MTR 00W0000097

MITRE TECHNICAL REPORT

FAA Data Registry (FDR) Concepts of Use and Implementation

September 2000

Dr. Nels A. Broste
Ronald G. Rhoades
Ronald A. Schwarz

MITRE
Center for Advanced Aviation System Development
McLean, Virginia

MTR 00W0000097

MITRE TECHNICAL REPORT

FAA Data Registry (FDR) Concepts of Use and Implementation

September 2000

Dr. Nels A. Broste
Ronald G. Rhoades
Ronald A. Schwarz

Sponsor: Federal Aviation Administration **Contract No.:** DTFA01-93-C-00001
Dept. No.: F062 **Project No.:** 02001208-2A

Approved for public release; distribution unlimited.

MITRE
Center for Advanced Aviation System Development
McLean, Virginia

MITRE Department
and Project Approval:

Frank L. Willingham
Program Manager, F062
ATM Systems

Abstract

As part of its plan to develop a National Airspace System (NAS)-wide Information Service, described in the NAS Architecture Version 4.0 [ASD], the Federal Aviation Administration (FAA) is in the process of developing a resource to manage NAS-wide metadata, data standards and, eventually, data element models. Called the FAA Data Registry (FDR), for FAA Data Registry[1], this capability will give FAA managers and staff across the NAS the ability to manage metadata from multiple systems in a single, integrated tool as well as to define, store, and disseminate its developing data standards. To be developed in concert with the Metadata Repository (MDR)[2], this effort is a fundamental step in the development of configuration management for NAS data elements over their life cycle. The creation of the FDR is a step towards configuration management of NAS data elements that support a wide range of interoperable systems. This activity is also expected to benefit modernization of the NAS and support international data harmonization efforts.

This document describes a concept of use, or vision for the FDR as a tool in the technical architecture evolution. The subject includes a brief description of the FDR and its functionality, how it would be deployed, fielded, and maintained, who its users might be and the anticipated benefits. In total, it will support the larger strategic goals of the FAA, facilitate work process improvement and provide benefits to the engineers and managers guiding change to the NAS.

KEYWORDS: data dictionary, FDR, configuration management, data registry, data standards, data warehouse, metadata, metadata repository, NAS-wide Information Service (NWIS), operational concept

---

1 Formerly known as the Aviation Data Description Registry (ADDR)

2 The Metadata Repository (MDR) is a related information system that will manage system information, eventually down to the data element level. The specific interface between the FDR and MDR is yet to be developed.

Acknowledgments

This document reflects the thinking of many colleagues at the FAA and at Center for Advanced Aviation System Development (CAASD) who have been proponents of a critical resource such as a data registry. In particular, we would like to thank Carol Uri, ASD-110, for her constant encouragement and her active efforts to create an agency-wide registry. We also thank Steve Bradford, ASD-110, for his critical and constructive suggestions.

The authors also thank Trish Palmer of MITRE/CAASD for helping us to produce the finished version of the document.

Table of Contents

List of Figures

List of Tables

Executive Summary

The Problem. Historically, systems at the Federal Aviation Administration (FAA), as in many large organizations, are developed to solve a particular problem. Typically, the addition of new systems increases the complexity of the overall system environment and makes the relationships among systems more difficult to manage. In particular, with regard to information systems management, system developments typically define their own data definitions and data structures. The net result is data inconsistency, more expensive system maintenance, more time needed for development (since there is little data structure reuse) and much harder and costly system interoperability.

Ironically, contributing to the extent of the problem is that there is an increasing need for systems to interoperate to share data. Interoperability is needed within National Airspace System (NAS) domains, across domains, between the NAS and business/administrative systems and for FAA systems and external systems, such as those managed by the airlines, aircraft manufacturers, regulators, and foreign aviation bodies, such as national civil aviation administrations (CAA) and Eurocontrol.

The Solution. A data registry is an information management product, common to most large-scale organizations today, that helps a variety of users to build, maintain, manage, grow and operate the collection of systems in that organization. The initial FAA Data Registry (FDR) will provide a single, standards-based registry for the collection, definition, documentation, and distribution of information (i.e., metadata) about data elements used widely in the NAS. The ISO/IEC Standard 11179, Information Technology—Specification and Standardization of Data Elements, Part 1, defines a data element registry as:

> *"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."*

The FDR described here will manage two types of information, namely:

- Data and information standards at micro and macro information levels to be used in data management. This will support data exchange, information interoperability, and collaborative decision-making, and system acquisition, development, and re-design. These standards will proceed from basic data element standards (initially) to more complex standards for related data elements, such as for flight components and aeronautical data (eventually).
- Information about current (legacy) data elements used in the NAS to the extent that these data are related to a data element standard. Such information will describe the context of legacy data elements (i.e., metadata) and be associated with a

corresponding data element standard. This information will also be related to equivalent information in the Office of Information Services/Office of Air Traffic Services (AIO/ATS)-developed FAA Metadata Repository (MDR), which will have additional metadata about the system and the context in which these data elements are used, including system ownership, distribution, access, security, and use, from NAS Management Directives (MD) and other official source material.

CAASD has completed a companion document entitled *FDR Software Requirements Specification (SRS)*, MTR-00W0000071, September 2000, that contains a detailed set of requirements to be used to build the FDR, including an ISO-standards-based data model that would serve as the structure of the database to manage the FDR metadata.

Rationale. Interoperability requires a sound 'back-office,' that is, a well-managed set of system resources, of which information management is the subject here. Information resources must be managed consistently from one system to the next. Moreover, there should also be an integrated, automated means for viewing the information resources managed by more than one system, by the systems within a domain, by the systems at a facility (there may be an exception here), or the NAS as a whole.

Today, it is difficult to obtain information about existing and planned NAS data in a rapid and automated fashion, especially when the question spans multiple systems. The concept for developing the FDR is motivated, in part, by the need to get answers to questions about such data in an efficient and timely manner. Providing a common view of the same data across systems can help answer questions such as:

- **Who** are the users of the data produced by a particular system? People and systems are users. Internally and externally.
- **What** are the data elements, how are they defined and formatted and what is the data architecture? Are data models being used to define and manage data relationships? What are the data dependencies among the data? What are the key fields?
- **Where** do the data originate and where are they used? Is the agency obtaining or buying the same data from duplicate sources? Often, the answer is yes.
- **When** in the system life cycle are the data needed? Can a common data service manage this function for multiple system needing similar data? Are algorithms and data transformations being developed redundantly and with different results?
- **How** will the data be needed in the planning process? Is the business side of the NAS (cost, schedule, performance) tied to the system development side (technology, data access, reliability, interoperability)?

Essential Concept. At the outset, it is essential that the reader distinguish the concept of *metadata* from that of *data*. The example in Figure ES-1 illustrates that distinction. The

attributes of Element Name, Definition, Format, and Source are metadata, i.e., descriptors of the data, just as someone's employee number and social security number are descriptors of a person. In the example, 'Airport_ID' is a value for the metadata item called 'Element Name' while 'DCA' is a value for the data item called Airport_ID.

A data registry manages metadata primarily. A data warehouse manages data values, e.g., the set of identifiers for the 18,000 or so airports in the US, of which three are 'LAX', 'DCA', and 'IAD'. Metadata values, such as the name of the data element (in this case, Airport_ID) need to be associated with the set of data values (the 18,000 airport IDs).

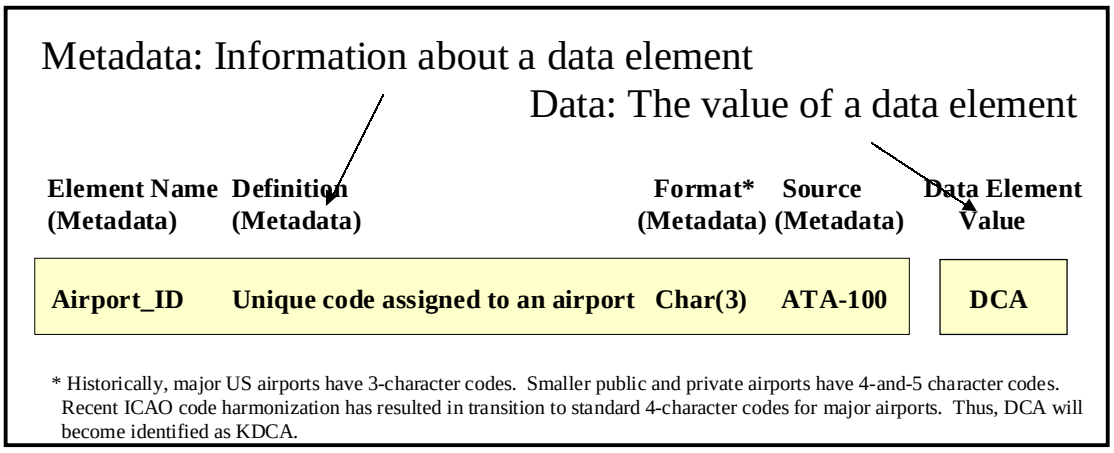


Figure ES-1. The Difference Between Metadata and Data

Context. There has already been some effort at FAA directed at building a metadata registry, primarily in the Architecture and Investment Analysis Division (ASD), AIO, ATS, ASY, and AOP organizations. These organizations are aware of each other's efforts and are examining ways to coordinate their development. Additional details about the work of these organizations are in the main document.

The FDR provides the following important benefits:

- **Data Quality and Access:** Data quality would be improved by reducing the ambiguity about similar data defined differently across systems. The current use of different geographic coordinate system standards and formats is an example. Safety is, in part, dependent on data quality.
- **Interoperability:** Today, system interfaces are customized between pairs of systems (rather than among many systems), expensive to build and maintain, and relatively inflexible. Data exchange is more efficient when systems exchanging data define data and data structures in common.

- **Cost Effectiveness:** Constrained budgets can be used more effectively when common services, in this case common data services, can serve multiple systems rather than when each system develops its own data services locally and in a non-standard way.
- **Flexibility:** Common data services developed with automated tools allows system-wide access to metadata and the data behind them more easily and efficiently. Maintenance to existing systems is also faster and cheaper.

Costs for managing the FDR are difficult to address in detail since current practices do not easily lend themselves to isolating costs at sufficiently low levels of system activities.[3] Also, some costs are not easily measured, such as the cost of *not* being able to access system metadata flexibly, on demand, and in electronic form. Given current practices and the redundancy in spending for similar information resource developments, it appears that significant cost avoidance is possible. At a minimum, considerably increased data quality and system quality can be achieved with current expenditures. However, it will require some direct experience and careful comparison to yield accurate cost and benefit values, some of which are qualitative, such as speed, accuracy, and flexibility of access to information.

The FDR will act as a bridge between old and new systems, between current and future system views, and between data standards and legacy data, as illustrated in Figure ES-2. While focused on data element standards, the FDR will also contain operational and system metadata about related data in current systems (e.g., information about legacy data elements), data in systems under development, such as the Center Terminal Automation System (CTAS), and redesign. The set of metadata selected for the initial build and the set of corresponding metadata services is yet to be determined.

Figure ES-3 implies a growth path for the FDR. Each histogram, moving from left to right, implies a growth path for metadata content and for corresponding services that could be offered. In the right-most four histograms, the data in the top box is the value-added from the histogram to its immediate left. In terms of data services,

- A Data Dictionary contains basic information about a data element standard, i.e., its name, definition, and format
- A Data Directory then adds 'location' information, that is, information about where and in what systems, data element standards and legacy data elements are used

---

[3] There is a current effort at the FAA in the ABA organization to restructure the way information is managed by the business systems and to link that information more closely with operational systems information for better accountability and control.

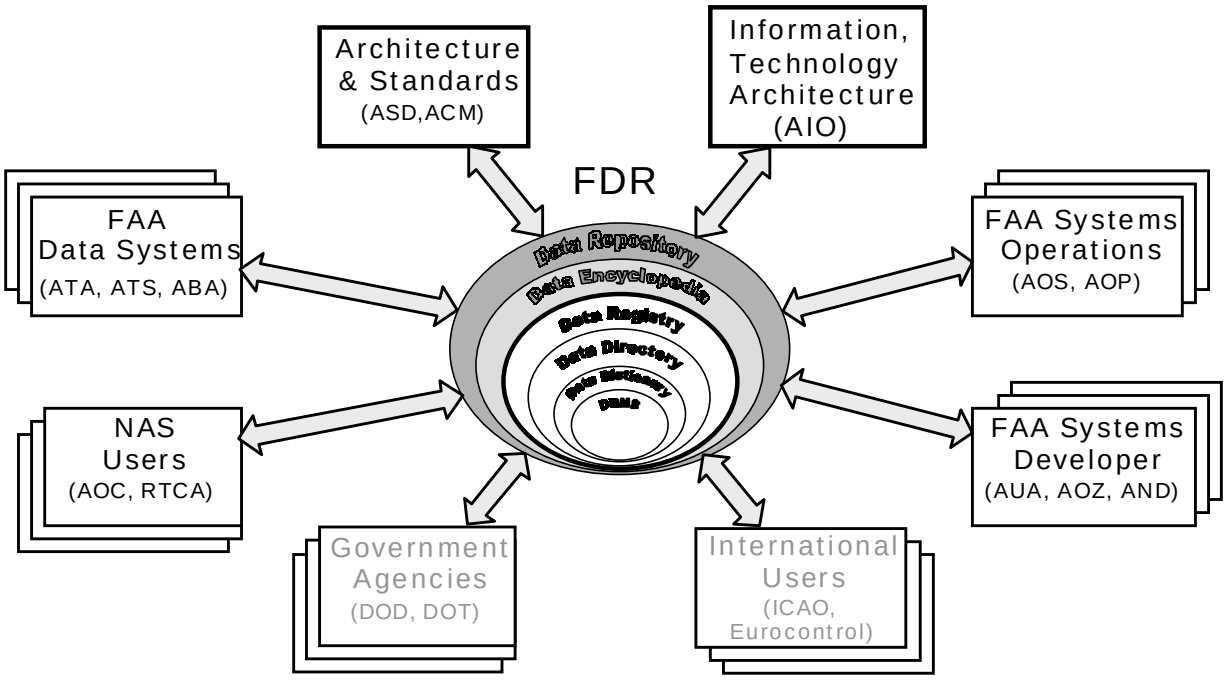


Figure ES-2. FDR Context

- A Data Registry then adds 'registration information, i.e., information about data stewardship and authority for managing that data element
- A Data Encyclopedia then adds information about the application and other contexts in which the data are used

A Data Repository then adds more higher-level information constructs in terms of application strategy, process, procedure, and budgets.

These services may be available now in some form within each application system, but not in a controlled, consistent, non-redundant manner in which the FDR could provide them. The FDR will also manage the NAS-wide view of data element standards information.

The FDR could also have a progression in the complexity of content. For example, rather than containing standard views of single data elements, the FDR could eventually contain standard views of *groups* of elements, such as would be contained in a flight data model. These groups, or data models, could be the basis for system acquisitions, leading to component reuse and improved system interoperability.

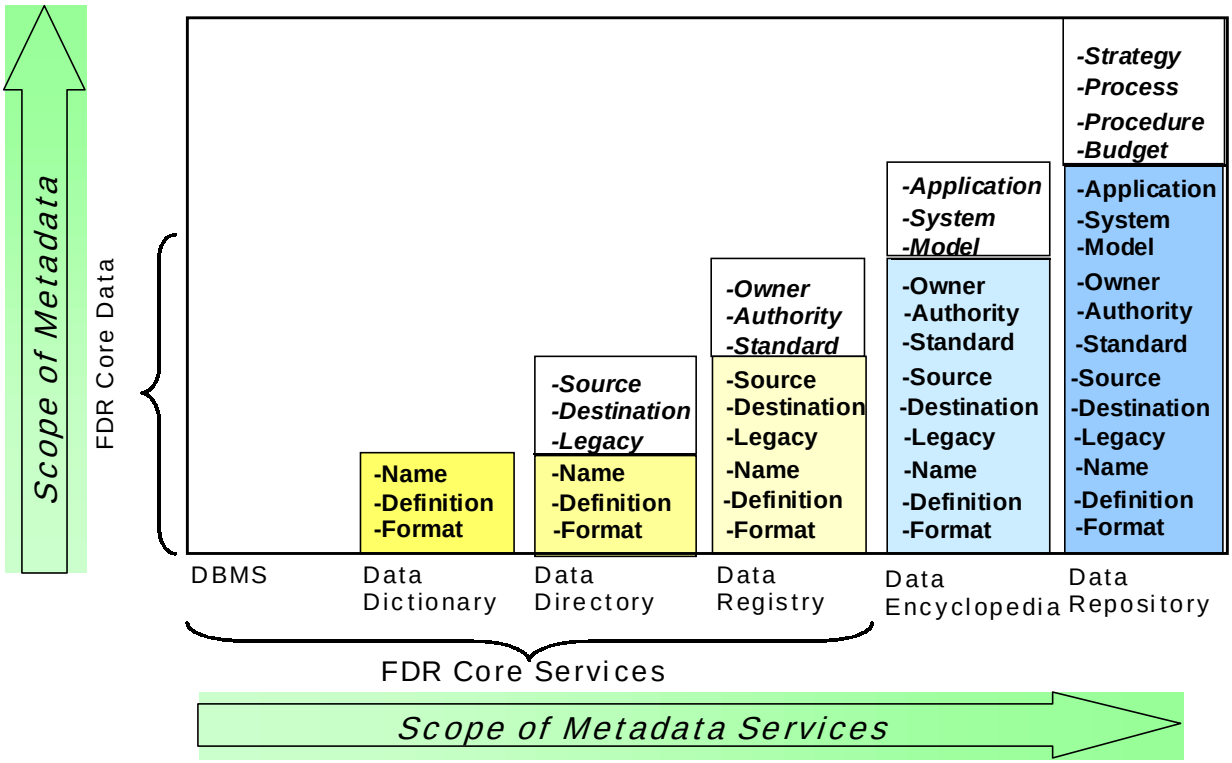


Figure ES-3. Scope of Initial FDR Contents and Functionality

Conclusions.

- Metadata management is a foundation step leading to integrated NAS information management.
- Based on industry 'best practices' and the increasing effort needed to make FAA systems interoperable, a case can be made for the FAA's need to manage agency-wide metadata in a consistent way.
- Without overall guidance and a clearly defined architecture and design, there will be a proliferation of metadata databases, each built to different operational concepts and designs. Integrating them 'after-the-fact' will be very difficult and expensive, as is already seen with operational application systems. The FAA's system inventory counts over 200 systems, each of which manages its data and metadata 'internally.'
- Managing agency metadata will yield access to important information that is not currently accessible or is not easily accessible.
- The software and technology needed to manage metadata are mature and affordable and are not an impediment to a successful outcome.

Recommendations. The following recommendations are made with respect to developing the FDR:

- The FDR should be developed to include the two main threads discussed in this document: data element standards and legacy metadata links to those standards.
- From a systems viewpoint, it would be useful to study current NAS applications and make a recommendation for how these systems should implement metadata management processes that complement the role of the FDR.
- Although extensive collaboration will be required among the many offices, systems, and activities that are engaged in metadata management, there should be a single point of responsibility to oversee and coordinate the effort.
- A data steward should be identified for each application program as a necessary step to organize the agency's metadata management effort. A high degree of collaboration will be required between the data stewards and the main coordination organization for the FDR.
- The roles and relationship between the FDR and the MDR, being built by AIO and ATS, should be made clear. Data exchange guidelines and content should also be addressed.
- The internet should be used as an access and dissemination point for metadata content since it is widely available and has become a fundamental vehicle for information management.
- To assure success, the roles and responsibilities described above for collecting metadata from application system stewards should be enforced in some manner. One option is through requirements applied by the Joint Resources Council (JRC) and the Acquisition Management System (AMS). Another is through the authority of an organization such as the AIO, whose mandate covers information architecture and information technology management in the FAA.
- Because there is currently some overlap in responsibility, written guidelines should be produced that clearly describe the stewards for the various agency metadata and the processes that will be established to manage the metadata.

