The Data Reference Model

Version 2.0

November 17, 2005

CONTENTS

LIST OF FIGURES

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:

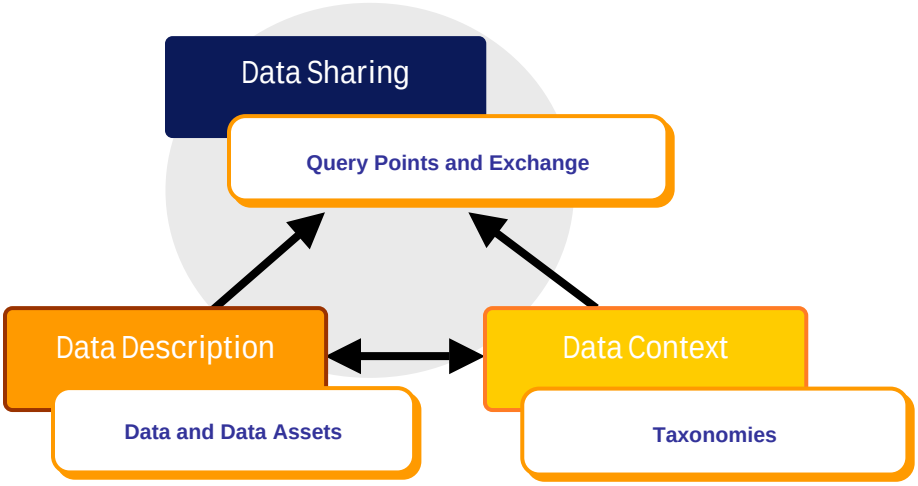


Figure 2-1 DRM Standardization Areas

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:

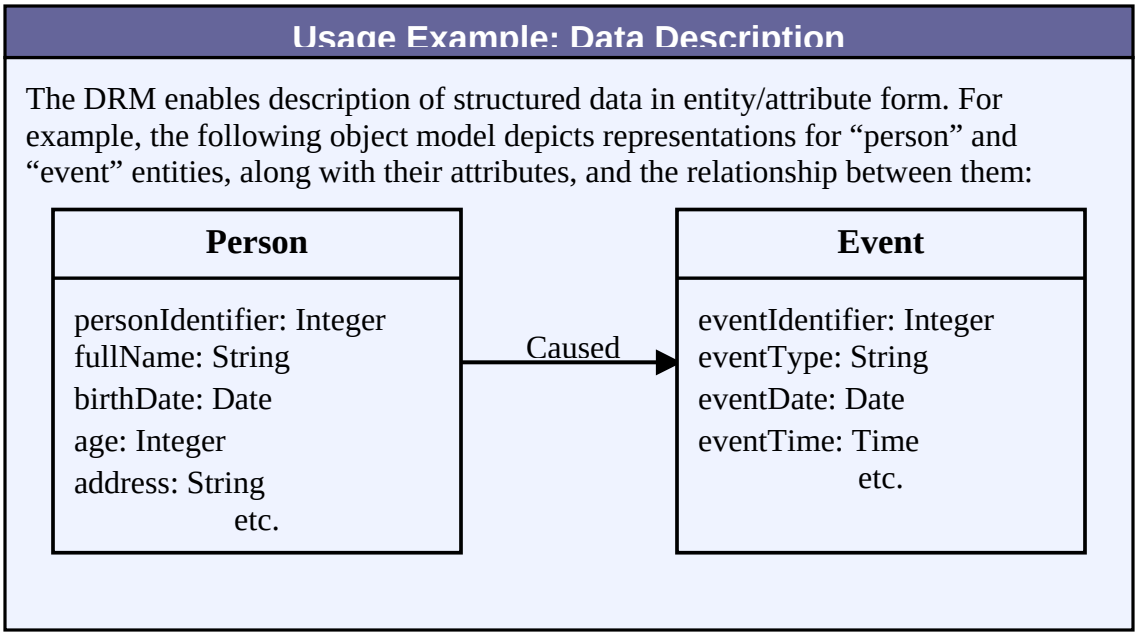


Figure 2-2 Data Description Usage Example

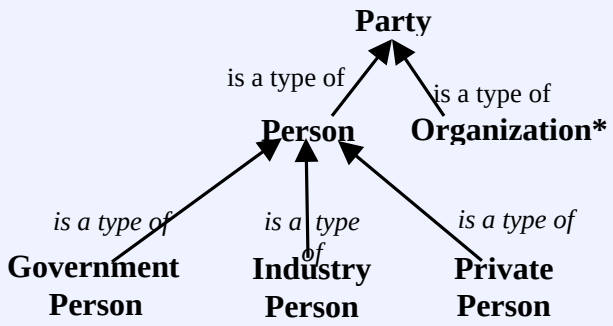
Data Context: The Data Context standardization area establishes an approach to the categorization of data assets using taxonomies and other descriptive information. In general, Data Context answers key questions about the data required within a COI and establishes the basis for data governance. Data 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.

