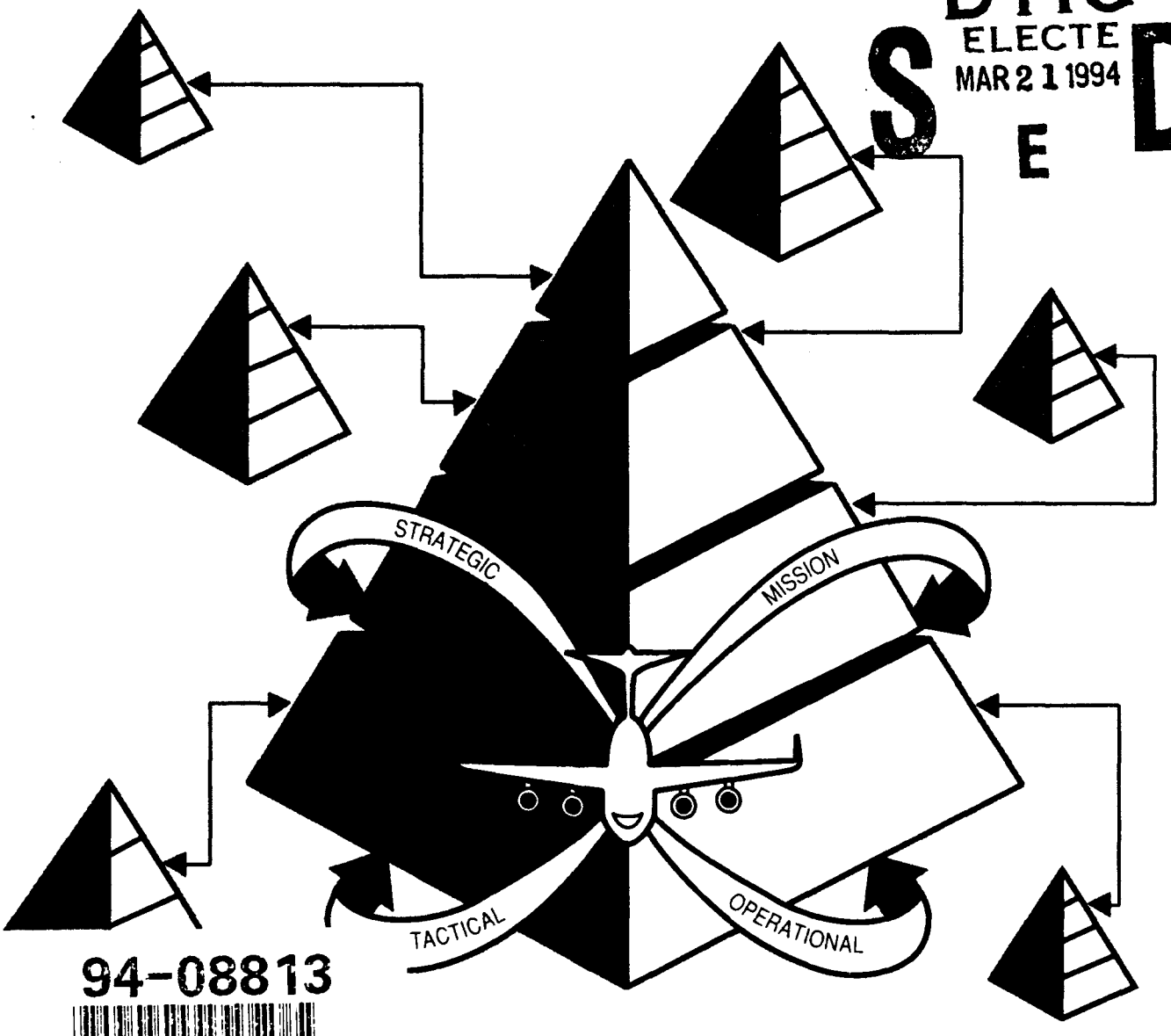
CORRECTING THE TOWER OF BABEL

A Tool For Data Standardization and Integration



Rosa L. Daniels, Major, USAF

Research Report No. AU-ARI-92-7

Correcting the Tower of Babel
A Tool for Data Standardization and Integration

by

ROSA L. DANIELS
Major, USAF

ARI Command-Sponsored Research Fellow
US Air Force Communications Command

Air University Press
401 Chennault Circle
Maxwell Air Force Base, Alabama 36112-6428

January 1994

Disclaimer

This publication was produced in the Department of Defense school environment in the interest of academic freedom and the advancement of national defense-related concepts. The views expressed in this publication are those of the author and do not reflect the official policy or position of the Department of Defense or the United States government.

This publication has been reviewed by security and policy review authorities and is cleared for public release.

Dedication

This publication is dedicated in memory of
someone who paved the way for most of
my life achievements

my sister

Zannie Mae Daniels Braico
6 June 1946–17 April 1991

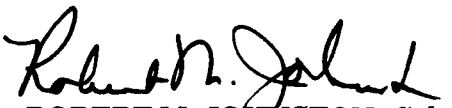
Contents

Illustrations

Foreword

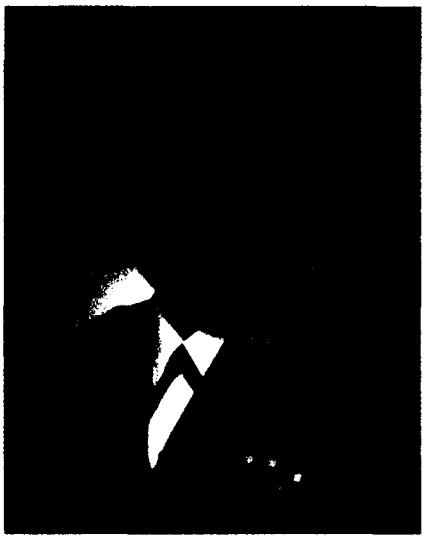
Maj Rosa L. Daniels examines the current Air Force Data Dictionary (AFDD) environment and finds multiple, independent data dictionaries which do not meet the need of the total force. The Air Force Data Dictionary historically has been unresponsive to user needs, having outdated capabilities, data redundancy, and lax regulatory requirements. Development and implementation of a comprehensive corporate data dictionary system is a *must* if the Air Force is to manage and control data effectively.

As we move into the twenty-first century, we will see significant changes throughout the Department of Defense (DOD) and the world. Major reductions in the defense budget and force structure will affect us most. We must find more responsive and economical means to provide for existing services. Major Daniels's recommendation for an Air Force "corporate" data dictionary will eliminate the need for multiple dictionaries with redundant data and high development and maintenance cost. It will also promote data shareability and provide the means for better management of Air Force information as a corporate resource.



ROBERT M. JOHNSTON, Colonel, USAF
Director
Airpower Research Institute

About the Author



Maj Rosa L. Daniels

Maj Rosa L. Daniels entered the Air Force in June 1979 and was commissioned a second lieutenant through Officer Training School (OTS). After completing computer OTS at Keesler Air Force Base (AFB), Mississippi, in April 1981, she was assigned to the 4501st Computer Services Squadron at Langley AFB, Virginia, as a joint systems test manager for the Joint Interoperability of Tactical Command and Control Systems program. In 1983 she became chief of operational software testing. In this capacity she provided technical engineering support for the Tactical Air Command systems, now called Air Combat Command.

In June 1984 she was assigned as a computer systems development officer in the only Air Force computer specialty code course at Keesler AFB. Later, she became an instructor-supervisor for the Communications-Computer Systems Operations Officer course. Here, she led instructors who guided students with any of the five communications-computer systems officer specialty codes for active duty, Reserve, and Air National Guard forces.

In October 1987 she became chief of software architecture while assigned to Headquarters Air Force Communications Command (HQ AFCC) at Scott AFB, Illinois. She developed concepts, policy, and guidelines to ensure that computer software and automated support systems met the needs of combat commanders. Before arriving at the Air University Center for Aerospace Doctrine, Research, and Education (AUCADRE), Maxwell AFB, Alabama, she served as communications-computer systems integration manager at the Technology Integration Center, Scott AFB. This paper shows the results of her research efforts as a research fellow at AUCADRE.

Major Daniels has completed the Communications-Computer Systems Staff Officer course, Squadron Officer School, and Air Command and Staff College. She holds a BA and an MEd degree in mathematics from the University of Florida.

Preface

I first became involved in the data dictionary world in the late 1980s. The Air Force Data Dictionary was quickly becoming obsolete, and the need to automate was imperative. In an era of high technological advances in the data automation arena, the Air Force must move toward providing corporate systems with more responsive and economical methods for conducting its business. Having a corporate data dictionary system has been an issue of major concern to HQ AFCC and DOD for quite some time. Since the concern continues, I have designed my research to provide information that will assist in the development and implementation of an Air Force corporate data dictionary.

The research and writing involved in this study would not have been possible without the support, encouragement, expertise, and patience of many people. First, I'd like to thank my sponsors at HQ AFCC, Maj Gen John S. Fairfield and Spain Hall, for their active interest and support of my efforts. Many thanks also go to the entire AUCADRE staff for lending their support and understanding and for making me feel at home.

Special appreciation goes to John Woodfin and Karen Singleton at the Standard Systems Center, Maxwell AFB, Gunter Annex, Alabama. They were a source of information and a sounding board and offered input and guidance. I am indeed thankful to my reading group chairman, Dr Bynum E. Weathers, for his guidance and patience; and especially my editor, Dr Richard Bailey, for believing in me, for helping me keep my research in focus, and for providing the "pep" talks. Also, special thanks go to those who assisted by reviewing certain parts of my research and providing input: Lt Col James G. Thompson, Headquarters USAF/Software Management Division; Bao Nguyen, SAF/Information Management Division; James DeGroff and Anthony Reiss, Technology Integration Center; Dan Lewis, Defense Information Systems Agency/Information Management SSC/XP and Programs; and many others.

Finally, and most of all, I am grateful to God for giving me the will to attempt such a tasking and to my family for being there for me, as always, loving, supporting, and understanding.

Rosa L Daniels

ROSA L. DANIELS, Maj, USAF
Research Fellow
Airpower Research Institute

Introduction

This study addresses current issues surrounding the Air Force Data Dictionary and serves as a means for providing direction and guidance toward resolving these issues. The failure of the Air Force to manage data as a corporate asset has created data dictionary data bases everywhere. These data bases carry the same data elements with different names, definitions, and lengths. Similar to the builders of the "Tower of Babel" in biblical days, these data bases do not share information. As a consequence, the Air Force has a variety of data dictionaries. Most were developed when Air Force directives and guidance were practically nonexistent. Unnecessary data duplication led to high development and maintenance cost and independent data elements that prevent information systems from sharing data. These are just a few of the problems associated with the AFDD environment.

The Air Force must exploit fully its information resources. According to the data management and standards regulation, the Air Force must change the way it treats data.[1] The Air Force must treat data as a "corporate" asset in much the same way it manages manpower, facilities, materiel, and financial resources. A study of the needs of corporate information management (CIM) indicates that information systems must communicate and share data throughout the Air Force, among services, and with other countries.

In his speech to the National Security Forum at Maxwell AFB, Secretary of the Air Force Donald B. Rice touched on the current dilemma and stated that "everywhere we can we will eliminate redundancy, stovepipes and stale thinking."[2] An Air Force corporate data dictionary addresses Secretary Rice's concern by providing an orderly migration from the Air Force's existing independent information systems environment to a shared data base environment. More than any other tool, a data dictionary provides for better management of Air Force information as a corporate resource.

In addition, Secretary Rice contended that "focusing on effects rather than simply analyzing quantifiable data will be critical in the future as our traditional measures of effectiveness are superseded by capabilities afforded us by advances in technology."[3] The increased demand for more efficient, reliable, capable, interoperable, and integrated systems mandates development of the data models and data management standards necessary for sustaining tomorrow's Air Force.