Decisions to be Made. Some decisions about the FDR require additional analysis, including those listed here:

- The MDR is currently managing system-level metadata. The FDR starts from a different point with respect to content by focusing on data element standards. An organizational and architectural decision is required regarding whether or not the MDR will also take on the role of managing data element standards and legacy

metadata or whether these capabilities will be managed by a separate, but related, system, such as the FDR.

- The agency's metadata, once assembled and automated, will be an important information resource. Decisions are needed about what portion of this information will be made available to system developers working for the FAA and to other interested aviation bodies, such as NAS users, and to the public.

Section 1

Introduction

This document describes a concept for an initial National Airspace System (NAS) data element registry and an integration and implementation approach that will provide near-term support to a number of NAS operational data management improvement processes. Based on an earlier operational concept [Schwarz] for the Federal Aviation Administration (FAA) Data Registry (FDR), this implementation and use document focuses on an FDR role with the NAS Configuration Control Board (CCB) data standardization process with a natural extension to related applications.

Increasing demands on the NAS are forcing changes in how the FAA develops and manages the airspace in collaboration with many aviation constituencies. Bringing greater structure and consistency to the management of NAS information are important elements of both NAS (systems) modernization and operations. The NAS Architecture Version 4.0 calls for the evolution of information services, specifically data standardization and interoperability across applications. [ASD, p. 19-2] Additionally, it calls for NAS-wide system integration with the NAS information architecture goal being attained through establishing NAS-wide data management roles and responsibilities. This document describes a capability that to manage NAS-wide data element standards and to assist domain application systems improve their interoperability by associating data element standards with equivalent legacy data elements.

As a metadata registry, the FDR will be used to record information about legacy and new standard data elements, permissible value domains, and associated attributes for operational data from many organizations. Some organizations will use the registry to record facts about how data elements are used in existing, baseline applications, while other organizations will use the registry as a repository of standard application-independent data elements to be used as models for data elements in application development. Actual data values are not recorded in an FDR but rather in a data warehouse, a product that is built upon the metadata and data structures defined in a registry.

At the outset, it is essential that the reader distinguish the concept of *metadata* from that of *data*. The example in Figure 1-1 illustrates that distinction. The attributes of Element Name, Definition, Format, and Source are metadata, i.e., descriptors of the data, just as someone's employee number and social security number are descriptors of a person. In the example, 'Airport_ID' is a value for the metadata item called 'Element Name' while 'DCA' is a value for the data item called Airport_ID.

A data registry manages metadata primarily. A data warehouse manages data values, e.g., the set of identifiers for the 18,000 or so airports in the US, of which three are ‘LAX’,

'DCA', and 'IAD'. Metadata values, such as the name of the data element (in this case, Airport_ID) need to be associated with the set of data values (the 18,000 airport IDs).

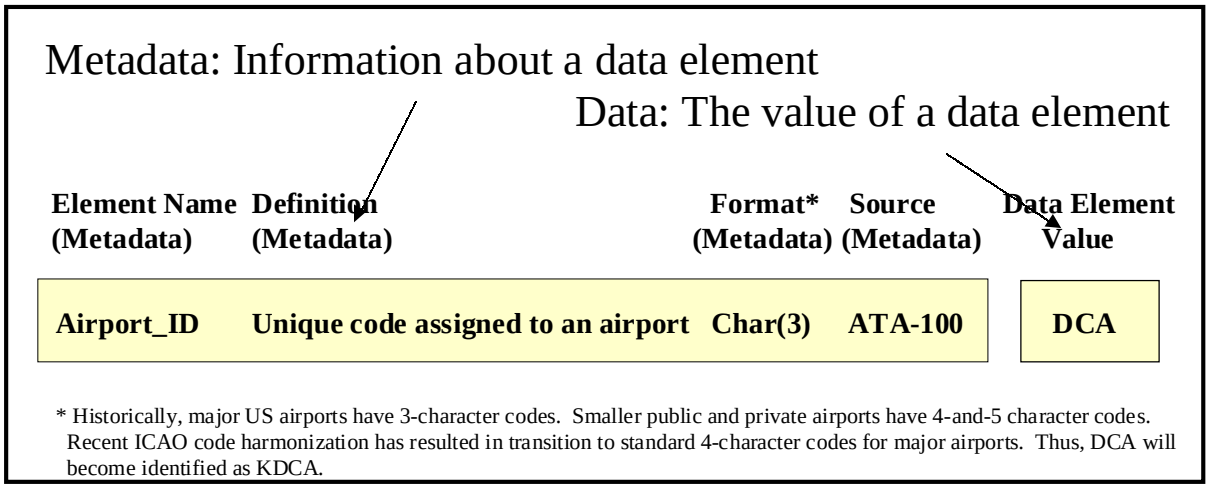


Figure 1-1. The Difference Between Metadata and Data

1.1 Definition of Terms

This document uses a number of information-based terms that have particular meaning. Unfortunately, reflecting a trend in the general culture, these terms are often defined by several variants, with no one recognized as authoritative. However, because it is important that the reader understand and distinguish these terms and how they are used in this document, they are defined here for convenience.

Table 1-1. Definition of Terms

Term	Definition
Data	Values assigned or taken on by data elements
Data architecture	A structure that defines data types, data formats, and the rules that govern their use
Data category (classification scheme)	This term corresponds very closely to that of 'classification scheme,' defined in ISO 11179 , part 3: A scheme for the arrangement or division of objects into groups based on characteristics that the objects have in common, e.g., origin, composition, structure, application, and function. A classification scheme may be a list, a taxonomy, a network, an ontology, or any other scheme for systematizing, where the categories are mutually exclusive.

Term	Definition
	Classification helps users find a single data element from among many data elements, makes data administration analysis of data elements easier and, through inheritance, conveys semantic content that is often only incompletely specified by other data element attributes like names and definitions. [ISO 11179 part 2]
Data warehouse	A collection of data values, organized in a structured fashion to assist the inclusion of additional values and their retrieval; the structure of the warehouse uses the metadata about the data elements whose values are being stored
Information architecture	A structure that defines the current or intended use and flow of data in an organization; it may also include information for managing the data, such as business rules, stewardship assignments, access controls, and security procedures
Metadata	Information about data or data elements; their characteristics and qualities, including its name, meaning, representation, and other contextual information
Metadata repository	An organized collection of metadata, usually stored in a database
Namespace	**Two definitions are offered.** 1. Namespace: The universe of discourse for which a specific Controlled Vocabulary and a specific Naming Standard are applicable. [ISO 2788] 2. Controlled Vocabulary or Thesaurus: A set of controlled terms in one or more languages and/or in one or more country specific languages used for a specific purpose and managed by a competent authority. [ISO 2788] **Context / Namespace** The context is the business system, standard document, or environment in which the taxonomy or data element was created. In the case of international aviation standards this will be identified as "International Standard" (and abbreviated as "Intl Std"). If a taxonomy is provided by a source system such as "ADREP" or ECCAIRS, the context would be the name of the system. In some registry designs, the context is referred to as a "name context" or "name space." [ASY, International Aviation Data Registry Phase 1 Software Requirements Specifications] [From ISO 11179-part 5]:

Term	Definition
	"Context: A designation or description of the application environment in which a name is applied or from which it originates." The context is used to manage the names of administered components in a metadata registry. Names shall be unique within a context and within each subtype of administered component (e.g. the name of a data element and a value domain cannot be the same within the same context).
Object class	A set of ideas, abstractions, or things in the real world that can be identified with explicit boundaries and meaning and whose properties and behavior follow the same rules.
Registry	A registry is a tool for the management of shareable data; a comprehensive, authoritative source of reference information about data. It supports the standard-setting process by recording and disseminating data standards, which facilitates data sharing among organizations and users. It provides links to documents that refer to data elements and to information systems where data elements are used. When used in conjunction with an information database, the registry enables users to better understand the information obtained. [Procedures for Achieving Data Registry Content Consistency, ISO draft December 1999] For some, a registry contains only standards-based information about individual data elements, independent of applications or systems that use the elements. For others, it contains system metadata and system engineering data as well. The concept adopted here takes the more expansive view.
Taxonomy	A scheme that partitions a body of knowledge and defines the relationships among the pieces. It is used for classifying and understanding the body of knowledge. [IEEE90] Here, it is a scheme for classifying aviation terms by type and subtype.

1.2 Organizational References

Many FAA organizational units are interested in and involved with this work. Again, for convenience, they are introduced here and referenced only by their 3-letter organizational code in the remainder of the document.

Table 1-2. Organizational References

Term	Definition
ABA	Assistant Administrator for Financial Services
ACM	NAS Configuration Management and Evaluation
AIO	Assistant Administrator for Information Services
AND	Director, Communications, Navigation, and Surveillance Systems
AOP	NAS Operations Program Directorate
AAF	Director of Airway Facilities Services
AOS	Operational Support Service
AOZ	Director, Free Flight Phase One
ASD	Director of System Architecture and Investment Analysis
ASY	Assistant Administrator for System Safety
ATA	Air Traffic Airspace Management
ATS	Associate Administrator for Air Traffic Services
AUA	Director of Air Traffic Systems Development

1.3 Purpose and Scope of this Document

This concept of use describes the FDR as a capability for managing application system metadata and NAS-wide data element standards. The approval of data element standards is to be accomplished through the NAS Configuration Control Board (CCB) via the NAS Information Architecture Committee (NIAC), as described in [Rhoades].

This document also addresses several related metadata efforts in the agency, especially those of the AIO [AIO 1, AIO 2], ATS and AOP organizations. This concept of use and implementation describes the FDR generally but also proposes what capabilities should be implemented initially, followed by additional capabilities that would appear in the mid-term and beyond.

Although some metadata management capabilities exist within various application systems, they are typically not coordinated outside of each system, and the concept of a metadata capability as a NAS-wide capability is new. Therefore, this concept of use discusses the main drivers that stimulate the need for such a capability. From an internal, FAA viewpoint, these drivers include information access, information consistency, system interoperability, more efficient system development, and cost reduction and avoidance. From a wider, strategic global perspective, additional drivers include the standards and templates already proposed and in place from the International Standards Organization (ISO) and other aviation bodies. Each of these drivers is discussed below.

1.4 Background

Over the last few years, there have been significant conceptual changes in the way that information is shared and needs to be shared in the NAS. In part, this is the result of several collaborative decision-making processes now in place that involve NAS users. As a consequence, there has been a shift in the data management paradigm from single Host-based, single Center-based operations to a multi-Center, Command Center, multi-system, NAS user collaborative environment. In addition, the need for information has grown in terms of responsiveness, quality, and breadth. This shift requires greater data synchronization than before and the availability of common information assets to all parties. Since the FAA's historical organizational and architectural thrusts have been to manage data assets *within* an organization and system rather than *across* organizations and systems, information functions that offer system-wide views in addition to local views are still the exception today, but that is changing.

The FAA has already established the conceptual framework for collaborative data management across organizations through several avenues:

- The ATS Concept of Operations, which describes operational views in the year 2005 for a collaborative environment that includes Free Flight.
- The NAS Architecture, Version 4.0 document produced by ASD describes NAS Information Services as the basis for data standardization, information sharing, and collaborative decision-making. ASD is now defining a future Technical Architecture as the infrastructure to better manage NAS systems.
- The NIAC, operating since early 1997, is a cross-cutting organization that is developing collaborative strategies to define and share information assets within the FAA and with related organizations, such as the RTCA, the aviation industry, and Eurocontrol.
- The AIO has produced two strategic documents in this area, an Information Technology (IT) Strategy and a Data Management Strategy, signed in September 1999 [AIO-1, AIO-2]
- There is an increased awareness of the need for coordination and cooperation by numerous FAA organizations. These include all of the organizations listed in Table 1-2 above, the FAA Technical Center (ACT), and others. These organizations cover a range of interests from policy and system evaluation to acquisition, implementation, operations, maintenance, and enhancement.
- NAS users are interacting with the FAA at many levels and through numerous organizations, such as the RTCA and ICAO (International Civil Aviation Organization), and through other working groups such as the CDM (Collaborative Decision Making) group.

- Eurocontrol is engaged in a system architecture exercise parallel to that of the FAA that has within it a significant information management thrust. The reference for their information component is SWIM (System-Wide Information Management). As in the FAA, it also reflects the need to have information strategy, policy, and some implementation occur at the broadest level of the organization while additional implementation occurs throughout the organization.

A cornerstone of sound information management is to define policies and information standards that can be used (and reused) throughout the agency. Reuse requires agreement on, and common understanding of, data names, semantics (i.e., meaning), and data representation, or format.

Given the diversity of current practice across the agency and the aviation community, there needs to be a way to manage information assets to reflect new policy and to be able to access the current status of information assets. The way that an organization accomplishes this is through a data registry with agency-wide scope. The name given here to the proposed, NAS-wide registry is the FDR. Its genesis is the NIAC, coupled with initiatives by several other organizations.

A registry is established in advance of a data warehouse, or data archive. The latter contains the data with which an organization carries out its mission and is the subject of a future study. A registry is the dictionary and directory of terms. It defines the semantic basis on which a data warehouse and any application system is built.

In describing a concept of use for the FDR, an earlier document on this subject [Schwarz] raised the following fundamental questions:

- What is a data registry?
- What functions will it perform?
- What benefits will it provide?
- Who are the potential users of this capability?
- What questions can it answer?
- How does it fit into the NAS Architecture and the NAS Information Architecture?
- How will it be developed and deployed?
- Who will manage it?

What is a data registry? The short answer is that it is an automated resource "used to describe, document, protect, control and access informational representations of an enterprise" [Parker 1]. For some, a data registry contains only standardized, application-independent, information. For others, it contains system metadata and system engineering

data as well. The concept adopted here takes the more expansive view of a data registry, but considers its development in stages, with capabilities added as needed, consistent with the definition referenced at [4].

Contrast these functions with that of a data warehouse, which the FDR is not. A data warehouse contains the actual data (i.e., operational data) that the registry describes. The FDR, as defined here, serves only registry and metadata repository functions. NAS operations would benefit from a data warehouse [5], but its structure and organization should be coordinated with and follow from, the data structures of the FDR.

The ISO has defined a widely accepted metadata registry standard in its ISO 11179 document [ISO]. This standard will become the basis for defining the FDR's content and structure.

Within the FAA, many people note that the FDR's functions are currently handled by the various application systems in the field. The problem created by such a independent or "stovepipe" approach is that each application system defines its own data — sources, definitions, relationships, and structures — creating multiple of views of the same, or similar, data elements. There are no NAS-wide data standards as there are many *de facto, system* standards. In addition, there is no guidance about data standards creation in the FAA, and information exchange today is an inefficient, expensive, and a potentially hazardous exercise, including its impact on safety. The risks to making changes are considerable, and the FDR will help to reduce costs, synchronize the data about information assets across applications and users (interoperability), and reduce those risks. The FDR is a key tool in this critical central information architecture requirement.

1.5 Purpose and Scope of the FDR

There are many reasons why any organization requires an FDR or its equivalent. Two of the more important reasons are that:

- Data are a corporate asset whose managment is as important as management of any capital investment.
- Data and database technology are of little benefit without semantic understanding of data over time and place.

---

[4] A web site with a rich set of links about metadata registries and their implementation is found at http://www.sdct.itl.nist.gov/~ftp/l8/other/coalition/Coalition.htm

[5] One important use of a data warehouse in the NAS would be to supply data for post-flight analyses, such as an analysis of delay, weather impacts, and patterns of congestion.

The FDR described here will manage two types of information, namely:

- Data and information standards at micro and macro information levels to be used in data management. This will support data exchange, information interoperability, and collaborative decision-making, and system acquisition, development, and re-design. These standards will proceed from basic data element standards (initially) to more complex standards for related data elements, such as for flight components and aeronautical data (eventually).
- Information about current (legacy) data elements used in the NAS to the extent that these data are related to a data element standard. Such information will describe the context of legacy data elements (i.e., metadata) and be associated with a corresponding data element standard. This information will also be related to equivalent information in the MDR, which will have additional metadata about the system and the context in which these data elements are used, including system ownership, distribution, access, security, and use, from NAS Management Directives (MD) and other official source material.

MITRE/CAASD has completed a companion document entitled *FDR Software Requirements Specification (SRS)*, MTR-00W0000071, September 2000, that contains a detailed set of requirements to be used to build the FDR, including an ISO-standards-based data model that would serve as the structure of the database to manage the FDR metadata.

The FDR will help to maintain traceability between the data in legacy systems and data standards defined for the NAS Information Architecture, e.g., for flight, surveillance, and aeronautical data. It will facilitate the sharing and reuse of information about NAS data with the many users of these data, including NAS users and system developers. The latter will use this information for developing new systems and for transitioning legacy systems.

The FDR will provide a mechanism for moving toward an integrated information management environment by defining data standards that redesigned and new NAS systems could or should adopt. In the interim, the new standards can be used as a transition vehicle for exchanging data between and among current systems. That is, data exchange would be done using standard definitions and formats found in the FDR. In addition, the FDR could offer the transformations needed to convert the exchanged data from the standard to formats acceptable to the legacy system. When a legacy system is upgraded or replaced and data standards are implemented, data translation would no longer be required.[6]

The core of the registry will contain the following types of information (i.e., metadata) for current and planned systems:

---

6 Data translation is more likely to be a repository/warehouse service than an FDR service but it may be included in the FDR.

- Data nomenclature and definition
- Data representation and structure
- Data ownership, stewardship, and configuration control
- Relationships among elements in the registry

Section 3 contains a more complete view of the types of descriptors/metadata that could be managed.

1.6 Assumptions

A number of basic assumptions have been made to bound the FDR in scope and development. They are described in this section.

For its initial phase, the FDR will address and contain references to NAS data only. The focus will be on those data and information items commonly shared across NAS operational systems now and in the future. To the extent necessary for the system to support international flight operations, there will be references to those interface data elements managed by other aviation organizations, such as NAS users. Administrative data will be addressed in a later phase, in coordination with the AIO and ATS organizations that already are working with those metadata.

The FDR will be based on national and international standards to the extent possible, especially the ISO 11179 standard and the comparable ANSI standard X3.285.

The FDR's scope will be NAS-wide but it will be built in increments to minimize risk, to offer some capability as soon as possible, to allow users to gain experience with it, to allow adjustments in scope and operation, to be consistent with level of current FAA funding, and to refine requirements for future builds. However, it will be designed and scoped for full implementation based on future anticipated growth and needs.

Once implemented, the FDR will support a number of related processes for managing metadata. In this capacity, it will be able to:

- Manage evolving NAS technical standards.
- Support process efficiencies in the FAA's Acquisition Management System by tracking sources and targets of data use, and by providing applications developers with FAA-approved standard descriptions of commonly-shared NAS data.
- Assist with improving system interoperability by reducing the quantity of Interface Requirements Documents (IRDs) and Interface Control Documents (ICDs) required and to help to re-format IRDs and ICDs for more efficiency and flexibility.

- Provide direct support to Integrated Product Teams (IPTs) in application development and system re-design.
- Greatly improve information access by organizing application system metadata and making it accessible via automated means.

1.7 Organization of this Document

This document contains six sections. This section contains an introduction, background material describing the context of a metadata registry, and its purpose and scope. Section 2 describes the need for such a capability and expands on how metadata information will support NAS-wide views of operations and system development and maintenance as well as expected benefits. Section 3 describes the functionality and content that the FDR will manage. Because the FDR needs to be integrated into the existing operational environment, Sections 4 and 5 describes how it might be integrated and implemented, in phases, over time. Starting with a basic capability, functionality can be added in a controlled way. Budgetary impact can be spread across several fiscal years. Section 6 offers conclusions and recommendations for proceeding.

