LOAN DOCUMENT

PHOTOGRAPH THIS SHEET

AD-A265 285

DTIC ACCESSION NUMBER

LEVEL

1

INVENTORY

Architecture Master Plan Study Data Dictionary
Concept Paper Draft Version 1.2 Vol. 29

DOCUMENT IDENTIFICATION

Oct 89

DISTRIBUTION STATEMENT A
Approved for public release
Distribution Unlimited

DISTRIBUTION STATEMENT

HANDLE WITH CARE

ACCESSION FOR
NTIS GRA&I ☒
DTIC TRAC ☐
UNANNOUNCED ☐
JUSTIFICATION

BY Per ltr
DISTRIBUTION/
AVAILABILITY CODES

DISTRIBUTION	AVAILABILITY AND/OR SPECIAL	
A-1		

DISTRIBUTION STAMP

DTIC
ELECTE
MAY 28 1993
S C D

DATE ACCESSIONED

DATE RETURNED

93

DATE RECEIVED IN DTIC

P5P 93-12064

REGISTERED OR CERTIFIED NUMBER

PHOTOGRAPH THIS SHEET AND RETURN TO DTIC-FDAC

DTIC FORM 70A JUN 90

DOCUMENT PROCESSING SHEET

PREVIOUS EDITIONS MAY BE USED UNTIL STOCK IS EXHAUSTED

LOAN DOCUMENT

UNITED STATES
DEPARTMENT OF DEFENSE
Computer-aided Acquisition &
Logistic Support (CALS)

October 1989

OSD CALS
Architecture Master Plan Study

DATA DICTIONARY

CONCEPT PAPER

DRAFT VERSION 1.2

Prepared by

U.S. Department of Transportation
Research and Special Programs Administration
Transportation Systems Center
Cambridge, MA 02142

EXECUTIVE SUMMARY

Rapid advances in information technology are changing the way technical data is created, stored, and used. These advances have created opportunities to reduce costs and improve productivity in both the administration of data and in the acquisition and operation of weapon systems. To realize these benefits, effectively administered data dictionary systems are required.

Data dictionary systems are databases that store and manage data related to the structure and use of information resources. These databases will facilitate the access and exchange of data within the DoD, within industry, and between the DoD and industry during the acquisition and operation of the weapon system. The access, exchange, and integration of data that the data dictionary supports will revolutionize weapon system engineering, manufacturing, and support.

To realize the full potential of data dictionaries, three requirements need to be satisfied:

- Establish an organization to administer and manage technical data and data dictionaries.
- Implement and administer data and procedural standards that will govern weapon system life cycle data and activities; and
- Establish a data management process that will provide needed services to technical data users.

The OSD must establish, charter, and support an organization empowered to administer technical data, data standards, and data dictionaries. This organization must be empowered to resolve conflicts over data standards, definitions, ownership, and other characteristics within DoD and between DoD and its contractors. Only through an effective data administration organization can CALS and data dictionary systems realize expected dividends.

Developing, adopting, and implementing standards is also critical for effective data dictionary implementation. Specifically, there is a critical need to standardize data and procedures, both within the Services and among contractors. Although standards that deal with how computer systems capture, store, and present information are important, establishing and enforcing standards that govern weapon system data is the most significant standards activity that OSD can undertake.

A data management process will be necessary to achieve a flow of consistent information given the diverse needs of the DoD community. In a truly integrated, heterogeneous environment, this process will provide the services needed to create, acquire, manage, distribute, and use technical information resources.

mb

CONTENTS

SECTION 1: INTRODUCTION

The purpose of this paper is to establish a common perspective within the OSD-CALS Directorate and Planning Groups regarding data dictionaries. The data dictionary is a technological tool that can be used to manage information within CALS. This concept paper focuses on:

- The "use" and "implications" of data dictionaries in managing information that supports today's weapon system life cycle (WSLC) activities; and
- Ways that data dictionaries will facilitate the CALS evolution from todays slow, inefficient, paper-intensive operations to more efficient electronic operations incorporating integrated data, unified WSLC functions, and heterogeneous computing environments.

Traditional discussions of data dictionaries have been from the perspective of technology and the data-processing organization. Indeed, many technical issues are associated with the use of data dictionaries, but management issues that pertain to the effective coordination of CALS-related information are of equal or greater importance. This paper examines data dictionaries and the associated issues of data administration and standards from a management perspective.

In addition to this introduction, this paper includes three additional sections:

- **Section 2, Basic Definitions and Concepts** - provides an overview of the major features and applications of data dictionary systems.
- **Section 3, CALS and Data Dictionaries** - discusses the evolution of the CALS environment and the role that data dictionaries will need to fill in this future.
- **Section 4, Recommendations** - presents recommended strategies to facilitate data dictionary applications and enhance data management within CALS.

SECTION 2: BASIC DEFINITION AND CONCEPTS

2.1 What Is A Data Dictionary?

A data dictionary is a tool that processes information about the structure and use of data. Data dictionary capabilities include mechanisms to:

- Identify, define, document, and control data;
- Enter and maintain information about data into a special-purpose database (i.e., a data dictionary); and
- Obtain reports concerning the descriptive information that is maintained.

The descriptive information about data (referred to as "metadata") that is maintained generally includes information regarding the data's scope, meaning, format, relationships, physical location, and application usage.

2.2 Why Data Dictionaries?

Interest and awareness regarding data dictionaries has grown along with the recognition that an organization's data can be a valuable and important resource. The philosophy of information resource management (IRM) is increasingly being subscribed to in both the private and public sectors. The IRM philosophy maintains that an organization's data resources should be managed with as much care and attention as other corporate resources.

Data dictionaries are tools needed to implement the IRM philosophy. They read and interpet descriptive information about organizational data. This descriptive information is needed to define what data is available, where it is stored, when it was (will be) updated, and how to access it. Effective implementation and administration of a data dictionary can:

- Prevent data redundancies, inconsistencies, and variant naming conventions;
- Help standardize data use and administration;
- Support prompt access to data regardless of its source; and
- Reduce costs and lead times for information system development and modification.

In addition to its direct impact on system development, the data dictionary tool can be used to enhance the overall system planning activity, to ensure that changes are being carried out through a systematic and orderly progression of tasks, and to assist in the integration of systems and applications.

Finally, data dictionaries act as directories, pointing to the appropriate location for data requests. A data dictionary defines where the data is physically located and how the data is to be accessed. A dictionary's directory function supports communication and coordination among automated data systems, defines mappings between different data meanings or user views, and facilitates database queries by users.

2.3 Who Uses A Data Dictionary?

There are two classes of dictionary users:

- People, and
- Systems.

Although both people and systems have many similar requirements (e.g., the meaning of data), the information needed to satisfy those requirements is often quite different. For example, a person defines a data element through a string of text, whereas a system defines a data element by its relationship to other data elements, and, in some instances, its method of computation.

The nature of data dictionary developments has been influenced by both classes of users and will continue to be influenced by these user needs. Since system technology is evolving at a faster rate than people, system technology changes will be the driving force for changes to data dictionary design and features in the future.

