NISTIR 88-3896

Proceedings of the Federal Information Processing Standards (FIPS) Workshop on Information Resource Dictionary System (IRDS) Applications

Alan Goldfine, Editor

U.S. DEPARTMENT OF COMMERCE
National Institute of Standards and Technology
(Formerly National Bureau of Standards)
National Computer and Telecommunications Laboratory
Gaithersburg, MD 20899

December 1988

NISTIR 88-3896

Proceedings of the Federal Information Processing Standards (FIPS) Workshop on Information Resource Dictionary System (IRDS) Applications

Alan Goldfine, Editor

U.S. DEPARTMENT OF COMMERCE
National Institute of Standards and Technology
(Formerly National Bureau of Standards)
National Computer and Telecommunications Laboratory
Gaithersburg, MD 20899

December 1988

National Bureau of Standards became the National Institute of Standards and Technology on August 23, 1988, when the Omnibus Trade and Competitiveness Act was signed. NIST retains all NBS functions. Its new programs will encourage improved use of technology by U.S. industry.

U.S. DEPARTMENT OF COMMERCE
C. William Verity, Secretary
NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
Ernest Ambler, Director

PROCEEDINGS OF THE FEDERAL INFORMATION PROCESSING STANDARDS (FIPS) WORKSHOP ON INFORMATION RESOURCE DICTIONARY SYSTEM (IRDS) APPLICATIONS

Alan Goldfine, Editor

This report consists of the user presentations at a workshop on applications of the Information Resource Dictionary System (IRDS), held at the National Bureau of Standards (now the National Institute of Standards and Technology) on March 24-25, 1988. Representatives of twenty Federal Government agencies discussed current and planned applications of the IRDS at their respective organizations.

Key words: data administration; data dictionary; data management; Information Resource Dictionary System; information resource management; IRDS.

TABLE OF CONTENTS

INTRODUCTION

On March 24-25, 1988, the Institute for Computer Sciences and Technology (ICST) of the National Bureau of Standards (NBS) hosted a workshop on applications of the Information Resource Dictionary System (IRDS). The workshop was attended by 32 people, representing 20 Federal Government agencies. Several agencies were represented by contractor personnel.

The workshop began with an overview, presented by Dr. Henry C. Lefkovits of AOG Systems and David T. Carpenter of Pansophic Software, of the status of the IRDS as an emerging American National Standard and Federal Information Processing Standard (FIPS) [1]. This overview is outside the scope of this report, but much of the material can be found in [2]. The workshop also included a presentation, by Judith Newton of ICST-NBS, on data entity naming conventions within the framework of the IRDS [3], and general guidance, from Dr. Margaret Law of ICST-NBS, on the application of the IRDS to support system development [4]. This report includes abstracts of the latter two talks.

A special feature of the workshop was a series of short presentations, by sixteen of the attendees, on the current and planned application of the IRDS at their respective agencies. These presentations, many of which amounted to brief case-studies in data administration and information resource management, revealed that the agencies varied widely in terms of data administration philosophy, areas of IRDS use, and sophistication in the application of dictionary software. Several of the agencies are actively developing their own IRDSs, others are planning to acquire commercial, off-the-shelf IRDS packages when such software becomes available. A common thread among many of the agencies was their use of the ICST-NBS IRDS Command Language Prototype [5].

The entire workshop was audio recorded. The user talks and related question-and-answer periods were transcribed, and the draft transcripts were sent to the respective speakers for revision. The bulk of this report consists of these revised transcripts.

The workshop was the eighth in a series of FIPS IRDS workshops that dates back to 1982. The seven previous workshops dealt with an IRDS whose specifications were still under development. The primary interest then was to discuss Federal agency requirements, and to consider and criticize drafts of the evolving specifications to ensure that the IRDS would support the agency requirements. No written proceedings were produced. This current workshop, on the other hand, focused on the application, rather than the development, of the IRDS. Therefore, the subject matter is of wide interest, and these proceedings are being published.

Readers wishing a general introduction to the IRDS are referred to A Technical Overview of the IRDS [2].

EDITORIAL NOTES

In August, 1988, the National Bureau of Standards became the National Institute of Standards and Technology (NIST), and the Institute for Computer Sciences and Technology was renamed the National Computer and Telecommunications Laboratory (NCTL). As the record of a prior event, this document uses the earlier names.

In the course of the workshop presentations, reference was occasionally made to specific, commercially available products. Identification of these commercial products does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the product identified is the best available for the stated purpose.

GENERAL REFERENCES

[1] ANSI, American National Standard X3.138-1989, Information Resource Dictionary System, American National Standards Institute, New York, 1989.

[2] Goldfine, A. H. and Konig, P. A., A Technical Overview of the Information Resource Dictionary System (Second Edition), NBSIR 88-3700, National Bureau of Standards, Gaithersburg, MD, January, 1988.

[3] Newton, J. J., Guide on Data Entity Naming Conventions, NBS Special Publication 500-149, National Bureau of Standards, Gaithersburg, MD, October, 1987.