Chapter 1 chronicles the AFDD from its infancy to its current state. It describes the dictionary's strengths and shortfalls in support of Air Force objectives and goals and its attempt to support the most recent data management and standards program. The chapter discusses the Air Force's search for a corporate system that will satisfy requirements and meet the needs of the Air Force as well as the Department of Defense.

Chapter 2 delves into the underlying system engineering and technical standards and concepts, and it details how these standards relate to overall system planning. It

addresses Air Force requirements for a data dictionary and provides DOD, Air Force, and organizational level guidance in the development and maintenance of a data dictionary. The key to identifying a certain system is the selection of a suitable architecture and verification that the existing infrastructure will support the system. Chapter 3, armed with requirements for a data dictionary as outlined in chapter 2, addresses the proposed system and resources required. It also introduces DOD efforts to develop a defense data repository system for use by all services.

Chapter 4 details how the major command and functional area dictionaries will migrate to the corporate structure, suspension of developments not consistent with Air Force strategic plans, and development of plans for data interchange. Chapter 5 addresses specific data dictionary major issues and concerns and makes recommendations based on the facts presented in the study. It also summarizes the findings of this study.

Notes

1. Air Force Regulation (AFR) 4-29, *Air Force Data Management and Standards Program*, 23 April 1990, 3.
2. Secretary of the Air Force Donald B. Rice, "Global Reach—Global Power: One Year Removed" (Speech presented to the National Security Forum, Maxwell AFB, Ala., 7 June 1991).
3. Ibid.

Chapter 1

The Air Force Data Dictionary History

Before I begin my discussion, I consider it appropriate to ask, "What is a data dictionary?" DOD Manual 8320.1M, "DOD Data Administration Procedures Manual" (draft), defines a data dictionary as a repository of information that describes the characteristics of data used to design, monitor, document, protect, and control data in information systems and data bases.[1] In other words, the dictionary contains information about data, not data itself. The Air Force Data Dictionary (AFDD) will serve as a central repository of information about data elements (metadata), records, files, systems, forms, reports, and other important entities. This information is critical, for without it there is no reference nor understanding of what the data itself means.

An AFDD also will assist in designing and managing data models. Over a period of time, it may evolve into an Air Force data repository or encyclopedia. The terms *repository* and *encyclopedia* are widely used in the same context as *data dictionary*. All three terms encompass the concept of "a store of information describing an enterprise's data, information, and the processes and organizations that act upon that data and information"; however, the terms *repository* and encyclopedia denote more robust systems and include full extensibility, versioning, security, and other special services.

Background

Since the advent of the automation era, the Air Force has been plagued with data processing and information management decisions: "What hardware should it use? Which software? What type of data base? What is the desired output? How much training is needed?" The list of concerns seems endless. Emerging technology has played a key role in creating these concerns and will continue to take its toll on outdated, incompatible, and nonstandard systems.

The history of the Air Force Data Dictionary begins with the Standard Systems Center (SSC).[2] The center was originally located in Washington, D.C., but moved to Gunter Air Force Base (AFB), Alabama, in the early 1970s as the Air Force Data Systems Design Center.[3] It was later redesignated as the Standard Systems Center. The early data dictionary system was developed in hard copy (paper) during the 1964 to 1974 time frame and later was transferred onto microfiche. The dictionary was created to serve as a central repository of descriptive data elements used by Air Force organizations.

In 1985, when the Air Force converted the 300-series manuals to 700-series regulations, the dictionary was realigned as Air Force Regulation (AFR) 700-20, *Air Force Data Dictionary (On-Line),* vol. 1. Many Air Force organizations, however, chose to create their own systems, as this one did not meet their particular needs. The system suffered in three key areas. First, there was only limited responsiveness to user needs (i.e., the microfiche was difficult to read, quarterly updates were not timely for accurate use by the functional areas, the addition and modification process for data elements was laborious and time-intensive, and there was no automated inquiry or search capability). Second, the system capabilities were outdated (i.e., since its software [L10] was unsupportable and undocumented, inquiries had to be uniquely written, content edits were done manually, management reporting was compiled manually, and the coordination process was labor-intensive and time consuming). And third, the system suffered from lax regulatory requirements (i.e., the existing data element standardization and management program was not enforced, users did not update the data base in a timely manner, and there was a lack of formal metadata-naming conventions).

Under the direction of the Air Staff and Headquarters Air Force Communications Command (HQ AFCC), SSC validated the requirement for an automated, on-line data dictionary system in October 1987. SSC also prepared a draft functional description, but the description was never approved. Meanwhile, other commands continued to develop dictionaries. The project gained momentum in April 1988 when the Software Management Division (SC) at the Air Staff directed an analysis of current and future major command (MAJCOM) data dictionary efforts. Results from these studies appear later in this chapter.

Current System Deficiencies

Early in 1989, due to the pending deactivation of the existing system and through guidance received from the Air Staff, SSC developed a prototype dictionary by using a relational data base (RDB) machine with a fourth generation language.[4] Based on the prototype, the Air Force purchased the RDB technology as its new dictionary hardware platform. The SC community at the Air Staff provided $350,000 to purchase the RDB and associated software. SC installed this new system in January 1990.

During this process, the Air Force drafted a new information management regulation, which defined an entirely new methodology for managing and standardizing Air Force corporate data.[5] AFR 4-29, *Air Force Data Management and Standards Program,* identifies standard data element attributes and data element standards. The new regulation establishes data element-naming conventions and highlights generic element structure and attributes, data element structure and attributes, data element aliases, and class words. For example, *elements* are named units of data, and attributes embody characteristics of these units. An *alias* is just an alternate label or name. *Class words*, often referred

to as class names, are words in a data element or generic element that identify the type of data being presented. The regulation includes a list of class words with definitions by category. This new regulation requires the participation of automated data system managers worldwide. Under these circumstances, the process could take years.

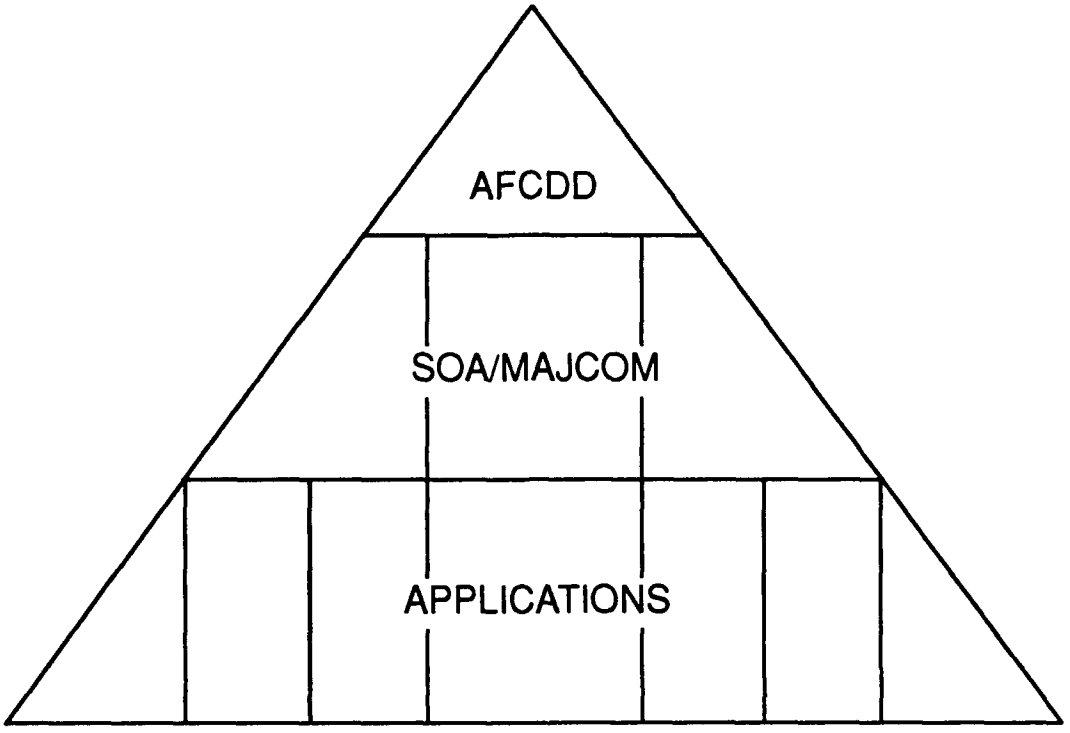
According to the program manager (PM) at SSC, the current system was on-line by 1 May 1990, but the lack of well-defined user requirements and the absence of an implementation plan hindered its use.[6] Two factors contributed to the lack of well-defined user requirements: the failure to baseline the existing system and the failure to keep pace with changing directives, such as the new data management and standards regulation, and other systems being developed by the Air Force. Inept management and poor documentation filing systems contributed to the failure to prepare an implementation plan. In addition, uncertainty, vacillation, and myopia at different levels resulted in contradictory directives and guidance.

A recent program management directive (PMD) led SSC to use the contractor-developed Military Airlift Command (MAC) data dictionary because of its adherence to the new data management and standards regulation and for referential integrity purposes.[7] The MAC Integrated Data Administration System (MIDAS) is a data-naming and standardization software application developed to meet the requirements of the MAC C-2 upgrade program.[8] MIDAS became operational at MAC on 1 May 1991. Since MIDAS did not have the capability to interface with the current AFDD data base, MAC had to change software and modify some of the current system. Once again the issues of poorly defined requirements and "instant" development of a dictionary to meet current needs surfaced. This system is not operational at SSC; however, MIDAS has been installed as the Air Force information resources dictionary system in the Office of the Air Force Data Administrator in the Pentagon.[9]

Department of Defense Data Dictionary Search

The Air Force has conducted several studies in search of a "corporate dictionary system." The Standard Systems Center performed one of the first studies in the spring of 1988.[10] Their study recommended three levels of dictionaries within the Air Force software community: applications level, MAJCOM and special operating agency (SOA) level, and Air Force level. Each level would have a passive or an active dictionary. In a passive mode, the dictionary stores metadata, but it does not interact with or control the computer environment external to the dictionary. When a dictionary operates in the active mode, it not only stores metadata, but it also interacts with and controls the environment without human intervention. The structure the study recommended resembles a pyramid with a passive Air Force corporate data dictionary (AFCDD) at the top, a passive MAJCOM and SOA data dictionary at the next lower level, and

an active applications dictionary at the lowest level. Figure 1 illustrates the pyramidal structure.



Legend:
AFCDD—Air Force Corporate Data Dictionary
MAJCOM—Major Command
SOA—Separate Operating Agency

Source: Standard Systems Center, "Standard Systems Center (SSC), Data Dictionary Study," Gunter AFB, Ala., 30 June 1988.

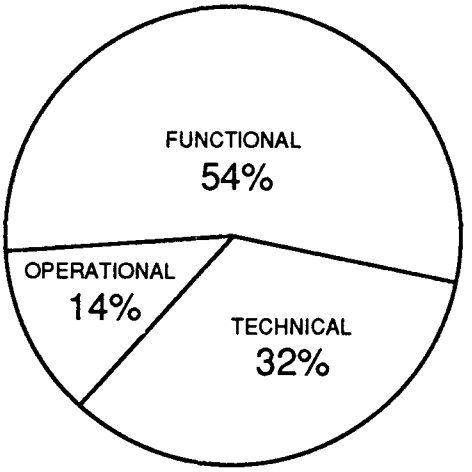
Figure 1. Data Dictionary Pyramid

The study by SSC recommended such generic capabilities for an Air Force corporate data dictionary as automation that uses a relational data base management system, 24-hour on-line accessibility, menu and structured query language inquiry capability, and an active thesaurus. The study also recommended the purchase of a data base machine with a front-end processor. Much discussion took place before a system was actually selected, thereby delaying the activation process considerably.