There is a broad range of user types among people. There are developers, administrators, and system users. Although system users are the most numerous, developers and administrators typically demand a broader range of services from a data dictionary than do system users. Typically, data dictionary services to developers lower costs, whereas data dictionary services to administrators increase the value of data by enhancing its integrity.

2.4 Data Dictionary Design Features

Many commercially available data dictionaries possess a variety of unique features and capabilities. Following is a brief discussion of three of the more significant data dictionary design features.

2.4.1 Integrated versus Freestanding Data Dictionaries

One of the basic distinctions among data dictionaries pertains to their relationship to database management systems (DBMS). Some data dictionaries are built within, or "integrated" with a specific DBMS. An integrated data dictionary depends on an existing database management system. It is normally developed and sold by a vendor to work with his system (although it may work with others to a lesser degree) and almost always relies on the

vendor's DBMS as the tool for its implementation. An integrated data dictionary can be particularly effective in enforcing data standards, and can typically leverage the vendor's underlying query processor, security, and backup/recovery features.

In contrast to integrated dictionaries, some data dictionaries have been designed as independent or "freestanding" products intended to provide dictionary services to a variety of DBMSs. Freestanding dictionaries are much more versatile and generally better suited for supporting a multivendor distributed environment that is typical in most commercial and government organizations today. Many freestanding dictionaries provide enhanced services, such as data element level security, and operate in coordination with integrated dictionaries provided by the DBMS vendor.

2.4.2 Passive versus. Active Data Dictionaries

Another distinguishing feature of data dictionaries is the manner in which they work with DBMSs and applications. Some dictionaries simply act as repositories for data and are "passive" in their involvement with the DBMS and applications. Passive data dictionaries do not provide metadata to system processes or components. Instead, passive dictionaries are used primarily to satisfy human requirements pertaining to documentation, security, etc. Freestanding dictionaries are nearly always passive in nature.

"Active" data dictionaries are involved with the DBMS or application and provide the necessary metadata for one or more system applications. Active dictionary systems are a powerful tool for initiating management controls and for ensuring compliance with defined data standards. They work with both human users and the "system" user to enhance development productivity, data integrity, and the coordination among application systems. Active dictionary systems often result in "user-friendly" systems but also involve performance trade-offs since all processes must work through the data dictionary facilities.

2.4.3 Data Dictionary Extensibility

A third significant design feature is the capability to add new data types. Some data dictionaries can "extend" the structure of data within a DBMS, expanding the scope of data; defining new, complex data types; establishing intricate or innovative data relationships; and creating new data operations/functions.

Traditionally, DBMSs and data dictionaries have handled basic data types, such as alphanumeric fixed fields, floating-point numerics, etc. These classic data types were designed to store and retrieve data for business and scientific applications. Extensible data dictionaries allow users to modify predefined information types and define entirely new types of information, such as screen, drawing, or icon. This capability, when applied to processes, offers a wealth of opportunities for new applications such as engineering design or image processing.

2.5 Data Dictionary Applications

While the basic data dictionary operations of defining, identifying, controlling, and documenting data remain valuable, the expansion of data dictionary operations to multiple applications substantially increases the dictionary's contribution within the organization. As illustrated in Figure 2–1, data dictionaries historically supported the operation of a specific DBMS implementation with a limited set of applications. Dictionaries performed their basic role of maintaining "data about data" for a reasonably limited information universe.

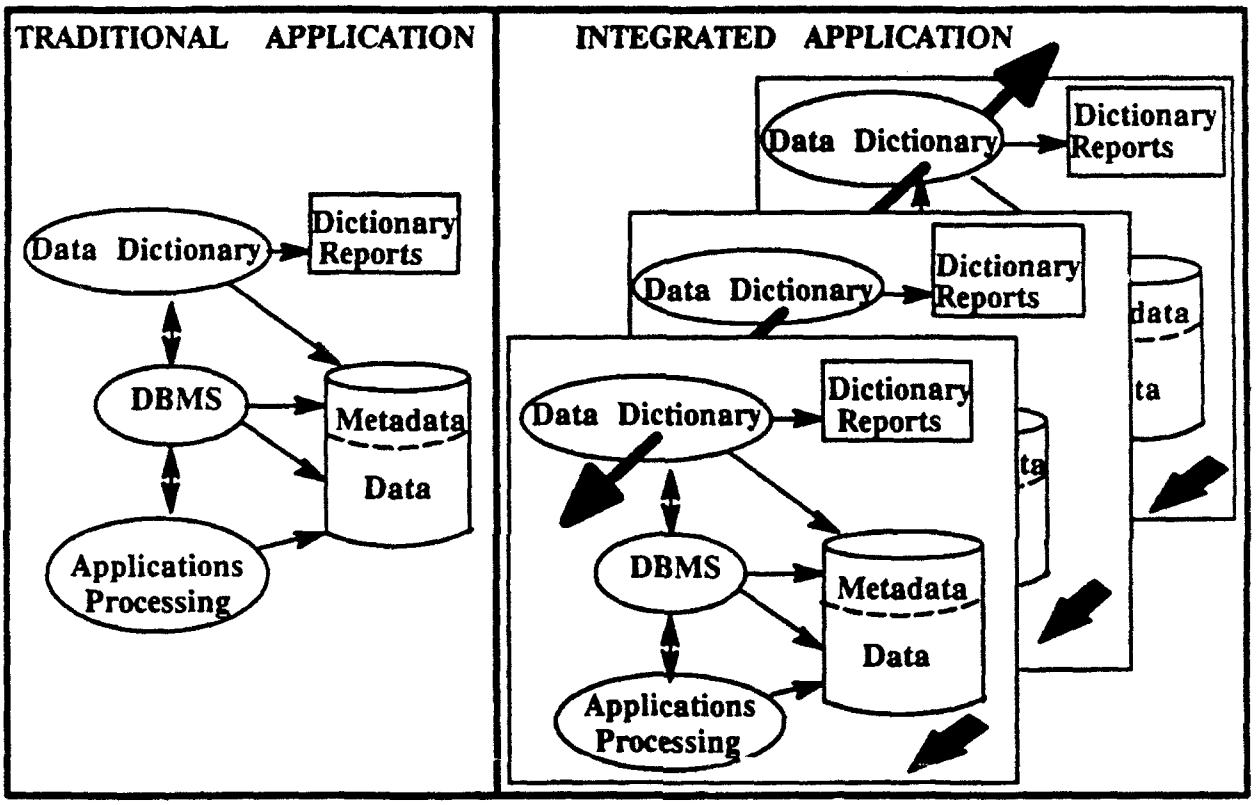

Figure 2–1 Alternative Data Dictionary Operating Environments

This traditional role has been generally sufficient to effectively manage the operational performance of the DBMS. However, this traditional implementation does not fully support major administrative goals of sharing data across applications, reducing the overall level of data redundancy, and providing a more integrated environment for data manipulation.

In recent years, data dictionaries have been used to capture and manage data about all of the information resources within an organization. Data dictionaries serve as tools for coordinating the transfer and use of data in multiple processing environments that incorporate hardware and software assets from many different vendors. Thus, data dictionaries are one key enabler of the open systems architecture concept.