Section 2

Need for the FDR

"A goal of NAS information services (in support of the CONOPS) is to share information seamlessly across these organizational boundaries; this requires data standardization." [ASD]

To support this goal an authoritative source of data standards within the NAS and the associated system configuration management program is needed. The FDR will be this source and it will facilitate the evolution of the air traffic management systems.

2.1 Activities That Require NAS Data Standards

A data standardization effort supported by the FDR will help mitigate the risks in making changes in the complex NAS.

2.1.1 NAS Change Management

Changes to the NAS are institutionalized through several recognized processes that need a source of NAS data standards:

- Acquisition Management System (AMS). The FAA acquisition process is currently burdened by the lack of data standards, resulting in: (1) requirements uncertainty and growth, and (2) unanticipated technical complexity in system development. Both of these problem situations impact the cost and schedule of delivering new systems to the NAS. An FDR containing data standards can provide a positive impact on the cost/schedule performance of the AMS.
- FAA Acquisition System Tool (FAST) set. The FAA, in seeking to make system-engineering resources available to Integrated Product Team (IPT) managers and their supplier-vendors, has created this tool set. However, it does not provide for data standards, a system-wide data dictionary, a way to understand the attributes of system data (i.e., metadata) or a reference to the form or model of the data to understand the context of its use. The FDR, used in conjunction with the FAST set, could mitigate these weaknesses.
- NAS Configuration Management (CM) Program. The FAA has created the NAS CCB to oversee the NAS Change Proposal (NCP) and CM program administered by the ACM organization. That program scope does not currently extend to data elements, but changes to that approach are in process. A data standardization effort, in conjunction with the NAS CM program will complete a vital, beneficial process. The FDR will be a tangible element of this effort.

- Integrated Product Team (IPT) Performance. The FAA employs IPTs to undertake research, development and engineering (RD&E) programs to bring new capabilities into the NAS. Data standards in the FDR will reduce the systems engineering complexity and reduce the data uncertainty risks for the IPT.
- FAA integrated Capability Maturity Model (iCMM) program. The FAA has initiated an agency-wide effort to promote improvements through the definition of maturity levels and organizational appraisals. Processes like quality assurance and configuration management of software and system engineering, which are standards-based work practices, are part of the iCMM evaluation. Data management and standards are intrinsic to the iCMM effort.

2.1.2 NAS Software Maintenance

Maintenance of software is complicated by the proliferation of data elements and information types used in a wide array of separately integrated computer-based systems in the NAS. The level of expertise associated with NAS software engineering and maintenance is high because of this complexity. This, in turn, drives the cost of maintenance to higher levels. It also drives institutional reluctance to change because of the perceived risk associated with changing a system that has become stable, though inefficient and costly. Data standardization will help reduce system complexity.

The number of data elements and information types used in a wide array of separately integrated computer-based systems contributes to the exponential growth of interface, or gateway, components. Each new interface creates a pair-wise need for data translation that must be documented for the maintenance program. Data standardization supported by the FDR will reduce the number and complexity of systems interfaces.

2.2 Expected Impacts and Benefits

The FDR will support numerous architectural and operational activities. These include facilitating data transformation, as needed, between systems. As a tool to help the Host computer re-engineering, it will permit design engineers to eliminate numerous patches and gateways in the data and information flows with the expectation of dramatically improving performance while reducing cost. Finally, the FDR is as a prerequisite to any data warehouse development.

Specific benefits include:

- **Interface Requirements Document (IRD) Cost Reductions**. Currently, the NAS physical architecture or systems architecture is a multi-layered computing and communications system that requires a large number of software engineers to design and maintain a multitude of interfaces. The overhead cost is high and justified by the flight safety imperative.

- **System Acquisition Cost Reductions**. The lack of standards opens the possibility for suppliers and venders to introduce unique solutions that can be declared proprietary or, though technologically compelling, require additional effort to become operationally viable. Uniqueness adds to the cost of engineering, testing, certification and operational evaluation.
- **System Design Improvements**. Currently, the NAS' behavior is driven by the physical operational architecture, which is characterized by complex, poorly documented legacy computer networks. This contributes to overall performance degradation that has marginally affected mission critical functions to date. As air transportation demand grows and facilities experience added stress due to congestion, the physical limits of the system will become more obvious and modernization will become more compelling.
- **Growth of the FAA Business Model with Data Warehousing**. The prospect for both enhancing the system performance and providing broader services to the air transportation industry and the flying public makes a clear case for establishing a data warehouse facility by the FAA. It will help operational decision-makers by organizing enterprise information and making it readily accessible. Organizing the metadata on which a data warehouse will be built is an essential step.
- **Process Improvement Analysis**. The ability to model NAS-wide data or information processes for the purposes of identifying areas of improvement is presently a complex task. It is expected that the existence of data models, metadata and data standards to improve quality and availability of operational data will support the analysis process. There is reason to believe that existing analytical tools, when provided with better operational data, will yield more efficient air traffic control systems.
- **Automation Tool Integration**. The complexity of air traffic management has grown and processing technology now supports the use of decision aids in many settings. The integration of these tools requires sophisticated transformation of diverse data formats and definitions. Establishing standards in the FDR will make information exchange among tools quicker and more accurate.

Section 3

Functionality and Content

The functionality and content of an initial build of the FDR described in this concept of use is a refined subset of the full FDR functionality described in an earlier paper [Schwarz]. The initial functionality and content will support the emerging NAS CCB data standardization processes, NAS legacy data element registration, and the NAS Architecture automation. Initial FDR capabilities provide the near term support required for these activities and at the same time provide a solid foundation for expanded metadata management processes. A number of options for the evolution and implementation of the FDR are described in Section 5.

The initial FDR will provide a single, standards-based registry for the collection, definition, documentation, and distribution of information (metadata) about data elements used widely in the NAS. The ISO/IEC Standard 11179, Information Technology—Specification and Standardization of Data Elements, Part 1, defines a data element registry as:

> *"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."*

A later ISO view offers the following definition:

> *"A registry is a tool for the management of shareable data; a comprehensive, authoritative source of reference information about data. It supports the standard-setting process by recording and disseminating data standards, which facilitates data sharing among organizations and users. It provides links to documents that refer to data elements and to information systems where data elements are used. When used in conjunction with an information database, the registry enables users to better understand the information obtained."* [Procedures for Achieving Data Registry Content Consistency, ISO draft December 1999]

The primary service of the FDR is to maintain the registry for all standardized, operational data elements approved by the NAS CCB and to provide a registry of operational legacy data elements exchanged in the NAS. As shown in Figure 3-1, the FDR registry service is built upon a number of core metadata management services and tools including a DBMS, data dictionary, and data directory. The registry service is also required to support a number of NAS-wide data management services, not in the initial FDR, such as data encyclopedias, data repositories, and data warehouses. The core services of the FDR,

including the data element registry service, are described in Section 3.1. Potential users of the FDR are described in Section 3.2. The information that needs to be stored in the FDR to support the tools and services provided by the FDR is briefly outlined in Section 3.3

3.1 Functionality

The FDR will contain the core software tools, shown in Figure 3-1, for the management of metadata [Parker 2]. Other tools may be added later. The core tools (DBMS, data dictionary, data directory, and data registry) are needed in the FDR to support the essential specification and standardization of NAS data elements. The core metadata management tools in the FDR can also provide registry services to the Capability and Architecture Tool Suite (CATS) supporting the NAS Architecture work in ASD.

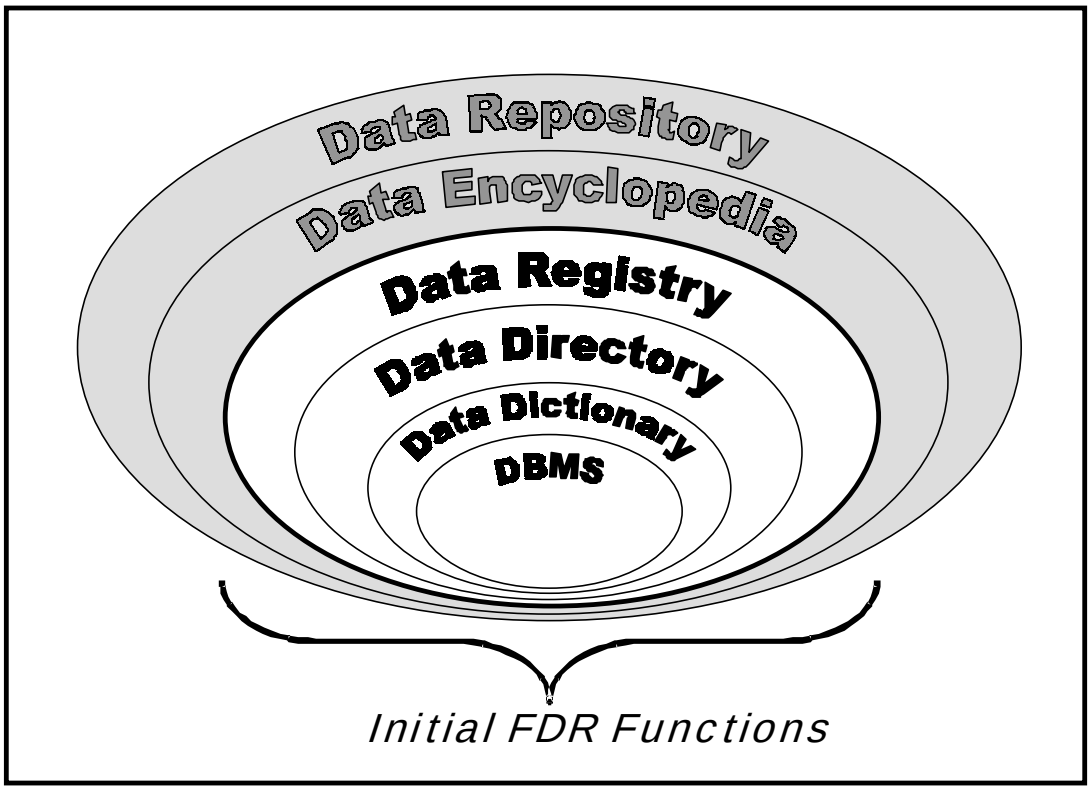


Figure 3-1. Metadata Management Functions in the Initial FDR

3.1.1 Database Management System (DBMS)

In the FDR context, a DBMS is the software environment designed to assist in the design, construction, operation, and use of databases containing metadata. Commercial-off-the-shelf (COTS) products are available to support database operations. The basic purpose of a DBMS is to organize, store, and access the instance values of the stored metadata in the FDR.

3.1.2 Data Dictionary

A data dictionary tool in the FDR is designed to assist in the documentation of data element syntax (representational form) and some of the semantics (usually only name and definition). The data dictionary may be part of a DBMS; part of a computer-aided software engineering (CASE) tool, or a stand-alone tool. Software designers and data administrators (DA) use the data dictionary.

3.1.3 Data Directory

A data directory service in the FDR is a software tool specifically designed to assist in the documentation of where data elements are used and where data sets using the data elements can be found. The directory may contain additional information on how to access instances, or values, of the data described in the directory. The directory, as stand-alone tool, provides the logical locations of data element instances that are not part of the direct FDR environment. If the FDR is expanded to maintain selected data sets, e.g., standardized (approved) domain value sets, in a DBMS, the directory could contain the physical locations where data elements are located in a database.

3.1.4 Data Registry

The core of the FDR will be a data element registry tool specifically designed to combine capabilities of the data dictionary, logical data directory, and contextual documentation capabilities of the data encyclopedia. Particular emphasis is placed on documenting the authority and responsibility for each registered data element. The purpose of the registry is to facilitate enterprise-wide understanding, standardization, and reuse of data elements used widely in the enterprise. The FDR data element registry will be a service to data managers within the FAA and to data managers in external organizations exchanging data with the FAA. The data registry will facilitate the interchange of NAS-wide data among the participants in the operations of the NAS.

3.2 Initial FDR Applications

The functionality to be supplied by the FDR does not currently exist as a NAS-wide, or even an FAA-wide, information management capability. An FAA-wide MDR is being implemented under the sponsorship of AIO and ATS. Initially, the FAA-wide repository

will capture information on all the Information Technology (IT) systems in the FAA. This systems level metadata will complement the operational data element level metadata in the FDR.

Some FDR-like functionality exists as isolated, stand-alone collections of metadata for a single system or program, and it is usually limited in scope. These existing metadata collections take many forms and are captured in various media, often as paper documents, e.g. interface control documents. The value of the FDR lies in its ability to support data and system metadata across all the NAS operational systems and programs. The accurate and efficient exchange of data and metadata across a wide spectrum of NAS users, systems, and service providers is essential to the successful implementation and realization of NAS modernization benefits

The FDR will act as a bridge between old and new systems, between current and future system views, and between legacy data and new data standards. It will be a source of current system information as well as a registry to help manage emerging information and system standards used to build future NAS systems. It will also be a tool that NAS users can use and contribute to, to represent the information and system views of their processes and to describe how these relate to corresponding FAA systems and processes.

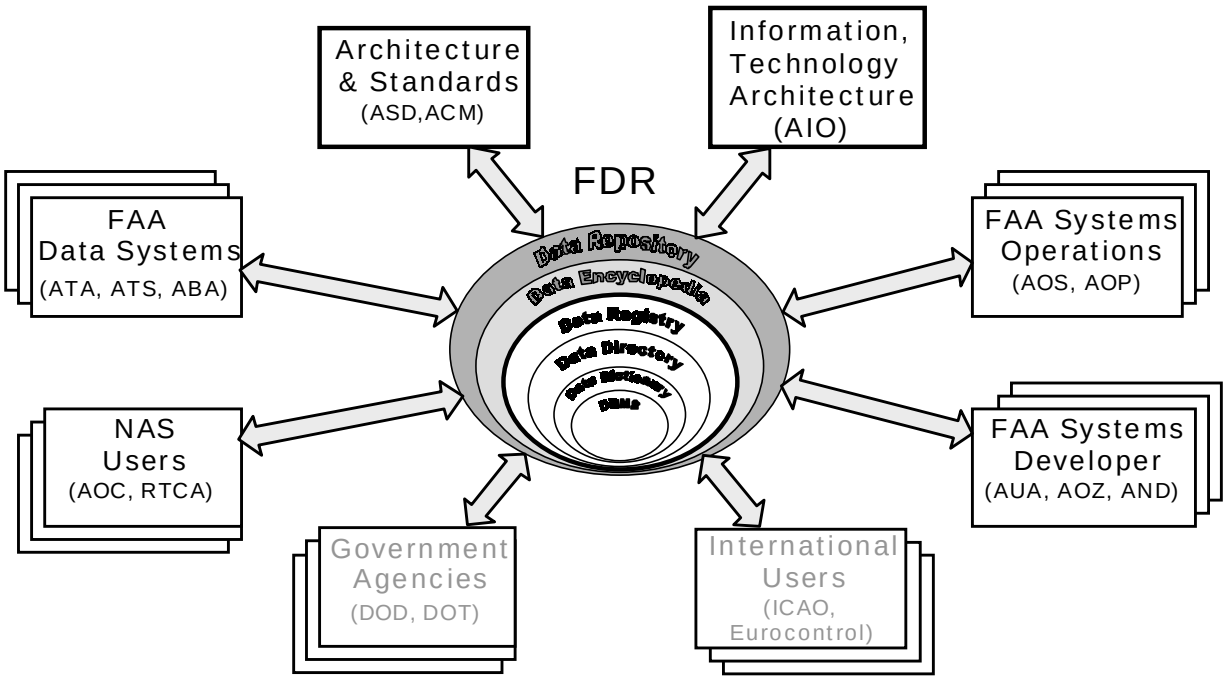


Figure 3-2. FDR Context

The many potential users of the FDR, indicated in Figure 3-2, include those inside and outside the FAA. The initial FDR will most directly support ASD, AIO, and ACM users and

those shown in dark print. The users shown in gray are more likely to be involved with advanced FDR implementations and the FAA-wide metadata repository, though they will most certainly collaborate with FAA in defining the information standards. Within the FAA, initial users include:

- ACM (NAS CCB) - working to establish data component standards
- ASD - working the future NAS Systems Architecture
- AIO – building an FAA-wide metadata repository
- AOS - working the site adaptation problem
- ATA - providing the authoritative source for NAS resources data

Within the wider aviation community initial users may include:

- NAS users - seeking to improve the collaborative exchange of aviation data.

The wide range of users, even for initial FDR functionality, indicates a requirement for easy access to FDR metadata. Options for providing this access for collection, processing, and distribution of FDR metadata are described in Section 5. The application of FDR functionality to support the various user activities is described in the following sub-sections. Because FDR functionality has not been available before, the application of the FDR to each of the activities is likely to evolve from that described below as experience is gained and as the FDR evolves to better meet user requirements. Detailed operational procedures and concepts of use will have to be developed and tested for each application.

3.2.1 Standards Development

The premier application of FDR functionality will be to support the NAS operational data standardization process being chartered under the NAS Configuration Control Board (CCB) in the ACM organization. Proposed and approved operational data standards will be recorded in the FDR's DBMS according to a specified registration process. A standard data element will be described using the basic registry attributes in Table 3-1. Administrative and identifying attributes have been added to the registry to support the NAS CCB processes (see Table 3-3).

All standardized data elements will have the same, unique value (e.g. “NASCCB”) for the “context” attribute.

All registered data elements will have a unique name (e.g., “NASCCB: FlightAsFlown_Position_Longitude_Degrees_Sexagesimal_Measure”) as derived using a specified naming convention (see Section 3.3.6.4).

Identification and prioritization of data elements in need of standardization will make use of a legacy data element baseline (see Section 3.2.2), which will also be developed and

maintained in the FDR database. Traceability from the standards to the legacy data elements can also be maintained in the FDR.

3.2.2 Legacy Data Element Baseline

The FDR will be used to establish and maintain a database of data elements as they are used in existing NAS components. Subsequent changes to usage, definition, administration, or representation of any baseline data element will be recorded in the FDR

The information on legacy data elements stored in the FDR is a key input to the analysis and prioritization that must be done to select candidates for the NAS CCB data standardization process. It will help identify the stakeholders who are most likely to be impacted by the shift to a standardized data element. These same stakeholders will be part of the collaboration needed to come to a consensus on a standard.

Even as the standardization process is working to improve and simplify systems integration, the access to the legacy data element baseline in the FDR will make the task of maintaining, upgrading, and re-engineering of NAS operational systems easier. It will make it easier for data user understand and accurately use the diverse data available in the NAS

3.2.3 NAS Architecture Development

The FDR will support the development of the NAS Technical Architecture. The Technical Architecture will capture, among many other system characteristics, the flow of information between both current and future components of the NAS architecture. For current and near-future components, the information flows will involve data elements that can be described and managed using basic data element attributes. For less well defined future systems, where detailed attributes of data elements may not be known, a less formal mechanism is needed to describe and manage the information flows between components. However, the FDR will also contain and manage a lexicon of well-defined terms from which element names, key words, or other identifying characteristics of information flows can be created to facilitate searches and maintain consistency between the FDR and Technical Architecture. (See Section 3.3.6).

3.2.4 Adaptation Process Improvement (API)

NAS adaptation data[7] represents a significant subset of the NAS data. Metadata about adaptation of current systems will be captured in the FDR legacy data element registry. The ongoing API process will use the FDR registry to record standard adaptation metadata.

---

7 Adaptation is defined as the conversion of a national system (data or software) so that it can be run in a local environment.

3.2.5 FAA-wide Metadata Repository Evolution

In the future, the FDR and the MDR will need to evolve to harmonize their content and data structures to support the exchange of metadata and the use of NAS-wide data standards. Experience gained in the use of the FDR in a NAS-wide application will provide valuable insights to the collaboration and analysis the AIO/ATS may use to establish FAA-wide registry standards.

