



Whitemarsh
Information Systems Corporation

Data Management Program:
Engineering

Whitemarsh Information Systems Corporation
2008 Althea Lane
Bowie, Maryland 20716
Tele: 301-249-1142
Email: Whitemarsh@wiscorp.com
Web: www.wiscorp.com

TABLE OF CONTENTS

| | |
|--|----|
| Acknowledgments | iv |
| 1.0 Data Management Program | 1 |
| 1.1 Data Standards Management | 1 |
| 1.2 Remainder of this Paper | 2 |
| 1.3 Additional Data Management Materials | 2 |
| 2.0 Data Management Terms and Concepts. | 3 |
| 2.1 Data Management. | 3 |
| 2.2 Data Management Planning | 3 |
| 2.3 Net-centric Environment | 4 |
| 2.4 Data Asset | 7 |
| 2.5 Data Asset Products | 8 |
| 2.6 Metadata | 8 |
| 2.7 Data Interoperability. | 9 |
| 2.8 Communities of Interest | 11 |
| 3.0 Data Standards | 13 |
| 3.1 Authoritative Data Sources (ADS). | 14 |
| 3.2 Enterprise Identifiers (EID). | 14 |
| 3.3 Information Exchange Standards Specifications (IESS). | 15 |
| 3.4 EXtensible Markup Language (XML) | 17 |
| 4.0 Communities of Interest Data Management Roles and Responsibilities. | 19 |
| 4.1 Enterprise Data Administrator Responsibilities | 21 |
| 4.2 Institutional Data Administrator Responsibilities | 22 |
| 4.3 Expedient Data Administrator Responsibilities | 23 |
| 4.4 Data Standards Producers Responsibilities | 23 |
| 4.5 Stakeholders Responsibilities | 24 |
| 5.0 Data Management Program Architecture | 26 |
| 5.1 Program Layer | 26 |
| 5.1.1 Program Architecture | 26 |
| 5.1.2 Data Management Program Governance. | 26 |
| 5.1.3 Data Management Program Assessments. | 27 |
| 5.1.4 Data Management Program Components | 27 |
| 5.2 Planning Layer | 28 |
| 5.3 Execution Layer | 32 |
| 5.3.1 Metadata Data Asset Projects. | 32 |
| 5.3.2 Data-Standard Data Asset Projects. | 33 |
| 5.3.3 Database Data Asset Projects. | 34 |



| | |
|--|----|
| 5.3.3.1 Data Warehouse Data Asset Projects. | 35 |
| 5.3.3.2 Original Data Capture Data Asset Projects. | 35 |
| 5.3.3.3 Subject Area Database Data Asset Projects. | 35 |
| 5.3.4 Data Asset Evolution, Maintenance and Support Projects. | 36 |



Acknowledgments

This paper is an evolution of two documents, AR 25-1 (paragraphs 4-7 through 4-12), and the DA PAM 25-1-1 (chapter 5). The primary contributors to these documents were Bruce Haberkamp, James Blalock, and Michael Gorman of the Office of the CIO, United States Army. The DA PAM material was significantly based on work by these three authors during the time frame, September 2003 through December 2004. The key constructs of this work was proven through a week long workshop during December 2003. The foundational components of this work has been favorably reviewed by subject matter experts within the U.S. Department of Defense.



1.0 Data Management Program

A data management program represents the established policies and procedures intended to control the production and applicability of data standards required to ensure data interoperability among information systems. The data management program addresses data standards creation, implementation, and evolution as it applies to automated systems, applications, data exchanges, databases, record and document management, and information presentation within and across business systems.

The data management program facilitates the dissemination and exchange of information among organizations and information systems throughout the enterprise. A common data management strategy is essential to allowing authorized users to access required information. Enterprise information is a resource and will be shared unless specifically limited by law, regulation, need-to-know, security, or privacy restrictions.

The data management program manages information requirements from data models and business rules within their mission, organization, and functional contexts down to data-element and data-value levels of detail per the policies.

The data management program facilitates interoperability through the standardization and use of common data standards.

The data management program facilitates the specification of standard data management services and conformance test requirements and represents these requirements to data management standards committees, as appropriate.

The data management program improves data quality and accuracy, and minimizes the cost of data production and data maintenance.

The data management program applies to any information system passing information through networks and/or IT assets in the information environment.

The ultimate goal of data standards is the specification, development, management and evolution of interoperable data assets. The process that achieves the necessary planning for interoperable data asset projects, called data management planning, is stratified across the enterprise from the CIO through to the data standards producers. This process is squarely set against data management goals, which are the measures of data asset success.

1.1 Data Standards Management

Data standards (expressed as authoritative data sources (ADS), information exchange standards specifications (IESS), enterprise identifiers (EID), and eXtensible Markup Language (XML) are



used to guide all data exchanges including those with legacy systems. Data management requirements are included in IT planning documents.

All organizations producing or using data standards (ADS, IESS, EID, XML):

- Ensure that only data standards are used in systems.
- Incorporates new data-based uses of the data standards in the appropriate part of the metadata repository, as needed.
- Provide input to data standards development, evolution, and use reviews.

Data standards are primarily enforced by the data administrator. They are in turn supported by community of interest data administrators, which in turn, are supported by data standards producers. The roles and responsibility of each is described in Section 4.

1.2 Remainder of this Paper

Section 2 provides basic terms and concepts involved in data management . These terms are generally presented in the order employed in the rest of the paper.

Section 3 describes the critical data standards. These standards are further described in supporting Whitemarsh books, courses, and methodologies on data management .

Section 4 presents roles within communities of interest for data management. These roles and responsibilities are also further described elsewhere, that is, in supporting Whitemarsh books, courses, and methodologies on data management .

Section 5 presents the three levels associated with data management . Included within each level are identified and briefly described projects appropriate for each level. Again, the projects associated with these levels are further described elsewhere, that is, in supporting Whitemarsh books, courses, and methodologies on data management .

1.3 Additional Data Management Materials

This paper is not intended to be either comprehensive nor exhaustive. There are many book, courses, methodologies, papers, and software products that accomplish what this paper provides in a summary form. For these other materials, please consult the Whitemarsh website, www.wiscorp.com.



2.0 Data Management Terms and Concepts.

The following provides a discussion of specific terms and concepts used in the context of this document.

2.1 Data Management.

Data management addresses the independent management of data shared by multiple applications. Data management supports data exchange and includes data dictionary, directory services, and database management systems (DBMS). DBMS support the definition, storage, and retrieval of data elements from monolithic and distributed DBMS. Information systems that employ commercial off-the-shelf DBMS must conform to the requirements of SQL:2003, Core.

2.2 Data Management Planning

The data management planning process facilitates the creation of data asset products so that, when implemented, the data asset represents the right data, produced at the right time, for the right user. In a "data-driven" enterprise, users not only transmit and receive data, they also produce, consume, and store data relevant to their own operations, and meet the data needs of higher and lower headquarters.

The data management planning process aids in the creation and management of products needed to define the data and metadata of data assets. Thus, data management planning identifies, plans, and manages projects associated with data assets throughout the entire life cycle of the data asset. data management planning projects produce products for future and improved data integration and reuse. data management planning enhances project scoping, responsiveness to business change, and management of systems development sequencing and prioritization.

Data management planning identifies, plans and manages the data asset projects that, in turn, collect, develop, and maintain the sum of all data asset products required by the enterprise to accomplish its mission tasks. data management planning engineers complete, accomplishable projects that include all metadata, measures, studies, plans, prototypes, models and databases necessary to enable subsequent and improved data integration and reuse. It also provides improved project scoping; responsiveness to business change; management of systems development sequencing and prioritization; and level of business involvement.