Given this expanded scope, the term data *"dictionary"* is being superceded by *"encyclopedia."* In addition to the traditional types of metadata information, virtually any type of information

entity is included within an organization's data encyclopedia. Examples of these enhanced types of information include:

- Process Data – including applications, programs, operating procedures, menus and screens;
- Information Resource Data – including the relevant hardware, software, and configuration information; and
- Integrity Information – including business rules governing data relationships and equivalent values, and information pertaining to information currency and update.

In addition to spanning multiple operating environments, the data dictionary tool is becoming a focal point for coordinating information management activities through the entire information system's life cycle (see Figure 2–2). Within the planning phase, the data dictionary can support data and activity modeling. It can also provide the central framework analysts use to build a conceptual model with consistent entities and terminology.

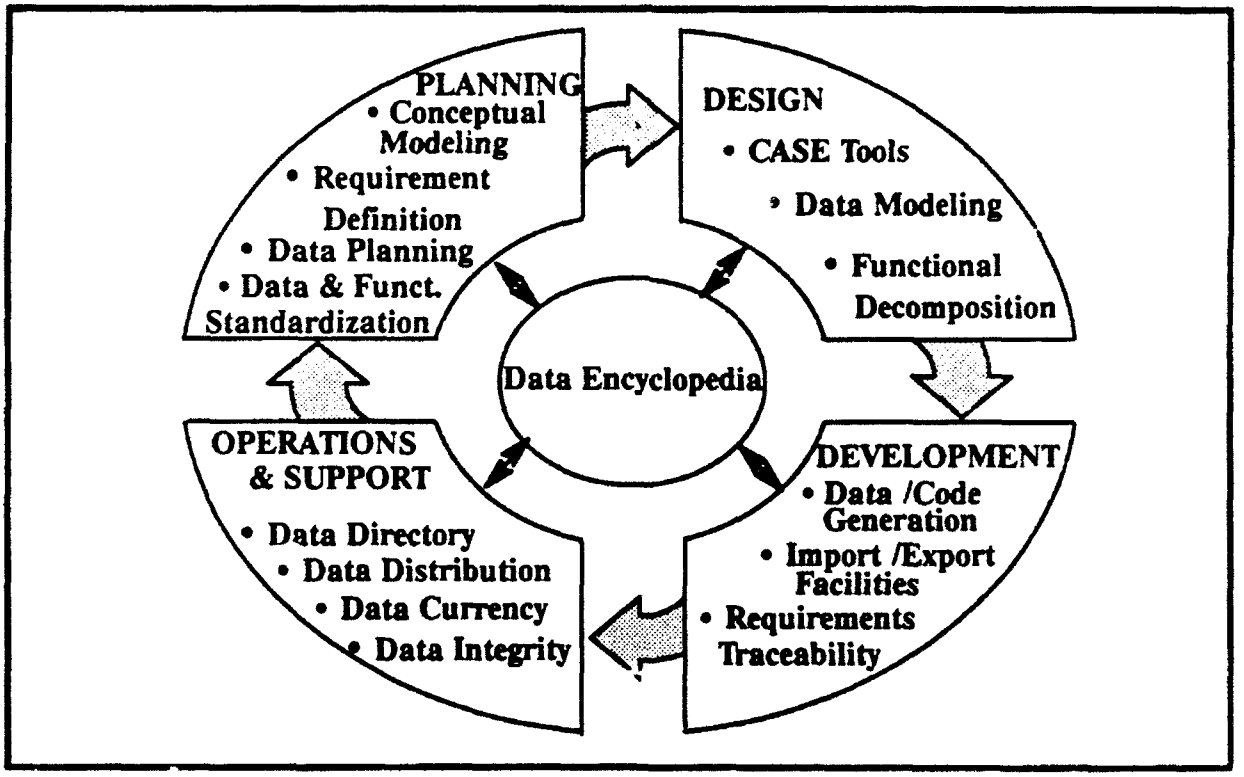

Figure 2–2 Data Encyclopedia – Life-Cycle Development Focal Point

In the system design and development phases, the data dictionary can serve as the primary reference point for defining system components and specifications, and act as the central data source supporting the use of a variety of computer-aided system engineering (CASE) technologies. In the operations phase, the data dictionary can serve as the heart of the user-friendly front-end interface for handling data access requests. It can be used to

coordinate with program processing activities to administer security and data integrity controls, and to perform various other administrative tasks.

In addition to providing a common source of information for all life-cycle phases, the data dictionary can provide the structure to catalog the data repository information according to the life-cycle phase to which it pertains. It can also serve to manage the control guidelines and ensure data integrity when working with different life-cycle phases.

Traditionally, organizations have often not taken full advantage of data dictionary capabilities because of the difficulties inherent in working with different data dictionaries across software systems, or from one phase of the system's life cycle to another. Along with these difficulties, a significant effort must often be undertaken in reentering and reformatting metadata information so that it can be used with different systems and software development tools. The establishment of an effective data administration organization and the implementation and adherence to a robust set of standards will typically alleviate these difficulties.

SECTION 3: CALS AND DATA DICTIONARIES

3.1 CALS Evolution

CALS initiatives will accelerate the evolution from manual, paper intensive activities to operations that are integrated and highly automated operations. Although the CALS environment will be changing continuously, it is convenient for discussion to focus on the four points pictured in Figure 3-1.

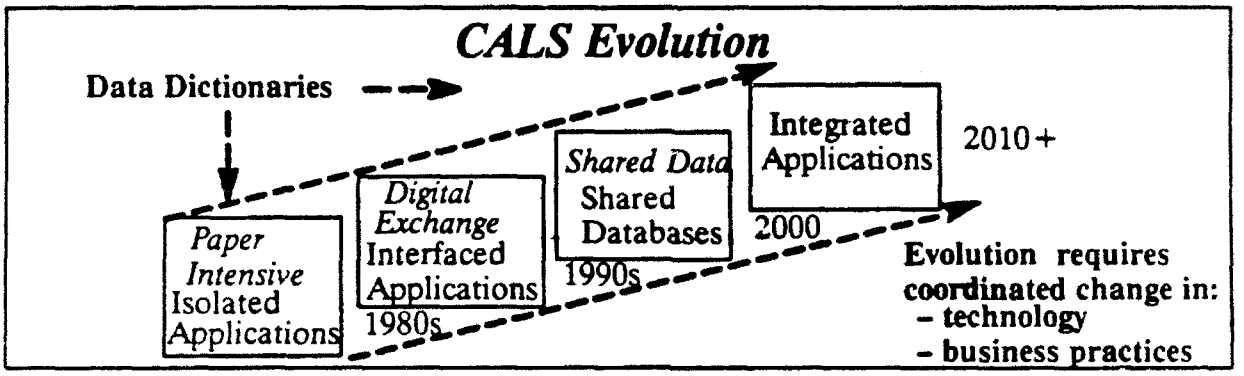

Figure 3-1 CALS Evolutionary Stages

3.2 Data Dictionaries within a Paper-Intensive Environment

The Armed Services have all invested large quantities of time and money to develop automated information systems. Still, much of the WSLC environment within DoD can be characterized as paper intensive. Many information systems within DoD are nonintegrated, stand alone batch applications with programs requiring extensive programming work for even modest changes. Typically, applications have been developed incrementally and without coordination. Applications are based on a wide variety of DBMSs, and have inconsistent data definitions, naming conventions, and indexing schemes. Both batch applications and data inconsistencies are barriers to the flow of information among applications and between contractors and the Services.