3.2.6 System Development and Acquisition

As the implementers and users of NAS standards, the systems development and acquisition offices such as AOZ, AUA, and AND will have a vested interest in collaborating on NAS standards, as outlined in Section 3.2.1. Once standards have been approved by the NAS CCB, these organizations will use the FDR metadata registry as the source for standards included in acquisition documents. As part of systems acquisition, the metadata for data elements used by a new system would be delivered in a form easily added to the FDR registry.

The development and acquisition organizations will also play an important part in the collection and validation of legacy data elements, as described in Section 3.2.2.

3.2.7 Collaboration With NAS Users

As collaboration between the FAA and NAS users is expanded, there will be an increasing demand for more and better data from the FAA. At the same time, NAS users will provide more and varied data to the FAA. It is essential that the metadata about the data being exchanged is also exchanged so that receiver, FAA systems and user systems can make full and accurate use the data received. The FDR will provide the mechanism for recording and distributing the metadata to both parties.

NAS users will also have an interest in the standards development processes described in Section 3.2.1.

3.3 Metadata Content

The initial FDR will contain operational and system metadata about data in current systems (e.g., information about legacy data elements), data in new systems under development (e.g., CTAS, STARS (Standard Terminal Automation Replacement System)), and data standards. FDR data elements, both standard and legacy, are described by a set of attributes and associated information. The ISO/IEC 11179 standard specifies a set of basic and additional, or extension, attributes shown in Figure 3-4 below. The set of attributes is independent of a data element's use in applications systems, databases, data interchange messages, data dictionaries, and data registries. This set of attributes may be extended to enable metadata management functions not included in the initial FDR. In Figure 3-4, source

and destination metadata extensions were included in the registry metadata concept prior to the FAA repository development. In addition to the data element attribute collection and dissemination, the FDR will require a set of rules and procedures for classifying, defining, identifying, naming and registering data elements.

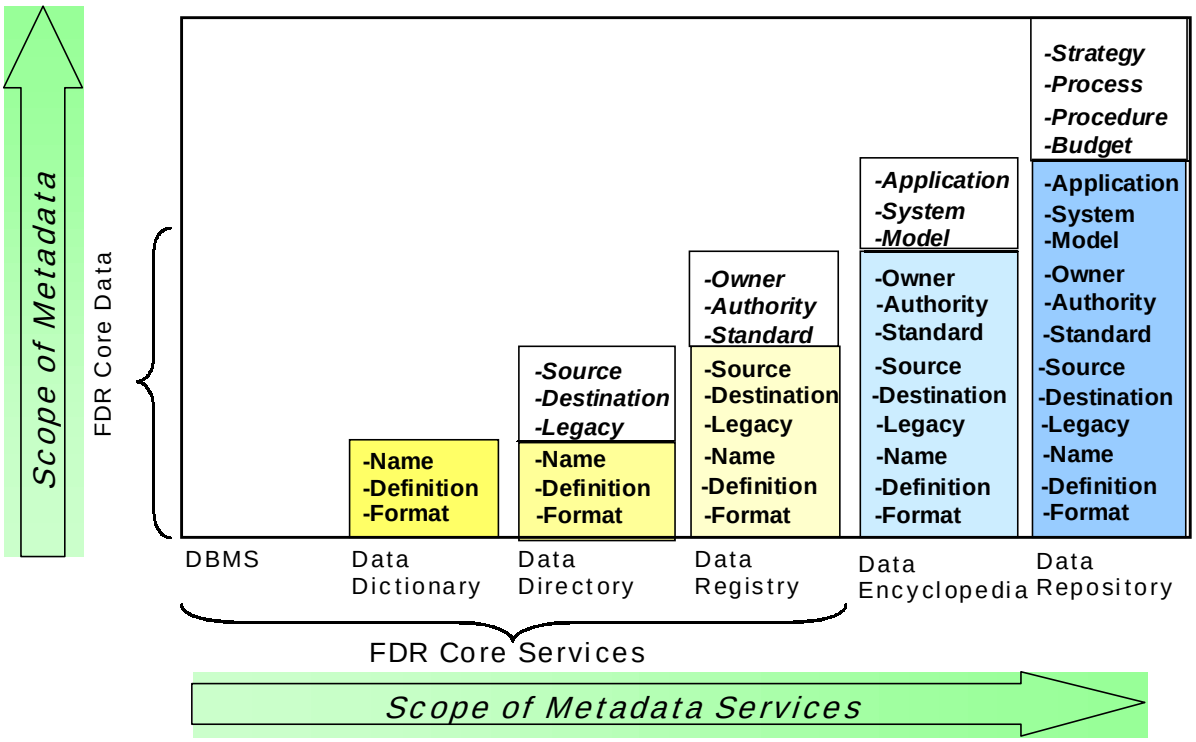


Figure 3-3. Scope of Initial FDR Contents and Functionality

3.3.1 Basic Attributes

Basic attributes are independent of a specific implementation. This does not imply that all basic attributes are required in all cases. A distinction is made between those attributes that are:

- **Mandatory:** always required
- **Conditional:** required to be present under certain conditions
- **Optional:** allowed but not required

Categories of basic attributes include:

- **Identifying:** attributes regarding the identification of a data element

- **Definitional**: attributes that describe semantic aspects of a data element
- **Relational:** attributes that describe associations among data elements and associations between data elements and classification schemes, concepts, objects, entities, applications, systems, or programs.
- **Representational:** attributes that describe form and format of a data element
- **Administrative:** attributes that describe management and control aspects of a data element.

Table 3-1 lists the basic attributes for each of the five categories shown directly above. Definitions of all attributes are summarized in Appendix A.

Table 3-1. Basic Data Element Attribute Descriptors

Category	Attribute Descriptor
Identifying	• Name • Name context • Data identifier • Registration authority identifier • Version • Registration authority • Synonymous name • Synonymous name context
Definitional	• Definition
Relational	• Classification scheme identifier • Classification scheme item • Keyword set identifier* • Keyword(s) • Related data references • Relationship type
Representational	• Representation class • Form of representation • Data type • Maximum character quantity • Minimum character quantity • Format • Permissible values • Examples
Administrative	• Responsible organization

Category	Attribute Descriptor
	• Registration status • Submitting organization • Comments

3.3.2 Additional Attributes

The basic attribute list can be extended to meet FAA administrative requirements. An FDR prototype developed in FY99 using the Ptech, Inc.'s Framework environment [AOP] identified a requirement for additional attributes beyond those suggested in ISO/IEC 11179. A combined list of additional attributes is shown in Table 3-2.

Table 3-2. Additional Data Element Attribute Descriptors

Category	Attribute Descriptor
Identifying	• Short name
Definitional	• Formula • Precision • Largest value • Smallest value • Unit of quantity • Data models (e.g., ERD, UML)
Relational	• Data element associations (composites) • Procedural information
Representational	• Valid usage category • Units • Quality
Administrative	• Date submitted to registration authority • Registration status • Security-protection category

3.3.3 Additional Standards Registry Attributes

Other attributes have been identified to support the proposed NAS standardization process, as listed in Table 3-3.

Table 3-3. Additional Attributes to Support NAS Data Standardization

Category	Attribute Descriptor
Identifying	•
Definitional	•
Relational	•
Representational	•
Administrative	• Case file number • NCP number • CCD number • Date case established • NAS CCB status • Valid from date • Valid until date
Reference	• DOORS File number • NAS Data Model Reference number • Activity Data Model number • Technical Architecture Key • CDIMS Collaboration Record • FAA-STD-xxx • General References
Usage	• Implementations • Equivalences • Transformations • Transition schedule

A number of the references in the table above are explained here:

- **Case file number.** A unique number assigned by the NIAC Executive Secretary to a data standardization task undertaken by a NIAC chartered working group (WG). It is a tracking aid and will point to a statement of work managed by the NIAC.
- **NAS Change Proposal (NCP) number.** A unique number assigned by the NAS CCB control desk to a case file upon acceptance of the case file for processing. It is a tracking aid and will point to a configuration control file managed by the NAS CCB.
- **DOORS File number.** A pointer to a Dynamic Object Oriented Requirements System (DOORS) file that hold the background requirements statements for the data element(s) now standardized in this effort. The DOORS file will physically be located with the Capability and Architecture Tool Suite (CATS).

- **FAA-STD-xxx.** A formal standard authorized by the declaration of a Configuration Control Decision (CCD) when the board (CCB) has approved an NCP. A draft of the standard is a part of the case file and the NCP.
- **Activity Data Model number.** A pointer to a file that shows the fundamental entity-relationship diagrams (ERD) for the standardized data elements. The data model may be created by use of one of the approved modeling tools, e.g., Erwin.
- **NAS Data Model Reference number.** A pointer to a file that shows the relationship of the activity data model to the larger NAS frame of reference. The NAS Data Model will be identified as to its version number and version date and coordinated with the NAS Data Model Steward.
- **System Namespace.** The attribute of the standardized data elements/concepts that shows the names of the systems that are applying the standard. The namespace is to be updated periodically by the NAS Data Steward to provide a current state of the NAS Technical Architecture.
- **NAS Data Steward**. The organization and name of the individual that is responsible for maintenance of the standard data element/concept. All subsequent changes must be coordinated with this party. Questions about the employment, waivers or other mitigating effects must be coordinated with this party.

3.3.4 Additional Legacy Data Registry Attributes

Table 3-4 contains additional legacy data attributes to support legacy registration.

Table 3-4. Additional Attributes to Support Legacy Registration

Category	Attribute Descriptor
Identifying	
Definitional	
Relational	• Legacy system name
Representational	• Representation description
Administrative	• Legacy system status • Data element standardization priority
Usage	• Legacy applications • Valid usage constraints • Activity data model • Quality of service

3.3.5 Additional Complex Data Element Attributes

The basic data elements derived from ISO 11179 are designed primarily to register simple or atomic data elements. The attributes in Table 3-5 were added to begin to support the registration of more complex data structures. It is expected that this may evolve rapidly in response to new initiatives from the standards working group responsible for ISO 11179. In the interim, the following attributes are proposed.

Table 3-5. Additional Attributes to Support Complex Data Structures

Category	Attribute Descriptor
Identifying	
Definitional	
Relational	
Representational	
Representational-complex	• Complex representation type • Description • Components
Administrative	

3.3.6 Controlled Metadata Values

A number of standards are emerging that may be applicable to the creation and operation of the FDR. As in many standards areas, there are overlapping and conflicting standards proposed for metadata management. Whatever standards are selected will be made accessible through the FDR. The FDR will contain metadata about itself. Standard aviation specific keyword lists and data categories will be needed to organize metadata in the FDR. Other taxonomies are in development as well. One of these is ICAO's Aviation Registry [ICAO].

Metadata registry standards are needed to facilitate the interoperability of the FDR with other metadata registries within and external to FAA organizations. Standards are needed in a number of metadata related areas. These are:

- Keywords
- Category
- Taxonomy
- Naming conventions
- Namespaces

For basic metadata registration, the ISO/IEC IS 11179, Information Technology - Specification and Standardization of Data Elements [ISO], is being applied to a growing number of government data management systems. The Object Management Group (OMG) has developed the Meta Object Facility (MOF) as a standard for the exchange of metadata about object oriented data structures [OMG]. The Meta Data Coalition (MDC) has adopted the Open Information Model (OIM) from Microsoft Corporation.

The FDR will need to contain lists of approved (registered) keywords, categories, and taxonomies, as needed, to enhance the creation and accessibility of metadata in the FDR. Existing keyword lists, categories, and taxonomies will be included in the FDR after collaboration with active FDR users.

The registry will contain NAS-approved names, at least for NAS CCB approved standard data elements, that bring consistency, clarity, and uniformity to naming conventions. What may seem like a fairly inconsequential matter at first glance may be extremely important for data names that are used in application systems in an operational setting.

3.3.6.1 Categories and Classification Schemas

To enhance the accessibility of metadata in the repository, each controlled element (e.g. legacy data element) can be assigned to one or more categories or classifications from one or more category schemas. Pairs of attributes ("classification schema" and "classification entity") will contain the name (unique identifier) of the classification schema and an identifier locating the data element within the classification schema. The initial version of the FDR will contain a mandatory, predefined categorization schema that must be used to assign at least one classification to every data element entered in the FDR.

Categories or classification schemas can be added to improve access and analysis of selected subsets of data elements documented in the FDR. The names of the classification schemas will be added to a controlled list of permissible values for the "classification schema" attribute. Procedures for adding or modifying classification schemas in the FDR are yet to be determined, but they should be similar to the data element registration procedures.

3.3.6.2 Keywords

The use of keyword attributes provides another mechanism for grouping and analyzing the data element descriptions in the FDR. The mandatory use of one or more keywords from a predefined, fixed set of key words (a lexicon) could be an alternative to the mandatory classification described in Section 3.3.6.1. Procedures for adding or modifying keywords in the FDR are to be determined.

3.3.6.3 Namespaces

A controlled set of values may be used in the "context" attribute to differentiate data elements with the same name but having different representations or definitions. The standardized data elements in the FDR will share a common namespace value indicating that they are unique and approved by the NAS CCB.

3.3.6.4 Data Element Naming Conventions

A naming convention will be established for use in the data standards process, and it will be applied in the FDR as well to provide consistently formed, unique names for the standard data elements entered into the registry. The same naming convention may be used to generate a supplemental name for the NAS legacy data elements as per Section 3.2.2. A naming convention proposed by ASD-130 [Uri] based on a Basic Semantic Unit (BSU) may be generalized, a process that should be explored for long-term data registry usage. A similar naming convention is outlined in Part 5 of ISO 11179.

Section 4

Integration

The FDR has a significant role in the modernization of the NAS as a resource and tool in the development of data and information standards. This section describes options for integrating the FDR with related information management activities.

4.1 Data and Information Standards

The data and information standards process in the NAS will be coordinated through the NAS CCB administered by the ACM organization. The NIAC will be chartered by the NAS CCB to recommend NAS information components for approval as standards.

The FDR will be a widely accessible resource of NAS legacy metadata so that agency life-cycle management interests are considered in the creation of new standards. It is expected that new standards will be coordinated with those parties responsible for sustaining systems that use the legacy data, and subsequent processes will coordinate transitions from legacy to new, NAS-wide standards. The transition process is expected to be an incremental change, in which technical refresh, system replacement, and new capability development will provide opportunities to introduce data standards.

4.2 Capability and Architecture Tool Suite (CATS)

The ARA organization of the FAA is creating a suite of computer-based tools and databases called CATS that is expected to support the maintenance and modernization of the NAS architecture. The CATS is designed primarily to support resource allocation from RD&E to procurement and fielding of upgraded or new hardware and software that enhance the capabilities of the NAS. It is intended to provide a timely snapshot of the NAS with life-cycle systems features at a detailed level. CATS provides two views of the NAS:

- A **logical architecture view** that portrays NAS systems with logical data and information flows and transaction flows. Connectivity is shown as lines with broad implications regarding the method and type of information that moves from one location to another.
- A **technical architecture view** that portrays systems at the detailed level. In this setting, details about data elements are known, to include specific interoperability issues. Modes, rates and other specifics on the movement of data and information are documented. In the longer term, the technical architecture will be a rich area for data and flow analysis, assessment and modernization.

The FDR is anticipated to be a fully accessible, database facility in support of CATS. Its integration into the suite of hardware and software has not been fully determined at this

writing, but it has the functional role of making data element descriptions available to CATS users when they drill down to the appropriate level of detail.

4.3 Acquisition Management System (AMS)

The FAA's Acquisition Management System (AMS) is a policy and process designed to address system life-cycle management. The FDR, once established as a standards registry, will be called for and applied to the systems acquisition process. Specifically, a suite of supporting tools called FAST, will be expanded to call out and make available this resource to those parties that are participating in the acquisition management process, be they agency procurement specialists or vendors composing a proposal for submission.

The FAST can be accessed online at http://fast.faa.gov/toolsets/ and it is expected that the FDR could, through a planned integration process, become a part of FAST. It may reside within CATS, but a graphical user interface (GUI) from a FAST viewpoint can be engineered to support acquisition management. It is expected that, over time, several AMS policy and planning documents will be updated to reflect the presence of FDR-based standards and their implementation. Specifically, the following policy documents may be modified:

- Mission Need Statement – Those missions that move information and data will be expected to articulate not only their information needs, but also the need for standard data.
- Investment Analysis Report – The life cycle costs associated with new or upgrades to systems will be expected to indicate the expenses associated with making changes to comply with standard data in the appropriate fiscal year.
- Acquisition Strategy Paper – The product lead will be asked to formulate a strategy or strategies that are based upon standard data. Alternative strategies may be necessary to achieve modernization along with the evolving standards process.
- Requirements Document – NAS-wide requirements will evolve to mandate standard data and these standards will necessarily be shown in the product requirements document for computer-based or digital communications systems. In those near-term situations where standards have not been specified, or with newly developed NAS functionality, the development of system requirements must detail down to the information level or data requirements as well as support the standards development process for the new requirements.
- Acquisition Program Baseline – A product team's capability, cost and schedule data must provide details to the standard data level and demonstrate conformance to the data standards, or else obtain a negotiated waiver of standards. Interdependency waivers must show standards conformance and non-conformance, so that future NIAC efforts can focus on achieving full conformance to standards.

- Integrated Program Plan – The plan must show where data standards are needed and the due date for standards approval to meet the overall flow and schedule. This document will aid NIAC's work prioritization and coordination.

4.4 Metadata Repository

The FAA's CIO and AIO have published documents proposing an FAA data management strategy [AIO 1] and an information technology strategy [AIO 2]. One of the features of their effort is the development and maintenance of the MDR. The data management strategy includes the development of a repository to manage current system information. AIO is collaborating with ATS in building this resource.

The FDR's NAS data element view complements the system-level view taken by the MDR. As described in Section 3.2.5, the metadata managed in these views must be harmonized so that, for example, FDR metadata associated with the system(s) in which a data element is used will allow linkage to the MDR metadata describing that system(s).

Section 5

Implementation and Deployment

This section discusses various deployment options for the FDR. One of the key architectural issues regarding deployment is whether there will be a single agency-wide FDR or whether there will a number of independent, but federated (linked in some fashion), registries. The following three paragraphs address this issue while the remainder of the section discusses implementation in a broader context.

In very large organizations, such as the Department of Defense (DoD), in which metadata are being managed by many sub-organizations with large missions, it has become evident that it is not realistic to create a single physical entity that encompasses all of the organization's metadata. Rather, in such an environment, a federated approach is recommended, in which metadata continue[8] to be managed within a sub-organization and interfaces are defined among them so that there is a single, virtual metadata view for the organization as a whole. Colleagues at The MITRE Corporation and at Boston College have described this environment at the DoD in a recent paper [Rosenthal] and draw conclusions about the most pragmatic and efficient options for sharing metadata in a large, diverse organization.

Circumstances at the FAA are different from the DoD in several respects. First, the FAA, although large and having multiple missions, is not as large or diverse as the DoD. Second, unlike the DoD, there is not currently an active metadata management activity in many parts of the agency. Third, several metadata management activities in the FAA are relatively new and still have considerable flexibility in how they will be implemented. All of these factors point to the possibility that the FAA may be able to implement a metadata capability for the NAS side as a single virtual capability whose physical representation should have a manageable number of interfaces. However, the challenge of this approach should not be underestimated.

Fortunately, there are workable alternatives to integrating metadata if an agency-wide standards approach is not feasible. As documented in [Rosenthal], there are variations that employ data mediation, data translation and transformation, interface management, and common schemas. They all require cooperation between a (lower level) development activity, usually seeking autonomy, and a (top-down) coordination activity, seeking to integrate information across the organization. In any case, a federated approach will likely

---

[8] Often, in speaking or writing that uses the term *data*, it is used in the singular form. Since the term *data* is the plural form of the term *datum*, a word that refers to a singular item of information, this document uses the intended plural form for *data* and its companion term *metadata*.

be necessary to interface NAS metadata with the FAA's business metadata, whose development is now underway in the ABA organization and with the MDR product, as mentioned.