Data management planning develops projects to structure and implement a sub-set of data structures for a specified purpose and use. A data management planning project incorporates within the proper scope of mission, organization, and function the existing and proposed: a) data model(s) (e.g., presented in entity-relationship format); b) business rules that govern the definition, production, storage, ownership, management, interchange formats, and replication of data, based on data categories (e.g., personnel data); and c) data standards/interoperability issues



and efforts that relate to accomplishing the organization's mission(s), and interrelates these to the various business events that instigate the business information systems that ultimately provide mission support data.

A data asset project is an organized set of activities used to solve problems related to the seven goals of net-centricity. Data asset projects produce one or more data asset products identified within the data management planning process.

2.3 Net-centric Environment

The core of the net-centric environment is the data that enables effective decisions. By net-centric, it is meant that the network (including its infrastructure of hardware and software) is at the center of information exchange rather than one or more database or information systems. It is however well recognized that the information exchanged over the network is valuable if and only if the databases and information systems that are the network's data sources exist in a high quality, reliable data management environment.

In the context of net-centricity, data implies all data assets such as system files, databases, documents, official electronic records, images, audio files, web sites, and data access services. One of the key goals is to populate the network with all relevant and appropriate. All data is advertised and available for users and applications when and where they need it. In this environment, users and applications search for and "pull" data as needed. Alternatively, users receive alerts when data, to which they have subscribed, is updated or changed (i.e., publish-subscribe). Authorized users and applications have immediate access to data posted to the network without processing, exploitation, and dissemination delays.

To facilitate the discovery of data assets, users and applications "tag" data assets with metadata, or data about data. Similarly, to facilitate the exchange of data assets, users and applications may also tag data-asset data. Users and applications post all data assets to "shared" space for use by the enterprise. In the net-centric environment tagging, posting, and sharing of data are encouraged through the use of incentives and metrics.

The seven net-centric goals and their applicability to data management are provided in the table that follows.



| Net-centric Data Management Support | | | |
|-------------------------------------|--|--|--|
| Net-centric Data Goal | Net-centric Definition | Data Management Support | Support Description |
| Visible | Users and applications can discover the existence of data assets through catalogs, registries, and other search services. All data assets (intelligence, nonintelligence, raw, and processed) are advertised or “made visible” by providing metadata, which describes the asset. | Identify data assets within natural contexts of mission, organization and function. Standardize taxonomies, ontologies and classification schemes that then enable viewing data semantics including where and how implemented through data assets. | The data management approach supports the net-centric goal by ensuring that data assets are set within their natural contexts of mission, organization, and function. Further, that they are all tagged with standardized taxonomies, etc., at a level and through a strategy that ensures consistent tagging of similar data assets. This is accomplished through use of ISO 11179 for data elements, and conceptual, logical, and physical data models. physical data models are at the level of data assets. The Data management approach also enables the automatic redetermination that data assets need to be re-tagged and the resultant metadata re-posted. |
| Accessible | Users and applications post data to a “shared space.” Posting data implies that (1) descriptive information about the asset (metadata) has been provided to a catalog that is visible to the Enterprise and (2) the data is stored such that users and applications in the Enterprise can access it. Data assets are made available to any user or application except when limited by policy, regulation, or security. | Include a metadata definition in every data asset. Standardize names, definitions and structures first through templates and then incorporate these standardization artifacts, where practical, within data assets or at least map these standardization artifacts to data assets. | The Data management approach supports the net-centric goal by enabling the automatic production of a comprehensive set of metadata that would enable users to thoroughly understand the semantics of a data asset. Included would be all data elements, their encompassing concepts, conceptual value domains, value domains, the conceptual entities within the data asset, the metadata for each database table, and the operative processes and business rules that govern data integrity. For each data object in the metadata (e.g., data elements, entities, tables, and business rules), there would be definitions, descriptive material, and precise formats. For application use of the data asset, there would be both XML schemas for the various standardized interfaces and SQL Views. |
| Institutionalize | Data approaches are incorporated into enterprise processes and practices. The benefits of Enterprise and community data are recognized throughout. | Standardize strategies for data definition. Create multiple layers to ensure define once use many times. Use ISO standards 11179 (part 3) for shared data elements and SQL for data models. | The Data management approach supports the net-centric goal by ensuring that there is a comprehensive set of metadata that is consistent within a data asset, harmonized across data assets within a community of interest, and harmonized across families of communities of interest. This is accomplished by ensuring that all data asset metadata, that is, data elements, conceptual, logical, physical and business rule models, are both vertically and horizontally harmonized. Further, the Data |



| Net-centric Data Management Support | | | |
|-------------------------------------|---|--|---|
| Net-centric Data Goal | Net-centric Definition | Data Management Support | Support Description |
| | | | management approach is supported by configuration management that enables knowledge of development, test, and iterations of production data assets. |
| Understandable | Users and applications can comprehend the data, both structurally and semantically, and readily determine how the data may be used for their specific needs. | Create standard vocabularies, commonly inherited semantics, commonly used data model templates, automatic names and definitions based on well defined words. Automatically create abbreviations where necessary. | The Data management approach supports the net-centric goal by ensuring that data assets are set within their natural contexts of mission, organization, and function. Further, that the metadata layer of data assets is able to be explored to ensure users that they not only understand a given data asset but also are able to browse among data assets of similar type, construction, or mission, functional, and organizational context. The syntactic and semantic metadata is engineered such that it can be readily employed by commonly available, industry best practice tool sets such as CASE tools, DBMSs, and XML readers and writers. |
| Trusted | Users and applications can determine and assess the authority of the source because the pedigree, security level, and access control level of each data asset is known and available. | Create consistent semantics, standard reference tables within ISO 11179 based shared data elements, and employ these to govern/map data models across multiple levels of abstraction. | The Data management approach supports the net-centric goal by standardizing both the semantics and value domains and meanings across all authoritative data sources. Additionally, the Data management approach enables the mapping of prior value domains and meanings to current value domains and meanings, and the mapping among uses of the same or semantically similar value domains and meanings. The Data management approach enables the quick determination of the processes that capture, update, and report data asset data. And finally, the Data management approach enables the allocation of various classes of security and access control. |
| Interoperable | Many-to-many exchanges of data occur between systems, through interfaces that are sometimes predefined or sometimes unanticipated. Metadata is available to allow mediation or translation of data between interfaces, as needed. | Standardize data structures. Establish well engineered data transactions, automatic XML wrapping of data, supported by accessible data definitions and contexts. | The Data management approach supports the net-centric goal by ensuring that regardless of localized naming and data types that the same data is commonly mapped to higher level data models and/or data elements. With the Data management process, data that is named the same but is different will be able to be discovered as easily as data that is named differently that is the same. The hierarchy of data models (data element to |



| Net-centric Data Management Support | | | |
|-------------------------------------|---|---|--|
| Net-centric Data Goal | Net-centric Definition | Data Management Support | Support Description |
| | | | conceptual to logical to physical to view) enables exchanges between and among systems to be quickly specified and managed. |
| Responsive to User Needs | Perspectives of users, whether data consumers or data producers, are incorporated into data approaches via continual feedback to ensure satisfaction. | Support reuse of already defined data assets metadata, central knowledge of all data assets and distributed access to same. | The Data management approach supports the net-centric goal by ensuring that data assets are set within their natural contexts of mission, organization, and function. Additionally, data assets that are either universal or that cross organizational boundaries are able to be interrelated. Specified in support of system exchanges are business cycles, calendars, and data value precisions. The Data management approach causes the What, When, Where, Why, Who, and How to be both captured and interrelated in support of ensuring satisfaction among data producers and consumers. |

2.4 Data Asset

A data asset is data in all forms: raw, processed, intelligence, non-intelligence, processes, applications and data sources. All data assets must be specified, created, and maintained through a data asset product.

A data asset may be an IESS data asset when it represents a consensus based data exchange standard determined by its user community. An IESS is a standard because it is subject to configuration management by its user community.

A data asset may be designated as an ADS when its value set is declared to be definitive. An ADS value set is subject to configuration management .