Presently, the Services rely on both manual and automated data dictionaries. Manual data dictionaries are paper-based repositories, recording data element meanings; the automated data dictionaries are most often simple data element listings, capturing descriptive information (metadata) about the meaning and values of data elements.

Some manual dictionaries act as functional standards while others support specific automated systems, capturing record layouts and formats. Although the dictionaries are usually coupled with specific applications, consistency among systems is limited. Inconsistencies include:

- Multiple definitions for the same data element;

- Multiple data computation methods for the same data element;
- Multiple physical representations for similar types of data; and
- Variant naming conventions for the same data element.

Additionally, data inconsistencies arise from procedures that are often ill suited to data management in automated operations. Update procedures are often cumbersome, and procedures to get information out to appropriate users are often ineffective.

3.3 Data Dictionaries within a Digital Exchange Environment

The current emphasis of CALS is on accelerating the digital exchange of data between contractors and the Services, and thus breaking down the barriers to digital information exchange (see Figure 3-2). Currently digital exchange of information is almost exclusively a one-way exchange, starting with the contractor and flowing to the Services.

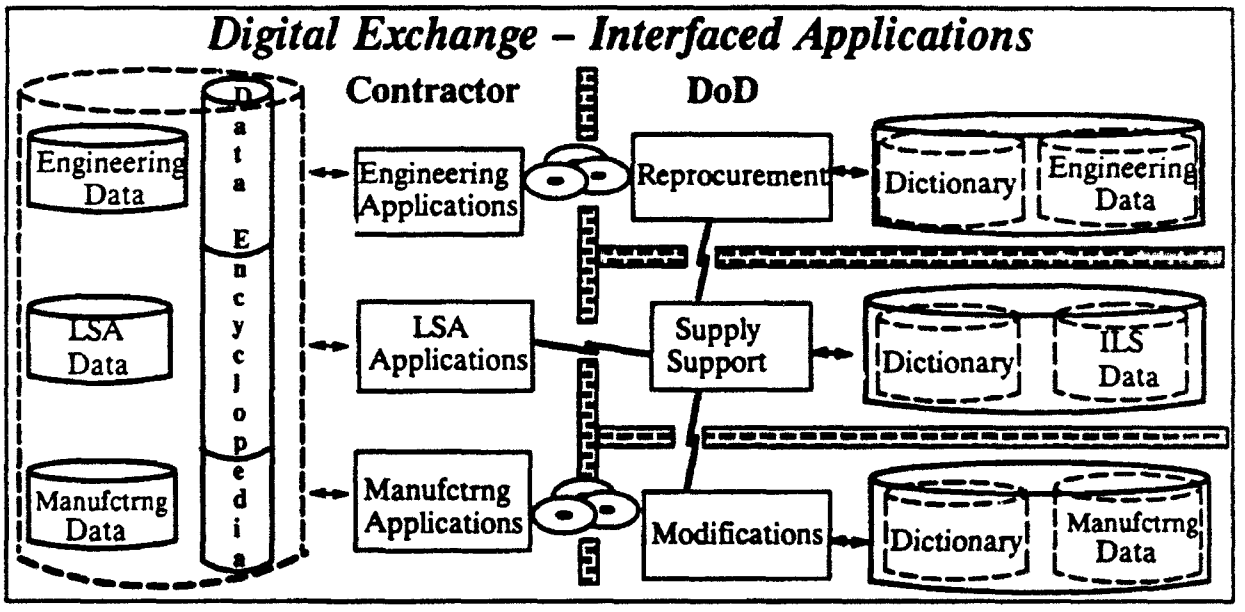

Figure 3-2 CALS Phase I Environment

The digital exchange of information is accomplished in many ways. In some instances, media - most often magnetic tape - is physically exchanged. In other cases, information is exchanged through remote access using telecommunications services.

Data dictionaries increasingly will play crucial roles in aiding the exchange of digital data by providing the means to read and interpret the information. For example, the exchange of media between contractor and DoD would be meaningless if the Services could not read and interpret the information. Some form of data dictionary, defining the data's format and meaning, is needed to process the data and to formulate ad hoc queries.

In this environment, data dictionaries are organization-specific, filling crucial roles for both the contractor and the DoD. Although many contractors have implemented corporate wide

data dictionaries, effectively establishing a single logical database for their organization. these dictionaries typically do not apply outside their organizations. In some instances, however, contractor teaming arrangements have forced the creation of common databases for the exchange of information on an application-by-application basis.

Data exchanges within the Services involve both digital exchanges and paper transfers. Limited interfaces among applications have, in some instances, been incorporated into the design of information systems. However, the variety of data inconsistencies previously described means that many interfaces still require paper exchanges or human intervention.

3.4 Data Dictionaries within a Shared Data Environment

Many transitional phases exist between digital data interfacing and total integration. One phase is the shared-data environment (see Figure 3-3). The integrated shared technical data, which is the central feature of this phase, includes both weapon system-specific information and information on common systems and parts (commodity information). Critical applications data are integrated with product, logistics support, and commodity data sets.