This section describes issues of implementing the FDR with a focus on the following topics, which are discussed in the subsections below:

- Metadata content and design (Section 5.1)
- Metadata acquisition (Section 5.2)
- Metadata access and dissemination (Section 5.3)
- Metadata architecture, including the distribution of metadata components across facilities and domains (Section 5.4)
- Implementation timing and schedule (Section 5.5)
- Management of the metadata capability: ownership and stewardship (Section 5.6)
- Risk management (Section 5.7)

5.1 The Process of Managing Metatada Content and Design

Section 3 above provides considerable detail about the proposed content of the FDR. The content derives from the purposes for which the FDR is designed, as discussed earlier. The data structure for storing and managing these metadata will be a relational, object or object-relational design targeted for implementation in a COTS database or metadata product. [9] The design of an implementation architecture is discussed in Section 5.4.

Section 3 above describes the range of metadata content that the FDR is capable of managing. That, together with the use of metadata standards, such as ISO 11179, will determine the specific information design required for the FAA. Beyond this design remains the issue of what organizations will contribute content to the metadata capability. On the assumption that this will be an FAA-wide activity, it follows that there are many organizations that will be involved in determining the FDR's metadata content. Some of these FAA organizations include:

- ATA: This organization operates the air traffic system, consisting of service providers, analysts, traffic management coordinators (TMCs), and other users of the application systems. It also is responsible for managing aeronautical data and

[9] Decisions on what type of technology to use and what specific products to use is not in the scope of this document. However, it is intended that information technology products will be used to manage this resource.

Notices to Airmen (NOTAMs) and the data standards on which these data are based. Along with many of the organizations listed here, it is involved in an effort to streamline the definition (e.g., metadata), organization, management and implementation of adaptation data in the Adaptation Process Improvement initiative (API).

- AUA: The Integrated Product Teams (IPTs) in this organization are responsible for the acquisition and development of the major NAS application systems, and this is where the majority of the metadata content exists. At present, the metadata for each system are managed locally within each system, and they are typically not visible or available outside of that system. In many cases, these metadata are not managed by a COTS product, but by customized code that is associated with the application. Today, these metadata are not routinely shared and no formal processes or automation exists for sharing them.
- AOS: As system maintainers, this organization manages a large amount of metadata, not only for the many application systems across the agency but also including the critical area of adaptation data. They, too, operate on a system-by-system basis with little capability to see across systems or to manage for commonality, consistency or reuse. This could change if they were to have access to system-wide metadata and could make a case for using common data standards to obtain life cycle cost and processing efficiencies.
- ASD: The system architecture organization, with its various system views and information products (e.g., the NAS Architecture, Technical Architecture, and the CATS) maintains considerable metadata (and data) about system services as well as about programs and budgets. The latter metadata could become part of a companion, administrative (non-NAS) metadata resource.
- AIO: In addition to its information policy role, this organization is building the MDR metadata resource in teaming with ATS. Their initial effort is directed at collecting and organizing the metadata efforts already underway and in compiling an inventory of metadata about application systems across the FAA.
- ATS: Over several years, this organization has built a metadata resource and data warehouse for a large amount of administrative metadata. With a new AIO partnership, they are expanding that scope to include metadata from a large number of FAA systems. That work will be coordinated with the effort to build the FDR.
- AAF: This organization manages a large amount of data about the equipment and facilities in the NAS, from radars to computers to communications systems. The metadata about this equipment and facilities data are required for a complete picture of the NAS operational data. Within AAF, AOP-600 is developing a metadata

resource describing the data to manage telecommunications resources. That effort will also be coordinated and integrated with the building of the FDR.

- ABA: Responsible for the business side of the FAA, this organization is currently involved in a major effort to document better that aspect of the agency's operations. The goal is to have stronger links between the operational side of the FAA and its corresponding costs. This effort has a large metadata component. When completed, ABA will establish links to the various application systems on the operational side. ABA is mentioned here, not because it has additional NAS metadata to contribute, but because the data and metadata that they manage are indispensable components of an FAA view. At some point, the two views will need to be coordinated or federated in an FAA metadata registry, as proposed by AIO.
- ASY: This organization operates the National Aviation Safety Data Analysis Center (NASDAC) and is a participant in establishing a International Aviation Data Registry. It is interested in a number of data element and data model standards, starting with an aircraft's manufacturer (i.e., 'make'), model, and series.

There are several important cross-organization efforts underway as well. These include:

- **Adaptation Process Improvement (API).** This activity is a cross-cutting one in that it affects numerous FAA organizational units, including system acquisition, development, and maintenance, operations in multiple domains, and airspace management. Adaptation refers to the need to customize operational data for specific functions, activities, and locations. The API initiative seeks to streamline the management of data used in adaptation in coordination with the application systems that require data to be adapted.
- **Integration of aeronautical data systems: ATA-100's NASR, AVN's IAPA, and NOS.** The NASR system manages NAS-wide aeronautical data and is a key supplier of such data to almost all systems in the FAA that require such data and also supplies such data to the global aviation community. The IAPA system manages instrument approach procedures and supplies this data to NASR. For many years, the automation branch of the National Ocean Service (NOS) in the Department of Commerce (DOC) has produced aeronautical charts under contract to the FAA using the data in NASR and its predecessor system. NOS will become a unit of the FAA in FY01.

  These three organizations and their respective products seek to improve the coordination of their processes and their management of data. Part of this effort requires a more integrated view of each system's metadata.
- **NASR/2000.** This program intends to coordinate a number of operationally-based systems to support airspace decision-making from real-time to post-event analytic.

Constituent systems include NASR, NOTAMs, OE/AAA, and a NAS archive of operational data.

- **Numerous AUA application program developments.** There are hundreds of systems in use and in development in AUA. Each is developed on its own with interoperability addressed after the fact. This process has allowed each system to define its own metadata. Most of the activities described here interact with these systems. AUA systems interact with one another. At present, there is no integrated view of AUA system metadata nor of metadata used within a domain. This has impact on development time and cost, on system interoperability, and on interoperability with NAS users and their systems. If scoped properly, the FDR will be able to assist AUA application systems to better manage their metadata, and by implication, their operational data.

The lines of interest and ownership of NAS and FAA metadata are complex in that the wider aviation community also has a large stake in how these data are defined and managed. Discussions with NAS users via the RTCA and the CDM group clearly indicate that NAS users are very interested in:

- A more comprehensive way in which they can access FAA data and metadata so that they can build their systems for greater accuracy and efficiency.
- Consistent use of data by FAA systems. That is, the industry wants the FAA to use standard data naming, definitions, and formats to minimize the ambiguity of the data they receive.

Access by NAS users and others in the aviation community can be controlled and limited so that airspace integrity and security are not compromised. Policies will be needed to govern metadata information sharing with NAS users. This process is also a two-way street. There are data and metadata managed by the aviation community that the FAA also wishes to access. Discussions about sharing such data have occurred over the past few years, but they haven't had a clear focus that begins with metadata. The development of the FDR can help clarify and focus those issues.

From this discussion, it is clear that developing a NAS-wide metadata capability will require the cooperation of many FAA organizational units. There must be agreement on the scope of the metadata to be managed and the process by which the data will be collected, managed, secured and disseminated. Operationally, there needs to be a process that converges and generates acceptance of the concepts described here for how a metadata capability will function. This integration and implementation plan will be supplemented by a set of requirements for the FDR. The NIAC, as a cross-cutting organization established to serve NAS-wide information management needs, is a prime candidate to host a discussion of the FDR and its role, as it has already been doing. However, many organizations, including the ones listed above, must play an active role for this activity to be successful.

Following an agreement on requirements, a design for a NAS metadata registry/ repository[10] can be started. This activity, too, needs to be coordinated across key FAA organizations, and the NIAC is proposed as the coordination organization, supported by the policy and oversight role of the AIO organization. Once an organization is selected to design, build, implement, and maintain the registry/repository, whether in-house or by a contractor, that organization will be responsible for managing the inevitable changes to the metadata content and the corresponding design of the metadata structure. After the metadata design is populated by agency metadata, it is anticipated these structures will become the basis for a data warehouse to manage the corresponding operational data associated with the metadata. This further capability will have direct impact on overall system interoperability as well as the many redesign activities soon to be underway, including especially a Host redesign, with its impact on so many allied systems.

It will be necessary to maintain a network of organizations with an interest in system metadata design as a way to feed recommendations for changes and improvements from this interested constituency. NIAC's FDR working group, already established, can fill this role.

Once the issue of responsibility, organization, and metadata structure has been covered, two other important operational aspects can be addressed. These are a) data acquisition and b) data access and dissemination, which are discussed below.

The design of the FDR metadata structure envisions two main threads: a) metadata that describe current NAS data elements in use, and b) metadata for data standards that will be proposed and approved via the NIAC and NAS CCB [Rhoades]. The data structures for these two threads will have some differences. For example, each legacy data element will be associated with the system and the system messages in which it is found while a standard data element won't have such associations.

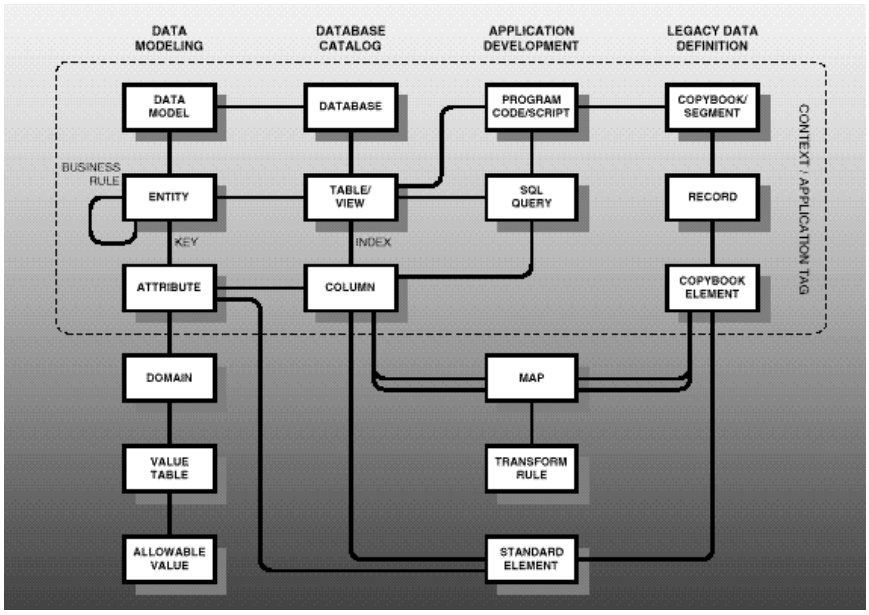
These two threads will be associated in that each legacy data element from an application system will be mapped to a standard data element, as these standard data elements become defined and approved.[11] This will involve numerous many-to-one relationships since there are presently many varieties of the same data element concepts in use throughout the NAS application systems. By associating all of the 'like' elements with a standard element, users will be able to search and find related data elements across all FAA systems.

---

[10] Recall that Figure 3-3 illustrates a growth path for the FDR. It must certainly act as a registry for data standards. Initial requirements may or may not also include a metadata repository capability. Ideally, they will.

[11] A linkage will still be created with a representative data element even before a data element is approved as a standard so that there can be a mapping from 'like' elements to a common element.

Figure 5-1 illustrates the relationship between legacy data and data that have been organized in a data model. It is not required that the data elements used in a data model be standard data elements, but that is an option. The figure indicates that each legacy data element will be mapped to a column in a database catalog, which will be associated with a particular modeled data attribute while a transformation routine will convert one form of a 'like' data element to a standard form. A standard data element will be associated with each legacy data element and linked to a data attribute in a data model. Over time, it is anticipated that these more complex data model views of collections of related data will also become common, or standard. For example, the FAA has developed or is currently developing data models for flight, adaptation, aeronautical data, geographic representation, communications and telecommunications facilities, and numerous other key subject areas. If these data models are to be representative and inclusive, they will need to be based on common and ideally, standard, views of data elements and their relationships.



Courtesy: Robert Seiner, Ciber, Inc.

Figure 5-1. Metadata Content and Relationship to Data in Legacy Systems

It is planned that each existing metadata element will be associated with a data category that represents the context of that element. To accomplish this, MITRE/CAASD has developed a fairly detailed data categorization scheme whose purpose is to offer a full

decomposition of NAS data into a set of mutually disjoint categories in such detail that a mapping for any existing NAS data element to a categorization will exist [Broste]. It will be part of the foundation for managing existing data elements and new data element standards. Figure 5-2 illustrates the process of mapping numerous 'like' data elements to a NAS-wide data element standard. This process was successfully applied to a large set of data for nine Traffic Flow Management (TFM) systems [Bolczak].

In Figure 5-2, consider variations of element A to be a flight identifier in Systems 1, 2, 3, and 4 (the left-most column, or rectangle, in the figure). Realize that today, the names of these data elements will generally be different, as will their formats. Possibly they will have the same meaning but not necessarily. In any case, each of these flight ID variants will map to the category for flight data and be associated with the data element representing a flight ID, shown in the middle column. Here, elements 1A, 3A, and 4A represent flight Ids from three systems. Finally, all three elements are mapped to the equivalent standard flight ID data element (element A) in the right-most column.

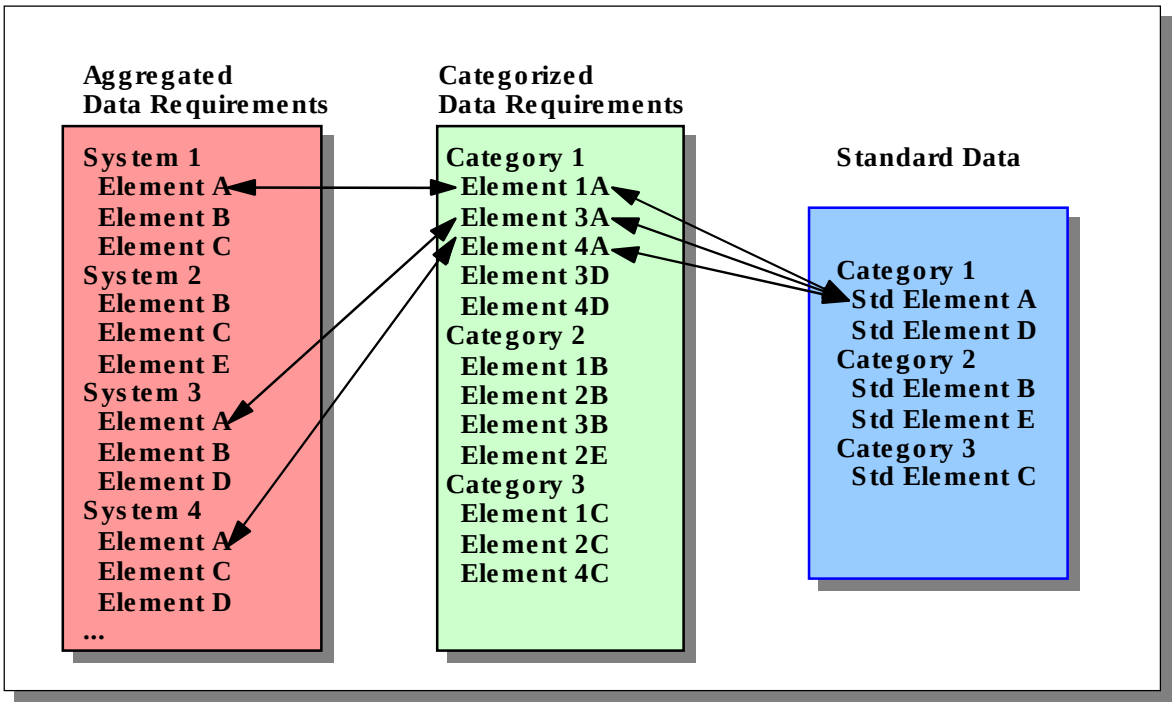


Figure 5-2. Mapping Legacy Data Elements to an FDR Data Category

Another process to consider is one that manages changes to the FDR's metadata schema. First, a schema is a dynamic product. The types of data one chooses to model and to manage

change. In this case, the metadata to be managed by the FDR will change or the descriptors of existing metadata may change. For this reason and others, the FDR requires a configuration management process to manage these structural changes. These changes must be recorded for audit and tracking purposes. Changes to the metadata schema will affect validation and edit routines and will impact the format and content in which metadata are sent to the FDR. These changes need to be conveyed to sending organizations so that they can modify their routines accordingly. Configuration management will ensure that changes to the schema and routines are coordinated and will take effect simultaneously, if that is required.

In summary, the following actions were discussed in this section and require resolution:

- An agreement is needed on metadata content.
- A mechanism (e.g., leadership, processes, computing and communications resources) is needed to obtain cooperation of numerous FAA organizations to define, collect, manage, and disseminate NAS metadata.
- A mechanism is needed for NAS users to participate in this process.
- NAS metadata and administrative metadata need to be interfaced, or federated, to obtain a full view of agency metadata.
- A decision is needed on an organization to coordinate this activity. The NIAC and AIO are candidate organizations for this role.
- The FDR will require a configuration management process.

5.2 The Metadata Content Acquisition Process

This section describes how metadata can be acquired from the many sources that own or manage metadata. The process is divided into two parts, the first being the initialization of the FDR (Section 5.2.1) and the second describes the ongoing process of collecting new and updated metadata (Section 5.2.2).

A formal requirements document will specify in a rigorous way what metadata will be managed by the FDR. It will be written after the completion and approval of this FDR implementation and use document. At this time, it is sufficient to note that the types of metadata to be managed are described in Section 3 of this document.

A key component of the metadata management capability in the FDR is a set of mandatory data categories as described in Section 3.3.6.1. Such a set of categories has been proposed by CAASD and requires review and discussion before acceptance. That process is occurring now. When approved, in whatever structure and format, these categories will become part of the FDR to help organize the vast amount of legacy information that exists in FAA systems today.

5.2.1 Initializing the FDR

There are a number of steps required to initialize the FDR and to populate it with existing NAS metadata and new data standards. These steps presume that the roles required to manage the FDR have been recognized and filled and that the FDR's logical (i.e., data structures) and physical (i.e., COTS software and computing capability) components are in place.

A prerequisite to implementing a physical database for the FDR requires that a logical and a physical database design has been constructed under the guidance of a data administrator (DA). To start, separate, but related, database designs should exist for legacy and standard data element views.

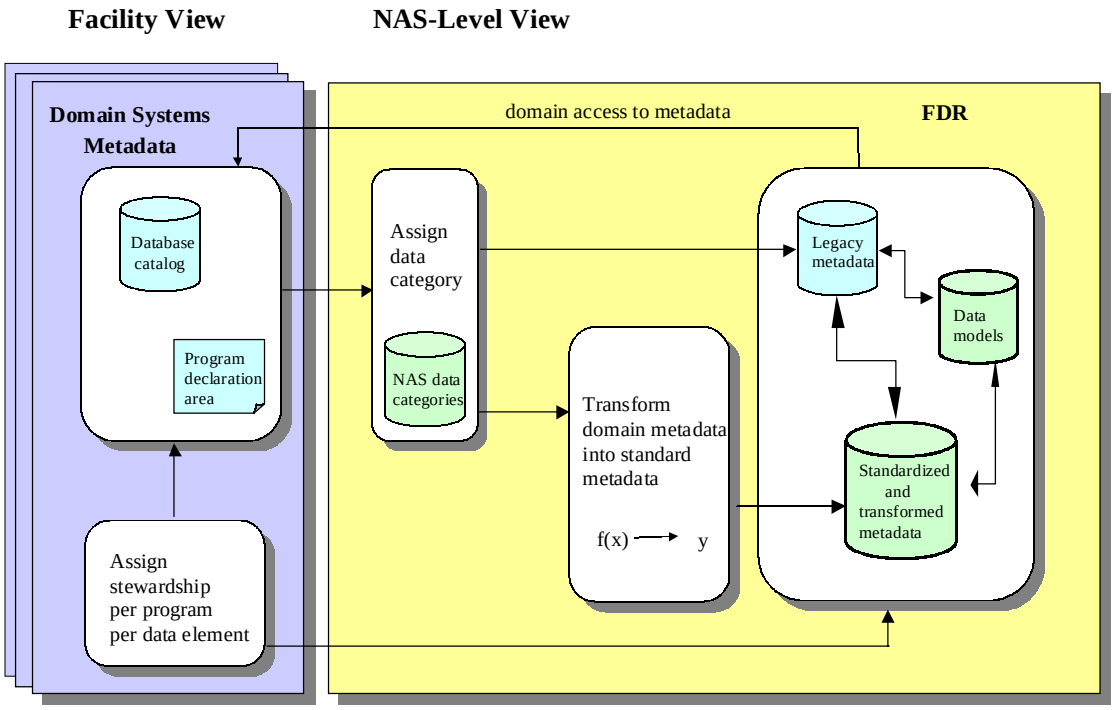


Figure 5-3. Initialization and Operation of the FDR

On the system and computing side, the physical database design, or schema, must then be instantiated in a software product, preferably a COTS product suited to managing metadata. The database must be sized for initial and ongoing volumes, and policies and processes must be in place to store, update, backup, recover, and secure the metadata database.