A data asset may have its data uniquely defined through EID so that EID-based assets themselves can be discovered, analyzed, and combined.

A data asset may be in either XML format or some other format.

Data assets within information systems must be periodically reviewed to ensure that data assets (exposed and unexposed) are available to the widest possible community. A data asset is considered exposed when it is tagged/registered with a proper metadata registry; registered in an enterprise searchable catalog; understandable; accessible to the widest possible community.



Within the context of a data asset, a data asset project causes both the data asset and the metadata intrinsically part of the data asset to be created. metadata is maintained and integrated with all other data asset metadata to ensure a completely integrated and seamless set of data assets across domains.

2.5 Data Asset Products

Data asset products are the data-specific inputs needed or outputs produced in data asset life cycle activities. Data asset products commonly organize and define interrelationships of data in support of each organization's missions, functions, goals, objectives, and strategies. These products give the basis for the incremental, ordered design and development of one distributed virtual database founded on successively more detailed levels of data specifications to build out the data asset product set. metadata repository identify the set of data asset products in each data asset project.

Data asset products provide the basis for the incremental, ordered design and development of one distributed, virtual database founded on successively more detailed levels of data specifications to "build out" the Data Asset Product set.

Certain data standards are defined within data asset projects, and become the underlying foundation of other data asset project products. For example, one data asset project may fully specify reference data standards as its data asset product, which, in turn, is then used in another data asset project as a critical component within its complete data asset specification. The functional and technical documentation about a data asset's structure, meaning, and use comprise the metadata for every data asset product. To achieve data exchange capabilities for information exchange requirements, implementing and enforcing data standards is mandatory.

To document and implement data standards, functional data proponents must be actively involved in managing information content. Functional proponents establish the business rules for data use and define the constraints governing how data are processed (e.g., referential integrity constraints for add, change, and delete transactions against records in a database). Business rule statements describe these constraints. To achieve consistency in data use, different data, obtained from different sources, must be combined (synchronized) in accordance with functional business rules defined by the functional process proponent.

2.6 Metadata

The term "metadata" refers not only to the set of definitions of the data in a data asset (i.e., products, parts, prices, etc.), but also to its formats, processing, transformations, and routing from source to target information system – everything, except the actual data's content, constitutes metadata.



Metadata allows the management of content. More importantly, metadata enables knowledge of the content, the environment within which content lives, interrelationship among content environments, and the ability for content to evolve. Simply put, high quality metadata management ensures flexibility, interoperability and evolvability.

In the context of net-centric conforming Data Asset data management planning, metadata necessarily also includes: missions, events, information systems, functions, and organizations within which the Data Views are defined and "live."

Metadata can be employed a variety of ways to enhance the value and usability of data assets. The traditional data administration approach used metadata to define data structures and relationships (e.g., data models) to support the development of databases and software applications. This "structural" metadata defines how data assets are physically composed and can include information that describes the relationship between specific parts of the data asset and what elements, or fields, are used in its definition.

In addition to supporting systems development, metadata can be associated with all data in the Enterprise for the purposes of "advertising" data assets for discovery. metadata that describes or summarizes key attributes and concepts of a data asset are used in the discovery process. This "discovery" metadata allows users and applications to quickly search through a wide range of data assets to identify those assets that are most valuable to support their needs. There are many other types of metadata including vocabularies, taxonomic structures used for organizing data assets, interface specifications and mapping tables.

Data Asset metadata exists at five levels: data elements, conceptual data models, logical data models, physical data models, and view data models. Data elements are the business fact templates for attributes of entities within conceptual data models. Conceptual data models are templates for use in building logical data models, which in turn are configured into appropriate physical data models that finally are intersected with AIS through view data models. There is a direct relationship between the quantity of levels of interoperable metadata and data asset interoperability. There is also a direct relationship between the quantity of metadata tagged data asset exchange instances and the quantity of metadata levels. Finally, quality metadata at the data element, conceptual, and logical, and physical data model levels can be interrelated, thus enabling semantic comparisons across and within the levels. Data asset interoperability begins with quality, interrelated, multi-level metadata repositories within federated environments.

2.7 Data Interoperability.

The data aspects of interoperability may be summarized by the term basic interoperability, which is the exchange of information that preserves the meaning and relationships of the data exchanged. In order for information to be fully understandable and interoperable, it is required at a minimum that --



- Semantics and syntax are well specified;
- Data elements are identifiable at the enterprise level;
- Authoritative data sources are well defined and managed;
- Exchange mechanisms are able to support current and future demands.

Interoperability consists of two parts: shared value streams and shared understanding. Both of these are created from within the Community of interest and are expressed via the IESS.

The role of EIDs within data interoperability is to support technology independent mechanisms to understand both metadata and values (both single value and value sets).

The role of ADS is to minimize the versions of the "truth". Additionally, an ADS enables the coordinated migration of "truth" from an originating value state through a chain of value states until the data source is either archived or deleted.

Finally, the role of XML is to take the value streams from an originating system and to transport them to an IESS or vice versa. Embedded within the XML stream are the EID tags that enable users to both understand the authority of the value sets and the supporting metadata. It is the primary function of a data model to provide a common specification of the meaning and relationships of information by which interoperability may be achieved. The IESS is a logical data model that represents the shared data of legacy system physical data models from members of a Community of interest. The Community of interest's end product is not only the IESS; it is also the mapping between the IESS's logical data model and the legacy system physical data models. To have consistent semantics across all the IESS logical data models, there needs to be two additional data model layers: enterprise data elements and shared data structure templates (i.e., conceptual data models).

- The enterprise data elements are fact-based, semantic templates for all the columns in the tables of the logical data model. These enable Community of interest logical data models to be interrelated. The enterprise data elements will come mainly from an analysis of existing databases and information systems with the goal of discovering these common semantics, regardless of differences in names, contexts, and data types.
- The shared data structure templates (conceptual data models) facilitate the "manufacturing" of data models' well-engineered collections of commonly employed enterprise data elements (e.g., materiel requisition or disposition, facility location characteristics, and person biographic information). Enterprise data elements are based on the information systems standard 11179, Part 3; and the conceptual, logical, and physical data models are based on information systems/ANSI standard SQL.



2.8 Communities of Interest

Communities of interest is the inclusive term used to describe a collaborative group of users who must exchange information in pursuit of their shared goals, interests, missions, or business processes, and who therefore must have a common vocabulary (names, meaning, schema/format) and common business rules for the exchange of information.

The purpose of a community of interest is to create a shared understanding of the terms, a vocabulary, used to describe and define the data asset. Simply put, a shared set of data semantics. Communities of interest will form in a variety of ways and may be composed of members from one or more functions and organizations as needed to develop the shared mission vocabulary.

Communities of interest always contain both data asset producers and consumers. Communities of interest may cross information system boundaries and organizational boundaries and sometimes functional boundaries as well. Subordinate communities of interest may be created when a part of the community of interest needs a specialized vocabulary for a subset of the community of interest subject area. New communities of interest will be formed as needed, and old communities of interest terminated when no longer useful.

The community of interest structure consists of three layers: Enterprise, Institutional, and Expedient

- Enterprise communities of interest focus on data harmonization and integration across the enterprise. This community of interest defines the central concepts that are used in total or in part by all other community of interest. Examples of such basic concepts are facilities, organizations, persons, and materiel. The enterprise community of interest focuses on those enterprise data elements and shared data structure templates constitute the primary data asset for the enterprise community of interest.
- Institutional communities of interest supervise the long-term development and configuration management of their respective functional area vocabularies, business rules, and authoritative data sources. Data schemas developed by an institutional community of interest would have to be conformant (in the ISO 11179 sense) to the enterprise schema as a starting point. Only then would each institutional community of interest extend its data specification to cover its functional area
- Expedient communities of interest are formed to address high-priority, capability deficiencies that must be addressed in a timely fashion to support command priorities and operations. Expedient communities of interest will typically be cross-functional and will need to coordinate capability gap analysis results with institutional communities of interest for long-term correction and management .

