MEMORANDUM FOR SECRETARIES OF THE MILITARY DEPARTMENTS
CHAIRMAN OF THE JOINT CHIEFS OF STAFF
UNDER SECRETARIES OF DEFENSE
DIRECTOR, DEFENSE RESEARCH AND EVALUATION
ASSISTANT SECRETARIES OF DEFENSE
GENERAL COUNSEL OF THE DEPARTMENT OF DEFENSE
INSPECTOR GENERAL OF THE DEPARTMENT OF
    DEFENSE
DIRECTOR OPERATIONAL TEST AND EVALUATION
ASSISTANTS TO THE SECRETARY OF DEFENSE
DIRECTOR, ADMINISTRATION AND MANAGEMENT
DIRECTOR, FORCE TRANSFORMATION
DIRECTOR, NET ASSESSMENT
DIRECTORS OF THE DEFENSE AGENCIES
DIRECTORS OF THE DOD FIELD ACTIVITIES

SUBJECT: DoD Net-Centric Data Strategy

In the December 2001 memorandum Interoperability and Data Management, the DoD CIO requested development of an enterprise level data strategy to advance the Department toward the goal of network centric operations. A Core Data Management Working Group, with representatives from across the Department, was established to develop this strategy consistent with other net-centric initiatives, such as the Global Information Grid (GIG) Enterprise Services.

The attached DoD Net-Centric Data Strategy provides a key enabler of the Department's Transformation by establishing the foundation for managing the Department's data in a net-centric environment. The key attributes of the Strategy include:

➤ Ensuring data are visible, available, and usable when needed and where needed to accelerate decision-making

➤ "Tagging" of all data (intelligence, non-intelligence, raw, and processed) with metadata to enable discovery of data by users

➤ Posting of all data to shared spaces to provide access to all users except when limited by security, policy or regulations

➤ Advancing the Department from defining interoperability through point-to-point interfaces to enabling the "many-to-many" exchanges typical of a net-centric data environment

The strategy also introduces management of data within communities of interest (COIs) rather than standardizing data elements across the Department. COI is the inclusive term used to describe collaborative groups of users who must exchange information in pursuit of their shared goals, interests, missions, or business processes and who therefore must have shared vocabulary for the information they exchange.

As outlined in the Strategy, a data directive and additional transition planning and implementation guidance will be forthcoming as the Department begins to implement the approaches identified in the Strategy.

Your involvement and assistance are critical to the success of the Department's net-centric transformation. We must work as a team to ensure the Strategy is accomplished. My point-of-contact is Mr. Anthony Simon at 703-602-1090 or anthony.simon@osd.mil.

John P. Stenbit

Department of Defense
Net-Centric Data Strategy

May 9, 2003

Prepared by:
Department of Defense
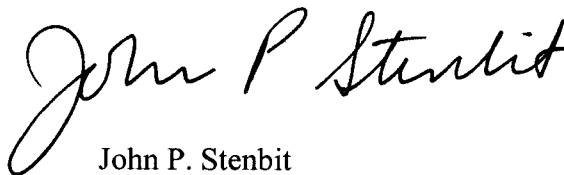**Chief Information Officer (CIO)**

FOREWORD

Across the Department of Defense, broad leadership goals are transforming the way information is managed to accelerate decision-making, improve joint warfighting, and create intelligence advantages. In support of these goals, the mission of the Department's Chief Information Officer (CIO) is to lead the Information Age transformation by building the foundation for net-centric operations through policies, program oversight, resource allocation, and value-added support.

The Department is taking an integrated approach to delivering a foundation for net-centricity. This approach incorporates network and communications enhancements to provide sufficient bandwidth for low-latency support, fusion tools to empower users and applications to pull multiple sets of data to create a current "picture," and information assurance and data strategies to enable trusted data for all users and applications.

This DoD Net-Centric Data Strategy outlines the vision for managing data in this net-centric environment. Net-centricity compels a shift to a "many-to-many" exchange of data, enabling many users and applications to leverage the same data—extending beyond the previous focus on standardized, predefined, point-to-point interfaces. Hence, the net-centric data objectives are to ensure that all data are visible, available, and usable—when needed and where needed—to accelerate decision cycles. In a net-centric environment, unanticipated but authorized users or applications can find and use data more quickly. One of the CIO's goals is to populate the network with all data (intelligence, nonintelligence, raw, and processed) and to change the paradigm to "post before processing"—allowing authorized users and applications access to data without wait time for processing, exploitation, and dissemination. Users and applications will post all data to "shared" spaces, increasing the amount of Enterprise and community data while minimizing private user or application data. All posted data will have associated metadata (i.e., data about data) to enable users and applications to discover, and evaluate the utility of, shared data.

The goals of net-centricity—empowering users through access to data and faster availability of data as a result of posting before processing—drive this Data Strategy. This Strategy builds on related net-centric efforts involving bandwidth enhancements and the development of nterprise services and capabilities to exploit data.

John P. Stenbit
CIO, Department of Defense

TABLE OF CONTENTS

List of Tables

List of Figures

1.   PURPOSE

This document describes the Net-Centric Data Strategy for the Department of Defense (DoD), including DoD intelligence agencies and functions. It describes a vision for a net-centric environment and the data goals for achieving that vision. It defines approaches and actions that DoD personnel will have to take as users—whether in a role as consumers and producers of data or as system and application developers. This Strategy will be followed by a subsequent directive and guidance on implementation details.

2.   INTRODUCTION

Net-centricity is the realization of a networked environment, including infrastructure, systems, processes, and people, that enables a completely different approach to warfighting and business operations. The foundation for net-centricity is the Department's Global Information Grid (GIG). The GIG is the globally interconnected, end-to-end set of information capabilities, associated processes, and personnel for collecting, processing, storing, disseminating, and managing information on demand to warfighters, defense policymakers, and support personnel. Net-centricity, by securely interconnecting people and systems independent of time or location, supports a substantially improved military situational awareness, better access to business information, and dramatically shortened decision cycles. Users are empowered to better protect assets; more effectively exploit information; more efficiently use resources; and create extended, collaborative communities to focus on the mission.

The approach to implementing the GIG uses communications, computing, and applications technologies but also recognizes that the cultural barriers against trust and data sharing must be addressed. To this end, the Department is using a comprehensive, integrated approach to deliver the foundation for net-centricity. This approach combines the overall Net-Centric Data Strategy, described in this document, and an information assurance (IA) strategy with the implementation of the layers of the GIG as indicated in Figure 1.