With these conditions satisfied, Figure 5-3 illustrates the initialization and operation of the FDR. In the figure, on the facility or domain/system side, i.e., at the location or organization that manages a system, there must be a mechanism to access and extract a

system's metadata into a well-defined format that can be imported by the FDR into its metadata schema. For application systems that are based on a COTS DBMS, it will be easy to access this information since it will be stored in a system table accessible by a SQL query. For non-COTS-based systems, it is very likely that an extraction routine will be needed to assist in the transfer. However, the values (i.e., metadata information) from each source system should be kept intact. These values must be accurately reflected in the FDR. For example, the FDR will contain the exact name, definition, and format used for each legacy data element from each application system. It will also contain additional types of information about the system, as shown in Table 5-1. For standards metadata, the standards registration search paths might include registration status, approval status, the cost and priority of standardization, NAS requirements, and traceability.

Viewed both as content and as possible search paths, these macro information subjects offer the scope of metadata that may be managed about NAS data elements. Note that this table takes an application system viewpoint, offering a broader systems view than would a data element viewpoint alone. Some of this information will be captured in the MDR. If needed, such information should be able to be accessed and extracted for use in other applications, including that of populating the FDR. The issue of federating various metadata applications, such as the MDR and the FDR, remains.

Table 5-1. FDR System Search Paths

Functions	Locations
Application code	Transformation rules
Database catalog	Subroutines
Messages	Users
Data sources	Data sinks/recipients
Data elements	Queries
Reports	Assumptions
Interfaces	Business rules
Databases	Performance goals
Data stewardship	Security features
Backup/recovery processes	Maintenance schedule
Domain	Cost and budget

Returning to Figure 5-3, it shows that another initialization task is to assign a data steward per system or per portions of a system as the person responsible to communicate new or changed data to the FDR. This administrative link ensures a point of contact and accountability.

On the FDR side of Figure 5-3, its database designs must be completed and installed either in a specialized metadata management tool or in a more generalized DBMS. Several databases will be established based on schemas to be designed, implemented, and maintained by the FDR. These are a:

- **Categories database**, to contain a set of data categories that span NAS data. Eight high-level categories have been proposed, with multiple levels of detail for each of the eight categories.
- **Legacy database**, to contain legacy metadata, as is, from each application system.
- **Transformed legacy database,** to contain legacy metadata that has been converted to a common view that is not yet a standard view. An example of one kind of transformation is the dissection of the data elements in the messages sent from an application system. These constituent data elements will be stored in a database by message name and system name. Then, similar data elements being used in all system messages could be easily located.
- **Standards database**, to contain the data element standards approved by the NAS CCB to which each legacy data element will map. [Rhoades] describes the process of defining and approving a NAS data element standard.
- **Data model database**, to contain information about legacy and standard data models, built from legacy and standard data elements.
- Basic semantics database, to contain a lexicon/controlled vocabulary of well-defined terms from which element names, key words, or other identifying characteristics of FDR components can be created to facilitate searches and maintain consistency within the FDR.

In addition, there are supporting activities involving data design that are also required of the FDR. These include establishing and enforcing:

- Data element naming conventions for category names and standard data element names
- Validation and edit routines for metadata sent from NAS application systems for entry into the FDR databases
- Transformation routines to associate legacy metadata with common and standard metadata.

In automating the process the metadata management, the discussion has focused largely on the automation that will occur at the FDR as the main coordination point, i.e., the place from which agency-wide metadata resources and database will be managed. However, a critical question is the extent to which the various contributing application systems will

automate their metadata management to make the integration of metadata management with the FDR a smooth process.

For example, it will not be efficient for application systems to send metadata information to the FDR validation process in any format other than electronic, in a format other than that to be specified, that is also easily read by an FDR preprocessing routine. This raises an issue of the extent to which NAS application systems will implement automation processes to produce metadata in such a format. Again, this should not be much of a burden for systems already using COTS database products, but it will require more effort for other systems. In addition to being required to interoperate with the FDR, there would likely be an immediate benefit to the application system in terms of improved metadata management, which in turn should improve its application maintenance and enhancement process.

This presupposes that a legacy application system already has “good” metadata that can be exported to the FDR. Simply because a system uses COTS database products doesn’t necessarily mean it has documented metadata beyond the absolute minimum required to define the structure of the database. Data definitions, business rules, and such are liable to be missing or out of date, and recreating or validating them is extremely costly. It also raises a question of the extent to which the FDR should address the inadequacies of legacy metadata. That is why initial FDR activities will focus on standardizing commonly shared NAS data since they have a higher likelihood of being useful for new application development.

An issue in data collection is to determine whether, from the viewpoint of an application system, the data collection process will be take a 'push' orientation, a 'pull' orientation, or a combination of the two. In 'push' mode, the data steward, or owner, of the metadata (in this case, the application system) will 'push' or send its metadata when one of two events occurs. These are either: a) a periodic cycle of the clock, e.g., monthly, or on a 56-day aeronautical data update cycle, or b) by event, that is, when there is a change to a data element's metadata or new metadata are introduced. If the former method is selected and only that method is used, all of the systems sending metadata should do it on the same cycle.

In 'pull' mode, the FDR would poll each application system to determine which metadata had changed and, in some way, access these changes. It seems reasonable that an application system would be most aware of changes to its metadata, and so would 'push' changes to the FDR.

Cycle Dates. It is important to distinguish metadata that are synchronized across the agency on a cycle. These metadata and their associated data become 'live' synchronously at the new cycle start date, such as the 56-day cycle for aeronautical data, the basis for aeronautical data in ATA-100's NASR database. Distribution of these data must be managed to the cycle and release dates must be controlled. Other metadata not on such a cycle can be distributed upon their changing.

In today's environment of active data access, users should be able to initiate access to agency metadata by a 'pull' process also. That is, query access should be available to metadata databases. It is the role of the DA to ensure that either a) only authorized and approved metadata are accessible, or b) metadata not yet approved are clearly described as such with a caution against official use. In the latter case, some users will have a need to appreciate what metadata changes are anticipated to be able to plan upcoming system changes. Access control policy and its implementation will cover these cases.

On the receiving end, the FDR must have procedures in place to acknowledge the receipt of the data per sender and to edit and validate the metadata before storing them in its legacy metadata database. One of the FDR's value-added processes will be to associate every incoming legacy data element with a data category from the data category database. Where no category exists for an incoming element, a new one will be established to fill the gap and the data categories database will be updated, assuming a process will exist to validate and approve such updates.

The system development effort to establish a fully implemented FDR will be significant. It must be sized for initial and ongoing data storage capacity in addition to establishing the databases and processes it will require. In terms of development, the developers of the FDR and its associated processes must decide what the design of the data collection and validation process will be. Ideally, it will be web-based, but until NAS application systems can load their system metadata into a database tied to the web, or intranet, early versions of the FDR will likely rely on traditional FAA communications networks to send and receive data.

Although not pictured in Figure 5-3, the query and access function needs to be developed as well. It needs to be available from any desktop in the FAA with a legitimate need for access to these data. Again, with a web-based process, efficient query-based development tools exist today that integrate with web site development. These tools support 'fill-in-the-blank' and 'drop-down menu' query building, so that a user can access metadata information fairly easily using a web-based browser that is already installed at desktops in the FAA. Without a web-based process, developers need to build application screens in a client/server environment in which a user can develop queries with the same general procedures.

In summary, the following actions were discussed either directly or implicitly in this section and require resolution:

- From an organizational view, the FDR needs to be staffed and administered. Who will take that role? Also, what will the FDR's authority be with respect to working with constituent FAA organizations?
- From a technology viewpoint, how will the FDR's computing needs be identified? If technologies are required, how will these be decided, procured, and managed? How will the FDR be physically integrated into the FAA computing network?

- How will relationships be established between the FDR and the numerous organizations that will be asked to send in their metadata representing the various NAS application systems?
- Who will be responsible for automating extraction of metadata from the various NAS application systems and sending them on to the FDR in a common format?
- Who will design, test, and implement the FDR database schemas?
- Who will identify, specify, and produce the pre-processing routines to edit and validate incoming metadata from application systems to be managed by the FDR?
- Who will develop a GUI a process to manage queries against the FDR database?
- Who will explore and implement the use of web-based technologies to manage the FDR?
- Who will setup and administer access controls needed to secure FDR metadata?
- Who will formulate a policy defining the role of NAS users and other aviation constituencies regarding their metadata and its relationship to the FDR?

5.2.2 The Process of Maintaining Metadata

The main issues in the ongoing process of collecting application system metadata are:

- Ensuring the cooperation of application system data stewards
- Establishing a mechanism to recognize changes in the status of application system metadata
- Creating a process to collect metadata from application systems for the FDR.

Ideally, each application system would be linked electronically to the FDR so that new or changed metadata can be extracted on a scheduled basis or via database triggers. Until that occurs, manual intervention is needed to extract the required data and to send it, in electronic form, to the FDR for processing and inclusion in the metadata database.

Another regular process, but one that will occur only occasionally, is the modification of the FDR metadata schema. This schema, or database structure for the FDR, describes the metadata information that the FDR manages. There will be changes made to it over time, and these changes will affect what information is required from the source application systems. Such changes will affect any automation and extraction routines that source application systems have written as well as validation and editing routines that the FDR has created to prepare incoming data for entry into the metadata database.

One of the important indirect impacts of the existence of the FDR will be to cause new processes to exist because of its agency-wide, cross-cutting scope. Organizationally, it will

require that there is a much more regular and systematic interaction among staff roles with complementary tasks. In the case of the data stewards associated with application systems, these staff must establish a process to identify, extract, and send new and changed metadata to the FDR. Since each of their systems will differ by content, their various extraction routines will be unique. However, each will need to format their metadata in a common way before they are sent to the FDR. In this way, there should be some uniformity to the process in a common format to be defined, probably by someone involved with FDR development.

A by-product of this activity should be an improved metadata management process at the application system level.

In summary, the following actions were discussed either directly or implicitly in this section and require resolution:

- The role of data steward for each application system is an important role and is the key liaison position to the FDR. One such person is needed for each application system.
- A product of a NAS-wide information service that includes the FDR should be the creation of formats and templates for software routines to be created by NAS application systems to send metadata to the FDR.
- The flow of metadata from NAS applications to the FDR should be as automated as possible, based on formats and templates to be defined, approved, and mandated.
- There should also be a process in place revise these formats and templates as well as a process that assists users in implementing revised formats and templates.

5.3 Metadata Access and Dissemination

The FDR is presented as an automated information resource whose content is readily accessible by those authorized to use it. How will it be used? The interesting thing about an information resource is that users will find creative ways to use these data that were not anticipated by the designers. Although an earlier study [Schwarz] contains a number of questions that the FDR will answer that cannot be answered easily or at all today, users will create additional novel and effective ways to use this resource. In such cases, this argues for a flexible *ad hoc* query capability to supplement queries and reports that are programmed or 'hard-wired.'

An effective presentation of the range of important systems questions that metadata can answer was made by Robert Seiner of Ciber, Inc. at the DAMA (Data Management) Metadata Conference, held 19-23 March 2000 in Arlington, VA. He organized and related numerous forms of metadata and presented a coherent picture of the diversity and uses of metadata. His categories are, in most cases, supersets of the metadata categories shown in Table 5-1 above. These broad metadata categories are:

- Database metadata
- Data model metadata
- Data movement metadata
- Business rule metadata
- Data stewardship metadata
- Application component metadata
- Data access and reporting metadata
- Rationalization metadata
- Data quality metadata
- Computer operations metadata

With authorized access, it is assumed that a legitimate user will be able to access the metadata database from his or her desktop, whether the FDR is developed as an intranet or internet application or as a client/server application. It is also assumed that the application will be delivered to the user with a GUI that offers a forms-based, fill-in-the-blank view.

For searching or querying, there will be many dimensions, or variables, by which to search the database. For one, metadata categories can be used as access, or search, paths for accessing metadata. Since much of the querying will focus on an application system, numerous access paths based on the categories in Table 5-1 will be supported. Although this list is focused on a system view, the FDR can also manage metadata that are unrelated to a particular system that is of more general interest.

An example of a system-based query is: "Display all system information (to be defined) and the subsystems within each operating system at Memphis (ZME) that manage flight information, and include forecasted flight position." In addition, each of the categories in the table will contain a considerable number of metadata elements.

An example of a data management-based query is: "Display all public and private sources from which the FAA acquires weather data and the arrangements by which these data are acquired." The term 'arrangements' can cover content, frequency, cost, users and other information about how the data are acquired, distributed and processed.

A query about data stewardship could return the names of the data stewards for each system or data element being tracked. It could also describe their responsibilities and could indicate where job openings still exist. As a final example, a query could also return information about a set of data elements and their use in application systems. Such sets, (e.g., flight, infrastructure, communications) can be organized by the data element categories

proposed by CAASD [reference] for easy retrieval at high or low (i.e., detailed) levels of data organization.

All of these queries are representative of the information needed by managers, analysts, system developers, and system maintainers -- information that is not available in one system in automated form today. The FDR will support queries based on all of these access paths. Users will be free to develop additional ones. In examining these access paths, one needs to be clear that the FDR will provide answers as metadata, not operational data.[12] Therefore, a query about data elements will not return the value of those data elements. That is the responsibility of a database or data warehouse. Instead, it might list the names and definitions of the data elements used by a system or by all of the systems at a particular location or in a specific domain.

This discussion has focused primarily on a 'pull' view of data access, i.e., where a user initiates a request for information that may be standard or may be customized for him or her. The obverse of such a request is a system that disseminates metadata information automatically, on a cycle date or upon a change in the status of some metadata. This model of information access could also be a part of the FDR's implementation, based on requirements yet to be established.

In summary, the following actions were discussed either directly or implicitly in this section and require resolution:

- Access to the FDR's metadata database will be universal, constrained only by legitimate access.
- Access will be supported by a parameter-based GUI, allowing queries to be constructed by selecting from the available attributes and their domains, i.e., values.
- Access should also be available by subscription, or by a 'push' option in which new or changed metadata information is automatically sent to those who have subscribed. Subscription should be available by profile, in which a user may indicate the type of metadata that should and shouldn't be sent.

5.4 Roles and Responsibilities

The FDR, as a new capability for the FAA, will require a collection of roles to be carried out, some of which already exists in the FAA and some of which do not. In an extensive study of data management for the Army, the MITRE Corporation has described the collection of roles required for an agency-wide data management capability [Boylan]. Some of those roles, plus additional ones, are among those referenced below. Some of these roles

---

[12] For some suggestions about the types of questions that metadata can answer, see [Seiner] and [Schwarz].

already exist in other parts of the FAA. Also, some of these roles can be combined in one position. These roles are:

- **Acquisition agent.** This role acts as a liaison between the system acquisition process and the FDR to ensure that data requirements reflected in the acquisition system are accounted for in the FDR.
- **Administrative planner.** This role plans data management and administrative activities at the tactical level and coordinates with the data administrator.
- **Data administrator.** This role will manage data throughout its life cycle. This includes managing data element definitions -- both for legacy and standard data elements -- that are part of the data models. This role also helps design the FDR data schemas.
- **Data model manager.** This role will be responsible for the design of the FDR's data structure and for assisting the data administrator with building and storing information about operational data models in the FDR.
- **Data requirements analyst.** This role will work with application specialists to track and manage data requirements in application systems and relate them to metadata requirements for the FDR.
- **Data security analyst.** This role ensures that information system developers and users abide by information security policies and procedures that have been established. This role will also help establish and implement access controls for individual users.
- **Database administrator.** This role is responsible for the operation of physical FDR installation and includes issues of backup and recovery, implementation of security, system capacity, and system reliability and responsiveness.
- **User.** This role, both at the FAA and in the NAS user community, will be one who uses the information in the FDR for the many purposes it is being established.
- **Policy representative.** This role will help establish policies under which the FDR will operate and how it will be positioned relative to other NAS systems application systems. A critical aspect of this role is to be aware of related FAA initiatives and to create a collaboration among them to maximize productive interoperability while minimizing unwanted overlap and redundancy.
- **System developer/programmer.** This role will develop custom routines to transform, parse, and validate incoming metadata before it is loaded into the FDR databases and will also help establish the system environment in which the FDR works. Ideally, this role will have an FAA-wide focus and will preclude the need for every program to build local versions of such agency-wide information services.

- **System representative.** This role will establish a point of contact between each application system and the FDR operation.
- **Tools administrator.** This role will evaluate and test COTS tools that may be used to operate and to improve the FDR.
- **Web database developer.** This role will establish links between a database and a web browser and will develop query and access capabilities for users to access the FDR's metadata using web technologies.

5.5 Proposed FDR Functional and Implementation Architecture

In brief, the architectural view of data management in the FAA today is that of independent systems that share operational data via pre-defined messages. This process suffers from a number of inadequacies. To cite three, the current process is inflexible in its ability to define new messages quickly, it is expensive to maintain, and it lacks *ad hoc* data access capability. [cite earlier CAASD document].

However, unlike what exists for operational data, there is no formal system of documenting or exchanging metadata in the NAS. Most metadata are managed locally within an application system as a way to maintain the software that runs that application. Typically, the metadata are of interest mainly to the FAA's maintenance organization (e.g., AOS) and to the development contractor. As a consequence, there is presently no agency-wide, multi-system, consistent metadata management. The activities described in Section 4 provide the start to developing an agency-wide metadata management capability. This concept of use suggests ways in which this can be accomplished.

A prerequisite to developing such a capability is to define an architecture in which this capability would operate. Affirming that the scope of the data in this concept of use is NAS operational metadata and not administrative metadata (at this time), the following are proposed as basic requirements for developing an agency-wide metadata capability:

- The scope of metadata management must be NAS-wide. The architectural design for developing this capability must not preclude its extension to non-NAS, i.e., administrative, metadata.
- Every FAA organizational unit using operational data must participate and all operational systems must be represented. NAS users that are responsible for a non-FAA system that contains metadata that are significant to the NAS should be considered partners within the scope of this effort.
- Access to the metadata must be NAS-wide, structured and automated, accessible by *ad hoc* query and by stored procedures (e.g., predefined queries), and come under access control.

- Detailed metadata information must flow up from specific systems and operations to a high-level coordination function, and summary data must flow down to, or be accessible by, every lower-level unit that wishes to see or access a broad picture of agency metadata.
- There must be an organizational plan for the 'care and feeding' of the metadata. This requires metadata owners and stewards, the latter being responsible for the metadata under their supervision being properly maintained and communicated to the high-level, coordinated metadata management function.
- At present, legacy metadata will continue to be managed within application systems. The metadata architecture and related processes must accommodate this but must also facilitate the delivering of metadata information from an application to the metadata management operation, i.e., the FDR, on a regular or as-needed basis.
- Data standards, to be managed by the metadata registry/repository, will be accessible (via 'pull'/on demand) and disseminated (via 'push') throughout the system.
- The FDR will be managed as an FAA intranet and internet application and access and dissemination will be managed using web browsers, database technology, and other technology products that link databases to web browsers.