In addition to these three classes of communities of interest there can also be cross-enterprise and cross-institutional. These two additional classes enable enterprises to dissolve stove-pipe



organizational structures and the stove-pipe information systems and databases that are their natural consequence.

Communities of interest form in a variety of ways and may be composed of members from one or more functions and organizations as needed to develop the shared mission vocabulary. Every Community of interest has a lead and a set of stakeholders supervising its operations. Communities of interest contain data asset producers and consumers. Communities of interest are most commonly cross-functional.

Expedient Communities of interest may be created when necessary. Their data asset products are incorporated into the parent's community of interest product set. New communities of interest are formed as needed and old communities of interest terminated when no longer useful.



3.0 Data Standards

The four data standards essential for data asset interoperability, that is, ADS, EIDs, IESS, XML. These are presented in Sections 3.1 through 3.4. There are, however, two additional classes of standards that frame the totality of data interoperability. That is: technology standards, and data wrapping standards.

Technology standards, effect the standardized expression required for certain content. For example, in the SQL command, CREATE TABLE Customer, CREATE TABLE is the SQL data definition syntax, and Customer is the content. There are currently three standards that address standard expression. The first, SQL, forms the language that is most commonly employed for expressing data content's definition, data content's access, and data content's protection within individual collections of data asset instances. The second, information systems Standard 11179 for data element metadata, is employed to control the engineering of the metadata surrounding unitary facts.

Data wrapping standards, most commonly typified by XML, which is promulgated by the W3C, is employed to wrap collections of data in tags such that these data collections can be transported between processes and environments in a technology independent manner. There are other technology standards such as SOAP, WSDL, and UDDI which facilitate transactional interoperability. These technology standards address only the mechanisms of data transport, not the content of the data transported. These are outside the scope of this chapter.

Standards for discovery and availability impact net-centricity. These standards include universal description, discovery and integration (UDDI), and Web services description language (WSDL). UDDI provides a conceptual phone book for Web Services. Organizations may register information about their Web services and types of services with UDDI. WSDL describes the operational information -- where the service is located, what the service does, and how to talk to, or invoke the service. These standards are important to the concepts of visibility and accessibility of data as addressed in the net-centric data Strategy.

Architecture development standards are needed because the semantic meaning and rules for information exchange need to be determined. It is important to remember that XML does not create semantics; it uses already created semantics. Semantics need to be captured and documented in the integrated architecture development process and products. In the context of data interoperability it is vital to focus on data-related architecture products and model those elements that help develop the Community of interest and Cross-Community of interest Ontology. Data centered ontologies include entities, relationships, properties, values, and axioms/rules.



The data standards, versus technology or data wrapping standards, effect standard constructions of data, that is, content, that can be employed within and/or that classify one or more data assets. The data standards are:

- Authoritative Data Sources
- Enterprise Identifiers
- Information Exchange Systems Specifications
- Extensible Mark up Language

XML is within this class because there needs to be standards around the construction of tags that provide the semantic meaning for value representations. E.g., Telephone Number is the tag and 1-703-602-6880 is the valued representation.

3.1 Authoritative Data Sources (ADS).

Authoritative data sources are those sets of data commonly used by or referenced by other databases and information systems. The most common form of ADS are reference data such as product codes.

Owners of reference data will make the coded data values available as an ADS to ensure maximum reuse and interoperability. These value sets will be stored in the metadata repository.

Data synchronization requirements will be identified and documented as part of the ADS documentation.

Data synchronization requirements will consider information flows and reference table value domains (including data transfers, system run cycles, management decision cycles, timeliness, and accuracy).

Data standards producers will:

- Create and maintain ADSs, such as reference table value domains, force structure decompositions, etc., whose values are shared among information systems.
- Synchronize ADS implementations with the uses of ADS data to ensure common and consistent use across all ADS employing organizations.

3.2 Enterprise Identifiers (EID).

Enterprise Identifiers are persistent values assigned to data assets such that the EID value is independent of any implementing technology and such that whenever that data asset is employed, its EID is also employed.



All data that is collected and maintained in certain designated databases will use globally unique enterprise identifiers (EIDs) to ensure full data integration, referential integrity, and data interoperability.

Data standards producers will:

- Use EIDs in both specified legacy and all new information systems to ensure maximum data integration and data interoperability.
- Use EIDs in legacy systems. Specified legacy systems will add EIDs to their physical schemas in a manner that best fits their fiscal constraints and user needs.
- Use EIDs in new systems. All new systems will add EIDs to their physical schemas in a manner that best fits their fiscal constraints and user needs.
- Use EIDs in commercial Enterprise Resource Planning (ERP) applications. As part of the contractual agreement with ERP application developers, provisions must be made in their physical schemas for the use of EIDs.
- Support and ensure that all pertinent data resources identified via globally unique EIDs will be maintained and registered to permit discovery and reuse within functional areas and at the enterprise level. Specifically, all reference data sets identified with EIDs will be documented and published in the metadata repository to facilitate exchanges by other users.
- Maintain a registry in the metadata repository of all the EID Seed users to provide optimal implementation oversight of the EID-based key management process.

3.3 Information Exchange Standards Specifications (IESS).

Information Exchange Standards Specifications (IESS) are those data structures that are developed through consensus, most commonly within a Community of Interest that act as a common are for storing and retrieving shared data.

To control the production and applicability of data standards required to ensure data interoperability for data exchanges among information systems, the participating systems must conform to standardized data exchange specifications. An information system is be deemed "conformant" with an approved Information Exchange Standards Specification (IESS) if the model of the particular information systems:

- Is based either on the entire IESS or on a subset of the IESS (not all attributes of selected entities need to be implemented).



- Has extensions of that subset but the extensions are not redundant with elements of the IESS itself; emerging extensions that could apply to a specific IESS will be proposed for general use in succeeding versions.
- Uses agreed data types and coded domains.
- Identifies points of contacts for generating instances of EID keys (to avoid redundancy and non-uniqueness).
- Has key attributes identical with or directly derivable from key attributes specified in the IESS. Alternatively, the IESS-conformant information system uses alternate keys, but the original IESS keys are preserved. To ensure fully faithful information transfer among databases, the IESS-defined primary keys of one database for any entity comprised within the IESS specification must be identical either to the primary or alternate keys of the same entity in any other IESS-conformant database. The primary or alternate key, in this case, will be based on the EID from the ADS.

All information systems will exchange data by specifying their exchanges within the metadata repository in a format that conforms IESS developed and agreed to by the Community of interest that supports the respective information systems.

Whenever database implementations identify data requirements not yet in a pertinent IESS, these will be shared with members of the Community of interest that owns the IESS so that the IESS can evolve to include all the core requirements.

Program Managers, Community of interest, and/or materiel developer will:

- Develop and maintain architecture models, data models, business rules, and other artifacts within the metadata repository.
- Will review and/or approve submissions to the metadata repository.
- Be responsible for integrating COTS software and ensuring interoperability with the existing metadata contained in the metadata repository.

Data standards producers will:

- Configure all data models through the ANSI Standard SQL language for their base set of data model artifacts and create necessary supporting business rules and processes as required by the metadata repository to specify all IESSs. To ensure maximum interoperability, the IESSs must be implemented through software that reliably conforms to the metadata repository.



- Add to the list of relevant tools any evolving structured languages for creating IESSs, if sufficient governmental and commercial support develops for them.
- Use only tools with non-proprietary extensions. IESSs will not be created with tools that use proprietary extensions for which there is no translation mechanism into and out of the metadata repository.
- Use information systems 11179 data elements already tested within Communities of interest whenever practical vice newly created information systems 11179 data elements. When selecting existing data elements for exchange, enterprise activities will adhere to the following order of precedence (highest to lowest) for selection:
 - ◆ ISO 11179 data elements from cross-functional Communities of interest
 - ◆ ISO 11179 data elements from enterprise Communities of interest mapped to those elements from cross-functional Communities of interest.
 - ◆ ISO 11179 data elements from other intra-enterprise Communities of interest mapped to those from cross-functional Communities of interest.