Figure 1. Integrated Approach for Delivering a Net-Centric Environment
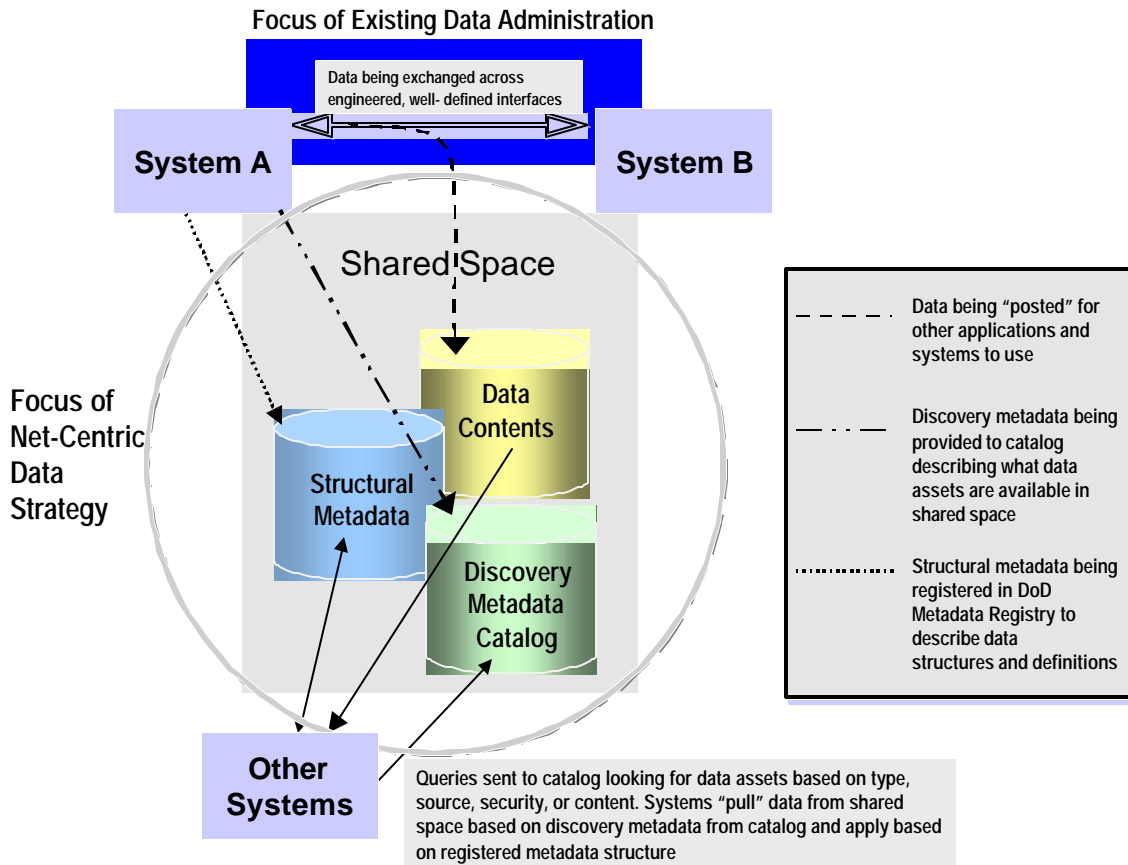
The traditional DoD approach to data has been data administration. That approach attempted to standardize and control data elements, definitions, and structures across the Enterprise, requiring consensus among and across organizations. Data administration was intended to promote interoperability through standardization of data elements, minimize duplication of data elements across the Department, and reduce the need for data element translation. However, the traditional approach, focused on standardizing data elements, has proved to be too cumbersome to implement across an enterprise of the scope of the Department.

This Net-Centric Data Strategy defines a modified paradigm for data management within the Department. This Strategy expands the focus to visibility and accessibility of data rather than just standardization. It also recognizes the need for data to be usable for unanticipated users and applications, as well as for those that have been predefined. This Strategy identifies approaches that will improve flexibility in data exchange, supporting interoperability between systems without requiring predefined, pair-wise interfaces between them. This flexibility will be essential in the "many-to-many" exchanges of a net-centric environment. While tightly engineered, predefined interfaces between systems will continue to exist (e.g., sensor-to-shooter systems), the objective in a net-centric environment is to increase the potential for many other systems to leverage the same data without having to anticipate this use in the development cycle. For example, tightly engineered and real-time systems can offer "exposure" services that work "behind the scenes" collecting real-time data, storing it, and providing access and discovery through an enterprise interface. Exposure services can be designed to have little or no effect on performance critical processes or predefined interfaces and still provide access to their data to unanticipated users. In an environment in which systems are continually being developed, deployed, migrated, and replaced, making allowances for unanticipated interfaces is essential. The Net-Centric Data Strategy continues to recognize the value of element standardization between tightly engineered, predefined systems but shifts the emphasis for standardization to subsets of the Department as needed. Figure 2 illustrates the expanded focus of the Net-Centric Data Strategy. The following section describes the vision and the concepts represented as the focus of the Net-Centric Data Strategy.

Figure 2. Scope of the Net-Centric Data Strategy

Focus of Existing Data Administration

Data being exchanged across engineered, well-defined interfaces

System A

System B

Shared Space

Data Contents

Focus of Net-Centric Data Strategy

Structural Metadata

Discovery Metadata Catalog

Data being "posted" for other applications and systems to use

Discovery metadata being provided to catalog describing what data assets are available in shared space

Structural metadata being registered in DoD Metadata Registry to describe data structures and definitions

Other Systems

Queries sent to catalog looking for data assets based on type, source, security, or content. Systems "pull" data from shared space based on discovery metadata from catalog and apply based on registered metadata structure

2.1   DOD DATA VISION

The core of the net-centric environment is the data that enables effective decisions. In this context, 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 CIO goals, as confirmed by the Deputy Secretary of Defense in Management Initiative Decision 905, is to populate the network with all data (intelligence, nonintelligence, raw, and processed) and change the paradigm from "process, exploit, and disseminate" to "post before processing." 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. Users and applications "tag" data assets with metadata, or data about data, to enable discovery of data. Users and applications post all data assets to "shared" space for use by the Enterprise. Figure 3 illustrates the shift from private data to community or Enterprise data as a result of increased data "sharing" in the net-centric environment. Tagging, posting, and sharing of data are encouraged through the use of incentives and metrics.

Figure 3. Increasing Enterprise and Community Data in a Net-Centric DoD
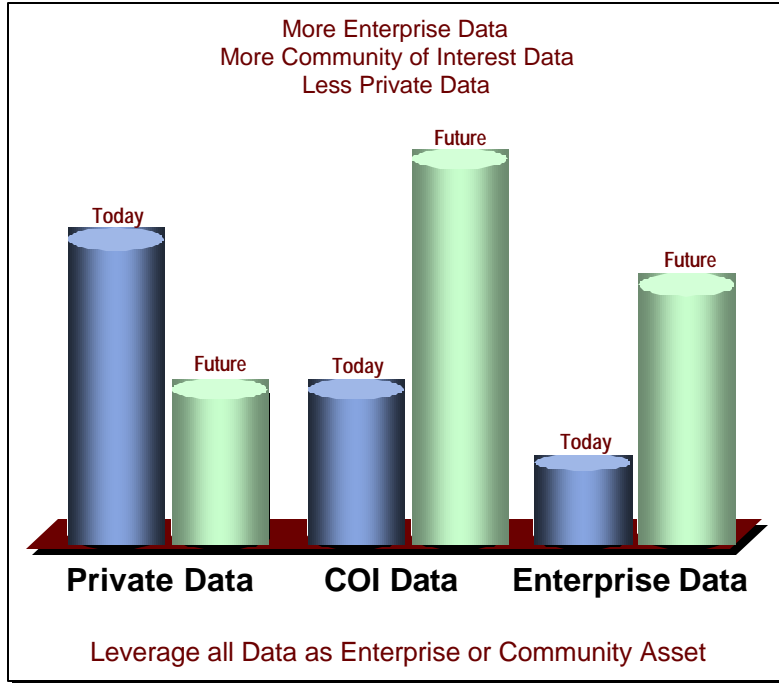
This data vision is predicated on several key elements:

(1) Communities of Interest to address organization and maintenance of data

(2) Metadata, which provides a way to describe data assets and the use of registries, catalogs, and shared spaces, which are mechanisms to store data and information about data

(3) GIG Enterprise Services that enable data tagging, sharing, searching, and retrieving.