Since development of such a system impinged on the integration arena, the then Air Force Communications-Computer Systems Integration Office (AFCSIO) became involved in the effort and conducted a brief analysis of the current data dictionary environment.[11] Its recommendation included the selec-

tion of a system with automated access to data elements currently documented in AFRs 700-19, *Computer Systems Authorization Directory (CSAD) FOUO (On-Line)*, and 700-20, *Air Force Data Dictionary (On-Line)*, vol. 1, and the implemention of the new data management and standards regulation (AFR 4-29). AFCSIO also recommended establishment of an Air Force working group to develop an implementation strategy and a transition plan for an Air Force corporate data dictionary. AFCSIO decided to delay further development of the data dictionary until requirements were formally identified, documented, and coordinated.

The newly activated Technology Integration Center (TIC) at Scott AFB, Illinois, conducted the most recent evaluation of the need for a data dictionary during January–May 1991.[12] Its study was the most in-depth and significant evaluation conducted to date. After an Air Force and DOD-wide identification of both complete and developing systems, a team of technical experts armed with evaluation criteria and profile factors analyzed each system. The experts grouped the factors into the three profiles depicted in figure 2. The functional



Source: "Technology Integration Center Evaluation Report," Headquarters, Air Force Communications Command/Studies Analysis, Scott AFB, Ill., May 1991.

Figure 2. Profile Factors and Weight

area of the profiles highlighted the ability of the system to meet functional requirements as prescribed in the appropriate regulation and represented 54 percent of the overall evaluation. The technical area measured the ability of the system to meet existing Air Force and other federal standards and represented 32 percent of the overall evaluation. The team designed the support portion of the profile to assess the life-cycle history and resources associated with system implementation, and it represented the remaining 14 percent of the evaluation.

Several data dictionary systems were in use or under development within the Air Force and other DOD components. The TIC team identified and evaluated six key systems: the Personnel Computer—War-fighting and Intelligence Sys-

tem Dictionary for Information Management (PC-WISDIM), the Air Force Logistics Command (AFLC) Command Data Dictionary (CD/D), DOD Logistics Data Resource Management System, the Army Data Dictionary System (ADDS); the Army Materiel Command Data Dictionary, and Headquarters MAC Integrated Data Administration System. The TIC team members conducted site visits and evaluated each of these systems. They based their evaluations on demonstrated capabilities, not on projected modifications and enhancements.

Their study did not identify a single system as outstanding. In fact, the study found similarities among four of the six systems evaluated. This discovery accounts in part for no system scoring significantly higher than any other in any of the three areas. At the time of the evaluations, the Army data dictionary was the only operational system that had been in existence for at least two years. The team initially recommended the Army's data dictionary as an interim solution. Later the study advanced MIDAS and PC-WISDIM for consideration as long-term solutions, since the team determined that either would be desirable if modified or enhanced.

Also during this time frame, the corporate information management (CIM) program brought about major organizational changes. One of the key and most valuable changes in the data dictionary/repository arena was the establishment of the Department of Defense Information under the Office of the Secretary of Defense (C^3I).[13] Paul Strassmann was appointed director of defense information. He greatly influenced the dictionary process for DOD and the Air Force by creating the Information Technology Policy Board (ITPB). As chairman of the ITPB, he immediately formed a joint data administration task force and began the search for a DOD data repository. Today, this endeavor is well under way.

Summary

This chapter has shown that selecting a viable corporate data dictionary is a long, tedious, and challenging process. Insufficient support for users, outdated capabilities, inability to share data, and existence of multiple and independent systems characterize the current environment. Corporate data management and standards programs are almost nonexistent, and they must be improved and enforced.

Notes

1. DOD Manual 8320.1M, "DOD Data Administration Procedures Manual," draft, 1 November 1991, A-4.

2. Standard Systems Center (SSC), "Data Dictionary Status Report," Gunter AFB, Ala., SSC/XPT, 1991, 1.

3. *Automated Data Processing Units, Organizational Changes 1967–1989*, Standard Systems Center, Gunter AFB, Ala., n.d., 1.

4. SSC, "Data Dictionary," 2.

5. AFR 4-29, *Air Force Data Management and Standards Program*, 23 April 1990, 1.

6. SSC, "Data Dictionary," 3.

7. PMD 9261(3)/PE 38610F, *Program Management Directive for the Air Force Corporate Data Dictionary*, 29 March 1991, 4.

8. *White Paper on Military Airlift Command (MAC) Information Resources Management System (MIRMS) Approach to Defense Information Systems Agency (DISA) / Corporate Information Management (CIM)*, Headquarters MAC/SC, 2.

9. Bao T. Nguyen, Document Review Notes, 28 January 1991.

10. Standard Systems Center, "Standard Systems Center (SSC), Data Dictionary Study," Gunter AFB, Ala., 30 June 1988.

11. "Air Force Communications-Computer Systems Integration Office (AFCSIO) Data Dictionary Analysis Report," Headquarters Air Force Communications Command/Studies Analysis, 3 July 1990.

12. "Technology Integration Center Data Dictionary Evaluation Report," Headquarters Air Force Communications Command/Studies Analysis, Scott AFB, Ill., May 1991.

13. Briefing, Technology Management School, subject: DOD Information Resource Management, 20 November 1991.

Chapter 2

Requirements, Planning, and Guidance

Before the Air Force implements a dictionary system, it must identify and define certain requirements. The Air Force must establish procedures to develop a data dictionary. To plan, identify, develop, coordinate, implement, document, and manage data effectively and efficiently, the Air Force must perform certain tasks. It must also develop data and business models to attain the information and related business processes needed to achieve the missions, functions, goals, objectives, and business strategies of the Air Force. This chapter discusses data modeling.

While DOD and Air Force standards and regulations dictate the development and implementation of a dictionary, the MAJCOMs and functional areas play a major role in this process. In a program update Bao T. Nguyen states that the functional users must be involved.[1] They must determine and clarify their data requirements and assist in the validation and verification of the overall logical data structure. Once data and processes can be viewed from an organization-wide perspective and placed in logical groupings, redundancies and inconsistencies can be identified and data and process sharing can be achieved. Since business processes change from time to time, this chapter focuses mainly on the data needed to achieve the mission of the organization.

Requirements and Development

An Air Force corporate data dictionary system will provide a modernized automated central repository of information about data in support of the Air Force data management and standards program. The objectives include, but are not limited to, the support of MAJCOM operations and decision making with data that meets the needs of the Air Force community for availability, accuracy, timeliness, and quality of information. Other objectives are to facilitate interoperability and data sharing, both horizontally and vertically; implement data standards; and develop an awareness of the value of data as a corporate resource. The long-term requirement evolves the dictionary into a repository to house more information and to provide more capability for managing data in a fully distributed environment to support standard and nonstandard communications-computer systems. The development of a corporate data resource repository is heavily influenced by emerging standards within the computer-aided software engineering (CASE) tools environment. These productivity tools improve development of software applications and the system's life cycle.

Systemwide Requirements

The goals of the Air Force Corporate Data Dictionary are met through requirements that offer concurrent support to unlimited multiple users. In general, these goals are to support different user types whose various needs are subject to environmental conditions and to provide user-friendly, menu-driven, graphical, and textual interfaces to support both novice and expert users. Specifically, they are to capture and store standard data elements and their attributes, provide convenient on-line data element documentation query and reporting capabilities, track the state of each standard element throughout its life cycle, and identify the effect of proposed changes in standard elements. The dictionary will provide support to other data management activities as it becomes the Air Force data repository. To meet these requirements involves a great deal of planning.

Planning and Development

The AFCDD must be implemented in a phased manner consistent with life-cycle management disciplines. While the phases may vary from time to time, they make it imperative that the tasks be performed in a prescribed order. A phase is no more than a "one-time slice of a development cycle."[2]

The initial phases must pertain to the acquisition of a data base and the development of the original on-line system to replace AFRs 700-19 and 700-20. This system serves as a baseline for the completed AFCDD. Subsequent phases will address the implementation of the baseline system and population of the system. Later chapters highlight the phased approach and the migration of the independent data dictionary systems and discuss future enhancements in more detail.

DOD Policy and Guidance

The key DOD document that promulgates procedures for data standardization and management is DOD Manual 8320.1M.[3] Currently in draft form, this document will provide the guidance necessary for managing data, data modeling, and integration and standardization. Also of great importance, the Information Resources Dictionary System (IRDS), a federal information processing standards (FIPS), was established to support development of automated tools to support the application of data administration standards and procedures.[4] The IRDS specifies a computer software system that provides facilities for recording, storing, and processing descriptions of an organization's significant data and data processing resources. It includes the functions performed by data dictionary systems and information repositories and promotes portability of valuable information resources within and among federal agencies. Both documents play a vital role in the development of an Air Force corporate data dictionary.

Air Force–Specific Requirements and Guidance

AFR 4-29 and the 700-series regulations are the key Air Force documents that support the goals and objectives of an Air Force corporate data dictionary. As I mentioned in previous sections, AFR 4-29 provides guidance for the activities that plan, design, model, synchronize, standardize, and control Air Force data at all echelons.[5] The 700-series regulations provide policy and guidance for life-cycle management of communications-computer systems and visual information systems.[6] AFRs 700-2 and 700-9 focus more on the area of concern. AFR 700-2 consolidates and integrates policy and responsibilities, defines planning processes, and establishes procedures and responsibilities for developing, using, maintaining, and implementing strategic plans and architectures for communications-computer systems.[7] AFR 700-9 sets policies and procedures for applying and developing information systems standards.[8]

The AFCDD will be developed to allow widespread sharing of data elements, reduce data element redundancy among Air Force systems, and improve the efficiency with which the Air Force conducts business. Overall, the Air Force needs the dictionary to provide information on where it does business, including business rules to indicate how it uses information and how that information flows to users at all levels within the Air Force. The community of users includes data administrators, software developers, and the Air Force community at large. The data administrators will use the AFCDD to aid in the standardization and coordination process. Software developers will use it to locate the standard definitions of data elements and codes and to register how the data are used. The Air Force community at large will use it as a source of common definitions, such as standard codes for aircraft, maintenance status, and personnel classifications.

Organizational Goals

Controlling and administering information resources in an organization is becoming increasingly difficult. Management plays a key role in this process. Called "management value-added" by Paul Strassmann, this role has value only if surrounded by the appropriate policy, strategy, methods for monitoring results, project control, talented and committed people, sound relationships, and well-designed information systems.[9] He advises the Air Force to "automate success, not failure." If top-level management does not play a major role in this process, then we can expect to fail.

Managers must develop data models that accurately and logically represent the data required to achieve functional mission goals and organizational business strategies. A data model represents a way to document our understanding of what data our organization (i.e., application and program) needs, how the

data are organized, and how the data are related to each other. Data modeling consists of planning, identifying, analyzing, and designing methods. Mission-based data models support the design of integrated information systems and data bases which meet organizational requirements. These models help an organization to identify its data and information requirements and the complex relationships those data and information have with other organizations. They also help standardize names and definitions for data elements, identify data redundancies and gaps, and point out data-sharing and interoperability opportunities.

There exist several types of approaches to data modeling. The Air Force may benefit through application of three basic models as cited in DOD Manual 8320.1M.[10] Figure 3 shows the interrelationship of the three models.