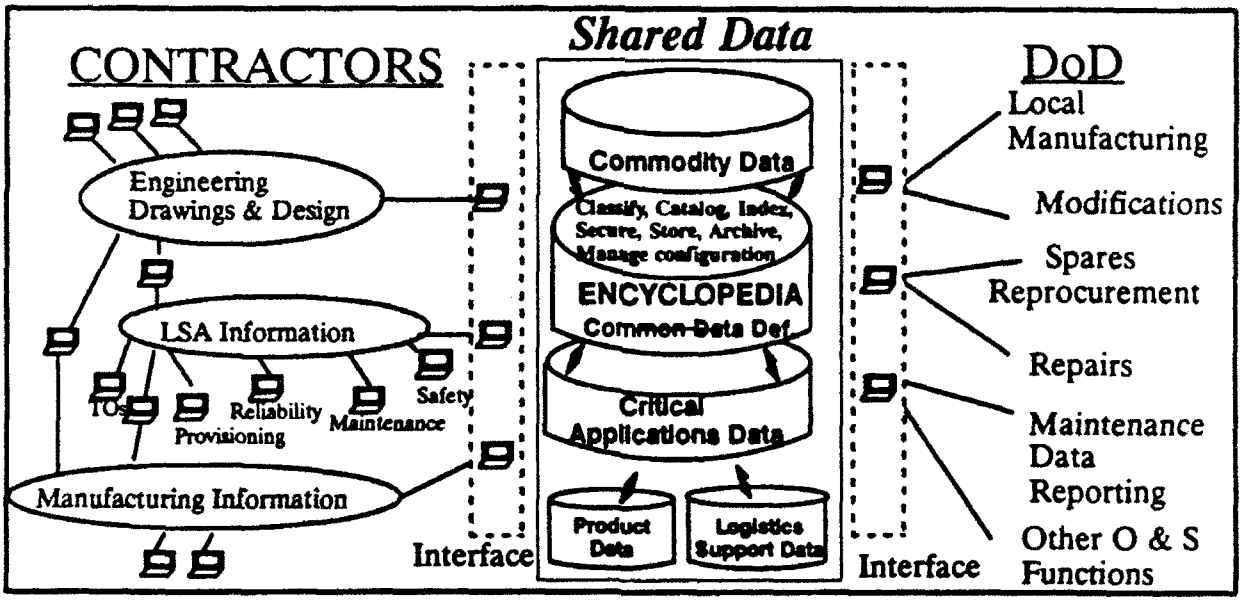

Figure 3-3 CALS Phase II Environment

The data sets will include a variety of data types, ranging from classic alphanumeric data to complex engineering constructs, video images, and audio information. Much of the data will be defined as data objects, involving expanded functionality within the basic definition of the data (e.g., relationships will be part of a data object definition). These data sets will be initially developed during system acquisition and will be organized and segregated on a weapon system/platform basis. As weapon systems/platforms become operational, information will be transferred to logically shared databases. Access will be available on geographically dispersed, heterogeneous hardware and software.

The data dictionary will need to be active, propagating standards and procedures to DoD and contractor users. Furthermore, data dictionaries will be enforcing agents, assuring that nonstandard data, relationships, tools, and other constructs are excluded from the database.

Dictionaries will be based on specifications that will support various implementations. These specifications will act like today's freestanding dictionaries, working with numerous DBMS products and hardware configurations; serving both contractors and the Services.

At this stage in CALS's evolutionary path, WSLC activities will remain essentially unchanged. The most significant change will involve the establishment of robust standards that will govern the administration of data. These standards will act as enablers in sharing data among organizations. Without these enablers, access to data will be limited by organizational boundaries, various technology configurations, and by information system boundaries. The importance of standards cannot be overemphasized. Technology advances will pay few dividends for CALS without standards and their effective administration.

Administrative requirements can be grouped into three areas. The first, data management, includes directory capabilities, glossary capabilities, and configuration management. The second, integration management, includes interface control and query control. Security management, the third area includes data classification and data access.

3.5 Data Dictionaries within an Integrated Environment

An extension of the shared data environment involves the integration of individual WSLC activities. This integrated environment, shown in Figure 3–4, will consolidate activities of the four basic WSLC functions; design, manufacturing, support, and operations.

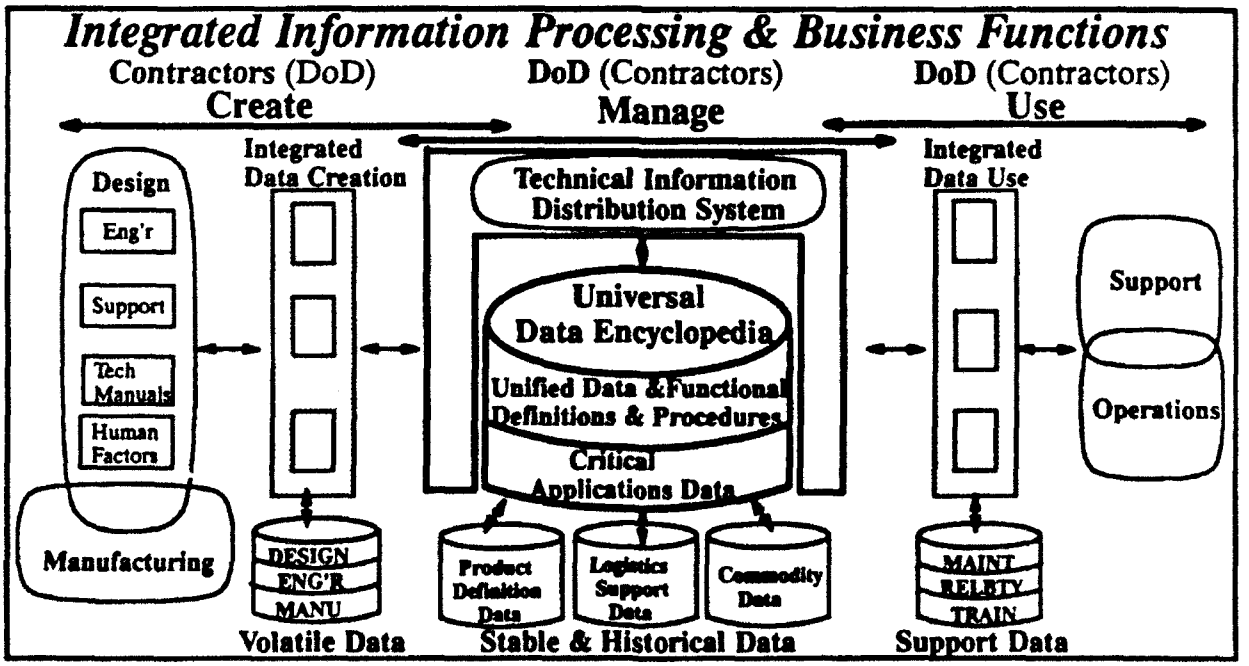

Figure 3–4 Advanced Stage of CALS

The scope and degree of integration during this phase of CALS will dramatically change not only how things are accomplished but what things are accomplished. Design activities and manufacturing processes not undertaken today will be facilitated by a near total integration of data with technical and information processing activities. Rules and procedures governing the creation, management, and use of information will also be standardized and coupled with WSLC functions. Standardized procedures, coupled with standard data specifications and dictionaries, will facilitate universal access (with appropriate authorization) on a real-time basis. Technical information for all phases of the WSLC will be readily available in an integrated open-system processing environment.

The data dictionary will be a critical element in facilitating this integration. As in the shared data environment, multiple data dictionaries will serve multiple data bases, with one acting as the *Universal Data Encyclopedia (UDE).* The functionality of the UDE will be greatly expanded interacting with information processing procedures and business functions throughout the life cycle. Complex rule based associations will govern technical activities ranging from design to manufacturing to maintenance. Execution of certain sequences will reinforce specified logic paths and thereby add to the encyclopedia's knowledge base.

3.6 Data Dictionaries and the Weapon System Life Cycle

The evolution towards greater integration involves changes in the CALS business environment and data dictionaries that parallel the CALS evolution (see Figure 3-5).