These elements, combined with the bandwidth enhancements and fusion capabilities being developed as part of the GIG, are critical to realizing a net-centric environment.

2.1.1   Communities of Interest (COIs)

COIs is the inclusive term used to describe collaborative groups of users who must exchange information in pursuit of their shared goals, interests, missions, or business processes and who therefore must have shared vocabulary for the information they exchange. Communities provide an organization and maintenance construct for data such that data goals are realized. Moving these responsibilities to a COI level reduces the coordination effort as compared to managing every data element Department-wide. For example, standardization and control of data elements, similar to the current data administration approach, can be done at the community level rather than requiring all data elements to be standardized across the Enterprise.

Communities 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. A community may have authority from explicit chartering (e.g., the Deputy Secretary of Defense tasking to

address a specific challenge) or implied authority as a result of existing command or organizational structures (e.g., a brigade commander leading a Joint Task Force threat assessment community). COIs in the Department will span a range of characteristics as illustrated in Figure 4.

Figure 4. COI Characteristics

	Functional	Cross-Functional
Expedient	Tactically driven, Implied authority, Formal processes modified for need, Relatively many entities (e.g., New Imagery Analysis capability for Damage Assessment)	Tactically driven, Derived authority, Ad hoc processes, Many entities (e.g., Forward deployed JTF planning New Threat Response)
Institutional	Explicitly recognized, Longer term, More formalized processes based on span of control, Relatively few entities (e.g., PSAs such as Logistics)	Explicitly or implicitly recognized, Longer term but priority driven, Blended processes resulting from agreements (e.g., JS area such as Battlespace Awareness)

Institutional COIs, whether functional or cross-functional, tend to be continuing entities with responsibilities for ongoing operations. They also lend support to contingency and crisis operations. Expedient COIs are more transitory and ad hoc, focusing on contingency and crisis operations.

COIs support users across the Enterprise by promoting data posting, establishing "shared" space, and creating metadata catalogs. Data within a COI can be "exposed" within the COI or across the Enterprise by having users and applications "advertise" their data assets by cataloging the associated metadata. These catalogs, which describe the data assets that are available, are made visible and accessible for users and applications to search and pull data as needed.

Although many of the COI functions will be similar regardless of COI characteristics, there will be some additional roles for institutional COIs. Institutional community members will collaborate to ensure that the necessary structures are in place to achieve the data goals. In particular, during the transition to net-centricity, institutional community members must take the lead in establishing COI-specific metadata structures, defining community ontologies, cataloging data and metadata, and having members post data. The COI-specific metadata structures provide an extended level of data definitions and structures, and the community ontology provides the data categorization, thesaurus, key words, and/or taxonomy. The COI-specific metadata structures and the community ontology serve to increase semantic understanding and

interoperability of the community data. These community ontologies and data structures are visible to the Enterprise—by increasing visibility, data "stovepipes" will be mitigated.

The institutional COI efforts may enable the expedient COIs to quickly become operational when needed. The users in an expedient COI not only pull and use data but also create and post data to the Enterprise. A member of an expedient COI may leverage the data structures defined by the institutional COIs. For example, when providing metadata for a new data posting, the member can provide the metadata already defined in one of the institutional COIs' schemas. However, expedient COIs may also create their own metadata structures, ontologies, and catalogs.

Based on the diversity of COI characteristics and roles, there will be a variety of operating processes and procedures that will be used by COIs to accomplish their data activities. Pilot activities with "trial COIs" will further refine the construct. More detail on COI functions will be provided in subsequent transition planning guidance.

2.1.2 Metadata

Metadata can be employed a variety of ways to enhance the value and usability of data assets. The traditional DoD 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. GIG Enterprise Services (GES) capabilities use metadata, in its various forms, to support data asset discovery and interoperability and to provide a richer semantic understanding of all data and metadata. Figure 5 shows an example of how some of these types of metadata are used.

Figure 5. Example Uses of Metadata
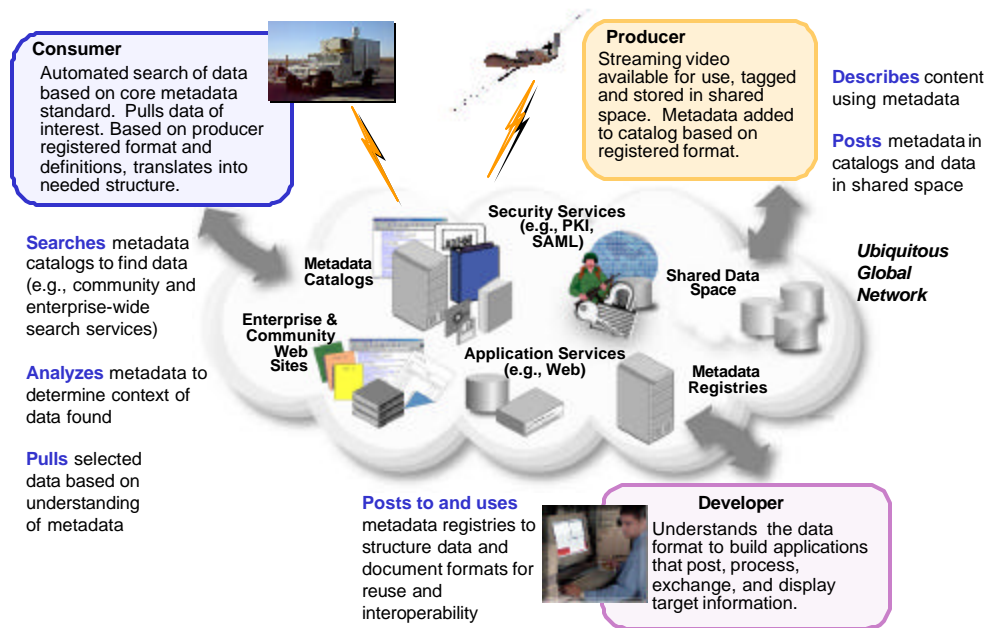
Various mechanisms are used to store and process the different types of metadata and data. Metadata registries, metadata catalogs, and shared spaces are three mechanisms used to store data and information about data to enable discovery, support interoperability, and enhance data asset understanding. These terms will be used throughout the approaches in Section 3. Hence, it is important to understand the use of each mechanism and the distinctions among them. Although some or all of these mechanisms will be provided as part of GIG Enterprise Services (see 2.1.3), they are defined here because of their importance to the Strategy.