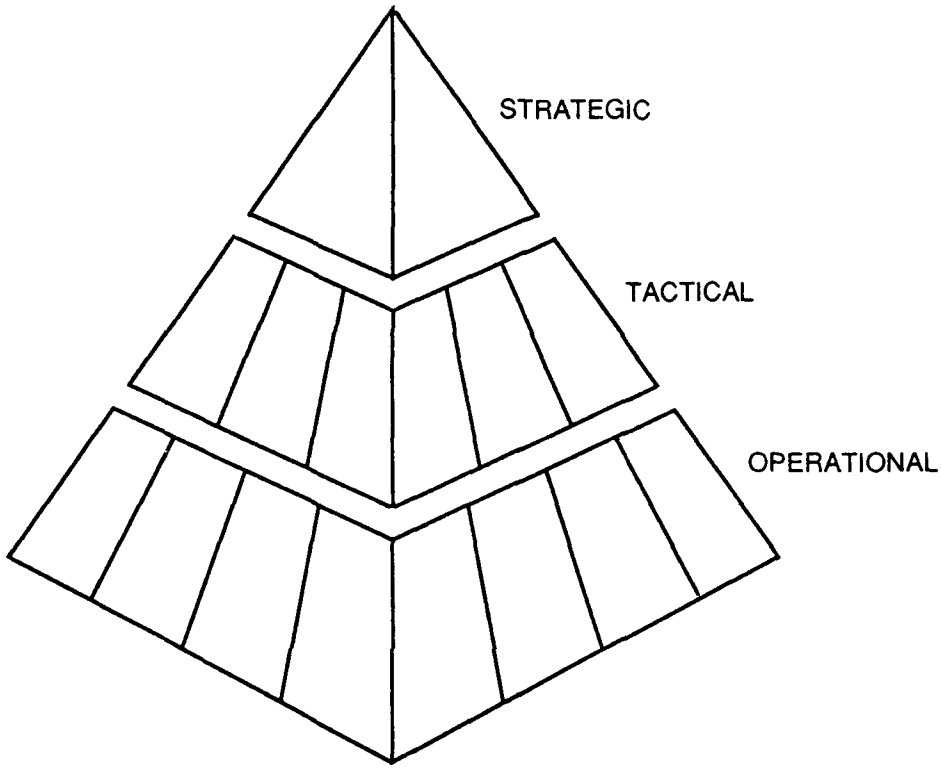


Figure 3. Interrelationship of Three Basic Data Models

The first model, the strategic data model, is the organization's highest level data map. Its entity list graphically represents the macro-level data, business processes, and data relationships needed to achieve the mission, goals, objectives, and business strategies of the organization. The model shows potential data-sharing opportunities throughout the organization. It is often referred to as the "enterprise" model.

The second model, the tactical data model, a subset or partition of the strategic data model, provides a more detailed data map with an entity list of the mission, goal, objectives, and business strategies of one functional area of the organization. It shows potential data-sharing opportunities within the functional areas.

A third model, the operational data model, a subset or partition of the tactical data model, performs at the operational subfunction level and offers the greatest detail on the data elements needed to achieve the mission, goals, objectives, and business strategies of the subfunction that in turn support the functional area and the mission of the organization.

Each pyramid or face of a pyramid becomes more complex or detailed the lower you go in the construct. For example, the strategic data model might consist of from eight to 10 or more functional or subject areas; each functional or subject area might break down into multiple operational or operational subject areas.

The ideal approach starts at the top and develops the strategic model first, the tactical second, and the operational third. This approach provides the required data elements which are also normalized for standardization. Standardization is the concept that the characteristics of each shared data element are defined uniquely and accepted by all data users across an organization(s). Data elements must be normalized before they can be standardized. That is, each attribute must be placed into the entity in which it belongs by associating it with the appropriate prime name.

Perhaps the most logical strategy involves a case of "reversed engineering." This approach identifies relevant data elements already documented, standardizes them, and documents existing data elements as data element aliases of the new standard data elements.

An Air Force data-modeling study was conducted recently to develop an Air Force strategic data model to identify high-level data entities or classes of data needed to operate and manage the Air Force, particularly Headquarters Department of the Air Force.[11] A series of workshop sessions were conducted, taking a top-down view of the business of the Air Force. The strategic model with documentation is contained in the referenced study. The model is not a static document but a work in progress. It can, however, be used as the starting point for developing future interoperable-by-design (rather than retrofit) management information systems and will assist in linking and interoperating existing information systems.

Summary

Much planning goes into the development of a data dictionary system. Although most formal policy and guidance is currently in draft form, a considerable amount of information, experience, and expertise is available. Numerous requirements are continuously being defined. Good life-cycle management must be employed. To achieve organizational goals and missions, data and process models must be developed. The use of data models provides the

organizationwide vision necessary for planning, designing, building, and maintaining future integrated and interoperable information systems.

Notes

1. Bao T. Nguyen, Atch 1 to SAF/AAID, letter, subject: Air Force Data Management and Standards Program Update, 10 July 1991, 3.
2. Philip W. Metzger, *Managing a Programming Project* (Englewood Cliffs, N.J.: Prentice-Hall, 1973), 5.
3. DOD Manual 8320.1M, "DOD Data Administration Procedures Manual," draft, 1 November 1991, 1-1.
4. Federal Information Processing Standard (FIPS) Publication 156, "Information Resources Dictionary System (IRDS)," draft, 8 May 1991.
5. AFR 4-29, *Air Force Data Management and Standards Program*, 23 April 1990, 1.
6. AFR 700-1, *Managing Air Force Communications-Computer Systems*, 28 February 1989, 1.
7. AFR 700-2, *Communications-Computer Systems Planning and Architectures*, 15 December 1987, 3.
8. AFR 700-9, *Information Systems Standards*, vol. 1, *Information Systems Standardization Program*, 15 March 1985, 1.
9. Paul Strassmann, *The Business Value of Computers* (New Canaan, Conn.: The Information Economic Press, 1990), 511.
10. DOD Manual 8320.1M, G-1.
11. "U.S. Air Force Strategic Data Model," study (Alexandria, Va.: Information Engineering Systems Corporation, 30 September 1991), 1–4.

Chapter 3

Architectures and the Data Dictionary

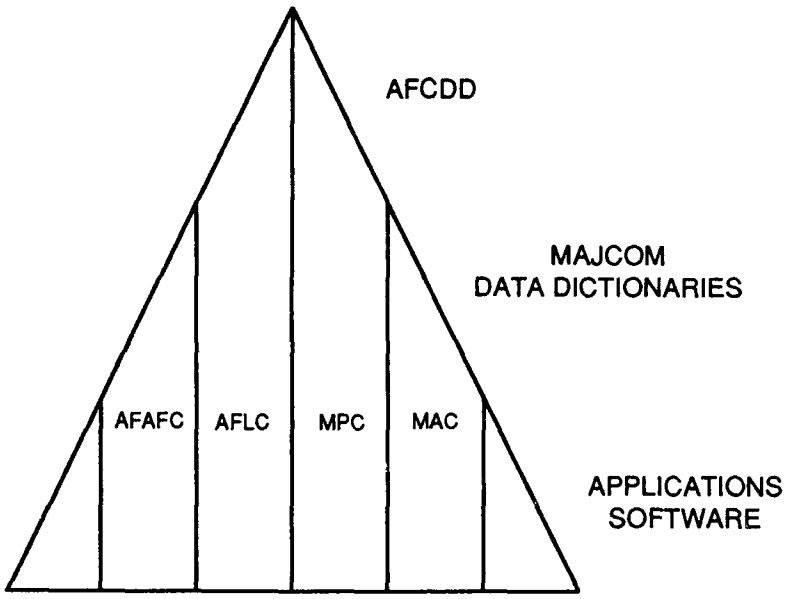
As previous chapters have shown, the current data dictionary environment offers insufficient support for users, unmet and undefined requirements, inability to share data between applications, lengthy schedules, and many other inadequacies. These weaknesses have hampered mission accomplishments and resulted in multiple independent dictionaries with increased systems development and maintenance costs. The latest communications-computer systems planning and architecture guidance challenges the Air Force to find new ways to use the resources and technology that exist today to satisfy our most urgent requirements.[1] This challenge does not involve the maintenance of nonvalue added oversight and documentation.

Designers must include in the Air Force Corporate Data Dictionary the data elements or data resources that mirror each organization's activities if the dictionary is to meet the information requirements of the entire Air Force in an accurate, controlled, and timely manner. The data dictionary will eliminate redundancy and provide control that will enforce security and standards.

If the data dictionary system is to eliminate variations in the Air Force's data that is shared and interchanged and simultaneously improve control and communications at all levels, it must have a sound structure, also known as an architecture. A number of sources help to define an architecture. *Webster's Ninth New Collegiate Dictionary* defines an architecture as "a unifying or coherent form or structure." John E. McRoberts, in "Establishing Control in a Data Dictionary Environment," defines it as an "orderly design or structure and the rules needed to create it."[2] This writer defines architecture as a road map, showing users how to implement or migrate data to the target system.

The architecture chosen will provide centralized control to facilitate data base standardization, while, at the same time, providing flexible and ready access to customers Air Force–wide. Major components of an architecture include, but are not limited to, applications, software, hardware and support software, data, networks, people, and organizations. Figure 4 shows an example of an architecture as presented in the Air Force Data Management Architecture (DMA).[3] The DMA describes the AFCDD as a hierarchy of data dictionaries, all compatible with one another, but used at different levels.

Many benefits can be gained through the use of a solid architecture. Three benefits of note include the improved use of stored information and improved decision making, the provision for an orderly phaseout of obsolete or ineffective systems, and the assurance of greater quality and reliability.



Legend:
AFAFC—Air Force Accounting and Finance Center
AFCDD—Air Force Corporate Data Dictionary
AFLC—Air Force Logistics Command
MPC—Military Personnel Center
MAC—Military Airlift Command

Source: Air Force Pamphlet 700-50, *Air Force Communications-Computer Systems Architecture Data Management Architecture*, vol. 3, 10 April 1990.

Figure 4. Data Dictionary Architecture

Developing a Data Dictionary

Development of a data dictionary is one of the most important and crucial steps in support of the Air Force Data Management and Standards Program. DOD Manual 8320.1M states that "it is the fundamental tool in support of data administration."[4] After selecting an architecture, planners and designers may use many approaches to develop the required system. In *Developing a Data Dictionary System*, J. Van Duyn outlines six major steps that will respond to the need of the enterprise for having complete control over its data resources.[5]

The first step establishes data-naming and definition standards and conventions. It includes standardization of data elements, data items, and data definition-naming conventions and program-naming conventions at a minimum.

The second step establishes a listing of standard abbreviations and acronyms. It includes standardization of abbreviations and acronyms and establishes the

rules to define a term the first time it is used in programs, documentation, and reports.

The third step identifies and defines "base data" data elements. It relates to identification and definition of "division or department" data of the enterprise.

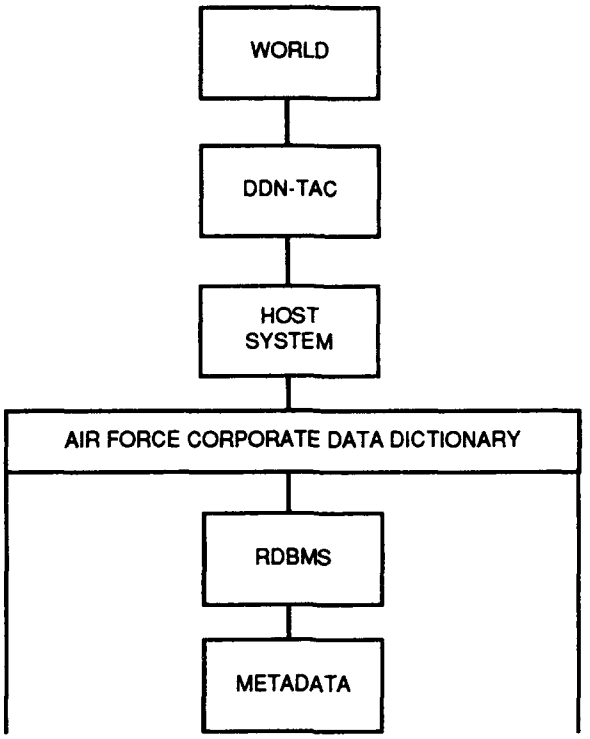
The fourth step identifies and defines code types of data elements. The fifth step identifies, defines, and standardizes input, output, update, and validation procedures. Step six identifies and defines the data characteristics.

