point provides the result set for an Exchange Package; (DRM usage). In SQL, a result set is a set of rows from a database, as well as meta-information about the query such as number of results returned and the column names. More generally, it is the data returned by any type of query (search).

RIDB

Recreation Information Database, Department of the Interior. RIDB is a warehouse of information about Federal recreation sites, with the ability to export that data to state tourism portals, recreation-related businesses in the private sector, etc. See Recreation.gov.

Schema

The structure of a data set, database, Exchange Package, etc. See also "XML Schema".

Security

Security defines the methods of protecting information and information systems from unauthorized access, use, disclosure, disruption, modification, or destruction in order to provide integrity, confidentiality and availability, whether in storage or in transit; (DRM usage). A condition that results from the establishment and maintenance of protective measures that ensure a state of inviolability from hostile acts or influences. With respect to classified matter, the condition that prevents unauthorized persons from having access to official information that is safeguarded in the interests of national security. Measures taken by a military unit, an activity or installation to protect itself against all acts designed to, or which may, impair its effectiveness.

Semantic Linking

A method of linking terms according to their meaning or meanings. (ANSI/NISO Z39.19-200x)

Semantic Web

A representation in two (or possibly three) dimensions of the semantic relationships between and among terms and the concepts they represent; (ANSI/NISO Z39.19-200x). The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries. It is a collaborative effort led by W3C with participation from a large number of researchers and industrial partners. It is based on the Resource Description Framework (RDF), which integrates a variety of applications using XML for syntax and URIs for naming; (W3 Semantic Web home page). Refers to a suite of technologies that aim to enhance the performance of the Internet for the functions of businesses, organizations and individuals by increasing capabilities to interpret and determine meaning in web-based data and information.

Semi-Structured Data (Resource)

Data that has characteristics of both structured and unstructured data, such as an e-mail (with structured data such as sender and subject, and unstructured text); (DRM usage). Semi-structured data is the term database theorists use to denote data that exhibits any of the following characteristics: numerous repeating fields and structures in a naive hierarchical representation of the data, which lead to large numbers of tables in a second- or third-normal form representation; wide variation in structure; and/or sparse tables; (C. M. Sperberg-McQueen) (More: series of ACM Queue articles.)

Service Component Reference Model (SRM)

A business and performance-driven, functional framework that classifies Service Components with respect to how they support business and/or performance objectives. The SRM is intended for use to support the discovery of government-wide business and application Service Components in IT investments and assets. The SRM is structured across horizontal and vertical service domains that, independent of the business functions, can provide a leverage-able foundation to support the reuse of applications, application capabilities, components, and business services. Service domains include: Customer Services, Process Automation Services, Business Management Services, Digital Asset Services, Business Analytical Services, Back Office Services, and Support Services; (FEA PMO site).

Service Oriented Architecture (SOA)

Expresses a software architectural concept that defines the use of services to support the requirements of software users. In a SOA environment, nodes on a network make resources available to other participants in the network as independent services that the participants access in a standardized way. Most

definitions of SOA identify the use of Web services (using SOAP and WSDL) in its implementation. However, one can implement SOA using any service-based technology with loose coupling among interacting software agents.

SQL-92

SQL is a database sublanguage that is used for accessing relational databases. The proper pronunciation is "ess cue ell," and not "sequel" as is commonly heard. SQL-92 was designed to be a standard for relational database management systems (RDBMSs) developed by the ANSI X3H2 committee. SQL-92 does not address objects in any way. Nevertheless, SQL-92 forms the basis for JDBC and other specifications. Depending on your source, "SQL" stands for SQL Query Language (recursive expansion), Structured Query Language, or nothing at all.

SRM

See “Service Component Reference Model”.

Standardization Area

The three aspects of data that the DRM addresses, namely Data Description, Data Context, and Data Sharing. The DRM’s standardization areas provide a foundation for agency data architecture initiatives to put forth requirements that can result in increased compatibility between agency data architectures; (DRM usage).

Structured Data Object

An entity within a data store. These entities, in turn, contain attributes that describe the object. Such objects rely on the structure and relationships defined in the data store to assign their meaning. Databases are examples of collections of structured data objects; (DRM usage).

Structured Data (Resource)

Data described via the E-R (Entity-Relationship) or class model, such as logical data models and XML documents. Structured data is organized in well-defined semantic “chunks” called entities; (DRM usage).

Subject Area

A topic of interest shared within a community. The full list of subject areas of interest to a community form the context for that community. A super type is a subject area that spans multiple COIs; (DRM usage).

Supplier

An entity (person or organization) that supplies data to a Consumer. Note that the Supplier may or may not be the original producer of the data. For this reason, the name “Producer” was not used; (DRM usage).

Synonym

A word or term having exactly or very nearly the same meaning as another word or term. (ANSI/NISO Z39.19-200x)

Synonym Ring

A group of terms that are considered equivalent for the purposes of retrieval. (ANSI/NISO Z39.19-200x)

Target Architecture

The set of products that portrays the future or end-state enterprise, generally captured in the organization’s strategic thinking and plans; commonly referred to as the "To-Be" architecture.