A metadata registry is a system that contains information that describes the structure, format, and definitions of data. Typically, a registry is a software application that uses a database to store and search data, document formats, definitions of data, and relationships among data. System developers and applications are the predominant users of a metadata registry. Defense Information Systems Agency (DISA) has established a DoD Metadata Registry in accordance with industry standards (see 2.1.3)

For example, libraries may use "cards" in a card catalog to describe information about each holding in the library. Metadata registries contain information that describes what information is required to be filled out on each card. Metadata registries do not contain the actual filled-out cards; rather, they simply store the format of the card (e.g., what information needs to be on the card and the format and definition of each field).

A metadata catalog is a system that contains the instances of metadata associated with individual data assets. Typically, a metadata catalog is a software application that uses a database to store and search records (or cards) that describe such items as documents, images, and videos. Search portals and applications would use metadata catalogs to locate the data assets that are relevant to their query.

For example, following the prior library analogy, a metadata catalog contains the actual filled-out cards that describe each of the holdings (i.e., the card catalog). In effect, the holding is "advertised" (i.e., made discoverable) by the existence of the card. Unlike the metadata registry, a catalog does not store information regarding the format of each card; rather, it contains the actual cards.

A shared space is a mechanism that provides storage of and access to data for users within a bounded network space. Enterprise-shared space refers to a store of data that is accessible by all users within or across security domains on the GIG. A shared space provides virtual or physical access to any number of data assets (e.g., catalogs, web sites, registries, document storage, and databases). Any user, system, or application that posts data uses shared space.

For example, continuing the analogy, the bookshelves in a library, or the library itself, are a shared space. A virtual library may be manifested as a repository that contains copies of, or links to, the actual holdings in the library. Registry content and catalog content are held in a shared space.

2.1.3   GIG Enterprise Services (GES)

GES enables the data goals by providing basic computing capabilities to the Enterprise. For example, GES must provide reliable identification and authorization services to assure the security of the data. In addition, users and applications exploit easy-to-use search tools and software agents that allow them to search metadata catalogs and "pull" data from across the various communities and the Enterprise. The pulled data may come from a variety of sources such as databases, files, electronic records, web pages, documents, and system services.

Another example of a GES capability is the DoD Metadata Registry. The DoD Metadata Registry, based on the International Organization for Standardization (ISO) 11179 specification for metadata registries, is available throughout the Enterprise. The Registry represents a "one-stop shop for developer data needs" and is a key component in achieving the Department's interoperability goals. All document formats, interface definitions, and exchange models used by systems will be stored in the DoD Metadata Registry. Developers can discover these metadata assets and utilize them to read, write, or exchange data that is made available throughout the Enterprise. All programs and COIs have a responsibility to support interoperability through active participation in the DoD Metadata Registry. The DoD Metadata Registry will provide capabilities to further support interoperability through the use of translation and mediation services and for the sharing and reuse of processes. For example, a COI may develop and share a process for calculating target coordinates for a specific weapon system. This process will be available to all users on the Enterprise, and its associated metadata (input/output format and connection information) will be registered in the DoD Metadata Registry. Through this capability, the DoD Metadata Registry is more than just a simple repository of data formats—it is a comprehensive source for supporting design, development, and execution of processes (e.g., business logic) in a net-centric, services-based data environment.

The DoD Metadata Registry currently incorporates a variety of existing metadata resources such as the DoD XML Registry, the Defense Data Dictionary System (DDDS), and commonly used data reference sets. Planned content enhancements will integrate other resources such as

messaging formats, symbology, ontologies, and transformation services. The expected contents of the DoD Metadata Registry are shown in Figure 6. Additional functionality will be added to the Registry, as required, to support implementation of the DoD Net-Centric Data Strategy. One such addition will provide the DoD Metadata Registry with functionality to support a "federated registry" concept. Federation allows multiple metadata registries to be integrated and synchronized into the virtual, central DoD Metadata Registry, thereby providing a single source for the discovery of all Department metadata.

Figure 6. Contents of the DoD Metadata Registry
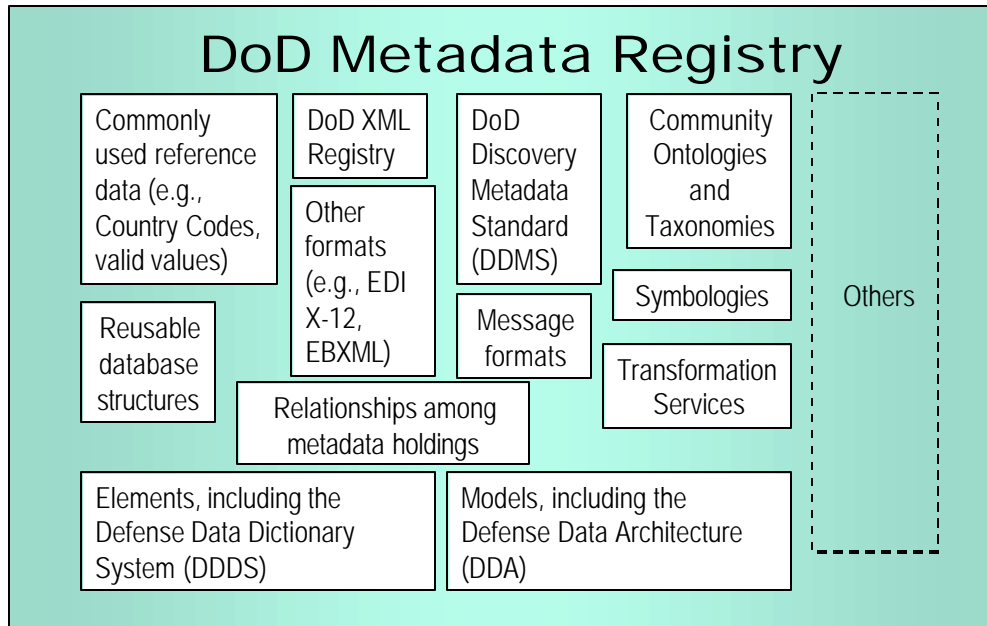
Mediation is a key GES capability in the net-centric environment that relies on availability of metadata. Mediation resolves differences in the name, structure, and representation of data. A range of mediation approaches is planned, including the following:

- Registration of translations and transformations in the Metadata Registry for use by developers and applications
- Using commercial mediators to provide transformation services
- Specialized mediation services offered by COIs
- Registration and publication of common schemas and other exchange models.

Systems should offer services that allow users and applications to further exploit data assets. For example, a system may provide a service that allows a user to query a relational database for specific content rather than requiring the user to understand how to develop an application that can search the database. In effect, the system provides an access service that "exposes" the information within the database. Community catalogs also contain "service metadata" that defines the capabilities of the service, the necessary inputs to use the service, and a description of what the service provides. By evaluating the service metadata, users can assess whether the service meets their information needs.

2.2   NET-CENTRIC DATA GOALS

This Strategy lays the foundation for realizing the benefits of net-centricity by identifying data goals and approaches for achieving those goals. To realize the vision for net-centric data, two primary objectives must be emphasized: (1) increasing the data that is available to communities or the Enterprise and (2) ensuring that data is usable by both anticipated and unanticipated users and applications. Table 1 describes the data goals in the context of these two objectives. These goals and the approaches discussed in Section 3 pertain to all legacy and new data assets, such as system files, databases, documents, official electronic records, images, audio files, web sites, and data access services, in the Department, including DoD intelligence agencies and functions.