3.4 EXtensible Markup Language (XML)

All XML tags for use in data exchanges will be derived from the pertinent IESS adopted by the Community of interest engaged in such data sharing and reuse activities. For ease of use, the physical data table and column names from the data models contained in the metadata repository will be used for the generation of XML tags. However, logical names from the metadata repository may also be used to enhance readability. If that is the case, an appropriate set of Extensible Stylesheet Language/Transformation (XSL/T) files to transform the tags into a form that facilitates the automated import into IESS-conformant databases will be provided and maintained by the Community of interest.

All data exchanges among information systems executed via web-based solutions will use XML as their transfer mechanism. The producers of the data will register their XML metadata and non-XML metadata (i.e., data models, message formats, database schemas) with the metadata repository.

Data standards producers will:

- Use World Wide Web Consortium (W3C) Technical Specifications holding a "Recommended" status to ensure maximum interoperability. A W3C Recommendation is a technical report that is the end result of extensive consensus building about a particular technology or policy. See <http://www.w3c.org> for further definition.



- Adhere to XML related standards promulgated by other nationally or internationally accredited standards bodies when developing applications within the domain that the standard addresses.
 - ◆ When a standard produced by one of these bodies competes with a similar product of the W3C, the W3C standard will take precedence.
 - ◆ XML implementations must not use proprietary extensions to XML-based specifications.

Actively participate in the work of appropriate XML and XML-related technical and business standards bodies. The data administrator will act as coordinator of such participation.

Use existing XML components whenever practical vice developing new XML components.

When selecting existing XML tags, activities will adhere to the following order of precedence (highest to lowest) for selection:

- Cross functional Community of interest IESS-based tags
- Enterprise Community of interest IESS-based tags
- Intra-enterprise organization Community of interest IESS-based tags
- The above recommended order does not preclude selection of a component with lower priority when other considerations, such as cost, implementation schedules, etc., would make the use of a component of higher ranking less defensible. All XML business standards will be at the enterprise level..

Leverage commercial practices, standards and products before creating unique ones



4.0 Communities of Interest Data Management Roles and Responsibilities.

Broadly speaking, communities of interest are communities of people who share a common interest. These people exchange ideas and thoughts about the given area, but may know (or care) little about each other outside of this area.

Communities of interest have many reasons for existence, such as human resource issues, logistics, manufacturing, technology standards, and community groups such as zoning, sports, or other civic areas.

Data management clearly functions as a community interest as a natural consequence of desiring to define and exchange data with maximum efficiency and minimum semantic conflict. Within such a community, the area of interest among the users is to exchange information in pursuit of their shared goals, interests, missions, or business processes, and who therefore must have a common vocabulary (names, meaning, schema/format) and common business rules for the exchange of information.

Each community of interest creates and maintains the community of interest's shared vocabulary, shared data spaces, metadata catalogs, and the registration of all pertinent metadata. During the accomplishment of the community of interest's program and scope of work, communities of interest:

- Create a shared understanding of the semantics used to describe and define the data assets.
- Execute data management planning through data asset projects that identify, create, and maintain data asset products.
- Capture the data asset specifications used by the community of interest. Work with architecture efforts in their problem space to ensure that data asset products stay aligned and integrated with all other appropriate products.
- Focus on that part of its subject area with the highest return on investment. High-priority information usually includes information required by a new or future capability and information that must cross organization or system boundaries.
- Ensure data assets are visible and accessible.
- Create and maintain a metadata repository of data asset products..
- Ensure that data assets that are to be interoperable are supported by EIDs, IESSs, and, as appropriate, XML to support asset identification and access, shared data exchanges, and exchanged data formatting.



- Identify data assets that are the ADS in the community of interest subject area, including the operational data owners supervising their management . Communities of interest may have to resolve potentially contradictory sources and coordinate with enterprise-wide governance bodies to reconcile/adjudicate authoritative source(s).
- Determine data owners and controllers to determine data creation and update cycles that govern the business rules related to data interchange.
- Develop data asset projects' plans, schedules, and funding. All community of interest participants and data owners update their planning, programming, and budgeting system processes and policies, as well as acquisition processes and policies, to reflect their participation in the Community of interest effort.

In general, communities of interest consist of the following staff types:

- Lead: an individual from a specific organization who has been tasked to "manage" the COI. Usually the organization that is leading the COI activity has committed to driving the COI to a solution and will ensure that agreements are implemented within the organization's plans, programs, and budgets.
- Governing authority: An organization that will review and adjudicate COI conflicts and will push for component implementation and support of those agreements.
- Stakeholders: Organizations or personnel who have an interest in the outcome of the COI effort. May not be active participants in the COI (i.e., a COI member) but will likely use and/or benefit from the capability.
- Developers: Personnel or organizations responsible for actually implementing the data sharing agreements.
- Data producers: Organizations, systems, programs, and personnel that create and maintain data assets
- Subject matter experts: individuals from participating organizations who's special knowledge enables them to be intensively engaged in defining semantics.

Within the scope of these communities of interest a key data management function is one of data administration. Paralleling the communities of interest, there are:

- Enterprise data administrator
- Institutional data administrator
- Expedient data administrators



Also there are cross-enterprise data administrators and cross-institutional data administrators. Finally there are the standards data producers, and the functional stakeholders

4.1 Enterprise Data Administrator Responsibilities

The data administrator is responsible for oversight and development of data standards policy, guidance, and procedures. The data administrator:

- Identifies enterprise communities of interest, Community of interest leads, and community of interest data administrators, who are responsible for data standards in their functional areas.
- Develops data standards strategies, implementation plans, and performance measures.
- Creates, deploys and maintains the metadata repository in support of the data management program.
- Establishes a governance structure to oversee data standards implementation, including processes and procedures, working groups, tools, training, and other resources.
- Acts as focal point for data standards activities, to include coordinating with external organizations.
- Develops and maintain a list of mandated approved data standards.
- Provides input on program milestone reviews as to compliance with data management policy.
- Coordination, integration and maintenance of IESS.
- Coordination of ADS across communities of interest.
- Coordination of enterprise identifiers across communities of interest.
- Development and coordination of technical implementation guidance.
- Coordination, management and integration of community of interest data-related architecture products with operational, system and technical architecture products.
- Coordination, integration, and maintenance of other related data standards that may be identified as critical to the success of data interoperability.
- Coordinate semantics across multiple enterprises as may be appropriate.



4.2 Institutional Data Administrator Responsibilities

The institutional data administrator is responsible for oversight and development of data standards policy, guidance, and procedures within a specific functional community of interest. The community of interest data administrator:

- Identifies community of interest data standards producers to carry out data management and standards actions for the organization and serve as liaisons between functional experts and technical personnel.
- Identifies funding requirements in support of the data standards producers, for their institutional community of interest.
- Develops data standards strategies, implementation plans, and performance measures.
- Creates, deploys, and maintains metadata repository content in support of the data management program.
- Reviews the data structure of assigned data standards in the metadata repository at each milestone and at five-year increments after system deployment.
- Creates or update metadata repository content that is either exchanged with or disseminated to any other organization.
- Manages uses of the ADS, IESS, EID, and XML data standards to ensure valid implementation.
- The community of interest leads will oversee the data management activities, identify appropriate governance strategies, and appoint a community of interest data administrator to carry out and implement data management actions for the community of interest.
- Community of interest leads will provide courses of action for the transition of legacy systems and data repositories to net-centric environments..
- Community of interest leads will coordinate with the data administrator to ensure that their activities, efforts, and products are integrated to the maximum level appropriate.
- Community of interest leads will ensure that their common vocabulary is harmonized with their community of interest ontology development and the mission area ontology efforts to enable visibility and accessibility of their data assets.
- Community of interest leads --
 - ◆ Develop the community of interest common semantics



- ◆ Capture community of interest operational/business.
- ◆ Register community of interest data schemas and models.
- ◆ Identify ADS within the community of interest.
- ◆ Promote data sharing across the enterprise.
- ◆ Supervise the community of interest net-centric migration plan.