As part of the standardization process, the DMA outlines several steps for standardizing data which are key in the development of a data dictionary.[6] The steps are similiar to the ones outlined above. The first step determines redundancies, groups identical data elements, and selects one of them as the data element (the others are related aliases). Next, planners determine standard groupings. This means combining the remaining data elements into groups that measure or record the same type of information. This process is often referred to as normalization. These groups provide candidates for creating generic elements. Another step assigns standard attributes and names generic elements. Planners then compare each data element to existing standards. If the Air Force can support that standard, it can create a generic element. The generic element can then be given both a long and a short name. After a generic element receives a name, it can give to a standard name and a mnemonic each data element in its group. The Air Force data administrator (DA) should discuss the recommended standards with representatives from the functional and technical areas before approval.

Current System Architecture

Initially in the process, as has been in the past, the AFCDD will be implemented at a central site. It will be used with a relational data base management system (RDBMS) and supported by the federal information processing standards (FIPS) software quality language.[7] The DMA describes the environment as having a data base, retrieval and analysis capability, management tools, and functional interfaces for users as shown in figure 5. The Defense Data Network (DDN) will provide the primary communications support between the system and remote users.

To make customer access to the dictionary easier and more economical, dictionary designers should migrate data to a distributed architecture consisting of replica data bases. For example, designers could improve customer access to the AFCDD by increasing communications capacity to and from the central site or by establishing replicated systems. Designers would locate replicated dictionary data bases at each regional processing center: one in the European theater, and one in the Pacific theater. Figure 6 shows what a distributed architecture might look like. It shows data generated and used centrally and at the local nodes. Also, it shows data shared between central and local nodes and data shared regionally.



Legend:
DDN-TAC—Defense Data Network, Tactical Air Command
RDBMS—Relational Data Base Management System

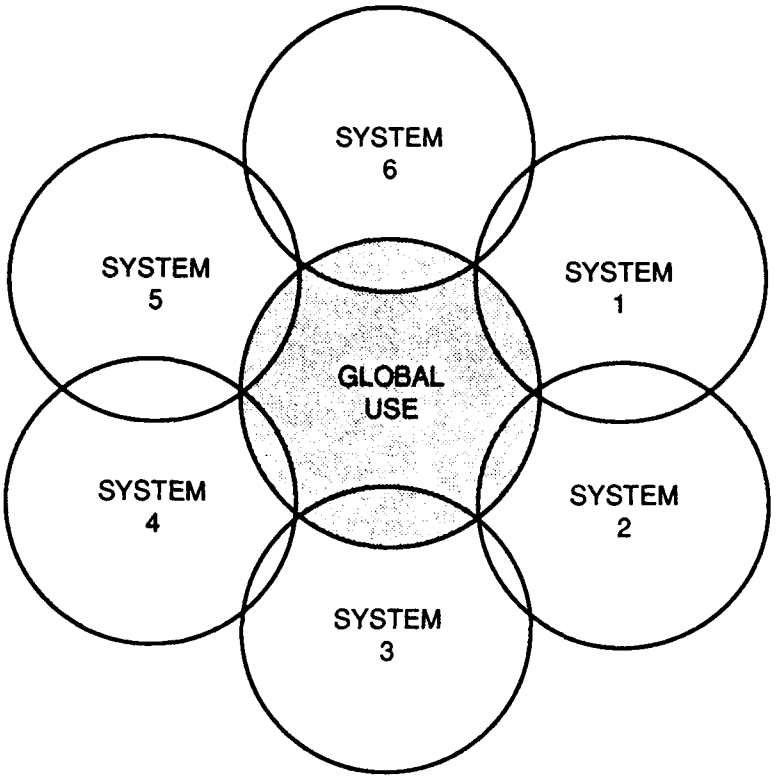
Source: Air Force Pamphlet 700-50, *Air Force Communications-Computer Systems Architecture Data Management Architecture*, vol. 3, 10 April 1990.

Figure 5. Data Dictionary Environment

Support Infrastructure

In addressing the support infrastructure dictionary, designers must assess the existing situation early on. They must make an inventory of the existing system and its data resources. The existing support structures, including DDN and allied support, must be in place. DDN provides communication between the system and the remote users, and allied support consists of personnel, equipment, and location.

The specific Air Force infrastructure referenced in this section is the emerging AFCDD, located at the Standard Systems Center. The information provided in the remainder of this section comes from a series of three documents maintained at the AFCDD program management office. Those documents consist of the AFCDD Concept of Operations, the Communications-Computer Systems Program Plan, and the AFCDD Security Plan.[8] SSC developed its data dictionary as the single source of standard data elements and codes for system software



Source: Data Analysis and Data Dictionary Briefing, 14 September 1988.

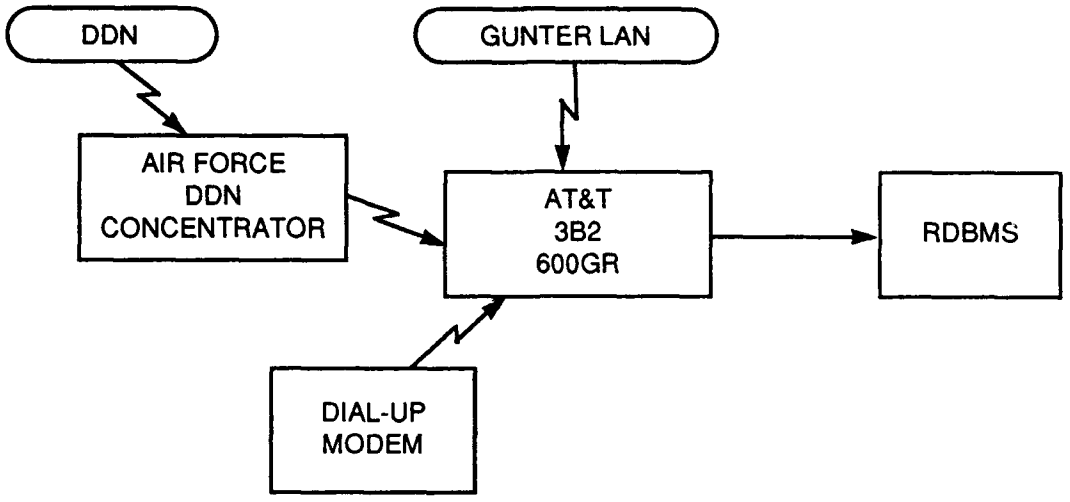
Figure 6. Distributed Dictionary Environment

developers and functional managers Air Force–wide. This dictionary serves as a passive data dictionary or repository of metadata. It is an on-line, menu-driven relational data base computer system designed for incremental implementation. Currently, SSC holds the development, implementation, and operational responsibility for the data dictionary.

Description

The AFCDD is a stand-alone computer system that operates on two hardware platforms: an Amperif relational data base machine and an AT&T 3B2 600GR computer. The AFCDD system consists of a set of data tables that resides on the RDB machine and is accessed through the AT&T 3B2 600GR front-end communications processor by way of the DDN, automatic-answer/automatic-dial modem, or the Gunter local area network (LAN). The data tables contain automated data system (ADS) identification and descriptive information, which will include (1) the assigned data system designator, (2) system code, (3) title, (4) acronym, (5) security attributes, (6) hardware equipment, (7) type of system, (8) responsible managers, (9) interfaces, (10) software and languages, and (11) associated costs. Standard data elements and codes and the unclassified

geographical locations of Air Force installations worldwide also reside in the data tables. The 3B2 600GR computer will store nonstatic project management documentation and perform front-end communications functions. Designers will use commercial off-the-shelf (COTS) software with the AFCDD. The AT&T 3B2 600GR relies on the UNIX operating system and Noah and FREEFORM interfacing software. The Air Force Corporate Data Dictionary is the only application that remains resident on the RDB. Figure 7 depicts the current system environment.



Legend:
DDN—Defense Data Network
LAN—Local Aera Network
RDBMS—Relational Data Base Management Systems

Source: Standard Systems Center/XPT 9001712-D1 Data Dictionary Briefing.

Figure 7. Current Data Dictionary Environment

Communications

Connectivity between the AT&T 3B2 600GR front-end communications processor and the RDB is the broadband LAN at Maxwell AFB, Gunter Annex. End users can access the AFCDD by connecting an intelligent terminal or through a dumb terminal that is connected to a host computer with DDN connectivity. By dialing into the DDN concentrator at Gunter, end users can reach the LAN and the AT&T 3B2 600GR. They can use an auto-answer or auto-dial modem as an alternative method of communication. Alternative communications paths available to the AFCDD include a direct connection from the DDN concentrator at Gunter to the AT&T 3B2 600GR and a direct connection between the AT&T 3B2 600GR and the RDB.

System Access

Passwords and user identifications enable users to access the system. The program manager approves registrants as authorized users and provides initial passwords and user identifications. Currently, SSC/Data Administration Division (XPSD) has to download files on floppy disks and mail them to requesting organizations. In the future, end users will probably download files to a floppy disk drive, a hard disk drive, or a printer.

System Availability and Security

The AFCDD operates 24 hours a day, seven days a week, except during routine maintenance periods. The program management office provides special instructions to schedule system downtime and to answer problem calls. The system will be taken off-line, however, during military contingencies or whenever it is not Air Force mission-essential.

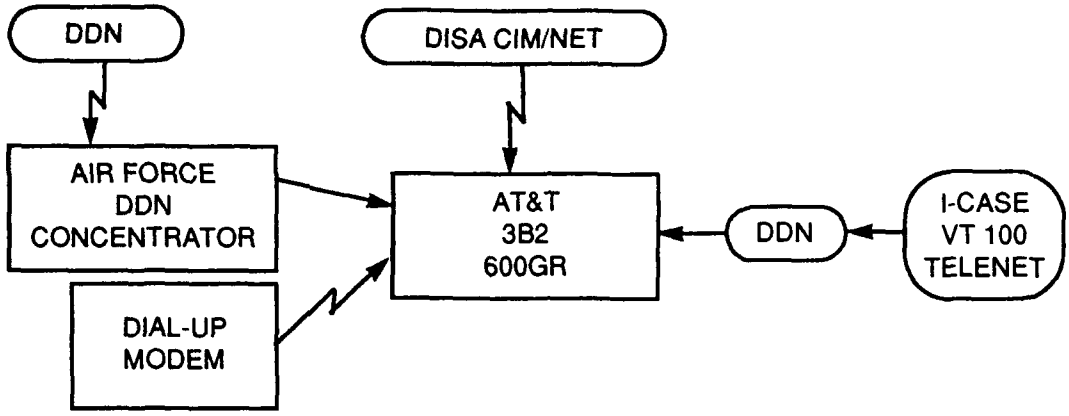
The AFCDD is an unclassified computer system. Information contained in the data bases is not classified, but it may be FOR OFFICIAL USE ONLY. Currently, dictionary designers have no plans to maintain or process classified information or information subject to the Privacy Act of 1974. Both the sensitivity and criticality of the AFCDD is CF5. Sensitivity relates to the information being processed, and critically identifies the mission the system supports and the degree to which the mission is dependent on the system. CF stands for "criticality factor." The number assigned describes the greatest effect on human life. The information contained in the system data bases does not directly affect approved DOD emergencies or war plans, does not directly jeopardize human health and well-being, and does not process sensitive unclassified data.

The Defense Data Repository System

While the Air Force has aggressively sought a corporate data dictionary system, the Department of Defense has plans to use a universal system called the Defense Data Repository System (DDRS). The DDRS is an open-systems data dictionary based on the Army Data Dictionary System (ADDS). The DDRS central repository will reside on an AT&T 3B2 600GR computer, using UNIX SV.4 (POSIX), "C," Ada, ORACLE RDBMS, ETIP, TCP-IP (X.25 and ETHERNET) located at the Kidwell Building in Vienna, Virginia. Data element naming conventions and the coordination and approval process are based on draft DOD 8320.1M.[9]

The DDRS is an open-systems data dictionary based on the WISDIM partitioning and query capabilities. Users will need to use a VT 100 terminal to access the DDRS through direct connect, dial up, or DDN. Designers will rely on I-CASE to define interface requirements for this system. They can do this by remote access, batch load, or through the DDRS extract interface. Remote access connectivity is over the DDN using telenet, where VT 100 terminal emulation is required. Batch load flat files are extracted from I-CASE into the

DDRS data base. The batch load files can be delivered by file transfer protocol over the DDN or on tape. As for DDRS extract, the DDRS acts as the data base server, and remote or local I-CASE clients will issue standard query language commands to the server. Connection to DDRS is established through ORACLE RDBMS SQLNET communication software. Figure 8 depicts the DDRS environment.