Table 1. Data Goals

Goal	Description
Goals to increase Enterprise and community data over private user and system data	
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.
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.
Institutionalize	Data approaches are incorporated into Department processes and practices. The benefits of Enterprise and community data are recognized throughout the Department.
Goals to increase use of Enterprise and community data	
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.
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.
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.
Responsive to User Needs	Perspectives of users, whether data consumers or data producers, are incorporated into data approaches via continual feedback to ensure satisfaction.

Two additional data properties are frequently considered: data quality and data accuracy. Data quality and accuracy will be improved as a consequence of the above data goals; making data more visible and usable across the Enterprise creates an incentive to produce quality and accurate data. Additional steps for improving data quality and accuracy in a particular system, application, or business process will be necessary but are not a part of this Data Strategy.

3.   APPROACHES TO ACHIEVE DOD DATA GOALS

This section identifies the approaches to achieve each of the data goals.

3.1   GOAL: MAKE DATA 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.

The following approaches achieve this goal:

3.1.1   Post Data to Shared Spaces

Users and applications will migrate from maintaining private data (e.g., data kept within system-specific storage) to making data available in community- and Enterprise-shared spaces (e.g., servers and services available on the Internet). These shared spaces will act as repositories where users and applications can submit, or post, data assets to the enterprise. The shared spaces will provide storage and serving mechanisms. Enterprise-shared spaces will be maintained, secured, and staged as necessary to support the Department's missions. Data that is posted to shared spaces will be advertised via the associated metadata and will be discoverable with enterprise search tools.

3.1.2   Associate Discovery Metadata With Data Assets

To facilitate discovery of data assets, users and applications will provide discovery metadata, in accordance with the DoD Discovery Metadata Standard (DDMS), for all data posted to shared spaces. The DDMS will provide a common set of structured attributes that support discovery of data assets using search tools. COIs and asset producers determine the desired level of discovery for a data asset, e.g., discovery of a database or a record within a database, discovery of a document or a paragraph within a document. The initial focus of the DDMS is to aid in the discovery of data assets as a whole; hence, the discovery metadata in the DDMS will not always be required for individual records or elements. For example, the discovery metadata will always indicate the existence of a database containing certain kinds of information but may or may not identify the contents of specific database elements. The DDMS does not preclude the use of other metadata processes or standards. For example, record-level database tagging and in-line document tagging are common practice today to support various Department objectives. These tagging initiatives will only have to enhance their existing processes to include the DDMS for Enterprise discovery.

The DDMS will be adopted and implemented across the DoD components for new and existing data assets. Figure 7 illustrates the logical layers and elements of the DDMS. This standard will be registered in the DoD Metadata Registry. Enterprise visibility of a data asset is promoted when its discovery metadata complies with the DDMS.

Figure 7. DDMS
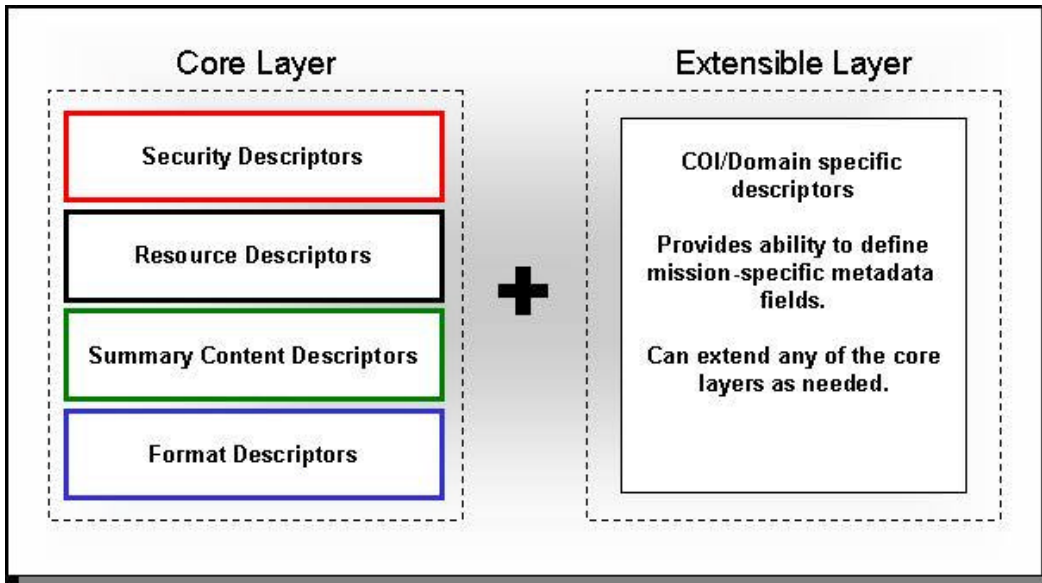
The core layer of the DDMS represents those attributes of a data asset that can be commonly described across the Department, regardless of the type of data asset or community to which it is applied. It is important to note that not all elements in the core layer are mandatory. The set of mandatory elements will be small yet robust enough to generate high value-added metadata to support Enterprise discovery. The extensible layer provides a mechanism for COIs to extend the core layer of the DDMS to support mission-specific metadata requirements.

3.1.3   Create and Maintain Catalogs

Metadata catalogs will advertise the existence of shared data and will contain information about all data assets contained in the associated shared space (including databases, system output files, web pages, documents, and access services). At a minimum, the mandatory discovery metadata elements in the DDMS must be represented within metadata catalogs for any data asset posted to a shared space. COIs will establish and maintain catalogs. Each catalog may be organized according to the community-defined ontology. An Enterprise catalog will be established that links to community catalogs, effectively creating a "catalog of catalogs." The Enterprise catalog will also contain metadata for data posted by users and applications without a specific COI affiliation.

Catalogs will be searchable by applications or through user-friendly, web-based interfaces. The web-based interfaces will have a consistent look and feel and will support posting of metadata to the catalog and data to the shared space. The catalogs will also be searchable, either manually or automatically via agents, through application programming interfaces. All metadata catalogs will adhere to Enterprise discovery interface standards to allow searches within a catalog or across catalogs.

3.1.4   Register Metadata Related to Structure and Definition

The DoD Metadata Registry will contain all metadata related to data structures, models, dictionaries, and schemas. The purpose of the Registry is to give developers and architects visibility into methods to compose and encode data and to share usage across the Department. Registration of such metadata is critical to achieve the data goals of interoperability and understanding by promoting semantic and structural understanding.

3.1.5   Inventory Data Assets

During transition to the new DoD Net-Centric Data Strategy, COIs may identify and prioritize key data assets and services within their domain. These data assets or services may already exist or they may be in development. In either case, COIs will identify the data assets and services that must be made compliant with the data approaches. This prioritized list of data assets and services will provide a focus for near-term COI initiatives to create metadata to advertise the data and ensure that the data is available in shared space.