Taxonomy

A collection of controlled vocabulary terms organized into a hierarchical structure. Each term in a taxonomy is in one or more parent/child (broader/narrower) relationships to other terms in the taxonomy. There can be different types of parent/child relationships in a taxonomy (e.g., whole/part, genus/species, type/instance), but good practice limits all parent-child relationships to a single parent to be of the same type. Some taxonomies allow poly-hierarchy, which means that a term can have multiple parents, and although the term appears in multiple places, it is the same term. If the parent term has children in one place in a taxonomy, then it has the same children in every other place where it appears. (ANSI/NISO Z39.19-200x)

Technical Reference Model (TRM)

A component-driven, technical framework used to categorize the standards, specifications, and technologies that support and enable the delivery of service components and capabilities. The TRM provides a foundation to categorize the standards, specifications, and technologies to support the construction, delivery, and exchange of business and application components (Service Components) that may be used and leveraged in a Component-Based or Service-Oriented Architecture. The TRM unifies existing Agency TRMs and E-Gov guidance by providing a foundation to advance the re-use of technology and component services from a government-wide perspective. Service areas include: Service Access and Delivery, Service Platform and Infrastructure, Component Framework, and Service Interface and Integration; (FEA PMO site).

Term

One or more words designating a concept. (ANSI/NISO Z39.19-200x)

Term Record

A collection of information associated with a term in a controlled vocabulary, including the history of the term, its relationships to other terms, and, optionally, authorities for the term. (ANSI/NISO Z39.19-200x)

Thesaurus

A networked collection of controlled vocabulary terms. A thesaurus uses equivalence (synonym), hierarchical (broader/narrower), and associative relationships. The expressiveness of the associative relationships in a thesaurus varies and can be as simple as "related to term," as in term A is related to term B. (ANSI/NISO Z39.19-200x)

Topic

A category within a Taxonomy. A Topic is the central concept for applying context to data. For example, an agency may have a Taxonomy that represents their organizational structure. In such a Taxonomy, each role in the organizational structure (e.g. CIO) represents a Topic. Topic is often synonymous with Node; (DRM usage).

Top Term

The broadest term in a controlled vocabulary hierarchy. (ANSI/NISO Z39.19-200x)

Transaction

An exchange of information between two or more services (or an entity and a service) in the performance of an operation or function; (DRM usage).

Transactional Database

A database that support transactions. A database transaction is a unit of interaction with a database management system or similar system that is treated in a coherent and reliable way independent of other transactions that must be either entirely completed or aborted.

Tree Structure

A controlled vocabulary display format in which the complete hierarchy of terms is shown. Each term is assigned a tree number or line number which leads from the alphabetical display to the hierarchical one; the latter is also known as systematic display or classified display. (ANSI/NISO Z39.19-200x)

TRM

See "Technical Reference Model".

UML

Unified Modeling Language (UML) is a non-proprietary, object modeling and specification language. As a graphical notation, UML can be used for modeling hardware (engineering systems) and is commonly used for business process modeling, systems engineering modeling, software engineering, and representing organizational structure. UML was designed to be used to specify, visualize, construct, and document the artifacts of an object-oriented software-intensive system under development. It represents an integrated compilation of best engineering practices that have proven to be successful in modeling large, complex systems, especially at the architectural level; (More: OMG's UML Resource Page). See also OMG.

Unstructured Data (Resource)

Data that is of a more free-form format, such as multimedia files, images, sound files, or unstructured text. Unstructured data does not necessarily follow any format or hierarchal sequence, nor does it follow any relational rules; (DRM usage). Unstructured data refers to masses of (usually) computerized information which do not have a data structure which is easily readable by a machine. Examples of unstructured data may include audio, video and unstructured text such as the body of an email or word processor document. Data mining techniques are used to find patterns in, or otherwise interpret, this information. Merrill Lynch estimates that more than 85 percent of all business information exists as unstructured data – commonly appearing in e-mails, memos, notes from call centers and support operations, news, user groups, chats, reports, letters, surveys, white papers, marketing material, research, presentations and Web pages; ("The Problem with Unstructured Data".)

Vocabulary Control

The process of organizing a list of terms (a) to indicate which of two or more synonymous terms is authorized for use; (b) to distinguish between homographs; and (c) to indicate hierarchical and associative relationships among terms in the

context of a controlled vocabulary or subject heading list. (ANSI/NISO Z39.19-200x)

Web Services

A software system designed to support interoperable machine-to-machine interaction over a network. It has an interface that is described in a machine-processable format such as WSDL. Other systems interact with the Web service in a manner prescribed by its interface using messages, which may be enclosed in a SOAP envelope, or follow a REST approach. These messages are typically conveyed using HTTP, and are normally comprised of XML in conjunction with other Web-related standards. (More: W3C Web Services Activity).