Legend:
CIM—Corporate Information Management
DDN—Defense Data Network
DISA—Defense Information System Agency
I-CASE—Integrated Computer-aided Software

Source: Defense Information System Agency for Corporate Information Management, May 1991.

Figure 8. Defense Data Repository System Architecture

Speaking at a recent I-CASE vendors' conference, Dan Lewis of the Defense Information System Agency for Corporate Information Management addressed the objectives, functions, and benefits of the DDRS. He identified the following as some of the objectives of the DDRS.

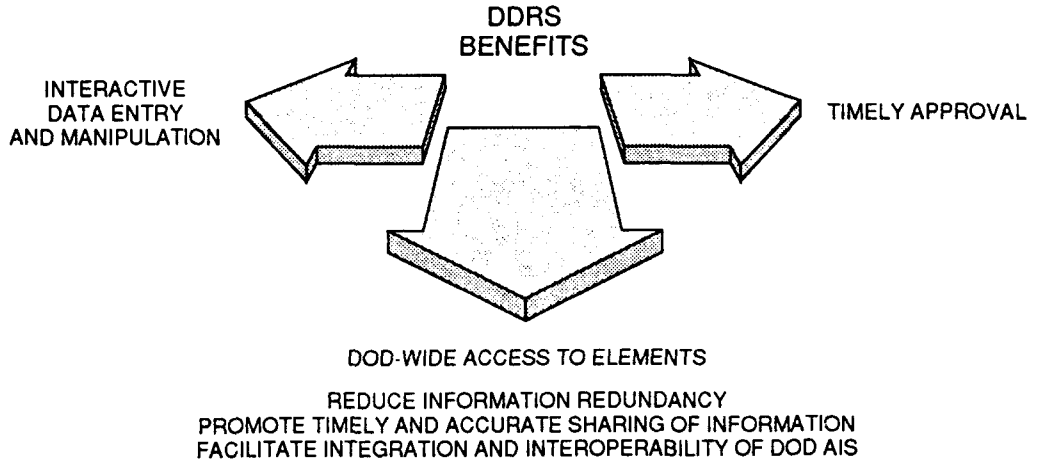
- Collect and store information about DOD standard data elements, including status information, as the element is tracked throughout its life cycle.
- Document organizations and processes using DOD standard elements.
- Provide DOD-wide access to on-line querying and reporting capabilities.

The DDRS will be a centrally controlled DOD-wide data repository to manage and store standard data elements, definitions, and associated metadata. Designers have developed it to support the DOD data administration program and procedures.

DDRS will support system developers and users, component data administrators (CDA), functional data administrators (FDA), and the DOD data administrator. System developers and users will employ the DDRS as a data retrieval tool for query of existing standard elements and users of those elements. Each DOD service and agency has a designated CDA. The data ad-

ministrators will use the repository system to review existing standard data elements and submit data elements developed under their purview for DOD approval. FDAs review data elements for functional consistency within a functional area such as logistics or personnel. They will use DDRS to exchange comments with the DOD DA during the scrutiny process. They may also identify new information requirements and use the DDRS to submit the elements for DOD approval. The DOD data administrator employs the DDRS to complete technical reviews of the elements submitted for approval. Technical reviews ensure that the candidate element complies with naming and structure conventions.

Figure 9 shows some of the benefits of DDRS. Because the DDRS is to be an automated, interactive system with automatic message capabilities, the approval of DOD standard elements will be completed in a most timely manner. Also, because it will operate in an interactive environment, accurate information will be shared quickly. Since the DDRS will maintain information about the systems and applications that use standard elements throughout DOD, the integration and interoperability of these systems is facilitated.



Legend:
DDRS—Defense Data Repository System
AIS—Automated Information System

Source: Briefing, Defense Data Repository System, Civic Center, Montgomery, Ala., March 1991.

Figure 9. Defense Data Repository System Benefits

Summary

This chapter has shown the significance of having a sound architecture in selecting, developing, and maintaining a data dictionary system. A sound architecture provides the framework necessary for decisions that have to be made today—informed decisions based on future vision.

The current AFCDD's architecture is a viable structure. Although their work is incomplete, developers plan to ensure the data dictionary will interface through I-CASE with the DDRS or through whatever system the Department of Defense selects. Enhancements may be necessary to meet additional or projected requirements.

Notes

1. Air Force Planning and Architectural Guidance, pt. 1, "From the SC," 1990, 2.
2. John E. McRoberts, "Establishing Control in a Data Dictionary Environment," Data Base Management (Livermore, Calif.: J. M. Auerbach, 1990), 2.
3. AFP 700-50, *Air Force Communications-Computer Systems Architecture Data Management Architecture*, vol. 3, 10 April 1990, 4.
4. DOD Manual 8320.1M, "DOD Data Administration Procedures Manual," draft, 1 November 1991, 4–5.
5. J. Van Duyn, *Developing a Data Dictionary System* (Englewood Cliffs, N.J.: Prentice-Hall, 1982), 56.
6. AFP 700-50, 8.
7. Federal Information Processing Standard (FIPS) Publication 127-1, "Standard Query Lanaguage," *Federal Register* 55, no. 23, 2 February 1990.
8. Air Force Corporate Data Dictionary, Concept of Operation, Communications-Computer Systems Program Plan, and Security Plan, May 1990, March 1990, and January 1990 respectively, Standard Systems Center, Plans and Program, Gunter AFB, Ala.
9. DOD Manual 8320.1M.

Chapter 4

The Transitioning Process

The Air Force Corporate Data Dictionary must support an orderly transition from the "stovepipe" information systems environment to a shared, standardized data base environment. Two of the key objectives listed in DOD Manual 8320.1M support this requirement.[1] One of the first objectives states that we must move from current directives governing data element and data code standardization to a DOD-wide data administration while preserving current investment in data systems, equipment, procedures, and trained personnel. Another key objective identifies mechanisms to interconnect diverse data management tools with DOD data administration tools, including an IRDS conformant data dictionary system that would preserve current investments and plans for using related data management tools.

Since the transitioning process occurs through a slow, long-term change and not overnight, the best term to describe this process is *evolutionary*. But for the sake of this chapter, the transitioning process will refer to an all-emcompassing change from the old to the new over a certain period of time. This applies to rules, standards, processes, and technological changes to meet the desired objectives. Some specific technological areas of change include hardware, networks, operating software and languages, applications software, data bases, user interfaces and procedures, and methodologies. A basic understanding of the transitioning process and of the means for accomplishing such a task is crucial to the success of a data dictionary.

General Areas of Transition

According to DOD Manual 8320.1M four major areas require transition: doctrine, organization, process, and technology.[2] With doctrine, we can no longer view data as front-end raw material or back-end finished products of a process; instead, we must view it as the "focus" by which we manage data as a corporate asset. In this way, planners and users focus on the data rather than on the process. Next, we must provide for an organizational transition.

Planners must also offer adequate resources to establish an organization of sufficient expertise and size to carry out the process. They must plan and schedule the transition carefully.

The process area refers to the administrative aspects of transitioning and includes standards, management, and training. Two of the most process-

significant transition areas are models and data element standards. Designers must offer a plan to move from the baseline to the new system.

The fourth key area of focus is technology. This aspect develops the baseline and perspective models, defines a technical transition path, and becomes a part of functional management and control boards. These boards will prioritize and control changes to the AFCDD requirements and capabilities.

Migration Strategy

As transitioning old data using new standardization procedures will be a major undertaking for several years to come, so will the process of transitioning from the old data dictionary to the new data dictionary. The old data dictionary and structure will need maintenance as long as they continue to serve the needs of existing systems or until a new system replaces them or until they are modernized and subjected to new procedures. When all existing standard data elements have been transitioned, the existing dictionary system will be decommissioned.

The migration strategy describes and prioritizes the capabilities needed to progress along an evolutionary path. The term *data migration* implies or suggests that one is moving or transferring data from existing files or data base structures to new data structures in a new subject data base as shown in figure 10. This migration includes an assessment of existing data structures or files. The object is to move away from files or data bases to individual applications and data that are highly redundant, then to proceed to an environment based on a foundation of nonredundancy and data which can be shared across applications.

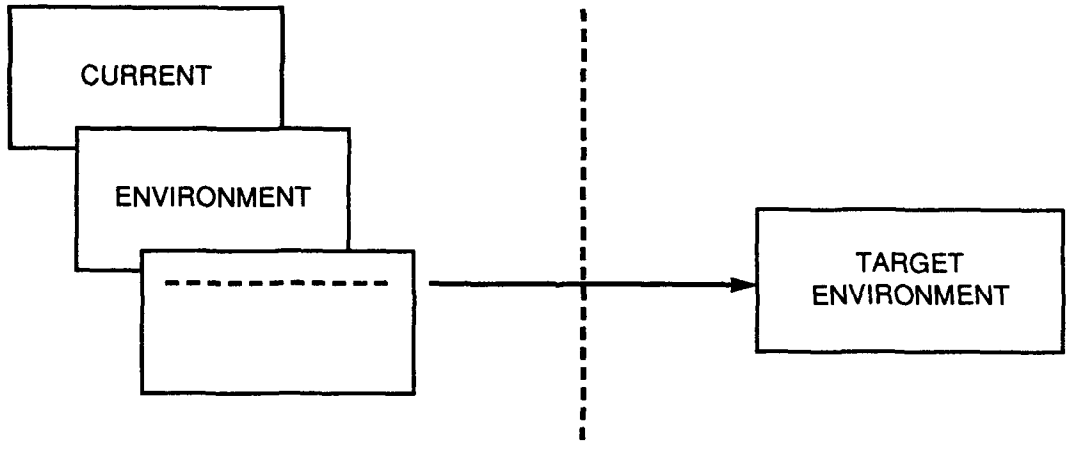


Figure 10. Migration Strategy

DOD Manual 8320.1M states that migration involves more than data movement or the transfer from old to new.[3] It also takes into account uploading, downloading, backloading, synchronous updating, parallel feeding, mirror image filing, periodic copying, refreshing, and data cloning. Some or all of these procedures may be involved in moving data from one system to another.

There are numerous approaches to migration. A recent system's architecture and integration seminar identified three such approaches.[4] First, the flash-cut approach moves data to the new system immediately, which may be risky. The second approach uses bridges to mix the old systems with the new ones. A bridge is a program which allows data to pass between old systems and new systems. The third approach places planning snapshots and scenarios along the transition path. This approach is more consistent with the evolutionary process, as it seeks to move in a phased fashion to the desired system.

Baselining of Existing Data Dictionaries

The Air Force must establish an inventory and baseline of all command data dictionaries and enhancement programs, along with the resources being expended. It also must suspend developments that are not consistent with Air Force strategic plans. The current AFCDD will serve as a focal point to collect and maintain a temporary data base of legacy systems and their data structures. As a start, this includes AFRs 700-19 and 700-20.[5]

As the AFCDD develops, it will house the required information; the need for the baseline will dissipate; and the system will be closed down. The temporary baseline data base will include, at a minimum, the identification of the migration system, data elements managed within each system, and data entities and key attributes.

System Selection

The system selected must meet Air Force and DOD-associated regulations and standards. Since a number of data dictionaries are already operational and do not adhere to these regulations, they must undergo the necessary changes to ensure compatibility with the AFCDD or Defense Data Repository System (DDRS). At a minimum, the dictionaries must be IRDS-compliant, capable of using electronic mail, and able to interface in a CASE environment and operate in an I-CASE environment. Rules and standards governing acquisition and collection of data must allow for varying data sources. The initial costs for using an IRDS dictionary will be the same as for converting to any new data base management system (DBMS) dictionary, but thereafter the savings should be significant.