4.3 Expedient Data Administrator Responsibilities

The expedient data administrator is responsible for oversight and development of data standards policy, guidance, and procedures within a specific expedient community of interest. The Community of Interest data administrator:

- Identifies community of interest data standards producers to carry out data management and standards actions for the organization and serve as liaisons between functional experts and technical personnel.
- Develops data standards strategies, implementation plans, and performance measures.
- Creates, deploys, and maintains metadata repository content within the program and scope of work.
- Creates or update metadata repository content that is either exchanged with or disseminated to institutional communities of interest.
- Report data semantics determinations to the appropriate institutional communities of interest and harmonize as appropriate.

4.4 Data Standards Producers Responsibilities

The data standards producer is responsible for the actual development of IT products that conform to data standards policy, guidance, and procedures within a specific functional community of interest. The data standards producer will:

- Use the metadata repository for the procedural storing, universal viewing, and selective reuse of (all, or parts of) architectures, data models, business rules and other artifacts of functional systems.
- Use the metadata repository content to perform technical reviews of functional data requirements. Information about data/meta data will be maintained and controlled in the metadata repository as part of the standard metadata documentation.



- Use data standards in conjunction with data standards documentation in information systems design documentation from the metadata repository.
- Use data standards in newly developed and redesigned applications and, when feasible, in existing systems.
- Submit candidate data standards for approval to the respective community of interest.

4.5 Stakeholders Responsibilities

Along with the user community, material developers, program managers, system owners and data producers make up the stakeholders of a Community of interest. The system development, acquisition and migration approach defined by the Community of interest will need to be planned for and executed by the stakeholders of the Community of interest. Stakeholders --

- Assist in the development and execution of the community of interest scope and programs of work.
- Ensure that the data and metadata conforms to the requirements of the community of interest.
- Ensure that the data and metadata conforms to the net-centric goals.
- Register metadata in metadata repositories.
- Plan and budget for services or capabilities to be exposed to the enterprise.

The above stakeholder efforts are associated with working within the community of interest construct. In the absence of community of interest activity, stakeholders and material developers can take the following actions to prepare for net-centric operations:

- Identify and prioritize shareable data assets within their individual systems
- Identify candidate ADS (those currently used by the system or considered for use by the system)
- Identify candidate services that the system may provide to the enterprise
- Plan for migration of their system/application to operate in a net-centric environment using Web services and associated protocols (for example, XML, SOAP, UDDI, and SDL)



5.0 Data Management Program Architecture

The data management program consists of three distinct layers: the program layer, the planning layer and the execution layer. The overall mission of the data management program is to manage and leverage information. This is accomplished through a program that involves its own architecture, governance, program components, and then specific projects to carry out the program. These three layers cascade one into the other.

5.1 Program Layer

5.1.1 Program Architecture

The overall architecture of the data management program is based on a top-down and bottom-up approach. The top-down components provide guidance and facilitation. The bottom-up components provides empowerment to the ultimate creators, owners and users of data asset products. Data asset products reside within communities of interest and are manifest in a federated metadata environment. The data asset products across communities of interest are then harmonized. These data asset products then are contributed, as appropriate, to the appropriate metadata repositories.

The net-centric data management program is accomplished through layered communities of interest throughout the enterprise to ensure both vertical and horizontal semantics integration. The data represented through the IESSs are made trustworthy through authoritative data sources for all critical reference data. The data is then guaranteed uniqueness through enterprise identifiers. Finally, data exchanges through highly managed XML definitions and exchanges.

This layer, the program layer consists of policy, guidance, procedures, methodologies, courses, workshops, and metadata repository software tool support is engineered, tested, and administered.

5.1.2 Data Management Program Governance.

The governance of the data management program is through policy, guidance, and communities of interest data goals assessments. Supporting the guidance are a myriad of workshops, white papers, seminars, and software tool sets. These collectively assist staff as they develop, publish, integrate, and maintain all data management program work products within and across communities of interest.



5.1.3 Data Management Program Assessments.

The net-centric data goals are the characteristics through which the data management program is assessed. These assessments are accomplished through a data management capability maturity model styled assessment that not only assesses the data management program projects for net-centric data goals but also for data management best practices.

Metrics (measures) are established for each goal, to assess the degree to which they have been achieved. Within each goal one or more objectives are established. Within each objective one or more strategies are established. Objectives and strategies characterize broad actions to pursue each goal.

5.1.4 Data Management Program Components

Data Management Program components include: process, technology standards, metrics, project management, technology components, data management planning, and training and awareness.

Process includes overall strategy and activities that are required to be performed to successfully accomplish the data management program. Process includes an overview of the activities within each of the data management program components, that is, the data management program process itself, technology standards, metrics, project management, technology components (e.g., EIDs, ADS, IESS, and XML (for data transport where appropriate)), the data management planning environment, and the necessary training and awareness. Examples of Process projects include this document, the creation, use, and evolution of the metadata repository, the creation of data asset product specifications, and the data management program methodology through which all data management program products are created, interrelated, employed, and evolved.

Technology Standards include, for example, information systems/ANSI Standard for the database language, SQL and the information systems 11179 standard of data element metadata. Described are each standard and the organizations that cause these standards to be created, modified and employed. Examples of technology standards projects include the creation and deployment of procedures regarding the proper use of these technology standards, the gathering of requirements to then advocate to the de jure standards bodies that develop and evolve these standards, and possibly the creation of technology standard use profiles that specifically identify subsets of features within these technology standards that are appropriate for use within the enterprise.

Metrics include different classes that should be involved with data asset product specifications. Examples of metrics projects include the development of metrics across all data asset products both for estimating the resources required for data asset product accomplishment, and also for evaluating their quality.



Project management includes the organization, planning, ongoing management, and evaluation of the various data management program projects. Examples of project management projects are the development of information systems plans, and the monitoring of multiple data asset projects across and within domains.

The technology components of the data management program include authoritative data sources (ADS), information exchange standards specifications (IESS), XML (where appropriate) for data transport, and enterprise identifiers (EIDs) to uniquely identify instances of data assets. Examples of technology component projects include the creation of the underlying design, implementation, and evaluation of the use of ADS, IESS, XML, and EIDs within specific data asset projects within and across domains.

Data management planning represent the structured approach wherein the complete set of data management program projects are identified, defined, and collected such that any resulting data asset can be judged compliant with the net-centric data goals. Examples of data management planning projects would be the determination of all the data assets projects required for a major enterprise program including the proper sequencing of those data asset projects, the determination, monitoring and evaluation of all data asset projects associated with a specific domain or data classification, and the determination of an impact of a changed requirement within a data view across all domains.

Training and awareness includes seminars, documents, white-papers, workshops, and web sites through which all the projects of the data management program can be understood and accomplished. Included in training and awareness are the activities associated with the various special interests, such as data stewardship, data modeling, XML schema creation, discovery metadata tagging, and the like. The data management program determines the proper membership to these special interest groups and then facilitates workshops, course, presentations, papers, and if necessary procedures to accomplish these critical data interoperability issues. Examples of training and awareness projects include the development and accomplishment of a data interoperability specification workshop, the proper use of the metadata repository system, and advanced techniques in assessing data asset product changes across a domain.

5.2 Planning Layer

The data management program applies direction for data management planning. Enterprise guidance takes a decentralized approach to funding by incorporating the management of data projects into the normal funding cycle and looks to the various functional proponents and enterprise components to champion their data projects.