3.2   GOAL: MAKE DATA 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.

The following approaches achieve this goal:

3.2.1   Create Shared Spaces and Data Access Services

Shared spaces—virtual and actual—will be created to provide a "store and serve" mechanism for data assets. In addition, data access services will be created to help facilitate access to database stores, business logic processes, and system data. Data access services are any mechanisms that help expose data that is not otherwise available to users and applications. For example, a data access service may be a registered, accessible software interface that allows users and applications to extract information from an inventory database.

3.2.2   Associate Security-Related Metadata

Security-related metadata will be provided for each data asset as defined by the Security Descriptors element set within the core layer of the DDMS (see Figure 7). Systems will be able to control access to assets based on classification metadata. GES will allow data that was typically inaccessible as a result of implementation barriers (e.g., distinct networks based on data classification and prearranged accounts for access) to be available to users and applications that have appropriate access needs. These services will permit access using security metadata, technologies such as public key infrastructure (PKI), and role- and permission-based access processes if adopted.

3.3   GOAL: INSTITUTIONALIZE DATA MANAGEMENT

Data approaches are incorporated into Department processes and practices. The benefits of Enterprise and community data are recognized throughout the Department.

The following approaches achieve this goal:

3.3.1   Govern Data Processes With Sustained Leadership

Best-practice studies have shown that cultural change is most effectively achieved when there is sustained leadership commitment and institutionalization of new processes. The DoD CIO will establish a DoD-wide net-centric governance process to promote and sustain successful data management practices across COIs by reviewing and sharing metrics, best practices, and incentive structures. This DoD-wide net-centric governance process will also provide oversight of net-centric infrastructure development efforts.

3.3.2   Incorporate Data Approaches Into Department Processes and Practices

DoD Components will recognize and fund implementation of data approaches (e.g., providing metadata, defining and registering metadata structures, and posting data). Planning, programming, and budgeting system processes and policies, as well as acquisition processes and policies, will be updated to reflect these approaches.

3.3.3   Advocate, Train, and Educate in Data Practices

Best practices show that new operating practices are assimilated more quickly when consistently promoted. The DoD CIO will continue to conduct the Data Strategy awareness campaign across the Department, promoting and educating all users about their responsibilities and the benefits of participation. The necessary tools to implement these approaches (e.g., DoD Metadata Registry, catalogs, and shared spaces) will be identified and implemented. As these tools become available, training and manuals will be provided.

3.3.4   Adopt Metrics and Incentives

The DoD-wide net-centric governance process will identify incentives and metrics. Users and system developers will be guided by incentives to encourage and foster participation. Incentives are intended to prompt users and developers to contribute to increased data visibility across the Enterprise. Incentives may include rapid or continued funding for initiatives that implement these data approaches (e.g., providing metadata in accordance with the DDMS). Incentives may also include Enterprise-visible "top products" lists that cite specific sources that have been especially valuable to users.

Metrics will be collected to track implementation and application of the approaches. Metrics will be helpful in evaluating usage to ensure participation across the Department. Metrics also serve as a means to evaluate the effectiveness of the overall Data Strategy. Measurement techniques will be developed to ensure that metrics are captured in a useful and consistent manner. Moreover, multiple publication methods such as metrics web sites and Enterprise- and

COI-specific metrics reports will be made available Department-wide to promote awareness of data management successes and areas requiring improvement.

3.4   GOAL: ENABLE DATA TO BE 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.

The following approaches achieve this goal:

3.4.1   Define COI-Specific Ontologies

COIs will be encouraged through metrics and incentives to develop an ontology that best reflects the community understanding of their shared data. Ontologies include data categorization schemes, thesauruses, vocabularies, key word lists, and taxonomies. Ontologies promote semantic and syntactic understanding of data. For example, taxonomies enhance discovery by providing a hierarchical means of searching for data while providing users and applications with additional insights about data assets by indicating their placement among other data assets. COI-developed vocabularies will define terms used in describing data assets, and the thesauruses will identify related terms to assist translation services. Any community that establishes an ontology will publish it to the DoD Metadata Registry to increase understanding across the Enterprise and promote possible reuse.

3.4.2   Associate Content-Related Metadata

The summary content descriptors element set of the DDMS (see Figure 7) is specifically aimed at providing "content-related" details about data assets. Content metadata provides topics, keywords, context, and other content-related information. Content metadata gives users and applications insight into the meaning and context of the data. Content metadata provides a basis for search engines to perform searches for data assets that address specific topics.

3.4.3   Associate Format-Related Metadata

The format descriptors element set of the DDMS (see Figure 7) is used to describe details pertaining to the format of the associated data asset. The format descriptors are useful when trying to understand the physical manifestation of an asset. For example, the format descriptors will provide information regarding the type of digital file (e.g., a Joint Photographic Experts Group [JPEG] image, or an Audio Interchange File Format [AIFF]). In addition, the format descriptors contain optional information that describes the extent of the asset, such as file size, bit rate, and dimensions. Format-related metadata allows users and applications to narrow down information searches and to select products that meet their particular operating constraints (e.g., a user who is able to view only Graphic Interchange Format [GIF] images would not want to pull a JPEG image).

3.4.4   Define COI-Specific Content-Related Metadata

To improve understanding, an extension of the discovery metadata standard is reserved for domain-specific, or COI-specific, metadata. This is represented as the extensible layer of the

DDMS in Figure 7. With this extension layer, COIs will be able to provide context relevant to their particular domain area and still be able to participate in Enterprise-wide search and discovery. COIs will be required to register their COI-specific content metadata requirements in the DoD Metadata Registry. These COI-specific metadata requirements may then be integrated into appropriate Enterprise and community services such as search and mediation.

3.5   GOAL: ENABLE DATA TO BE 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.

The following approaches achieve this goal:

3.5.1   Associate Data Pedigree and Security Metadata

The Resource Descriptors elements of the DDMS (see Figure 7) allow identification of the author, publisher, and sources contributing to the data, allowing users and applications to assess the derivation of the data (i.e., data pedigree). This metadata allows users and applications to select data from known sources. Reliable and quality sources will become more widely used, enhancing overall data quality throughout the Enterprise as more data sources become visible.

The Security Descriptors elements of the DDMS (see Figure 7) allow security and privacy markings consistent with established standards where applicable. For information assurance (IA) and security, GES will provide auditing tools that can track access, by individual user, of each data asset. GES may also provide access control to data assets based on security markings in the metadata.

3.5.2   Identify Authoritative Sources

COIs may identify authoritative sources for key data assets in their domain. The community will publicize their identified authoritative sources to the Enterprise, thus allowing users and applications to evaluate and understand the community-implied authority of data sources. COIs may have to resolve potentially conflicting sources and, where appropriate, coordinate with the DoD-wide governance body to identify authoritative source(s).

3.6   GOAL: SUPPORT DATA INTEROPERABILITY

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.

The following approaches achieve this goal:

3.6.1   Register Metadata

Registration of metadata (e.g., eXtensible Markup Language [XML] components, database segments, and data dictionary elements) is an important activity to support interoperability in a net-centric environment. COIs will register their metadata components in the DoD Metadata

Registry. Registering metadata components to the DoD Metadata Registry supports many-to-many interoperability by providing system architects and developers with insight into existing data schemas that they can employ and extend.