Under these conditions, an architecture for the FDR can be quite straightforward, especially when connectivity issues are solved by using internet technology, such as web browsers, that eliminate operating system and other potential incompatibilities.

5.5.1 Design

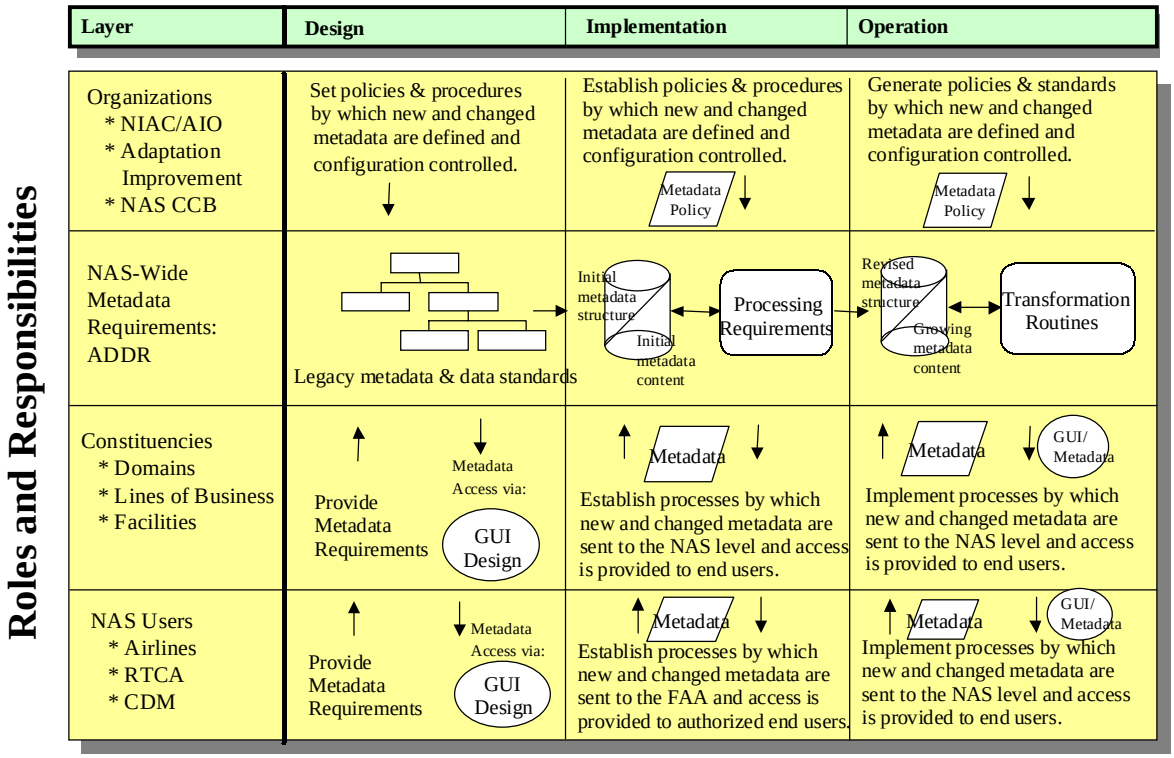
Operationally, the FDR needs to be designed and implemented as a single, virtual, coordinated capability under quality control and under configuration management control. Its design will accommodate legacy metadata in current systems as well as data standards that will be defined over time. In addition, links will be established in the FDR between legacy data elements and the standard data elements, as these are defined and approved by the NAS CCB. The design of the FDR is critical. It must be robust enough to manage a wide range of NAS metadata and to include the type of data represented in Table 5-1 and more.

Figure 5-4 provides an overview of the multi-faceted aspect of the FDR. As a NAS-wide capability, its scope will span NAS metadata and involve all operational systems, and it can extend, if desired, to include some NAS users' metadata.

Reading down the first column in Figure 5-4, one sees that the FDR will span various policy, standards and process-based organizations in the FAA (top row), the domains and lines of business across facilities (third row), and extend to NAS users (fourth row), who may be suppliers and users of the metadata in the FDR. The NAS-level (second row) will act

as a metadata coordinator across the FAA. It will likely be where the FDR design is established and where policy, security, backup, recovery, access control, content definition, query management, and a number of other management processes are defined and implemented.[13] Most of these processes will be defined in consultation with organizations at the many other layers of the FAA and NAS user community.

The third row in the figure includes domains, facilities, and lines of business. This layer will be designed with a GUI to enable data entry and data access. For data entry, the GUI can be customized to work with the applications associated with this domain, facility, or line of business. Rather than maintaining a manual process between an application and the FDR data entry GUI, an automated process should be developed to scan each application's data catalog, or equivalent, for its metadata content. Then, this process should write that content to a file and supply it to the GUI for transmission to the FDR coordination point as new or updated metadata. This process should be activated periodically, at a minimum, but could also be activated as needed, e.g., when some metadata changes.



[13] Another option is for the FDR to be housed within a particular domain or program as long as it operates as a NAS-wide capability although this option is not preferred.

Figure 5-4. Scope and Development of the FDR

Several organizations will be involved with design but will not be suppliers of operational metadata. The NIAC, AIO and the NAS CCB are three such organizations (first row). In addition, selected activities and processes that operate within a time window or until a specific mission is accomplished may also have a bearing on the FDR's design. One such activity is Adaptation Improvement Program (AIP), under the sponsorship of the Software Engineering Resource Center (SERC). Its role is to restructure the way adaptation information is managed today — largely a manual and costly process — to gain efficiencies through better process definition, automation and information management. Projects such as this one will be integrated with the FDR and may represent, as this one does, a number of application systems across several domains and organizations.

Probably the most unusual feature of this implementation concept is the recommendation of the involvement of NAS users. Typically, when systems are built, interfaces with non-FAA organizations are developed at arms' length. These organizations aren't usually seen as an integral part of an FAA system, but that view is changing as the FAA and NAS users build bridges to one another. They now realize that their worlds are indeed closely aligned and that the fluid and efficient exchange of data between them is required.[14]

In this light, the design of the FDR should accommodate the metadata used in NAS user systems, especially when the FAA is considering acquiring operational data based on NAS users' metadata. In addition, the FAA must also be aware of the metadata used by organizations such as NASA and major FAA system developers so that the NAS Architecture is complete with respect to the major participants that supply and use NAS data and metadata.

In addition to satisfying the requirements shown above for the FDR, its design will allow for the classification of data elements according to NAS data categories now being defined by MITRE's CAASD in its work for the ASD organization. These categories are meant to be a 'covering' set for the NAS, to include the set of data types used across all applications. A CAASD document describes these categories and how they will be applied in some detail [Broste]. Figure 5-5 illustrates the top level of these categories.

---

[14] An additional extension that will be considered in the future is the global aviation community, starting with Eurocontrol and European State aviation bodies. There is considerable dialog already with Eurocontrol regarding issues of safety and efficiency based on consistent international aviation data standards.

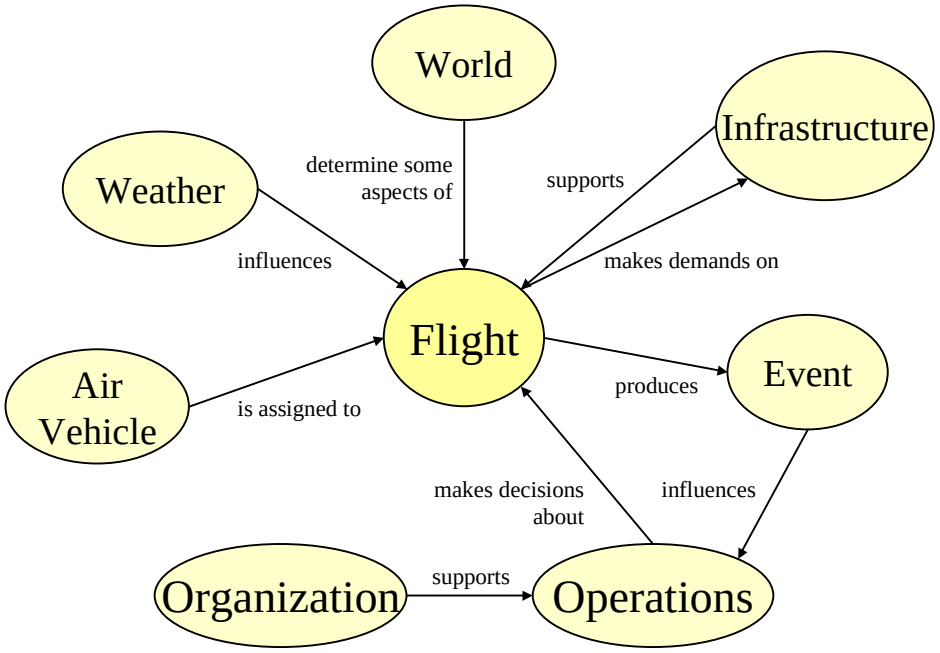


Figure 5-5. Data Categories, High-Level View

5.5.2 Implementation and Operation

One of the implementation steps is the design of the information that the FDR will manage. This involves capturing system metadata and relating them to data standards, as these are defined, and to data models that either exist or that will be developed to model various sets of NAS data.

In operation, the FDR will offer an integrated and consistent view of metadata across the NAS. Figure 5-6 illustrates a high-level functional view of the capabilities the FDR will be able to offer, from a view adapted from a presentation by Ciber, Inc.

Functionally, the core of a fully implemented FDR is the Coordinated Metadata Repository. It will store and manage NAS data standards, and as described above in Section 3.1.4, it will "document the authority and responsibility for each registered data element." Complementing this view are various information resources, i.e., databases, that store and manage legacy data information and data standards information. This information is delivered by viewing the FDR as a functional resource, or system, that integrates the components depicted in Figure 5-6. These functional resources include:

- **Legacy data definitions:** A database of information about each data element currently used in each NAS application system

- **Data movement tools:** Information about the sources that supply data and the recipients of that data. This database includes internal and external sources that supply source and derived data to every system, such as third party weather data.
- **Data standards:** Standards that have been proposed, approved, superseded, etc.
- **Data models and data modeling tools:** Data models that reflect key subsystems of the NAS, such as flight. The component elements may be standardized or not, e.g. a model could have a message, table or object whose attributes are not all standardized yet, but all of which have been assigned to one or more categories for querying purposes.
- **Database directory:** A database of the information about NAS systems, in considerable detail, including system functionality, location, interfaces, and data elements, essentially the descriptors found in Table 5-1.
- **Reporting and query tools:** Stored procedures, canned queries, and query builders with which users can formulate *ad hoc* queries to extract data from any of the FDR-related databases pictured in the figure.
- **Decision support tools:** Eventually, the FDR could provide analytic tools, such as data mining and knowledge engineering tools, to assist users in finding patterns among the metadata and the data.
- **End user access:** This function will extend FDR capabilities beyond FAA organizational boundaries to allow authorized members of the aviation community to access information in the FDR.

The view in Figure 5-6 is intended to illustrate a capability that can exist over time. It will not exist initially but can grow to add a number of the capabilities shown. What is not shown, but is of interest to various FAA decision-makers, is a companion data warehouse that will contain the operational and administrative data that corresponds to the metadata managed as described in Figure 5-6.

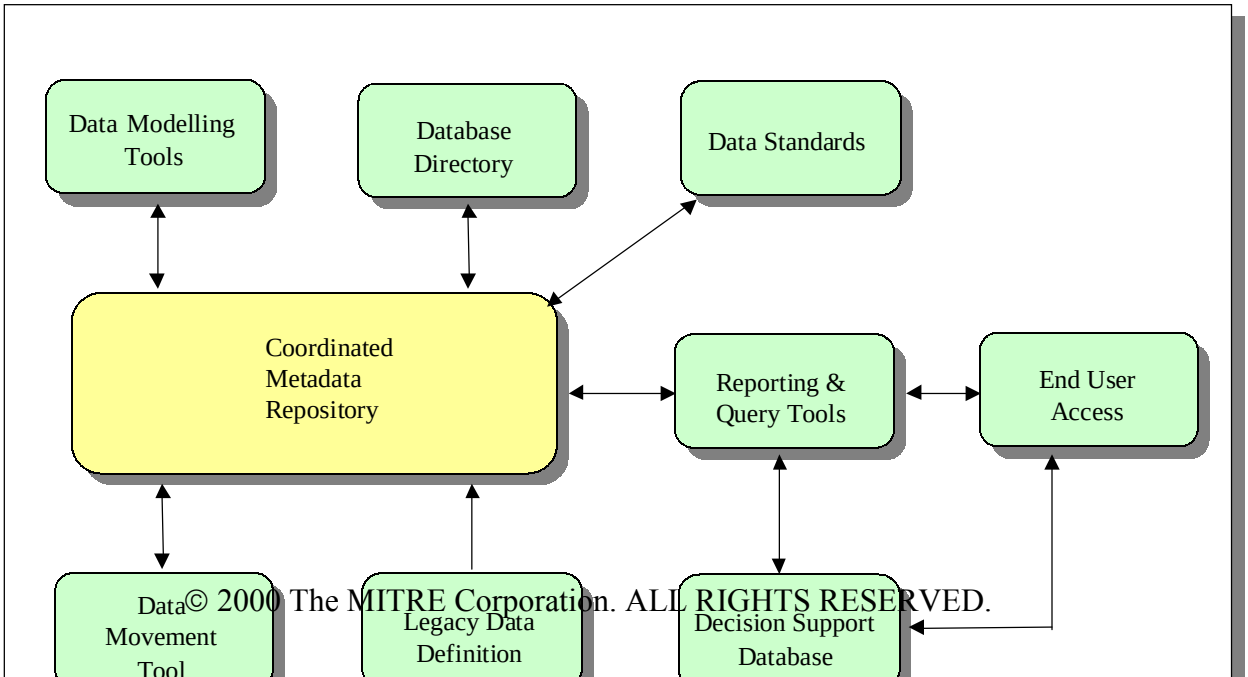


Figure 5-6. High-Level Functional View of FDR Capabilities

There are many organizational, managerial, and technical aspects of an FDR capability. Figures 5-7 and 5-8 illustrate two aspects of FDR implementation. Figure 5-7 describes a progression of content, from a registry that focuses on data element standards (bottom of the view) to a fully functional metadata repository of system information that also includes data element metadata (top of the view). Figure 5-8 describes basic FDR capabilities and the sequence in which they might be implemented. It describes three core functions, shown in the rows. It also describes three timeframes, each indicating a greater tendency toward compliance with the standards and practices of the FDR. These two views, taken together, offer implementation choices and growth paths for the FDR, depending on requirements and the resources available to build and manage it.

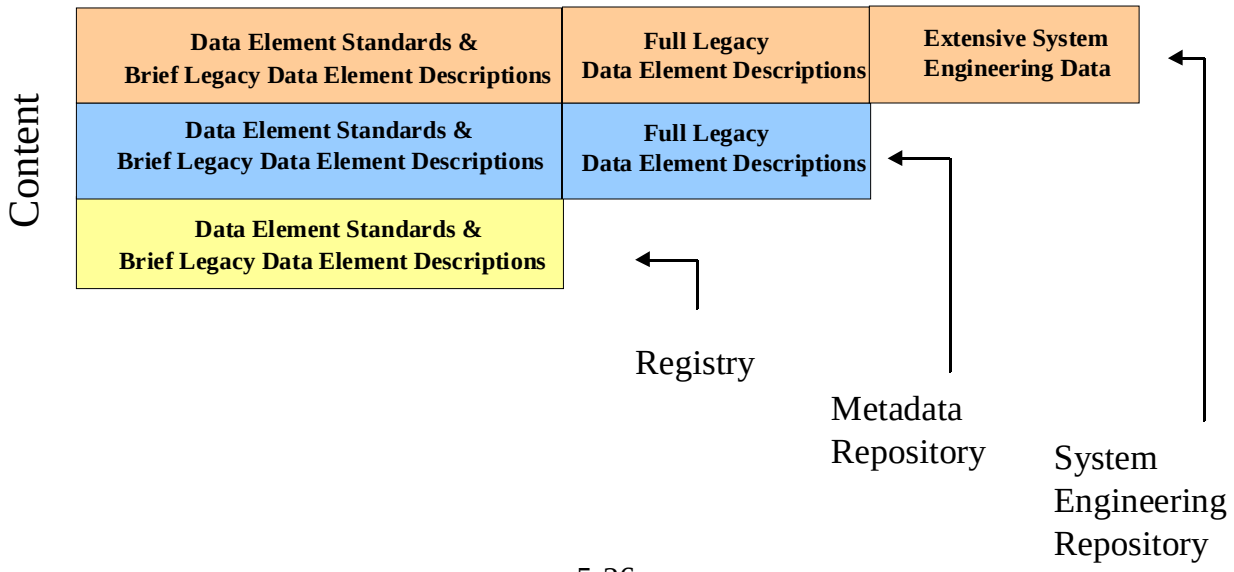


Figure 5-7. Progression of FDR Content

Among the several important decisions to be made regarding the FDR, one that concerns an organizational and managerial issue for the FAA, is to decide who will build and manage the FDR. The context for this decision is that there are several viable candidate organizations interested and available to manage the FDR (in whatever form it takes). In addition, there are several ongoing initiatives, each of which is constructing an FDR-like capability. At a minimum, these activities must be coordinated to achieve maximal metadata consistency and quality while minimizing unwanted redundancy and expense.

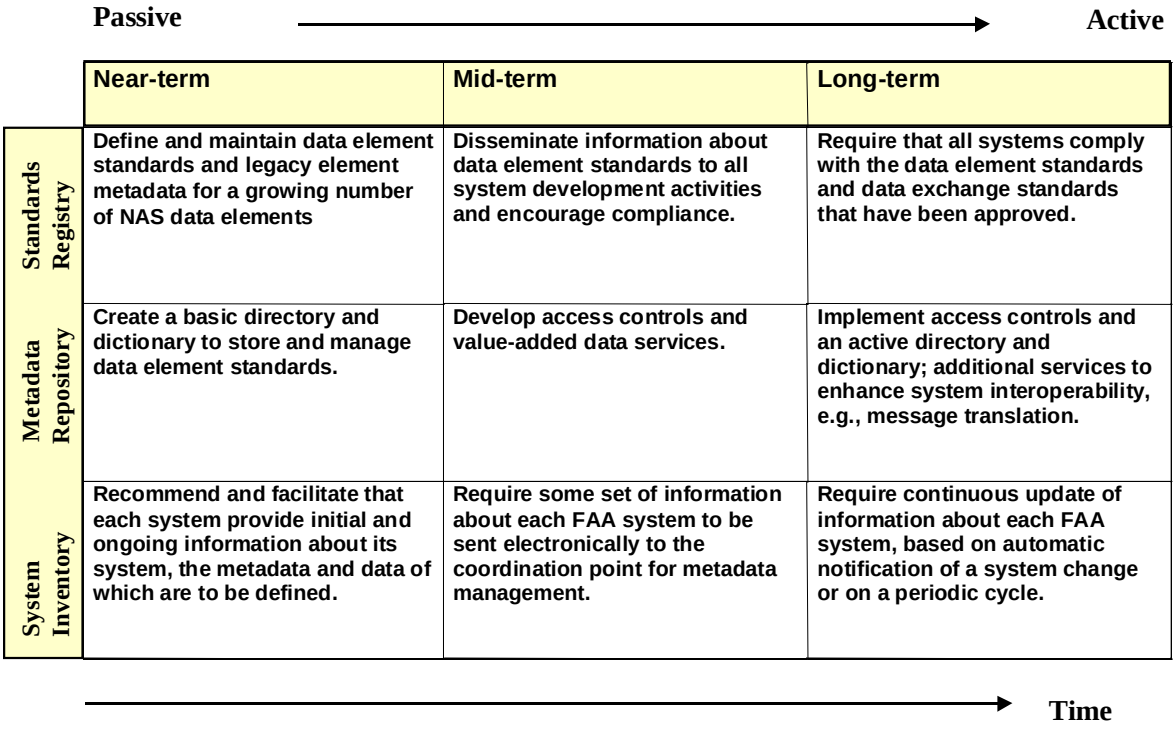


Figure 5-8. FDR Capability Areas and Growth Over Time

5.6 Implementation Risk

There are numerous risks with any new initiative having high impact potential, such as the one described here. It is important to be aware of these risks as these ideas about metadata management move from concept to design to development to implementation and then to maintenance and enhancement. The list below describes some of the more significant risks, most of which are common to any development effort, but the comments given consider the particular environment in which the FDR will operate.