Data management planning collects, develops, and maintains all information about data required by the enterprise constituency of organizations to accomplish its mission tasks. It produces



complete, implementable projects that include all metadata, measures, studies, plans, prototypes, models and databases necessary to:

- Enable subsequent and improved data integration and reuse; and
- Provide improved project scoping, responsiveness to business change, management of systems development sequencing and prioritization, and level of business involvement.

Data management planning ultimately results in the development of the data asset project action plans that set out the plans to achieve the net-centric data goals for data.

Data management planning projects are identified to implement strategies that achieve objectives and attain capabilities. data management planning projects are accomplished through supporting action plans. Measures identified for the supported goal are incorporated into the action plan for the specific project.

Data management planning projects do not actually effect the creation and/or evolution of data assets. Rather data management planning projects identify, plan, and manage data asset project accomplishment. Data asset projects thus develop and/or use the data asset product specifications necessary to create or evolve a specific data asset, that is, the existing and proposed:

- Data model(s);
- Business rules that govern the definition, production, storage, ownership, management, exchange formats, and replication of data, based on data; and
- Data standardization/interoperability efforts and issues that relate to accomplishing the organization's mission(s).

To eliminate the possibility of stovepipe data assets, a certain class of data asset projects may need to create products in the areas of mission, event, information systems, function, and organization so that the various data assets can be set within their proper contexts thus ensuring properly integration across all other data assets of the enterprise, within a community of interest and/or across communities of interest.

To illustrate, a primary focus of a data asset project within data management planning may be to model data and create the specification of how data will be stored, transferred and managed. Thus, a data model may state that a person's name is related to the person's Social Security Number (SSN). Just a data model is not sufficient to specify that the SSN is Privacy Act data and must be protected from unauthorized access, nor will it say how the Person Identifier is created, assigned, physically stored in databases, and formatted for data exchange. These additional rules are part of other data asset products that might be created in other data asset projects. Because there exists an integrated metadata repository that would contain all these products they could be available for use by all. Simply put, the purpose of data management planning is to create and



manage an environment within the enterprise that enables the development of data assets that are flexible, interoperable, and evolvable. Anything short of that is failure.

Data management planning projects begin with a problem statement that is then taken through a rigorous functional and technical analysis process, resulting in a feasible solution for implementation at the appropriate level, ranging from specific, pair-wise, database-to-database exchanges up to enterprise level data sharing in a "virtual, distributed single database" environment. data management planning projects result in the identification of specific data asset projects. The steps in every data management planning project are:

- Developing the problem statement, which an external statement of requirement, or deficiency to identify the need, or the identification of the deficiency to be addressed, the new need to be addressed, or the need for improved performance to be addressed.
- Performing the requisite analysis includes employing enterprise missions, functions, and organization to frame or focus the analysis. If possible, refine this analysis to the point of specifying the problem space to the appropriate level of detail. Develop an initial data management planning project problem statement. Identify the specific objectives that must be achieved to support mission performance. This is a statement of desired results, with specific, quantifiable, and measurable outcomes that contribute to achievement. Use existing and legacy data models to produce a characterized (completely documented) normalized data model representation of the current data-for-exchange environment. Use existing database documentation and meta-data to develop a data dictionary for the current data environment. Use information about the current environment and extend the data dictionary meta-data to capture this information. Review the information with user representatives to verify the "as is" or current environment documentation. The product of this review is a verified specification of the current environment.
- Configuring relevant data asset projects includes their work breakdown structures that are the task statements that support the achievement of the data asset project. The data asset project also includes the various data performance goals and objectives that are to be achieved. Alternative approaches are created and evaluated through Proof of Principle Projects that validate the recommended alternative (risk reduction) or validate the specifications (for issuance to the developer).
- Managing the data asset projects includes the accomplishment of the actual data asset projects. Additionally, it includes monitoring and evaluating the proper use of a created data asset, that is, cycling back lessons learned reports, issue resolution reports, or refinement of performance metrics that were employed in the development effort. In the data asset project, the nature of the technical solution, evaluations via the functional (user) and technical environments are accomplished. Within the constraints of priorities the data exchange requirements, the data use environment, the existing models and data standards, the current legacy data, and the associated meta-data are developed to a sufficient level to begin work on establishing the new or improved capabilities.



The specific types data management planning projects include:

(a) Data management program projects that are to be accomplished in support of some aspect of the data management program overall program. Thus, there are projects related to the data management program program's architecture, governance, and components. Each project has a firm goal, specific objectives, a workplan, metrics, deliverables, and a schedule that must be accomplished. All deliverables must fit within the overall set of all deliverables from all the other data management program projects.

- Component Architecture projects which are specifically targeted at the engineering, design, deployment, and long-term evolution and maintenance of the components within the data management program. Components of the data management program include its overall process, specific standards such as SQL, XML, 11179 Data Element metadata, metrics, project management, the technical components of data management program such as EID, ADS, IEES, XML (as appropriate for data transport), and finally the metadata repository system that creates, holds, and interrelates all data asset products within and across all communities of interest. Each component architecture project would be scoped, its work plan developed, resource loaded and staffed, and managed during its accomplishment.
- Concept of operations projects that address either the data management program in general, or specific components of the data management program architecture. Thus, there would be concepts of operations documents addressing process, standards, metrics, project management, data management program technical components, and the federated metadata repository environment. Each project would be scoped, its workplan developed, resource loaded and staffed, and managed during its accomplishment.
- Data management planning environment projects are generally in three areas: metadata management conformance, the ongoing evolution of the environment's functionality, and extensions to the data management planning functionality.
- The training and awareness projects are those that cause the creation of various presentations, workshops, courses, and support services such as hot-line, on-line tutorials, and the like.
- Methodology projects are created so that different groups of persons, whether contractor or enterprise produce the same set of deliverables from the same or similar requirements. The methodology is to be at least one or more levels more detailed than the actual management of the work. Methodologies have well engineered deliverables and metrics for work efforts. Methodologies are accompanied by training, workshops and as needed, consulting. Methodologies may address any aspect of the data management program effort. Ultimately, methodologies are procedural guidance that enables quality products to be developed.



- The technical support projects are those engineered to make experts available to those performing a project.

5.3 Execution Layer

Data management program data asset projects are the specification, development, and evolution of interoperable data assets that are identified. In general, there are four classes of data asset projects: metadata, data standards, databases, and data asset evolution, maintenance, and support. In general, every data asset project includes: data asset creation, data asset evolution and maintenance, and data asset supports such as creating and/or updating enterprise level data elements, enterprise layer semantics, and the creation of enterprise level data model templates that collectively accelerate the creation of data assets, improve their quality and greatly enhance data interoperability.

5.3.1 Metadata Data Asset Projects.

Metadata data asset projects are divided into three classes: metadata contexts, data element metadata, and conceptual data model projects.

Metadata context projects are those that cause the creation of the necessary and sufficient metadata to support contextually accurate data asset projects. Included are the creation of mission, organization, function, AIS, and business event metadata, as well as required supporting metadata such as data types, DBMS, classes of data asset projects, and the like.

Data element metadata is created either top-down, bottom-up, or through time, evolved. Data element metadata consists of all the metadata components from the information systems standard 11179 for data element metadata, for example, concepts, conceptual value domains, value domains, data element concepts data element classifications, derived and compound data elements, and all associated data stewardship information. An example of a data element metadata project is the mining of the existing metadata to then represent a large collection of business fact templates.

Conceptual data models, also called standard data segments represent engineered collections of attributes within entities within narrowly constructed subjects. Conceptual data models can be used as data model templates for use within logical data model construction. Examples include the set of attributes for a person's name, for personnel skill information, for a standard requisition, disposition, or payment.



5.3.2 Data-Standard Data Asset Projects.

A data-standard data asset project is one in which the constructed data asset is primarily employed by database projects. For example, an EID project causes the creation of a data asset that creates and stores enterprise identifiers and the necessary data to fully represent the data asset that is unambiguously represented by the enterprise identifier regardless of "where" that data asset resides.