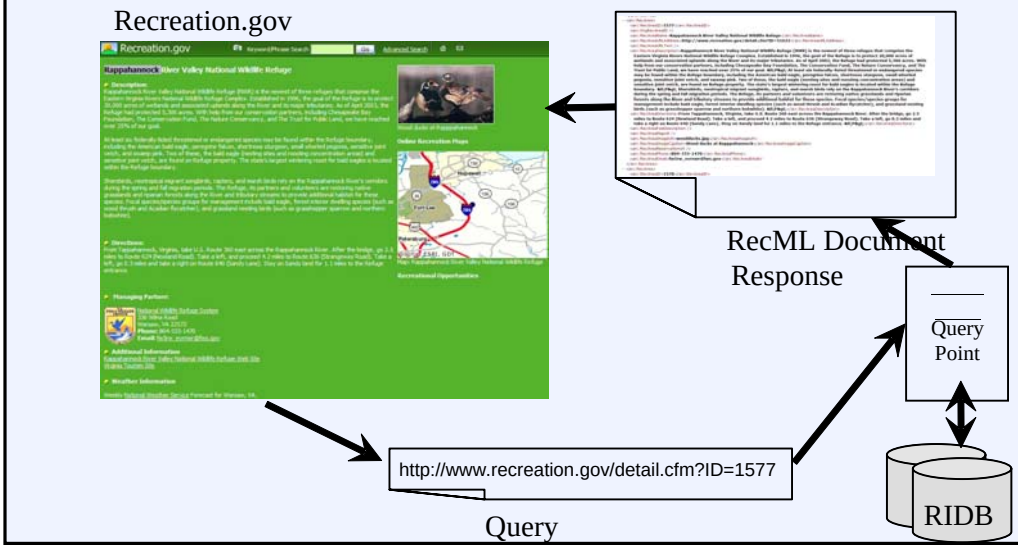


Figure 2-4 Data Sharing Usage Example

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.