3.6.2   Associate Format-Related Metadata

Users and systems can employ the elements of format descriptors to specify the extent (e.g., size and dimension), type, and physical manifestation of assets. The format descriptors element set of the DDMS (see Figure 7) supports interoperability by allowing systems and users to determine the physical manifestation of data assets, which in turn helps to identify which tools and capabilities are required to use the asset.

3.6.3   Identify Key Interfaces Between Systems

Interface engineering in an environment of many-to-many exchanges requires an unrealistic degree of interface control and an enormous commitment of resources. To facilitate interoperability within a community, COIs can determine the appropriate focus and level of data standardization within their community. This decentralized, distributed approach to interoperability ensures that key interfaces and data structures are controlled when tightly engineered interfaces are required. COIs will register metadata that results from interoperability activities in the DoD Metadata Registry. Mediation services will use the registered metadata to facilitate system interoperability between unanticipated interfaces as needed.

3.6.4   Comply With Net-Centric Interface Standards

Developers will be responsible for adhering to published net-centric interoperability standards, including data standards where applicable. Successful discovery and interoperability of data assets depend on compliance with metadata standards (i.e., DDMS) and data exposure standards (e.g., GES discovery interface standards). For example, data assets that are maintained by the Department's Records Management functions must provide a means for the Enterprise discovery capability to query the inventory of their stored records. In doing so, these records management applications should employ the DDMS to respond to Enterprise discovery queries.

3.7   GOAL: BE RESPONSIVE TO USER NEEDS

Perspectives of users, whether data consumers or data producers, are incorporated into data approaches via continual feedback to ensure satisfaction.

The following approaches achieve this goal:

3.7.1   Involve Users in COIs

As described in Section 2.1, institutional COIs in particular are focused on ensuring the implementation of these approaches. Institutional COIs will take the lead in creating catalogs, defining ontologies, and developing COI-specific metadata. To adequately reflect user needs, these COIs must engage a range of known users and developers in these activities.

3.7.2 Establish a Process To Enable User Feedback

COIs, under the DoD CIO Enterprise-wide net-centric governance process, will establish processes to evaluate and refine the user experience. Users may provide ratings for data sources, catalogs, or services, and content metadata usability. Ratings may include factors such as ease of use, applicability, or quality. These ratings will be published Enterprise-wide and used to promote participation in posting, identifying, and sharing data assets. Overall, this Department-wide feedback and ratings process, coupled with improved data asset visibility, will increase the integrity and quality of data. In addition, the feedback process allows COIs and data producers to identify previously unanticipated users and applications.

To improve Enterprise data visibility, the process may allow users to identify needed data by publishing a "data want ad" to a community or Enterprise collaboration space. In some cases, the data may be available but not currently visible or accessible. Hence, the source may choose to make it visible or accessible to the user or application. In other situations, providing the data may not be available or cost-effective, and the user's "want" will remain unfilled.

4.   NEXT STEPS, CHALLENGES, AND CONCLUSION

This document identified the key goals and approaches for net‑centric data across the Department.
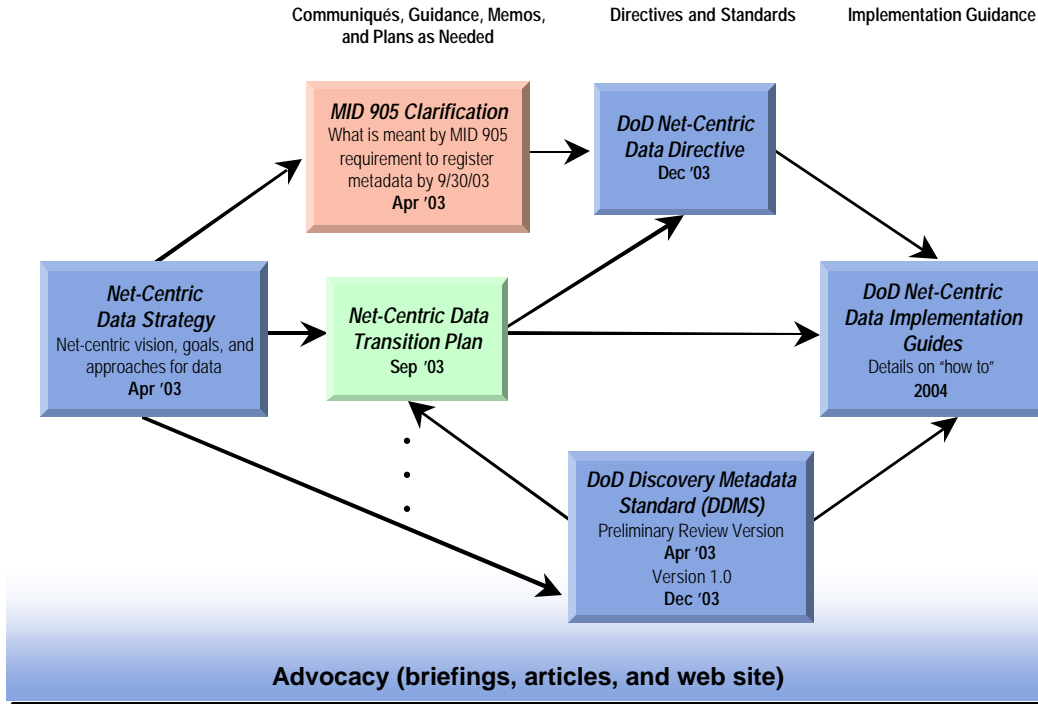
4.1   NEXT STEPS

To achieve these goals, the DoD CIO will evolve the Net-Centric Data Strategy. In particular, the DoD CIO will provide the following:

- Communiqués and guidance memorandums to clarify and provide additional details on Departmental directions concerning data approaches.

- A draft data directive for the Deputy Secretary of Defense's signature. This directive will codify execution of this Strategy.

- The DDMS that identifies the elements of the Enterprise discovery metadata represented in Figure 7. The DDMS, currently in a preliminary review version, will be put under the configuration control of the GES Metadata Working Group (GES MWG).

- A *Net-Centric Data Transition Plan* that describes the action plans, management constructs, and sequencing of activities required for implementing the concepts in the Data Strategy. This plan will be coordinated through the GES steering committee and GES executive board.

- Implementation guides that address specific areas of the Strategy such as the functionality of the federated DoD Metadata Registry, governance processes for use by COIs, identification and application of metrics and incentives, and transition of legacy systems to accommodate the data approaches.

- Continuation of the ongoing awareness campaign to promote the data goals for the net‑centric environment.

Figure 8 illustrates the evolution of the Net-Centric Data Strategy.

Figure 8. Evolution of the Net-Centric Data Strategy

Communiqués, Guidance, Memos, and Plans as Needed | Directives and Standards | Implementation Guidance

MID 905 Clarification
What is meant by MID 905 requirement to register metadata by 9/30/03
Apr '03

DoD Net-Centric Data Directive
Dec '03

Net-Centric Data Strategy
Net-centric vision, goals, and approaches for data
Apr '03

Net-Centric Data Transition Plan
Sep '03

DoD Net-Centric Data Implementation Guides
Details on "how to"
2004

DoD Discovery Metadata Standard (DDMS)
Preliminary Review Version
Apr '03
Version 1.0
Dec '03

Advocacy (briefings, articles, and web site)

In addition, the data approaches discussed in this Strategy are being piloted through a number of related activities. These pilot activities, such as Horizontal Fusion Enterprise Services (HFES) and Global Combat Support System (GCSS) Logistics test, will be demonstrating discovery, mediation, and posting capabilities. The results of these pilots will be used to refine the data approaches and the DDMS, as well as assist development of planning and implementation guidance. Figure 9 provides a roadmap for refining and implementing the data approaches.