Although many data dictionaries are commercially available, they do not fully meet Air Force or DOD requirements and standards at this time. Developers have made significant progress in this area and soon should have a data dictionary available that meets Air Force standards.

Plans for Consolidation and Communication

Major command plans must be developed to migrate command-unique data dictionaries to the standard AFCDD or DDRS software or to enhance command dictionaries to use the standard for data dictionary interchange. Passing information among separate systems requires that the information be converted. Changes made to one system can play havoc with other systems. Often, the passing of information between incompatible systems involves complex procedures. Communication in an integrated environment ensures that data flow is greatly simplified.

Data Interchange

Interface services play a major role in the transitioning process. These services access the IRDS contents by external functions, applications, data base, and people. They import and export data and metadata from external DBMSs; they also provide the means for CASE tool and programming language interface.[6]

The National Institute of Standards and Technology, Gaithersburg, Maryland, developed and approved a series of standards for federal use.[7] These standards port data and applications for access and reuse from one system to another. The Government Open Systems Interconnection Profile (GOSIP) permits communications across many different vendor-computing environments. With a common definition of data elements, digital data interchange and applications sharing can be facilitated. Figure 11 shows an example of interfaces with software and users of the data dictionary.

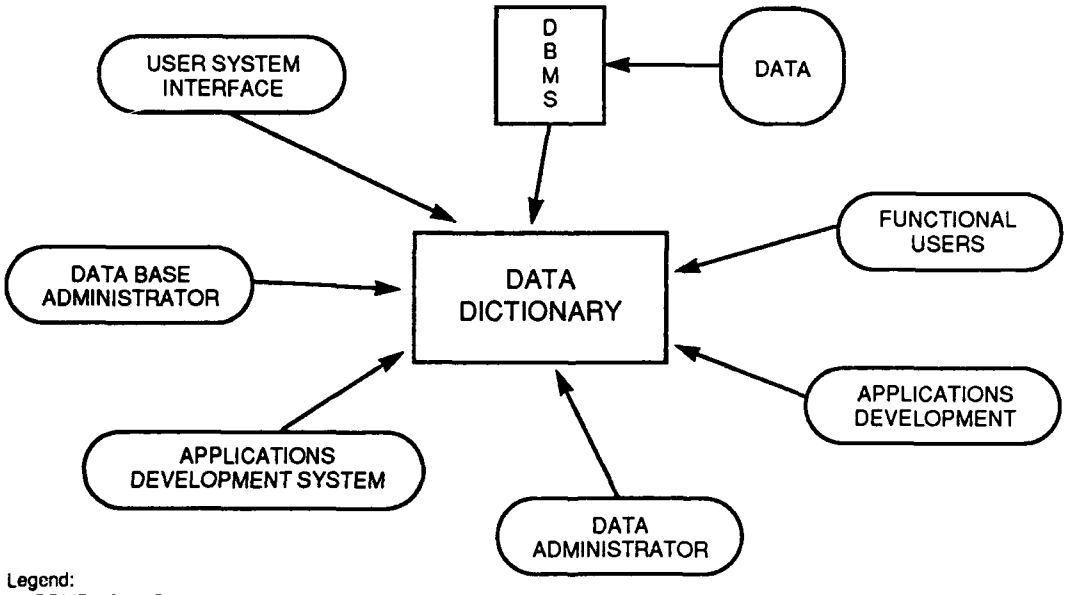


Figure 11. Interfaces with Software and Users

Implementation Road Map

The AFCDD seeks to support liaison among creators and users of Air Force data, system developers, information class proponents, data administrators, and AFCDD system administrators. AFR 4-29 positions this responsibility with Air Force heads of agencies, commanders, and the assistant secretary of the Air Force.[8] Consolidation and standardization of existing data elements throughout the Air Force will require time and consistent MAJCOM support. This should be done in concert with modernization, regionalization of existing systems, and development of new applications.

Implementation of the AFCDD will be done in phases. Some components of the environment have more importance than others in the different developmental stages. AFR 4-29 states that data management policies and procedures must be incorporated in the development and implementation phases of the AFCDD.[9] To achieve operational status, the program must provide implementation procedures to functional users and data processing personnel. Prior to each implementation, planners must develop and conduct initial training programs for system administrators, functional proponents, and users. These training requirements include both AFCDD rules and administrative system

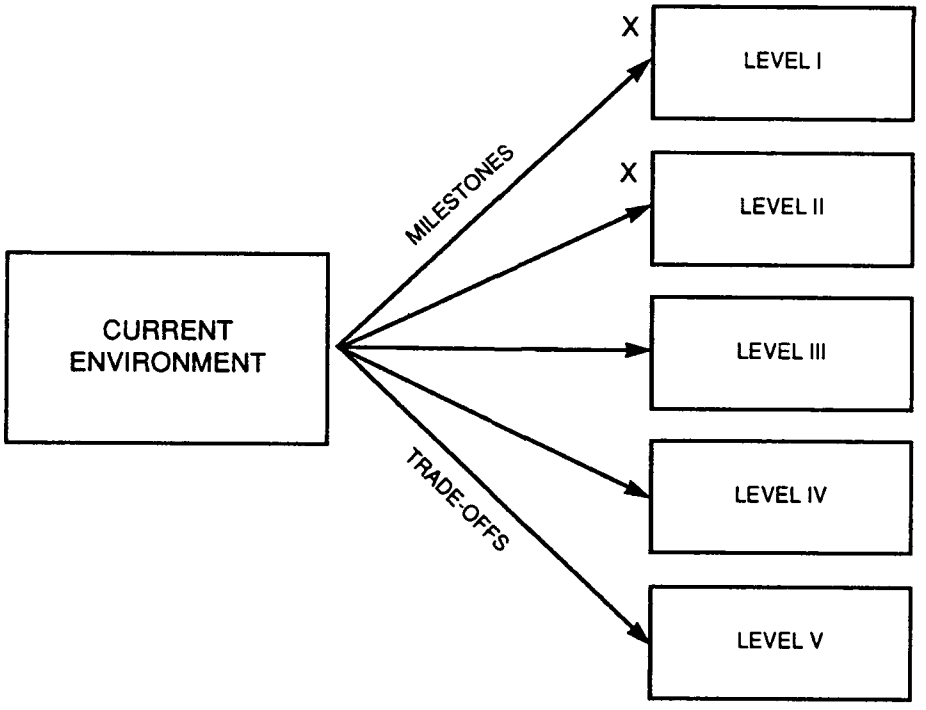


Figure 12. Different Levels to the Desired Environment

operating procedures. The AFCDD system should include an on-line training facility for unfamiliar users.

Levels to the Target Environment

Transitioning must take place in a stepwise fashion; however, several technically related steps must take place to support the transition from existing models, applications software, libraries, and dictionaries. An evolutionary path from the baseline to the desired environment must be established. Figure 12 shows this concept.

The evolutionary path describes the key target components comprising the desired environment. These components will evolve and mature with usage. As each component evolves, milestones are set. Table 1 shows a projected milestone chart.

Table 1

Projected Data Dictionary Milestone Chart

Level 0 (2–3 months)	*Level I (3–6 months)*	*Levels II and III (6–10 months)*	*Levels IV and V (6+ months)*
Long-term planning	Data and process modeling	System implementation	Population of data base
Defining goals	Test specifications and procedures	Evaluation of the system in operation	Enhancements
Cost justification	Establish baseline	Begin data base population	Migration of Air Force data dictionaries
Identify and define requirements	Establish control		
	Standardize equipment, manpower, and data		

Two designers, Edward Yourdon and Larry Constantine, couldn't have put it better when they remarked, "We rarely can have our cake and eat it, too."[10] Along this same evolutionary path, they indicate that we will encounter trade-offs which include risks, costs, and benefits. As we increase one parameter, we almost always decrease another. If we opt for more efficiency, we frequently sacrifice ease of maintenance. Similarly, we usually gain execution speed at the expense of memory storage. We must be aware of what we trade off and select the balance that best reflects our overall goals.

The following levels might be used as a basis in the development and implementation stages of the data dictionary. These levels reflect some similarity to the program management directive (PMD) for the current AFCDD, but they should be used as part of the new PMD revision process.[11] (See table 1.)

- Level 1 establishes the baseline and offers a place for the development of the current AFCDD. This development should take from two to three months.
- Level 2 initiates implementation of the baseline and population of the dictionary with existing AFR 700-20 elements. Implementation should require from three to six months.
- Level 3 populates the dictionary with standard data elements from the MAJCOMs and functional areas as they are approved. This continuous process has no specified time frame for completion.
- Level 4 distills the enhancement of the dictionary and allows users to identify the areas of their dissatisfaction.

Guiding and Monitoring the Implementation Process

To guide the building of the long-term target environment and day-to-day decision making, users must develop certain policies for guidance and monitoring of the implementation process. They must have a way to gauge compliance and adherence to the standards. A retirement policy for the existing infrastructure must be in place. Detailed procedures for transitioning from legacy dictionaries also must exist.

The IRDS states that a suite of automated validation tests for implementation is under development.[12] Waivers will be granted only when compliance with a standard would adversely affect the completion of the mission of an operator of a federal computer system, or when compliance causes a major adverse financial impact on the operator that is not offset by government savings.

Implementation Checklist

In supporting the existing infrastructure during migration, planners should develop a checklist. J. Van Duyn, one such planner, believes a number of things should be considered in the checklist, which includes, at a minimum, the following:[13]

- validation rules for correctness
- conformity to the rules of usage
- frequency of use of the applications processed
- accessing paths that the applications follow
- test data to be produced
- permissible ranges of values for data elements
- detailed descriptions of data structures
- consistency checks to ensure that the new data is complete and correctly formatted
- access control

Van Duyn maintains that validation rules help to ensure quality control. These rules allow the data administrators to check the adequacy and correctness of the system's validation procedures. In addition, conformity to the rules of usage ensures that programs use proper and allowed data. How often particular applications are used and the accessing paths they follow indicate how the applications use the system. This information then may be analyzed by the data administrator for more detailed predictions of machine resources required.[14]

Summary

The transitioning process is the most important subject in this study. Regardless of what target environment planners desire, they gain nothing if the means and plans for moving to it are nonexistent. Designers must develop the all-important formal transition plan. The transition plan should develop and document a phased approach as we move from the old environment to the new one. Designers often refer to this scenario as a migration strategy. Capabilities must be described or defined and prioritized. Doctrine, organization, process, and technology were four general areas of transitioning discussed. This writer placed a great deal of emphasis on the technology area due to its major role in the process.

Once designers have developed a data dictionary, they can migrate data more quickly because they would have considered a number of factors. One such factor is the approach the implementation of the system will take. Certain tools for guiding and monitoring the implementation process such as policies and checklists may be extremely beneficial. Also, the DOD and the Air Force are quickly changing to standards that promote hardware independence and software portability, integration, and interoperability. This change will help the Air Force to achieve a dictionary environment that is more responsive to user needs and to increase productivity, eliminate redundancy, and reduce development and maintenance costs.

Notes

1. DOD Manual 8320.1M, "DOD Data Administration Procedures Manual," draft, 1 November 1991, 1-3.
2. Ibid., 12-1.
3. Ibid., 12-8.
4. "System's Architecture and Integration Seminar," US Professional Development Institute, Orlando, Fla., January 1992.
5. AFR 700-19, *Computer Systems Authorization Directory (CSAD) FOUO (On-Line)*, February 1990; and AFR 700-20, *Air Force Data Dictionary (On-Line)*, vol. 1, February 1990.
6. Federal Information Processing Standard (FIPS) Publication (Pub) 156, "Information Resources Dictionary System (IRDS)," draft, 19 October 1988, attachment to 30 March 1989 notice, 7.