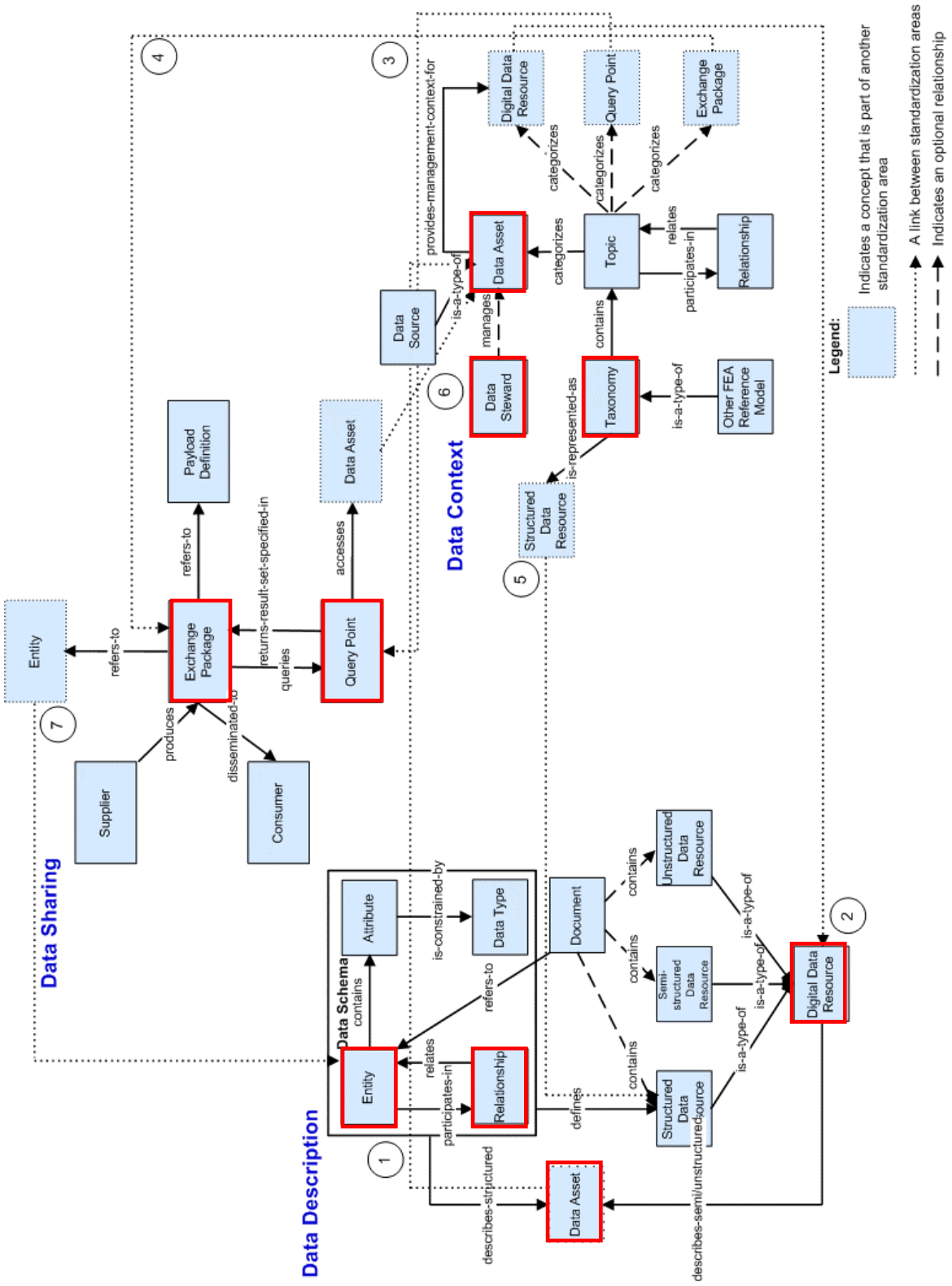


Figure 2-5 DRM Abstract Model

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.

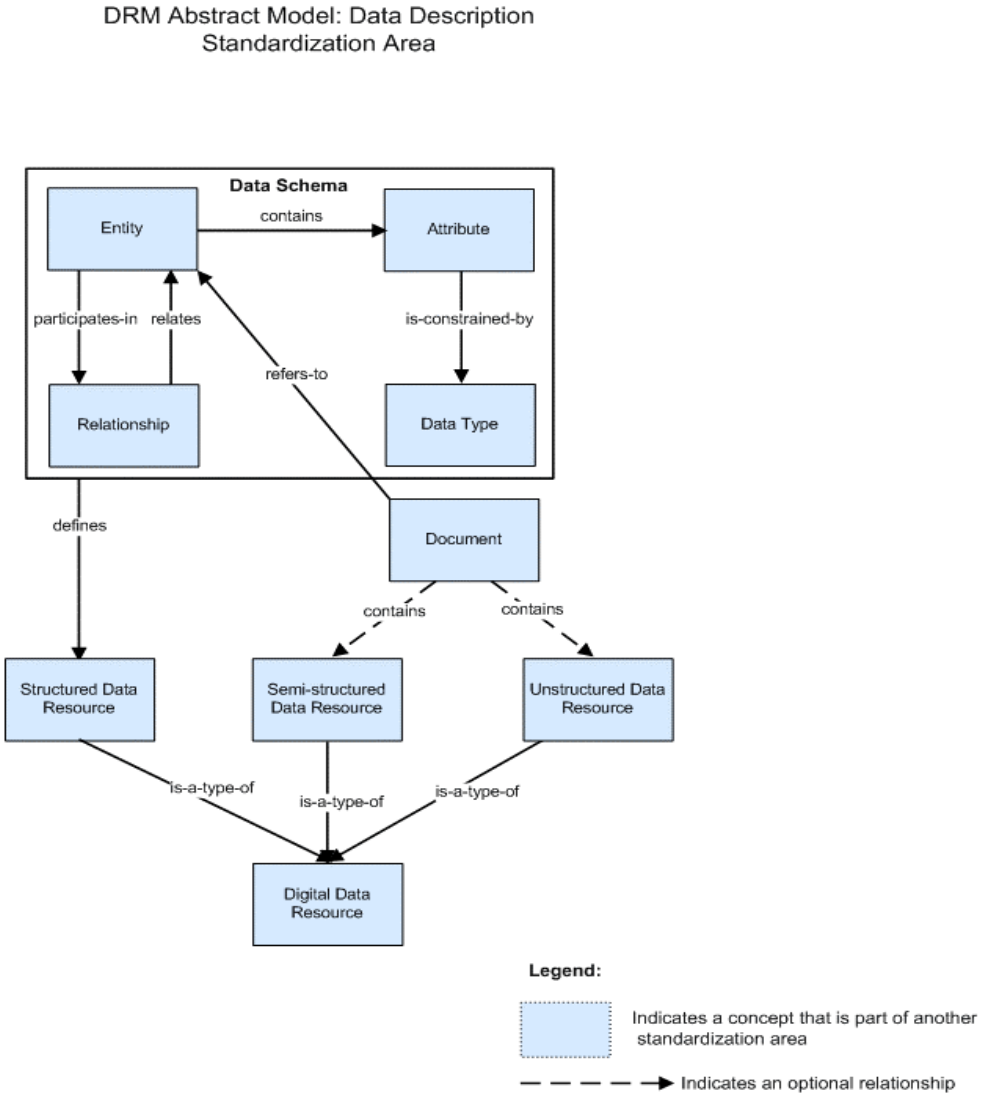


Figure 3-1 DRM Data Description Abstract Model

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:

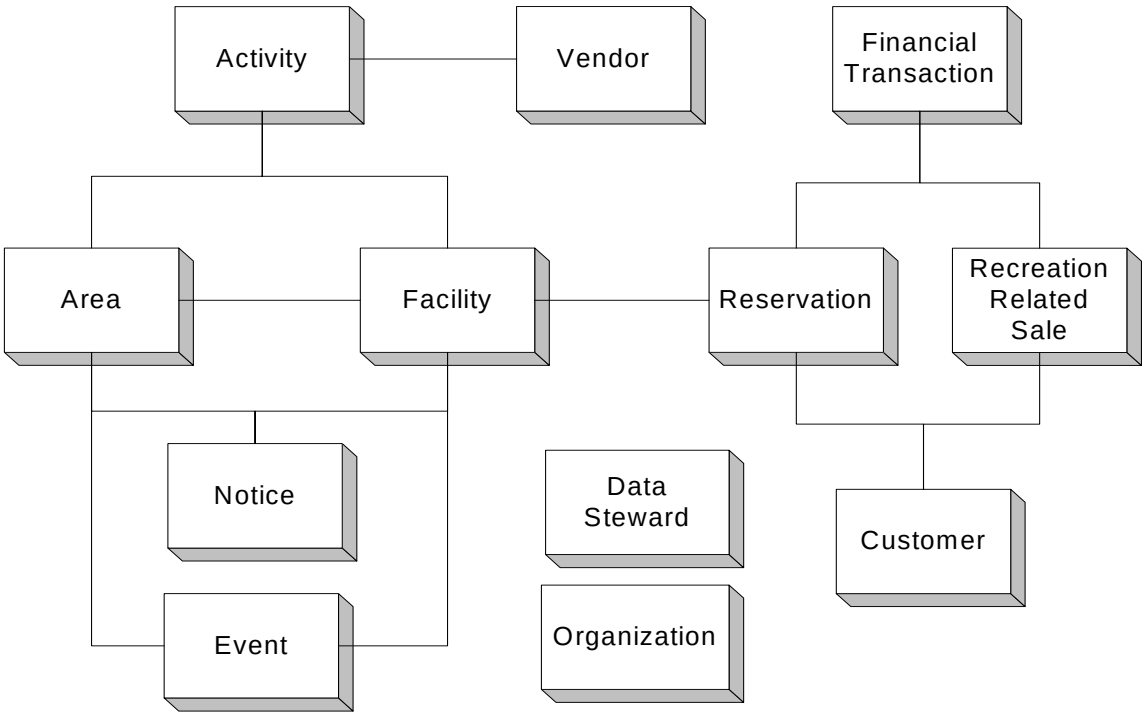


Figure 3-2 Recreation One Stop Information Classes

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

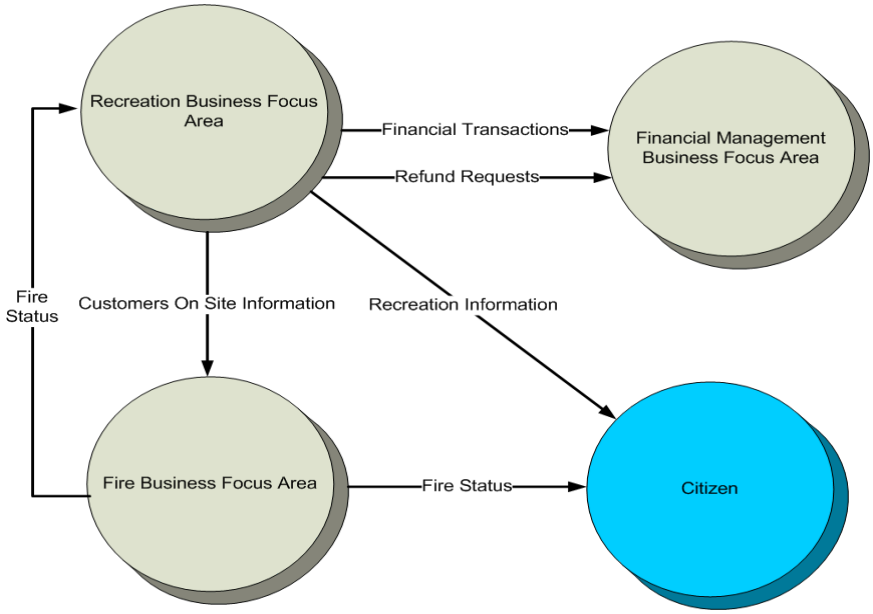


Figure 3-3 DOI Three Business Focus Areas

[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).