X.500

A series of computer networking standards covering electronic directory services. The X.500 series was developed in order to support the requirements of X.400 electronic mail exchange and name lookup. ISO was a partner in developing the standards, incorporating them into the Open Systems Interconnect suite of protocols, ISO/IEC 9594. X.509, the portion of the standard providing for an authentication framework, is now also widely used outside of the X.500 directory protocols for public-key certificates. Because of the complexity of the protocols, a simplified alternative, known as LDAP, was developed implementing only a subset of the protocols.

XML

Extensible Markup Language has at least two distinct meanings: 1. A set of generic syntax rules to enable the creation of specialized markup languages that follow similar conventions. 2. An ever-growing collection of standard, de facto standard, and special purpose languages based on XML syntax (e.g., XSLT, UBL, ebXML, XML Schema, XHTML, RDF, OWL, SVG, etc.). Sometimes the term "XML" is used incorrectly when really "XML Schema" is intended. (More: W3C XML home page).

XML Document

A storage unit (i.e., a file) containing XML markup and content; (DRM usage). A data object is an XML document if it is well-formed, as defined in this specification. A well-formed XML document MAY in addition be valid if it meets certain further constraints; (XML 1.0 Recommendation, Third Edition).

XML Registries and Repositories

An XML *registry* is an information system that securely stores XML artifacts (e.g., XML schemas, data elements, etc.) and non-XML artifacts (e.g. other e-

business objects), as well as details (metadata) about the artifacts. The storage facility (e.g., a file system or database) that holds registered objects is known as a *repository*, while the part of the information system that maintains the metadata for the registered objects is known as a registry; (ebXML document). (More: XML.Gov XML Registries page; DoD Metadata Registry; and OASIS ebXML Registry Technical Committee).

XML Schema

Defines the vocabulary (elements and attributes), the content model (structure, element nesting, and text content), and data types (value constraints) of a class of XML documents. When written with a capital 'S', the term refers specifically to the XML Schema Definition (XSD or WXS) language developed by the W3C. However, when written with a lowercase 's', the meaning is more generic, referring to any of several schema languages for use with XML, such as DTDs, RELAX NG, Schematron, etc. In both cases, an XML schema is used to validate XML instances, to verify that the instances conform to the model that the schema describes.

XPath

XML Path Language (XPath) is a terse non-XML syntax for addressing portions of an XML document. A path expression is written as a sequence of steps to get from one set of nodes to another set of nodes. XPath also allows more conventional expressions, involving arithmetic and boolean operators and a range of functions to perform string manipulation, etc.

XQuery

XML Query (XQuery) is a query language with some programming language features designed to query collections of XML data. It is semantically similar to SQL and is being developed by the XML Query working group of the W3C. The work is closely coordinated with the development of XSLT 2.0 by the XSL Working Group; the two groups share responsibility for XPath 2.0, which is a subset of XQuery 1.0. At the time of this writing, XQuery is a W3C Candidate Recommendation, although dozens of implementations are available in various states of completeness. In addition to XPath addressing, it provides SQL-like FLWOR expressions based on five possible clauses: FOR, LET, WHERE, ORDER BY, RETURN. Note: XQuery 1.0 does not include features for updating XML documents or databases. It also lacks full text search capability. These features are both under active development for a subsequent version of the language. (More: W3C XML Query page).

XSLT

XSL Transformations (XSLT) is an XML-based, declarative language used for the transformation of XML documents. The original document is not changed; rather, a new XML document is created based on the content of the original document. The new document may be serialized (output) by the processor in standard XML syntax or in another format, such as HTML or plain text. XSLT is most often used to convert data between different XML Schemas or to convert XML data into web pages or PDF documents. It can also be used to extract portions of an XML document; (More: W3C XSL page). Note: "XSL" stands for Extensible Stylesheet Language, which includes XSLT, XSL-FO (XSL Formatting Objects), and XPath.

XSLT Stylesheets

A transformation expressed in XSLT is called a stylesheet. A stylesheet contains a set of template rules. A template rule has two parts: a pattern which is matched against nodes in the source tree and a template which can be instantiated to form part of the result tree. This allows a stylesheet to be applicable to a wide class of documents that have similar source tree structures.; (More: W3C XSLT 1.0 Recommendation).

Z39.50

National Information Standards Organization (NISO) Z39.50 Information Retrieval Protocol (Z39.50/ISO 23950), a computer protocol that can be implemented on any platform, defines a standard way for two computers to communicate for the purpose of information retrieval. A Z39.50 implementation enables one interface to access multiple systems providing the end-user with nearly transparent access to other systems; (NISO Z39.50 Resource Page) Despite a common misconception to the contrary, Z39.50 is not simply used by libraries. The standard specifies a client/server-based protocol for searching and retrieving information from remote databases. (More: Z39.50 Maintenance Agency Page at the Library of Congress.)