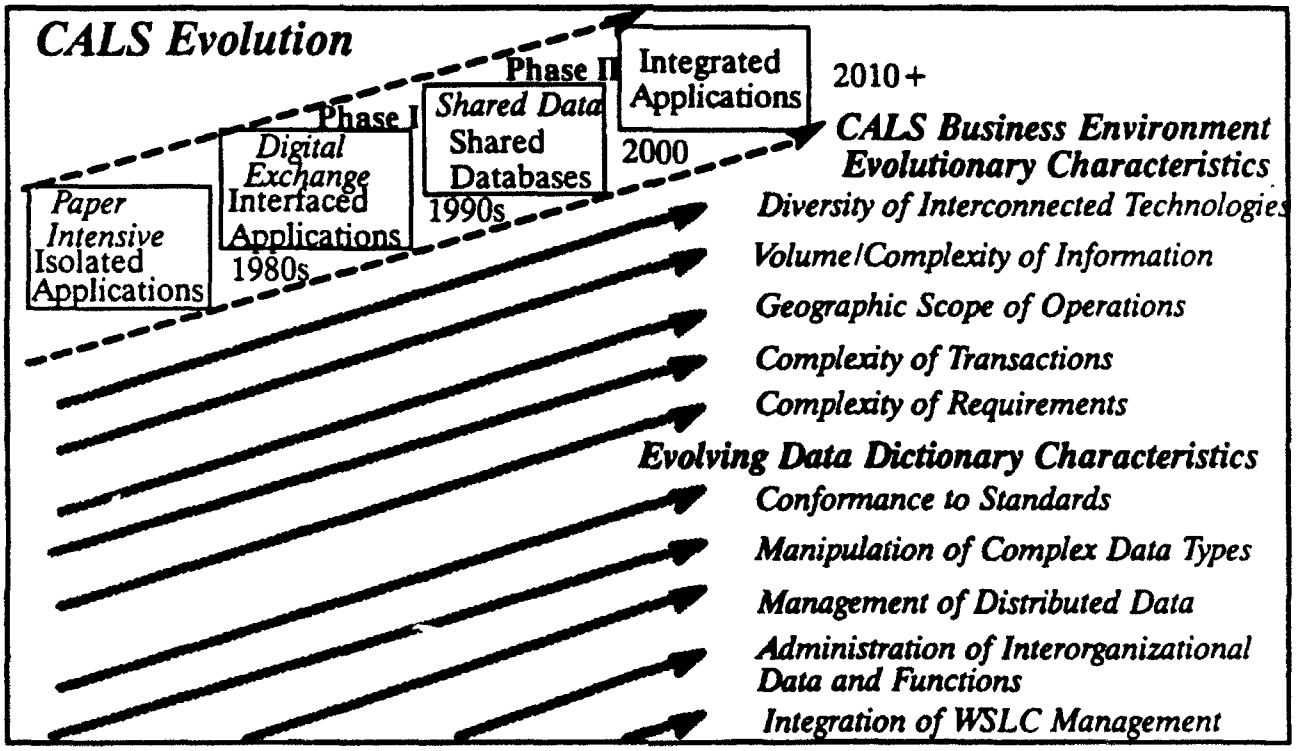

Figure 3-5 Evolving Trends for CALS, the Business Environment and Data Dictionaries

Changes in the business environment will shape how data dictionaries function just as data dictionaries will influence how a weapon system is acquired and managed throughout its life cycle. The trends pictured in Figure 3-5 involve a variety of relationships. For example, the growth in volume and complexity of information will stimulate the development of a data dictionary capability to manipulate complex data types. In general, each characteristic of the business environment is complemented by a corresponding data dictionary characteristic. Note that both sets of characteristics will experience extended growth during the 1990s.

As the diversity, complexity, and scope of the business environment grows, the data encyclopedia functionality will expand. The UDE will administer weapon system engineering and development processes as well as the data associated with a weapon system. All phases of the WSLC will be administered by the UDE. It will also be the key link among various organizations involved with weapon system acquisition, operation, and modification (see Figure 3-6).

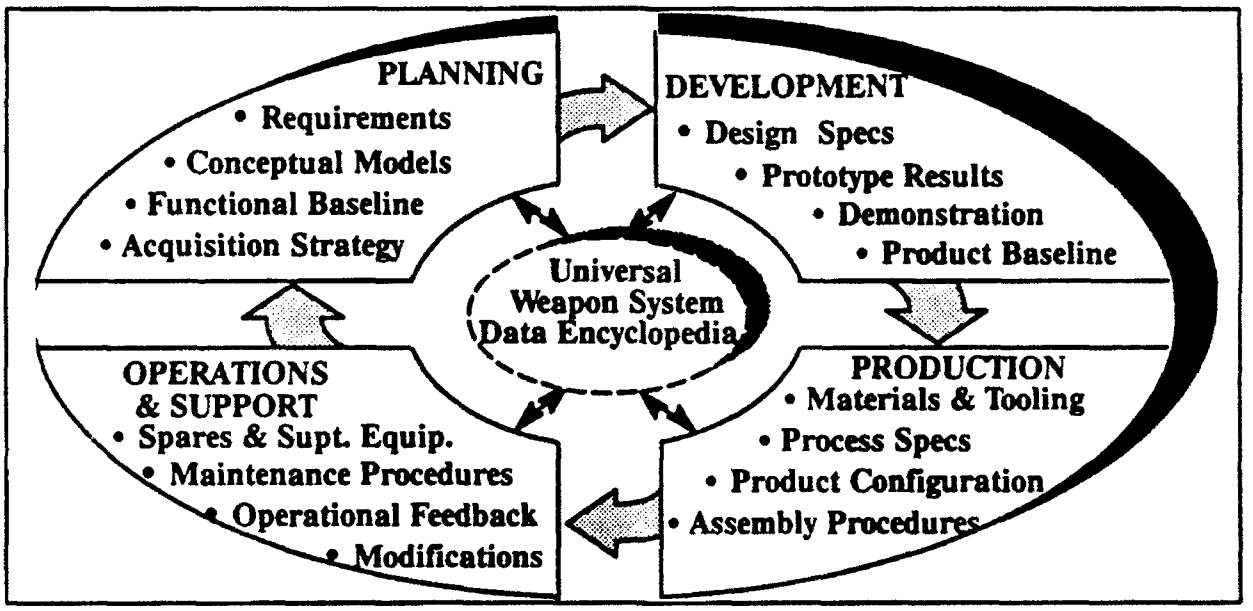

Figure 3-6 Data Encyclopedia - Weapon System Life Cycle Focal Point

The UDE will administer a CALS environment populated by a growing number of interconnected systems. These systems will be increasingly diverse, involving complexity far exceeding today's systems. This growth in interconnected technologies will depend on robust standards. The adoption of such standards, specifically data dictionary standards, will accelerate system interconnection.

As greater conformity to standards is established, the volume and complexity of digital data will increase. The processing of data types that is currently only at an embryonic stage will become a common occurrence. The growing volume and complexity of information during a weapon system acquisition will include complex data objects, audio and video information,

and graphic images (including text and contractual documents). The UDE will need to operate upon these data types just as numeric data is added or updated today.

Another trend is the geographic expansion of the business environment. Contractual arrangements among businesses will result in increasingly dispersed work sites. Information processing transactions will often involve several of these sites. Given the variations that will exist across many computer configurations and organizations, the UDE will need to manage their distributed operation and administer interorganizational data and processing functions.

3.7 Summary - The Data Dictionary Evolution Within CALS

In analyzing the role of data dictionaries within CALS, it becomes evident that a complementary business environment is necessary for the effective utilization of data dictionary capabilities. As shown in Figure 3-7, an evolving business environment is associated with current and evolving data dictionary characteristics. In some instances the data dictionary will enable changes in business practices; in other cases, enhanced business practices will enable full utilization of data dictionary capabilities.