- Risk 1 — Organizational Will and Management Support. One of the biggest risks to implementing the FDR is the organizational will to accomplish it. As a new capability, the FDR must be justified in terms of contributing to the FAA's mission and to NAS user benefits. As a new capability, staff and resources must be identified to design, build, and staff it. Upper management support is required for success. As a cross-cutting initiative, the FDR will require high-level support from many FAA organizational units. The fact that this capability may be directed by one or more

FAA organizational units makes its implementation that much more of a challenge, as discussed above.

- Risk 2 — Budget and Resources. Any capability requires budget and staff resources. The FDR can be started as a fairly modest capability, but, as a NAS-wide capability, it needs to grow into a robust, agency-wide tool. This will require that the FDR is part of the FAA network for data acquisition and dissemination. It will also require continuous maintenance and upgrading. Since every program is explicitly or implicitly managing their metadata anyway, a business case can be made to develop the FDR using contributions from current program budgets in their own interests and for the benefit of NAS-wide operations.
- Risk 3 — Data Contribution. One of the critical aspects of a metadata capability is a continual feed of metadata by the owners and stewards of that metadata to the data collection and coordination point, the core FDR. There must be a sufficient incentive for the many organizational units that own or manage these data to convey this information to the metadata coordinating body. Ideally, these will be positive incentives, but the contributors of metadata must also perceive and receive tangible benefit to their organizations as users of the metadata.
- Risk 4 — Integrating Operational Metadata with Operational Data. To be of greatest benefit, the FDR capability should be integrated with the operational systems from which they originate. However, NAS operational systems are currently not well integrated from a data or functional viewpoint, and that makes this task more difficult. On the other hand, the FDR capability will be a foundation tool to help integrate operational data across the NAS.
- Risk 5 — Integrating Operational Metadata with Business Metadata. This operational concept focuses on operational metadata, but there are simultaneous activities in place that are building a metadata capability for managing the business side of the FAA. These two aspects of agency-wide metadata must be coordinated and synthesized to obtain maximum benefit from each activity.
- Risk 6 — Integrating FAA Metadata with NAS User Metadata. One of the capabilities that NAS users rate very high is that of being able to access critical FAA-generated operational data. Along with this need is their need to access FAA-generated and managed metadata. That is, NAS users want to see how FAA data and data models are defined and formatted so that they can coordinate their information management processes more closely with those of the FAA. There are considerable time and cost savings available to NAS users to be able to do this.

Section 6

Conclusions, Recommendations, and Decisions to be Made

This document describes how an FDR should operate. This section offers some conclusions and recommendations. Also, there are still some decisions required regarding its content, development, and ownership. Two of these are described below.

Conclusions.

- Metadata management is a foundation step leading to integrated NAS information management.
- Based on industry 'best practices' and the increasing effort needed to make FAA systems interoperable, a case can be made for the FAA's need to manage agency-wide metadata in a consistent way.
- Without overall guidance and a clearly defined architecture and design, there will be a proliferation of metadata databases, each built to different operational concepts and designs. Integrating them 'after-the-fact' will be very difficult and expensive, as is already seen with operational application systems. The FAA's system inventory counts over 200 systems, each of which manages its data and metadata 'internally.'
- Managing agency metadata will yield access to important information that is not currently accessible or is not easily accessible.
- The software and technology needed to manage metadata are mature and affordable and are not an impediment to a successful outcome.

Recommendations. The following recommendations are made with respect to developing the FDR:

- The FDR should be developed to include the two main threads discussed in this document: data element standards and legacy metadata links to those standards.
- From a systems viewpoint, it would be useful to study current NAS applications and make a recommendation for how these systems should implement metadata management processes that complement the role of the FDR.
- Although extensive collaboration will be required among the many offices, systems, and activities that are engaged in metadata management, there should be a single point of responsibility to oversee and coordinate the effort.
- A data steward should be identified for each application program as a necessary step to organize the agency's metadata management effort. A high degree of collaboration

will be required between the data stewards and the main coordination organization for the FDR.

- The roles and relationship between the FDR and the MDR, being built by AIO and ATS, should be made clear. Data exchange guidelines and content should also be addressed.
- The internet should be used as an access and dissemination point for metadata content since it is widely available and has become a fundamental vehicle for information management.
- To assure success, the roles and responsibilities described above for collecting metadata from application system stewards should be enforced in some manner. One option is through requirements applied by the JRC and the AMS. Another is through the authority of an organization such as the AIO, whose mandate covers information architecture and information technology management in the FAA.
- Because there is currently some overlap in responsibility, written guidelines should be produced that clearly describe the stewards for the various agency metadata and the processes that will be established to manage the metadata.

Decisions to be Made. Some decisions about the FDR require additional analysis, including those listed here:

- The MDR is currently managing system-level metadata. The FDR starts from a different point with respect to content by focusing on data element standards. An organizational and architectural decision is required regarding whether or not the MDR will also take on the role of managing data element standards and legacy metadata or whether these capabilities will be managed by a separate, but related, system, such as the FDR.
- The agency's metadata, once assembled and automated, will be an important information resource. Decisions are needed about what portion of this information will be made available to system developers working for the FAA and to other interested aviation bodies, such as NAS users, and to the public.

List of References

[AIO 1] *Federal Aviation Administration Data Management Strategy, Version 1.0,* Information Management Division, Office of the Assistant Administrator for Information Services and Chief Information Officer, Federal Aviation Administration, Washington, D.C., 21 September 1999.

[AIO 2] *Federal Aviation Administration Information Technology Strategy: FY2000-FY2002, Version 1.0,* Office of Information Services and Chief Information Officer, Federal Aviation Administration, Washington, D.C., 21 September 1999.

[AOP] *FDR Demonstration as the NIIS,* Office of NAS Operations, AOP-600, Federal Aviation Administration, Washington, D.C., October 1999.

[ASD] *NAS Architecture Version 4.0,* Federal Aviation Administration, November 1998, Washington, D.C.

[Bolczak] Bolczak, C., Broste, N., Schwarz, R., Stalnaker, S., Truong, L., *Preliminary TFM Information Architecture Steps,* WN980000020, The MITRE Corporation, McLean, VA, January 1998.

[Broste] Broste, N., Rhoades, R., Schwarz, R., *NAS Data Categories,* MTR-00W0000067, The MITRE Corporation, McLean, VA, September 2000.

[Broste2] Broste, N., Rhoades, R., Schwarz, R., *FAA Data Registry (FDR) System Requirements Specification (SRS*), WN-00W0000071, The MITRE Corporation, McLean, VA, September 2000.

[CAST] *Information Technology Issues for the Delivery and Use of Common Taxonomies* (Draft), Information Aviation Data Registry Sub-Team of the CAST/ICAO Common Taxonomy Team, August 1999.

[Feinberg] Feinberg, D., *Strategic Database Management Conference 1 (SDM1): Creating the Right Data Architecture,* The Gartner Group, 28 July 1995.

[Finkelstein] Finkelstein, C., Aiken, P., *XML and Corporate Portals,* 1999, http://www.ies.aust.com/~visible/papers/xml-portals.htm

[ICAO] *Establishment of an ICAO Aviation Data Registry,* Interagency Group on International Aviation (IGIA), Department of Transportation, Federal Aviation Administration, Washington, D.C., 20591, Meeting IGIA 73/4.42, June 9, 1999; presented by the National Transportation Safety Board (NTSB).

[ISO] The ISO/IEC 11179 Metadata Registry Implementation Coalition; http://www.sdct.itl.nist.gov/ftp/18/other/coalition/coalition.htm

[MDC] *The Open Information Model*, http://www.mdcinfo.com/OIM/index.html

[OMG] *Meta Object File (MOF) Specification*, http://www.omg.org/cgi-bin/doc?ad/97-08-14

[Parker 1] Parker, B., *A Process Improvement Approach to Data Management*, briefing, The MITRE Corporation, McLean, VA, May 1995.

[Parker 2] Parker, B., *About Data Dictionaries and Such*, http://grouper.ieee.org/groups/scc32/DD.white.SUCH.html, April 1998.

[Rhoades] Rhoades, R., Broste, N., Schwarz, R., *NAS Data Standards Process*, WN-00W0000xxx, The MITRE Corporation, McLean, VA, September 2000.

[Rosenthal] Rosenthal, A., Sciore, E., Renner, S. *Toward Unified Metadata for the Department of Defense*, IEEE, 1997.

[Schwarz] Schwarz, R.A., Broste, N.A., *A Proposed Concept for Managing NAS Standards and System Metadata: The FAA Data Registry (FDR)*, MTR-99W0000101, The MITRE Corporation, McLean, VA, September 1999.

[Seiner] Seiner, R., Ciber, Inc., *Questions Metadata Can Answer*, DAMA Metadata Conference, Arlington, VA, 19-23 March 2000.

[Uri] Uri, Carol, *Data Classification and Naming*, Office of System Architecture and Investment Analysis, ASD-130, Federal Aviation Administration, Washington, DC, 23 November 1999.

Appendix A

Attributes of a Data Element

Table A-1. Definitions of Basic Data Element Attributes

Data Element Attribute	*Definition*
Identifying	
Name	A single or multiword designation assigned to a data constructed in accordance with the TBD naming convention. This name is unique within a single registry context.
Context or Namespace	A designation for the application or discipline in which a Name or Synonymous Name is applied or originates from. NAS Standard data elements registered in the FDR are assigned a default context, e.g. "FDR"
Registration Authority	The Organization authorized to maintain a data registry
Registration Authority Identifier	An n character code assigned by the International Registration Authority that uniquely identifies the registration authority, e.g. NIAC, with in an international registry context. One of three attributes that together uniquely identify a data element description in an international registry context.
Data Identifier	A TBD character number assigned by the NAS Registration Authority that uniquely identifies the data element with in a registry. One of three attributes that uniquely identify a data element description in an international registry context.
Version Identifier	An m digit number assigned by the NAS Registration Authority. One of three attributes that uniquely identify a data element description in an international registry context.
Short Name	A short name of the data element
Synonymous Name	Single or multi-word designation that differs from the Name but represents the same data element concept. Synonymous Names may be familiar names in other applications environments. In this case, a

Data Element Attribute	*Definition*
	Synonymous Name Context must be provided.
Synonymous Name Context	Required if Synonymous Name is from another context.
Definitional	
Definition	A statement that expresses the essential nature of the data element that permits its differentiation from all other data elements.
Precision	The smallest meaningful difference between two quantitative data elements.
Largest value	The largest quantitative value a data element can have.
Smallest value	The largest quantitative value a data element can have.
Unit of quantity	The units of measure for quantitative measures, e.g. hours, meters, nautical miles, knots, etc.
Relational	
Classification scheme identification	The name or identification of a category schema that groups data elements that have common characteristics, e.g. origin, composition, structure, application, function, etc.
Classification scheme item	The designator identifying the group or groups from the Classification scheme that can be assigned to the data element. [At least one Classification scheme, Classification scheme item pair will be generated from a mandatory scheme.] [Multiple Classification schemes and Classification scheme items can be used.]
Keyword set identifier	The name or identification of a set of key words maintained in the registry. Keywords are significant words used to access data element descriptions. Keyword sets are controlled by the registry administrator. [Multiple Keyword Sets can be used]

Data Element Attribute	*Definition*
Keyword(s)	Words selected from the Keyword Set.
Related Data Reference	A reference to another data element that is related to this data element.
Relationship Type	An expression (e.g. 'qualified by', 'part of', 'is derived from') that characterizes the relationship between this data element and the related data element. [Related Data Reference and Relationship Type attributes occur in pairs.]
Representational-simple	A simple data element is based on a single data concept that can be represented with only one type of data representation unit, e.g. character, integer, bit, byte. No subset of the data units has any meaning.
Representation Category	The type of symbol, character, unit, or designation used to represent the data element, e.g., characters, bar codes, graphics, voltages.
Form of representation	Name or description of the form of representation for the data element, e.g. 'quantitative value', 'code', 'text', 'icon', 'image'. (The full list of approved class words must still be assembled.)
Datatype	A distinct term describing the data representation units, e.g., character, integer, real, bit, byte.
maximum character quantity	The maximum number of data representation units (of a given data type) used to represent the data element value.
minimum character quantity	The minimum number of data representation units (of a given data type) used to represent the data element value.
Format	The layout of characters in a data element represented by character string.
Permissible values	The set of representations of permissible instances of the data element. The set can be specified by name, by reference to a source, by enumeration, or by rules for generating the instances. This may include a description of algorithms for computing values for an element derived from other data elements.
Example	An example of a typical instance of the element, in quotes, if it can

Data Element Attribute	*Definition*
	be represented as a printable character string.
Representation Description	A textual description providing representational characteristics of the data element not included in other data element attributes.
Representational-complex	A complex data element is composed of or derived from more than one registered data element linked by one or more relationships that together represent a single data concept.
Complex representation type	The notation or methodology to be used to create the description of a complex data element, e.g., UML, XML-DTD, E-R, text.
Description	The description of the components and relationships making up a complex data element.
Components	A summary list of registered data element components used in the complex data element.
Administrative	
Responsible Organization	The organization responsible for the contents of the mandatory attributes of this data element: - Organization Name - Organization Symbol - Organization e-mail - Organization mail address - Organization Representative Name - Organization Representative e-mail - Organization Representative mail address
Submitting Organization	The organization or unit within an organization that has submitted the data element for addition, change, or cancellation to the Registration Authority - Organization Name

Data Element Attribute	*Definition*
	- Organization Symbol - Organization e-mail - Organization mail address - Organization Representative Name - Organization Representative e-mail - Organization Representative mail address
Date Submitted to Registration Authority	The data the Registration Authority first receives a data element description from the Submitting Organization.
Date Case Number Established	The date on which the NAS CCB begins the data element standardization process.
Valid From Date	The date on which the NAS CCB designates that the data standard may be used.
Valid Until Date	The date after which the NAS CCB designates that the data standard may no longer be used.
Case File Number	Identifier assigned by the NAS CCB
NCP Number	NAS Change Proposal Number assigned by NAS CCB. [Details of the data element standardization process are in process.]
CCD Number	Configuration Control Decision identifier assigned by the NAS CCB. [Details of the data element standardization process are in process.]
Registration Status	A designation of the position in the registration cycle of a data element, e.g. Incomplete, Recorded, Certified, Standardized, or Retired. [Details of the data element registration process are in process. See ISO 11179-6.]
NASCCB Status	A designation of the position in the NAS CCB data standardization cycle of a data element. [Details of the data element standardization process are in process.]
Data Element Standardization Priority	Score sheet Value from the NIAC SOP for defining data standardization priority based upon Capability Importance, Infosec Classification and Use Weight
Security – Protection Category	A designation that identifies data elements whose distribution is sensitive. Categories will be maintained in the FDR. [Security

Data Element Attribute	*Definition*
	designation may be dependent on the usage context and not inherent in the data element]
References	Examples of supplementary documentation that may be required to support the NAS CCB data standardization process. . [Details of the data element standardization process are in process.]
Requirement(s)	Primary source document followed by DOORS registration index
Requirement, Other	Alternate source document followed by DOORS registration index
NAS Data Model	Pointer to NAS Data Model(s) that include this data element.
Technical Architecture Key	CATS Diagram location reference for this element.
CDIMS Collaboration Record	File number and Date entered into CDIMS & assigned Moderator Name
ICD(s)	A list of existing or proposed Interface Requirement Documents (IRD) and Interface Control Documents (ICD) related to the data element.
General References	A list of other documents, objects, diagrams, and other media providing additional information about the element.
Usage	
Implementations	Summary of systems and applications implementing this data element.
Equivalence	An assessment of how closely a standardized data element is implemented is a specific usage context.
Transformations	A description of a transformation or procedure that will map an implemented data element to a standardized data element.
Legacy Applications	Summary of systems and applications using data elements with the same or similar definition or data element concept. This could be generated from an inventory of legacy systems data elements.

Data Element Attribute	*Definition*
	- Systems/Application Name - Responsible Organization - Impact of converting - Applications is source or consumer data values for this data element.
Transition Schedule	Pointer to a schedule detailing when and where the proposed data standard will be phased into use.
Valid Usage Constraints	Identify Certification, Safety Critical, etc related issues.
Activity Data Model	Pointer to Activity Data Model Diagram Index; the FDR address will reveal the responsible organization and name.
Quality of Service	Descriptions of constraints on the quality of the data represented by the data element within specified operational context. - Uncertainty in measured quantities. - Probability that data is corrupted - Latency or delays - Synchronization of distributed elements [Other qualitative measures may be defined. Detail meaning and use of this attribute is open for discussion]
Comments	Any text that can improve the understanding and usage of the data element.

Appendix B

Registry Data Model

A full view of a Registry data model may be found in a companion document entitled *FAA Data Registry (FDR) System Requirements Specification (SRS)*, [Broste2] published in September 2000.

Glossary of Acronyms

ADDR	Aviation Data Description Registry
AIO	Office of Information Services
AIP	Adaptation Improvement Program
AMS	Acquisition Management System
API	Adaptation Process Improvement
ASD	Architecture and Investment Analysis Division
ATS	Office of Air Traffic Services
BSU	Basic Semantic Unit
CAA	Civil Aviation Administration
CASE	computer-aided software engineering
CATS	Capability and Architecture Tool Suite
CCB	Configuration Control Board
CCD	Configuration Control Decision
CDM	Collaborative Decision Making
CIO	Chief Information Officer
CM	Configuration Management
COTS	Commercial-off-the-Shelf
CTAS	Center Terminal Automation System
DA	data administrator
DBMS	Database Management System
DOC	Department of Commerce
DoD	Department of Defense
DOORS	Dynamic Object Oriented Requirements System
ERD	Entity-Relationship Diagrams
FAA	Federal Aviation Administration
FAST	FAA Acquisition System Tool

FDR	FAA Data Registry
GUI	Graphical User Interface
ICAO	International Civil Aviation Organization
ICD	Interface Control Documents
iCMM	integrated Capability Maturity Model
IPT	Integrated Product Team
IRD	Interface Requirements Documents
ISO	International Standards Organization
IT	Information Technology
JRC	Joint Resources Council
MD	Management Directive
MDC	Meta Data Coalition
MDR	FAA-wide Metadata Repository
MOF	Meta Object Facility
NAS	National Airspace System
NASDAC	National Aviation Safety Data Analysis Center
NCP	NAS Change Proposal
NIAC	NAS Information Architecture Committee
NOS	National Ocean Service
NOTAM	Notices to Airmen
NWIS	NAS-wide Information System
OIM	Open Information Model
OMG	Object Management Group
SERC	Software Engineering Resource Center
SOP	Standard Operating Procedure
SRS	Software Requirements Specification
STARS	Standard Terminal Automation Replacement System
SWIM	System-Wide Information Management

TFM	Traffic Flow Management
TMC	traffic management coordinator
WG	Working Group
ZME	Memphis