7. The US Department of Commerce, National Institute of Standards and Technology, Gaithersburg, Md.

8. AFR 4-29, *Air Force Data Management and Standards Program*, 23 April 1990, 4.

9. Ibid., 5.

10. Edward Yourdon and Larry L. Constantine, *Structure Design* (Englewood Cliffs, N.J.: Prentice-Hall, 1990), 14.

11. PMD 9261(3)/PE 38610, *Program Management Directive for the Air Force Corporate Data Dictionary*, 21 March 1991, 2.

12. FIPS Pub 156, 8.

13. J. Van Duyn, *Developing a Data Dictionary System* (Englewood Cliffs, N.J.: Prentice-Hall, 1982), 107.

14. Ibid., 47.

Chapter 5

Summary, Recommendations, and Conclusion

It is a well-known fact that even the best designed and most comprehensively tested computer system is subject to changes almost as soon as it is implemented. Errors, dissatisfied users, inefficiencies, and technological advances are some of the reasons for the changes. By the time this study is published, a number of changes will perhaps be in progress at the strategic planning level. This chapter addresses specific concerns and makes recommendations on present data dictionary issues as we move into the twenty-first century.

Summary

To quickly recap this study, chapter 1 has shown that the Air Force has a major investment in not only one but several data dictionaries which do not meet the need of the total force. The search for a corporate data dictionary has been a continuous process. The requirements for a corporate system will remain in a flux for the near future. The true requirements, the needed system, will continue to evolve.

Chapter 2 dealt with requirements, policy, and guidance. It focused mainly on data modeling. We found that the problem today is not so much in the specific data base design of the system but with the preceding work—that of designing the models. Data modeling is the starting point for understanding complex interrelationships among business processes, information requirements, and organizational elements for the entire Air Force. According to James Martin, whenever data is consolidated in data base systems, data modeling holds the key to success.[1]

Chapter 3 addressed the structure of the data dictionary and the importance of having a stable architecture. The architecture is the framework that portrays relationships among all data and process components identified in models. The ideal structure of a data dictionary will support metadata used for information and data architectures, corporate data control, and applications implementation. Trends in technology, economics, standards, and regulations are just a few of the issues that affect the current architecture. The most recent efforts by the Department of Defense to develop a corporate data dictionary for all services is encouraging. A recent study conducted by the Joint Data Administration Task

Force to select a DOD standard dictionary-repository system was also discussed in this chapter. Once operational, this system, referred to as the Defense Data Repository System, is projected to reduce information redundancy, promote timely and accurate sharing of information, and facilitate integration and interoperability of DOD-automated information systems.

Finally, chapter 4 paved the way for understanding data migration. A number of dilemmas surround the movement of data from one environment to another. In terms of migration, environments are classified from static—that is, requiring no changes—to those requiring possible restructuring, reengineering, or "scrapping" and rebuilding.

Recommendations

The information presented in this study has made it clear that planners must address numerous concerns and issues. Though not always easily identified, there are solutions or improvements for many problems. This section addresses major issues and concerns and offers recommendations. The order in which this section addresses them does not reflect their importance or their degree of severity.

Issue 1: Elimination of Multiple Dictionaries

The Air Force must reduce the number of its independent data dictionary systems. The needs of corporate information management prescribe that information systems communicate and pass data throughout the Air Force and other services. This cannot be accomplished with several independent, incompatible data dictionary systems. AFR 4-29 states that an Air Force corporate data dictionary will provide an orderly migration of information from the existing stovepipe environment to the objective shared data base environment.[2] Having a corporate data dictionary will save the Air Force time and money and meet Air Force goals. Current DOD cutbacks will take away the sparse funding available for continuous operation and maintenance of separate systems. Something needs to be done immediately. Just recently, the 7th Communications Group employed an IBM Corporation repository-compatible extensible data dictionary to switch its data bases.[3] Group personnel found that the data dictionary was incompatible with their environment, but they were still able to transfer their data bases with it. Along this same path and of major concern is the transferring of the AFCDD from the Standard Systems Center Secretary of the Air Force (SAF)/AAID (Information Management Division). The AFCDD is currently located at SSC, but plans are underway to relocate it to SAF/AAID under the control of the Air Force data administrator. The relocation will involve quite an undertaking for an office that would also provide the policy and guidance for Air Force automated information systems. Currently, SSC has

more than 300 customers who rely on the AFCDD. The Air Force will maintain this data dictionary to support SSC's mission and goals regardless of where "the" AFCDD resides, and a new system at SAF/AAID will add yet another Air Force data dictionary.

Issue 2: Standards for Defining Data Elements

DOD policy mandates the use of standard data elements in all DOD information systems within any DOD component. Although the National Institute of Standards and Technology has issued guidelines on data-naming conventions, I have found that most of the documentation for defining data elements is in draft form and not readily available. Using a dictionary and sharing data resources require standard specifications for such items as names (length, format, and abbreviations), attributes (picture size, range, and data class), and utilization assignments. In a recent *Defense News* interview with Paul A. Strassmann, the magazine asked what percentage of DOD's total data definitions had been developed thus far.[4] His response was, "We don't know." He stated there are hundreds of data dictionaries in various degrees of completion and integrity and offered a guess of between 5 percent to 10 percent completion. Planners must avail to system users a data element standardization approval process. To rectify this situation, they must finalize key standards and regulations and make them available immediately.

Issue 3: Management and Training

Regardless of its efficiency and sophistication, a data dictionary does not eliminate the need for good management, clearly defined objectives, and qualified data processing personnel. Success of an Air Force data dictionary requires active support by data administrators and managers at every level throughout the Air Force. The effectiveness and efficiency with which Air Force corporate and functional managers carry out their mission responsibilities depend heavily on their ability to manage the data needed by and resulting from the business processes they control. Strassmann contends that management is the business value of computers.[5] A well-structured and properly implemented and maintained data dictionary system eases the tasks of users and provides system analysts, designers, and data administrators with an effective tool to control and manage corporate data resources. As for training, aggressive education and training are necessary to ensure success of the data dictionary. It will not help the support of the mission if adequate training is not available to take full advantage of its capabilities. A lack of training could cause a misuse of tools, which could lead to many future support problems with the associated mission degradation. Planners must develop and conduct training programs for all systems administrators, functional proponents, and users.

Issue 4: Migration of MAJCOM Data Dictionaries

Planners have expressed much concern with the migration process. According to the draft of DOD Manual 8320.1M, these concerns deal with the scope and quality of data to "migrate," technical constraints, the depth of historical data required for migration, migration from old to new data creation routine, technical problems of data concurrency and integrity during transition, and administrative and operational problems during transition.[6] Planners must address these questions: Which files and data elements should we retain, and which ones do we get rid of? Can we programmatically or mechanically convert data from existing files to the new data base? Can we transfer the historical or archived data to the new data base? How and when can we eventually sever the umbilical cord from the old data bases?

The migration will affect the most developed data dictionaries more because of the amount of time and dollars invested. Still, this will not be an easy process. The key advantages emanate from the use of a single software, reduced operation, and less maintenance. Planners must design future systems and redesign less-developed systems for ease of migration. Also, the emerging I-CASE purportedly holds the key to the migration of legacy systems. Planners believe I-CASE can completely evolve systems, thereby constantly improving their design and regenerating code.

Issue 5: Applying Proper Security Measures

Users must maintain the proper level of security for the data required and produced by the systems. The Air Force has not classified the current data dictionary; however, if we move to a distributed or joint environment, security becomes critical. Some of the DOD systems may be unclassified individually; however, in aggregate, these systems may process classified information which will require some type of multilevel security. Most of tne Air Force architectures show migration towards standardized multiuser, multiapplications, multilevel security, openly interconnected communications-computer systems environments, *but* the absence of effective safeguards makes these systems vulnerable to exploitation and disruption.[7] In effect, the absence of standards will detrimentally influence the flow of data and will negatively impact data security and control measures. A student at a recent Executive Forum on Communications-Computer Systems seminar indicated that viruses, increased personnel usage, and connection to nationwide nets pose a constant threat to computer security.[8] Security compromise, virus infection, and copyright violations are never-ending threats. In *Mind Children*, Hans Moravec argues that today's computer systems resemble bodies with skins but without immune systems—walled cities without police; that is, they can deflect some external attacks but are defenseless once an intruder has entered.[9] Data security and control procedures are currently being developed as part of DOD Manual 8320.1M.[10] These procedures must be in place before the data dictionary system becomes operational. Additional security guidance may be found in AFR 205-16, *Computer Security Policy (FOUO).*[11]

Conclusion

When the uncertainty has ended, the central Air Force Corporate Data Dictionary will evolve and interactively support the various Headquarters Air Force–managed components of a corporate data information resource repository regardless of the location or technological platform. The dictionary will be more than an automated documentation tool; it will serve as a design aid, control and administration tool, widely accessed (meta) data base, and a means of communication. Planners envision standard data elements, metadata, and codes; a programmatic data base; a software reuse library; a specifications data base; and information and data models. The data dictionary will support program managers, system managers, data administrators, software designers, developers and testers, design agencies, MAJCOM and functional area planners, and Air Force Headquarters staff.

This study has argued for the essentiality of a central corporate data dictionary system. Is the Air Force prepared to correct its Tower of Babel–like situation? Have we indeed developed a system to meet the needs of the corporate Air Force or DOD? With the information and recommendations provided, we should be off to a good start.

Notes

1. James Martin, *Information Engineering, Book I* (Englewood Cliffs, N.J.: Prentice Hall, 1991), 61.
2. AFR 4-29, *Air Force Data Management and Standards Program*, 23 April 1990, 6.
3. "AF Uses Data Dictionary to Manage Database Transfer," *Government Computer News*, 2 March 1992, 46.
4. "One on One Interview with Paul Strassmann, Director of Defense Information," *Defense News*, 17 February 1992, 54.
5. Paul A. Strassmann, *The Business Value of Computers* (New Canaan, Conn.: The Information Press, 1990), 511.
6. DOD Manual 8320.1M, "DOD Data Administration Procedures Manual," draft, 1 November 1991, 11-2.
7. Ibid., 12-7.
8. Executive Forum on Communications-Computer Systems, Student Issues and Concerns, Class 92A (Maxwell AFB, Ala.: Ira C. Eaker Center for Professional Development, 13 November 1991).
9. Hans Moravec, *Mind Children: The Future of Robot and Human Intelligence* (Cambridge, Mass.: Harvard University Press, 1988), 129.
10. DOD Manual 8320.1M, C-3.
11. AFR 205-16, *Computer Security Policy (FOUO)*, 28 April 1989.

Glossary

AAID	Information Management Division
ADDS	Army Data Dictionary System
ADS	Automated Data System
AFCDD	Air Force Corporate Data Dictionary
AFCSIO	Air Force Communications-Computer System Integration Office
AFDD	Air Force Data Dictionary
AFLC	Air Force Logistics Command
CASE	Computer-aided Software Engineering
CDA	Component Data Administrators
CD/D	Command Data Dictionary
CIM	Corporate Information Management
COTS	Commercial-off-the-Shelf(Software)
DBMS	Data Base Management System
DDN	Defense Data Network
DDRS	Defense Data Repository System
DMA	Data Management Architecture
FDA	Functional Data Administrators
FIPS	Federal Information Processing Standards
I-CASE	Integrated-Computer-aided Software Engineering
IRDS	Information Resources Dictionary System
ITPB	Information Technology Policy Board
LOGDRMS	Logistics Data Resource Management System
MIDAS	MAC Integrated Data Administration System
PM	Program Manager
PMD	Program Management Directive
RDB	Relational Data Base
RDBMS	Relational Data Base Management System
SC	Software Management Division
SOA	Special Operating Agency

SQL	**Software Quality Language** **Standard Query Language**
TIC	**Technology Information Center**
WISDIM	**War-fighting and Intelligence System Dictionary for Information Management**

We welcome your comments on this research report or opinions on the subject matter. Mail them to: CADRE/RI, 401 Chennault Circle, Maxwell AFB AL 36112-6428.

Correcting the Tower of Babel

A Tool for Data Standardization and Integration

Daniels