Figure 9. Data Roadmap

	FY03				FY04				FY05			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Formalize Data Strategy		Strategy Released ▲		Transition Plan ▲	▲	Directive		▲	Implementation Guidance			
Discovery Metadata		DDMS Review Version ▲		Establish GES MWG ▲	DDMS V1.0 ▲							
Data Pilots - **HFES, GCSS Logistics, others**												
GES Coordination - **Discovery** - **Includes DoD Metadata Registry** - **Mediation** - **Shared Storage**			Pilot COIs ▲		Capability definition, interface standards, architecture development							
Net-Centric Advocacy Efforts		Articles, briefings, website										

4.2   DATA CHALLENGES

A number of challenges must be mitigated to enable the Department to achieve the goals described in Section 3. These challenges and mitigation measures were considered in developing the approaches for achieving the data management goals and will be reflected in subsequent implementation guidance. Table 2 identifies the primary challenges, along with mitigation measures.

Table 2. Data Management Challenges and Mitigation Measures

Challenge	Mitigation
1) Promote culture change to encourage data sharing	• Use awareness campaigns and incentives • Ensure organizational and leadership awareness of net-centricity, data posting, and sharing • Commit to data posting and sharing by incorporating approaches in processes, procedures
2) Financially support the implementation of data approaches	• Ensure that data responsibilities are recognized and suitably resourced so that approaches are adopted • Incorporate appropriate approaches in planning, programming, and budgeting system (PPBS) processes and acquisition processes and policies
3) Promote the use of metadata	• Adopt the DoD Discovery Metadata Standard for all data in the Enterprise • Use automatic generation of as much metadata as possible • Minimize the mandatory components in the discovery metadata standard

Challenge	Mitigation
4) Implement security structures that enable data visibility and accessibility for appropriate but unanticipated users and applications	• Coordinate with IA activities • Ensure inclusion of security classification in the discovery metadata standard • Suitably resource and develop cross-domain solutions that use metadata labeling to enforce access at the appropriate classification level • Provide a strong identity management infrastructure (such as DoD PKI) to enforce security, roles and accesses, author identification, and audit trails
5) Ensure that infrastructure development and operations achieve data goals	• Coordinate with GIG infrastructure development efforts • Provide guidance and user's and developer's manuals
6) Transition legacy systems to implement data approaches	• Use COIs to prioritize ordering of system and data asset transition • Implement guidance that accommodates incremental opportunities to accomplish data goals

7) Implement a data governance structure that—
  • Streamlines data processes
  • Maximizes the synchronization with existing related governance structures
	• Adopt a COI approach that supports coordination among users with a common mission, business processes, or interest but mitigates the need for Department-wide coordination on every data asset
	• Eliminate redundant data governance structures and cancel former data policies and processes
	• Coordinate and synchronize with related GIG governance structures

4.3   CONCLUSION

Implementing the approaches outlined in the *Net-Centric Data Strategy* will require the active participation of all DoD Components in collaboration with the DoD CIO. There is much more work required to develop a net-centric security approach, to operationalize the COI and GES governance process, and to deliver the necessary Enterprise Services that make the goals of this Strategy achievable. Realizing a net-centric environment involve the commitment of the Department, particularly as Components migrate legacy systems to this net-centric approach. In executing this Strategy, the DoD CIO will continuously seek to refine the approaches and maintain communications to ensure that the Department can realize the benefits of net-centricity.

APPENDIX A. TERMINOLOGY

Terms used in this Strategy are further explained for reference.

- *Communities of Interest (COIs)* is the inclusive term used to describe collaborative groups of users who must exchange information in pursuit of their shared goals, interests, missions, or business processes and who therefore must have shared vocabulary for the information they exchange.

- *Data asset* refers to any entity that is composed of data. For example, a database is a data asset that comprises data records. In this document, "data asset" means system or application output files, databases, documents, or web pages. "Data asset" also includes services that may be provided to access data from an application. For example, a service that returns individual records from a database would be a data asset. Similarly, a web site that returns data in response to specific queries (e.g., weather.com) would be a data asset.

- *Extensible Markup Language (XML)* is a tagging language used to describe and annotate data so it can be consumed by human and system interactions. XML is typically arranged hierarchically using XML elements and attributes. It also uses semantically rich labels to describe elements and attributes to enable meaningful comprehension. An example of XML data describing an element named "Person" appears as follows:

```
<Person>
  <FirstName>John</FirstName>
  <MiddleInitial>H</MiddleInitial>
  <LastName>Doe</LastName>
</Person>
```

- *Global Information Grid (GIG)* is the globally interconnected, end-to-end set of information capabilities, associated processes, and personnel for collecting, processing, storing, disseminating, and managing information on demand to warfighters, defense policymakers, and support personnel. (See the complete definition of GIG in DoDD 8100.1, September 19, 2002.)

- *Metadata* is descriptive information about the meaning of other data. Metadata can be provided in many forms, including XML.

- *Metadata Catalog* is a system that contains the instances of metadata associated with individual data assets. Typically, a metadata catalog is a software application that uses a database to store and search records that describe such items as documents, images, and videos. Search portals and applications can use metadata catalogs to locate the data assets that are relevant to their queries.

- *Metadata Registry* is a system that contains information that describes the structure, format, and definitions of data. Typically, a registry is a software application that uses a database to store and search data, document formats, definitions of data, and relationships among data. System developers and applications are the predominant users of a metadata registry.

- *Net-centricity* is the realization of a networked environment (including infrastructure, systems, processes, and people) that enables a completely different approach to warfighting and business operations.

- *Ontology* includes data categorization schemes, thesauruses, vocabularies, key-word lists, and taxonomies. Ontologies promote semantic and syntactic understanding of data.

- *Schema* is a diagrammatic representation, an outline, or a model. In relation to data management, a schema can represent any generic model or structure that deals with the organization, format, structure, or relationship of data. Some examples of schemas are (1) a database table and relationship structure, (2) a document type definition (DTD), (3) a data structure used to pass information between systems, and (4) an XML schema document (XSD) that represents a data structure and related information encoded as XML. Schemas typically do not contain information specific to a particular instance of data.

- *Shared space* is a mechanism that provides storage of and access to data for users within a bounded network space. Enterprise-shared space refers to a store of data that is accessible by all users within or across security domains on the GIG. A shared space provides virtual or physical access to any number of data assets (e.g., catalogs, web sites, registries, document storage, and databases). As described in this Strategy, any user, system, or application that posts data uses shared space.

- *Web services* are self-describing, self-contained, modular units of software application logic that provide defined business functionality. Web services are consumable software services that typically include some combination of business logic and data. Web services can be aggregated to establish a larger workflow or business transaction. Inherently, the architectural components of web services support messaging, service descriptions, registries, and loosely coupled interoperability.