CALS STAGES	BUSINESS ENVIRONMENT	DATA DICTIONARY CHARACTERISTICS
PAPER-INTENSIVE ENVIRONMENT	• Nonstandardized data • Stand-alone applications • Manual data exchange	• Passive • Integrated/freestanding • Automated/paper
DIGITAL-EXCHANGE ENVIRONMENT	• Limited data standards • Interfaced applications • Digital media exchange/ limited remote access	• Active/passive • Integrated/freestanding • Automated/paper • Nonextensible • Organization specific
SHARED-DATA ENVIRONMENT	• Standardized tech. data • Integrated applications • Extensive remote access • Dist. heterogeneous processing/logically unified databases	• Active • Freestanding • Automated • Extensible • Interorganizational
INTEGRATED TECHNICAL DATA ENVIRONMENT	• Standard data specs • Standard applications/ unified WSLC • Universal real-time access • Integrated open-system processing	• Active • Freestanding • Automated • Extensible • "Universal"

Figure 3-7 CALS Business Environment/Data Dictionary Summary

Currently in DoD, business is conducted both paper-intensively and through digital means. In this setting, the lack of comprehensive standards is the most significant factor inhibiting the use of data dictionaries among individual organizations.

As data standards mature, data sharing among diverse organizations will be feasible. Shared data operations will result in interorganizational data dictionaries. This trend will necessitate the careful administration of freestanding, active data dictionaries.

Effective data administration, open-systems architecture, and freestanding, active data dictionaries will act as catalysts in standardizing WSLC activities. These trends will be drivers, enforcing standardization among functions as well as data. This standardization will revolutionize both how things are done and the type of activities that are undertaken. Designs will be more intricate, product creation more exacting, and maintenance and operations far more complex.

4.0 RECOMMENDATIONS

4.1 Scope of Requirements

As DoD and industry systems evolve toward greater levels of integration, the scope and complexity of CALS-related activities will progressively increase. Attempts to connect larger, more diverse systems will rely on data dictionaries that can be logically interrelated. Data dictionaries will be the essential tool in linking organizations and integrating functions, as well as in capturing and storing large volumes of information.

Although the data dictionary is a technological approach at eliminating integration barriers, most CALS data dictionary requirements are non-technological in nature. Two primary requirement areas, shown in Figure 4-1, are essential for effective CALS data dictionary applications. These are;

1) data administration, and

2) dictionary related standards.

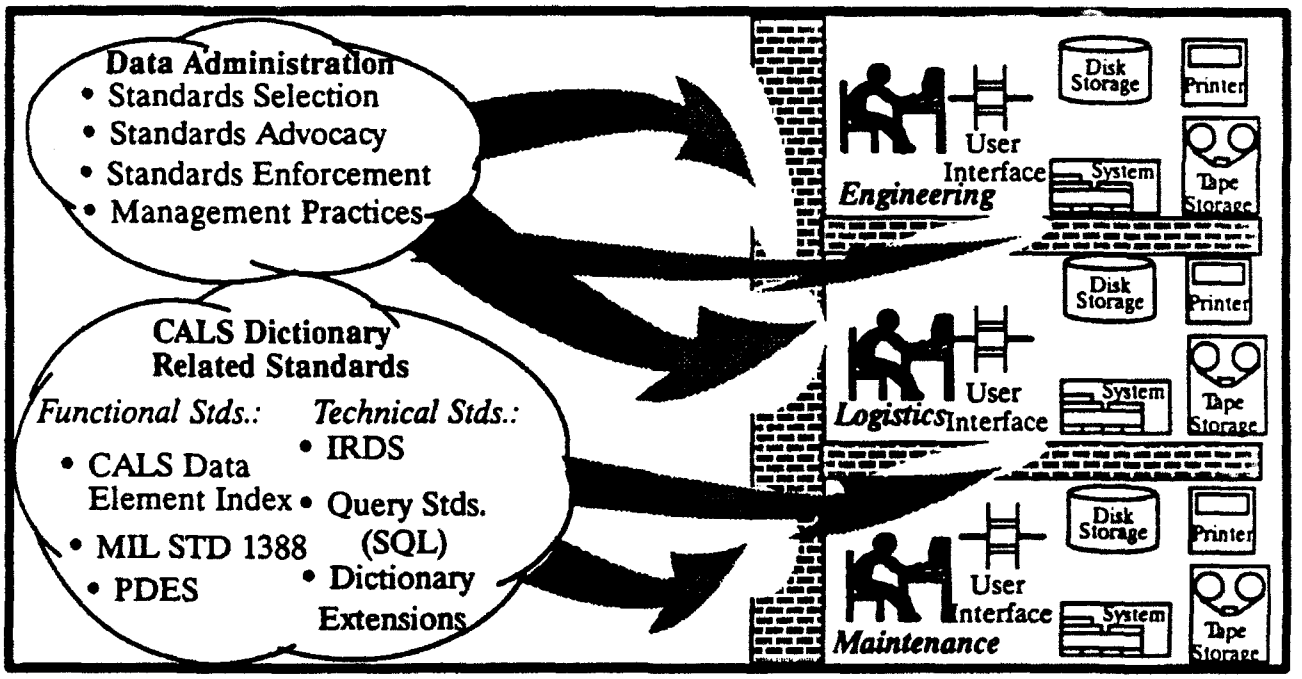

Figure 4-1 Scope of Data Dictionary Requirements

4.2 Recommendations - Data Administration

Data administration is the critical enabler for effective use of data dictionaries within CALS. Data dictionaries are merely tools; as such, they will only provide benefits if they are used effectively. Data dictionary benefits can be significantly increased by administering procedures and policies that efficiently govern data creation and use.

A successful data administration function has three critical components. First, an organization needs to be established that will administer and manage data. Second, a set of standards needs to be adopted that will govern the mechanics of how data is created and used. Third, a data management process needs to be established. This process needs to provide supporting services for the acquisition, management, and distribution of data and related technologies, and to resolve conflicts associated with creation and use of data.

RECOMMENDATION #1 OSD should establish a CALS Information Resource Management Organization (CIRMO).

A strong, centralized data administration organization needs to be established within DoD to coordinate and standardize weapon system technical information. This organization, CIRMO, will need to encourage not only technical standardization, but also mitigate the limitations that are created by fragmentation of organizational/business practices in an organization as large as DoD. A Joint Services group should be established to charter, organize, and oversee the initial operations of the CIRMO. In addition, the CALS industry steering group should recommend direction, identify requirements, and specify design considerations for the CALS Dictionary and other CIRMO responsibilities.

CIRMO needs to be focused upon the selection and adoption of data management standards, the implementation and administration of a data management process, and the resolution of conflicts that arise concerning data standards. CIRMO needs to be established at a level in DoD that will empower it to effectively carry out these activities, specified below as recommendations 2, 3, & 4. By focusing on these activities, awareness levels regarding standards will be heightened, and a central frame of reference will be available to nurture and promote common approaches in data management and data dictionary applications.

RECOMMENDATION #2 Identify, define, and adopt those standards necessary to effectively manage and use digital technical information for DoD weapon system acquisitions and operations.