Data-standard data aAsset projects broadly include: EID projects, ADS projects, IESS projects, and/or XML projects. Once a data asset is complete it conforms to the net-centric goals, and represents an interoperable data asset. Certain data asset projects are accomplished completely through communities of interest. These data asset projects are most commonly the IESS data assets. Other data asset projects are accomplished within the program by material developers.

- Enterprise identifier (EID) Projects. Enterprise identifier projects involve the identification of all the materials that represent the definition of the problem, the requirements for an effective solution, and then the materials that define the problem's solution addressed by enterprise identifiers. These projects include materials related to various designs, alternatives, analysis of the alternatives, and the selected solution. Identified also are the classes of data structures that should and should not be supported by enterprise identifiers. Contained as well are the policies and procedures for the creation and deployment of enterprise identifier seeds and incrementers. Contained too are the detailed procedures for locating enterprise identifier sources and their effective use. Finally contained are any IT assets that assist developers in the proper use of enterprise identifiers within IT systems and database, and database management systems.
- Authoritative Data Source (ADS) Projects. ADS projects are designed to create the existence of data structures that represent data that is authoritative in nature. These data structures likely will contain an EID so that the authenticity of the data's use can be tracked. Authoritative data sources are commonly one of two types: reference data or multi-column data structures.
 - ◆ The reference-data project is generally employed to represent coded values and meanings. Reference data may, however, be more robust and thus represent entire tables or collections of tables. For each, the data structure requirements, and the requirements for value re-casting over time must be created, including establishing the precise mapping strategy from one value set to the next. All reference data level data elements must be interrelated with enterprise level data elements.
 - ◆ A multi-column data structure is one or more tables of data from a data asset that has been identified as an authoritative data source. For example, the authoritative value set of a person's name, address and telephone number or the definitive specifications of a weapon system's component. In either case, the overall process



causes the creation the mission, organization, function, data element, conceptual, logical, physical, and view models. Built also are the various information systems and event models so that a complete specification can be generated. The SQL DDL is also created so that a DBMS can create the actual database. Authoritative data source data can either be centrally stored or distributed under a very controlled update environment.

- Data Transport (e.g. XML) Projects address the creation of the necessary set of XML schemas and/or XSLT that enable XML tagged data from databases to be exchanged with other databases. The overall process creates XML objects that are squarely based on view data models generated from physical database schemas. Managing the XML Data Transport layer is the overall process of creating, evolving and maintaining a high quality and effective environment of data transport. Included in data transport are all forms which range from technology bound data exchanges through to generalized data exchanges such as those included in XML.
- IESS Projects, that is the information exchange standards specification, is a data model that represents a collection of data that is commonly shared by multiple database applications.
 - ◆ The overall process presumes the existence of mission, organization, function, data element, conceptual, logical, physical, and view models for an existing set of database applications. From the physical databases the process causes the discovery of shared data structures. From this discovery, a common schema at the logical data model level is created that then maps to the shared data of the database applications. Then information systems are specified including events set within business and calendar cycles so that users of the information exchange standards specification can know when the IESS would be updated by the "put" information systems.
 - ◆ Business rules are contained in the IESS if and only if they are common across all uses of the IESS. Business rules that are common across a subset of business information systems are specified in a SQL View on that IESS. Finally, business rules unique to a particular business information system are specified in the business information system . Regardless of the containment layer for a business rule, all business rules must be centrally defined and managed to ensure they are non-redundant and that there are zero semantic conflicts across the business rule set.

5.3.3 Database Data Asset Projects.

A database data-asset project is one that is primarily focused on the collection, access, and evolution of enterprise data, that is, data related to the enterprise's infrastructure business, and/or



external product mission areas. Of necessity, database data asset data may include values from data standard data assets. For example, a logistics database would contain a data element called Supply Condition Code. The value representing the Supply Condition code is drawn from a data-standard data asset. There are commonly the following kinds: Original Data Capture, Data Warehouses, and Subject Area Databases.

5.3.3.1 Data Warehouse Data Asset Projects.

The data warehouse data asset class is a type of database that is critical to be managed across the enterprise. For each, the data structure, and the organizational structure that is responsible for its data structure creation and maintenance, is established. Established also are the sources for all data contained in the data warehouse, all the data loading, update, and quality control mechanisms. Established too are the processes, schedules, and the required supporting information systems that retrieve data for the data warehouse and that populate it. Data warehouses are commonly not authoritative. Data warehouse data structures must be mapable to either subject area database data assets and/or original data capture data asset database data assets. Additionally, all data warehouse level data elements must be related to enterprise level data elements.

5.3.3.2 Original Data Capture Data Asset Projects.

The original data capture data asset class is a type of data asset that is the very first captured representation of data. Activities in this are must identify the various classes and types of original data capture databases that are critical to be managed across the enterprise. For each, the data structure is established, and the organizational structure that is responsible for its data structure is created and maintained. Established then are the sources for all data contained in the original data capture database, all the data loading, update, and quality control mechanisms. The processes, schedules, and the required supporting information systems that acquire and perform ongoing maintenance of original data are then established. In the event that the original data capture database is completely contained within an Enterprise Resource Planning system, it is likely that the Original Data Capture database is an ERP-independent database that functions like an IESS database. This ensures that the enterprise is independent of the ERP Vendor's proprietary database designs. Finally, all data warehouse level data elements must be related to enterprise level data elements.

5.3.3.3 Subject Area Database Data Asset Projects.

A final class of data asset is the subject area database. Activities have to identify the various classes and types of subject area databases critical to be managed across the enterprise. For each, the data structure, and the organizational structure that is responsible for its subject area database creation and maintenance is established. Established as well are the sources for all data contained



in the subject area database, all the data loading, update, and quality control mechanisms. Then, established are the processes, schedules, and the required supporting information systems that retrieve data for the subject area database and that populate it. Subject area data structures must be mappable to original data capture data asset database data assets. Additionally, all subject area database level data elements must be related to enterprise level data elements.

5.3.4 Data Asset Evolution, Maintenance and Support Projects.

Participation in data asset evolution and maintenance efforts include both the evolution and maintenance of all the metadata associated with specification of the data asset, that is, the modification of its design, implementation, etc., the management of the data asset itself, that is, the quality of its content, and finally all the associated discovery metadata associated with the data asset. Consequently, a quality engineered data asset consists of its view data models, its physical data models, its logical data models, the source data model templates, and the source enterprise level data elements. All these must be completely interrelated across all classes of data assets.

The evolution and maintenance of data asset discovery metadata includes determining the requirements for such data and then instituting the policies and procedures that ensure that all data asset metadata is created from the same overall basis in taxonomy and expertise. It is critical that all data asset metadata be discriminating among data assets, be reliably produced via independent terms mechanisms, and validly represent the data asset.

The evolution and maintenance of specific data assets include the discovery of new and/or changed requirements either on a specific data asset or across an entire class of data assets, the modification of any data structures, the identification and modification of any data acquisition and/or maintenance systems, the modification of any information system ancillary supports, and the like.

Managing data asset quality and content involves use of standards established for all data asset contained data elements, timeliness of all relevant data, and quality across all value domains. Data asset quality is focused both on the form of construction (e.g., third normal form for tables, or SQL view derived XML schemas), and the quality of the content (e.g., only two gender codes and that all stored person gender data conforms to these values).

Data assets do not exist in isolation. Supporting them includes the development of all the metadata associated with the Data Element layer via the information systems standard, 11179. Included as well would be any taxonomy that would support data asset discovery.

This data asset support effort also includes the creation of the enterprise level data element component of the data management program that in turn is required for branding all data asset level data elements with enterprise level semantics. This is essential to then define and control data interoperability.



This data asset support also includes the creation of the enterprise level data model templates of standard data segments that can be employed within and across data asset tables or other data asset data structures as a way of standardizing the specification and granularity of data within the enterprise's data assets.