There are numerous functional standards that impact the management and administration of data. As part of the data administration function, CIRMO, or the OSD CALS office, must undertake a comprehensive review of existing standards and select those standards that are appropriate for inclusion within the CALS initiative. Based upon this review, modifications, where appropriate, will need to be proposed and guidance, concerning standards adoption, provided to the various Service Branches.

In addition, CIRMO, or the OSD CALS office, will need to work actively and participate fully with the standards community to ensure that technical standards are consistent with CALS objectives. This is particularly true for data dictionary standards and query

standards. IRDS, the emerging dictionary standard, only addresses a small portion of the functionality that future data dictionaries will need to deliver. SQL, the accepted query standard, may require significant enhancements to work effectively in an heterogeneous, object oriented, distributed database environment.

RECOMMENDATION #3 Establish a data management process that reflects sound management practices, that complements appropriate data standards, and that satisfies legitimate contractor and Service needs.

Since CALS will involve large user populations, geographically dispersed, and representing diverse Service and contractor viewpoints and interests, achieving consistency of information resources will not be easy. In addition to adopting a set of functional and technical standards, a process that supports the acquisition, management, and retirement of information resources (i.e., data and technology) needs to be established.

As a first step, a CALS data management process needs to be defined, with its component services identified (e.g., information versioning, data security, data communications and distribution, data storage and archiving, etc.) and responsibilities defined. Once the CALS data process has been scoped out, requirements and procedures for each component service needs to be specified and integrated with other component services. Data management service requirements will need to reflect both contractor needs (as elicited from the CALS industry steering group) and DoD needs.

RECOMMENDATION #4 Administer and resolve conflicts pertaining to data characteristics, including its definition, ownership, and currency.

The CIRMO will also need to resolve conflicts pertaining to data – e.g., ownership of the data documented in the dictionary. Conflicts will need to be resolved in areas where similar assets must be made identical and where seemingly identical assets must be differentiated. Conflicting standards between contractor and DoD, among different elements within the DoD, and between "CALS standards" and standards sponsored by such organizations as the National Institute of Standards and Technology [NIST] will also need to be resolved.

RECOMMENDATION #5 Work with individual Service Branches in establishing Service-wide IRM organizations.

The data administration function needs to be implemented at the Service level. Therefore it is critical that IRM organizations be established for each Service and be empowered to administer and enforce standards and procedures throughout their respective Service. Specific authority and responsibilities may need to be tailored for each Service but at a minimum each Service IRM should establish a data management process and be empowered to resolve conflicts pertaining to data.

4.3 Recommendations -- Dictionary Related Standards

The second crucial data dictionary enabler involves the development and implementation of a rigorous set of standards. Although standards can be viewed as a general CALS requirement, a specific set of standards is crucial for effective CALS data dictionary applications. Developing and implementing technical and functional standards will be necessary for the efficient exchange and integration of metadata.

Technical standards involve the definition of how a computer system (or specifically a data dictionary) operates and interacts with other assets. Functional standards deal with the business environment and address business procedures and information. Although functional standards and technical standards are often interdependent, they need to be specified separately to ensure the maximum degree of flexibility possible. Functional standards are the standards that OSD can have the most impact on in the short and intermediate term and will require the most work effort.

RECOMMENDATION #6 Establish a CALS data standard.

A glossary needs to be established defining all CALS-related technical data as the first step in establishing a CALS data dictionary. This glossary will need to establish broad classes of data within which specific instances (data elements) can be defined. The glossary should build upon ongoing data standardization efforts and incorporate a standardized notation to represent the various relationships among data elements.

Three efforts, dealing with the standardization of data, are under way that need to be coordinated and their development and implementation accelerated. These efforts are:

- Mil Std 1388-2B, and
- CALS Data Element Index,
- PDES

In order to facilitate data standardization, OSD needs to define the high-level CALS information requirements and then determine how these three efforts relate to those information requirements. Currently there is no clear definition as to the relationship of these three efforts to CALS requirements or to each other.

Mil Std 1388-2B is focused on logistic support data that is created and acquired during the acquisition phase of the life cycle. Mil Std 1388 represents one attempt to standardize on a specific set of data. To be effective, this standardization needs to go beyond the logistic support analysis activity of the acquisition phase and include other activities within the acquisition phase of the life cycle, as well as other phases of the WSLC.

OSD (or CIRMO) needs to modify 1388 so that it assures data element traceability to all related design and engineering disciplines during a weapon system acquisition. In

addition, the Services need to use 1388 data elements during the operations and support phase of the weapon system. If the Services indicate that 1388 data elements are inappropriate, then the standard needs to be amended to incorporate the appropriate elements.

As a supplement to 1388, the CALS Data Element Index is an attempt to consolidate integrated logistic support (ILS) information that is created and used during the WSLC. Development of the CALS Data Index needs to be accelerated and then adopted throughout the Services. The completion, adoption, and faithful adherence to this standard will require effective data administration organizations at both the DoD and Service levels.

PDES is attempting to create common product definition data models. Unlike 1388-2B and the CALS Data Index, the exact identification of the data included in PDES is still being defined. OSD can accelerate the development and ultimate adoption of PDES by sponsoring prototype PDES implementations on acquisition programs of limited scope. Although these prototypes will only involve a subset of the PDES data models (typically mechanical aspects), prototype implementations will help refine the logical models and provide needed experience with physical implementation.

In the area of technical standards, CALS needs to adopt emerging standards instead of actively developing standards. Two technical standards, dictionary standards and query standards, are crucial for an integrated CALS environment. DoD needs to support their development and enhancement, and actively encourage their adoption and use within the CALS community. Although other standards, such as the seven layer ISO networking standard, are also integration enablers, they do not directly impact data dictionaries and so are not discussed here.

***RECOMMENDATION #7** Identify and define CALS data dictionary requirements.*

Data dictionaries need to evolve toward greater design consistency as they become effective tools in a heterogeneous systems environment. The Information Resource Dictionary System (IRDS) standard, which has been proposed by NIST, clearly dominates in establishing the guidelines for coordinating the design and interfacing of data dictionary systems. While the details of the features of the IRDS standard are discussed in other documents, general consensus within the data-processing community indicates that IRDS concepts will need to be extended to handle future data dictionary designs. These extensions will need to support object oriented data base systems, and distributed heterogeneous data bases that manage a variety of complex data types (including video and audio). OSD needs to specify desired functionality, participate and monitor the standards approval process, and at the appropriate time require IRDS-compliant dictionaries to support digitized weapon system acquisition data.

RECOMMENDATION #8 *Identify and define enhanced capabilities for a CALS data query capability.*

Data dictionaries facilitate communication between different systems. This communication is often accomplished by queries to the database. The Standard Query Language (SQL) is the standard for database query. The adoption of the SQL standard is significant when considering data dictionaries because information is captured by a query through the data dictionary. The data dictionary stores data location and data meanings – parameters needed to satisfy a query.

DoD needs to assure SQL compatibility with CALS databases, so a single query interface exits for CALS data dictionaries. For example, SQL capabilities to query object oriented data bases that capture complex engineering constructs needs to be evaluated. SQL functionality may need to be broadened to capture these information types.