[4] Law, M. H., Guide to Information Resource Dictionary System Applications: General Concepts and Strategic Systems Planning, NBS Special Publication 500-152, National Bureau of Standards, Gaithersburg, MD, April, 1988.

[5] Goldfine, A. H. and Kirkendall, T., The ICST-NBS Information Resource Dictionary System Command Language Prototype, NBSIR 88-3830, National Bureau of Standards, Gaithersburg, MD, August, 1988.

DATA ENTITY NAMING CONVENTIONS

Speaker
Judith Newton
National Bureau of Standards
Institute for Computer Sciences and Technology

Naming conventions are guidelines for the format and content of data entity names, and are enforced by the organization's data administrator. They help to establish consistency of data throughout the organization. This results in greater efficiency through reduced data handling as the number of discrete data elements is reduced, and a reduction in confusion among both staff and management, as communication is enhanced. Guidance for developing and applying naming conventions is found in Guide on Data Entity Naming Conventions [2].

At first glance, data entity names may seem no different from natural language nouns. But they differ from nouns in the same way programming languages differ from natural languages: by the constraints imposed upon them by hardware, software, and human users, and by the possibility for the expression of the organization of the data itself.

Data entity names can reflect the organization of the data both logically, through prime words, and associatively, through class words. Prime words represent the logical groupings of data, such as all information which describes the concept **employee**; class words describe the basic nature of a class of data, such as **name**, **code**, or **date**. Data elements, one type of entity, may need a set of class words to fully describe all elements, while other entities such as **file** or **record** may need only one. Modifiers, which establish uniqueness of the data entity name, are the third name component.

While there may be many rules to be established for a set of naming conventions, there are a few guiding principles to follow while writing those rules:

Clarity - names are as clear as possible to a casual user.

Brevity within uniqueness - names are short while still maintaining uniqueness within the database.

Conformance to rules of syntax - each name is in the proper format. If there are too many names which cannot be made to fit the naming conventions, the rules may be too rigorous.

Context-freedom - each name is free of the physical context in which the data entity is implemented.

The IRDS provides a framework for establishing the structure of the names of each entity and the names' relationships to each other, i.e., the metanaming structure. There are three types of names for each entity: access name, descriptive name, and alternate name.

The access and descriptive names are functionally identical, but by providing two names, the IRDS allows them to share the burdens of the guiding principles of clarity and brevity. The access name may be terse, with abbreviations and acronyms but no connectors allowed (for example, EMPLOYEE-NAME), while the descriptive name allows for a longer and more discursive style (NAME OF EMPLOYEE). A user familiar with the database may want to use the access name for retrievals, while a more casual user would prefer the descriptive name. The alternate name may encompass any number of contingencies, such as physical name(s), report header name, and form input name. The majority of this discussion about names is concerned with access name grammar and usage.

The content component of naming grammar has been discussed above; the other component is format. Establishing format rules completes the process by which naming consistency is achieved. For instance, if the prime word is always the first word in the name and the class word last, there is no ambiguity in their identification. Searching by logical group (prime word) or basic nature (class word) is greatly simplified. See Figure 1 for examples of this naming scheme.

Application of naming conventions assists the data administrator in the analysis of data by (for instance) facilitating identification of coupled data elements and their decomposition into atomic data elements; and restructuring data names in which data is mixed in with metadata.

A hierarchy of data elements can be developed based on class words (Figure 2). A "kernel" of class words can be used to form a set of standard or generic elements. These determinate elements consist of a class word and modifier combination. Full data elements, called application elements, can then be formed with the addition of a prime word and any extra modifiers as needed. For instance, an application element EMPLOYEE-BIRTH-STATE-NAME is formed of the kernel class word NAME, which is contained in the generic element STATE-NAME; the prime word EMPLOYEE; and the modifier BIRTH.

Descriptive names are derived from access names by casting the access names into natural language grammar and adding connectors as needed. It is important to retain the prime and class words. For instance, EMPLOYEE-BIRTH-STATE-NAME becomes NAME OF BIRTH STATE OF EMPLOYEE.

Like most design activities, the effort expended in advance of the application of data entity naming conventions will pay off over the life of the enterprise.

PRIME WORDS - LOGICAL GROUPING

CLASS WORDS - BASIC NATURE

	EMPLOYEE	PURCHASE	ORDNANCE ....
AMOUNT	EMPLOYEE-SALARY -AMOUNT		
CODE	EMPLOYEE-STATUS -CODE		ORDNANCE-PHASE -CODE
COUNT		PURCHASE-ORD-MONTHLY-COUNT	
DATE	EMPLOYEE-START-DATE	PURCHASE-INIT -DATE	ORDNANCE-PHASE -DATE
NUMBER		PURCHASE-ORD -NUMBER	. .
NAME	EMPLOYEE-NAME		
PERCENT			
. . . .			

Figure 1

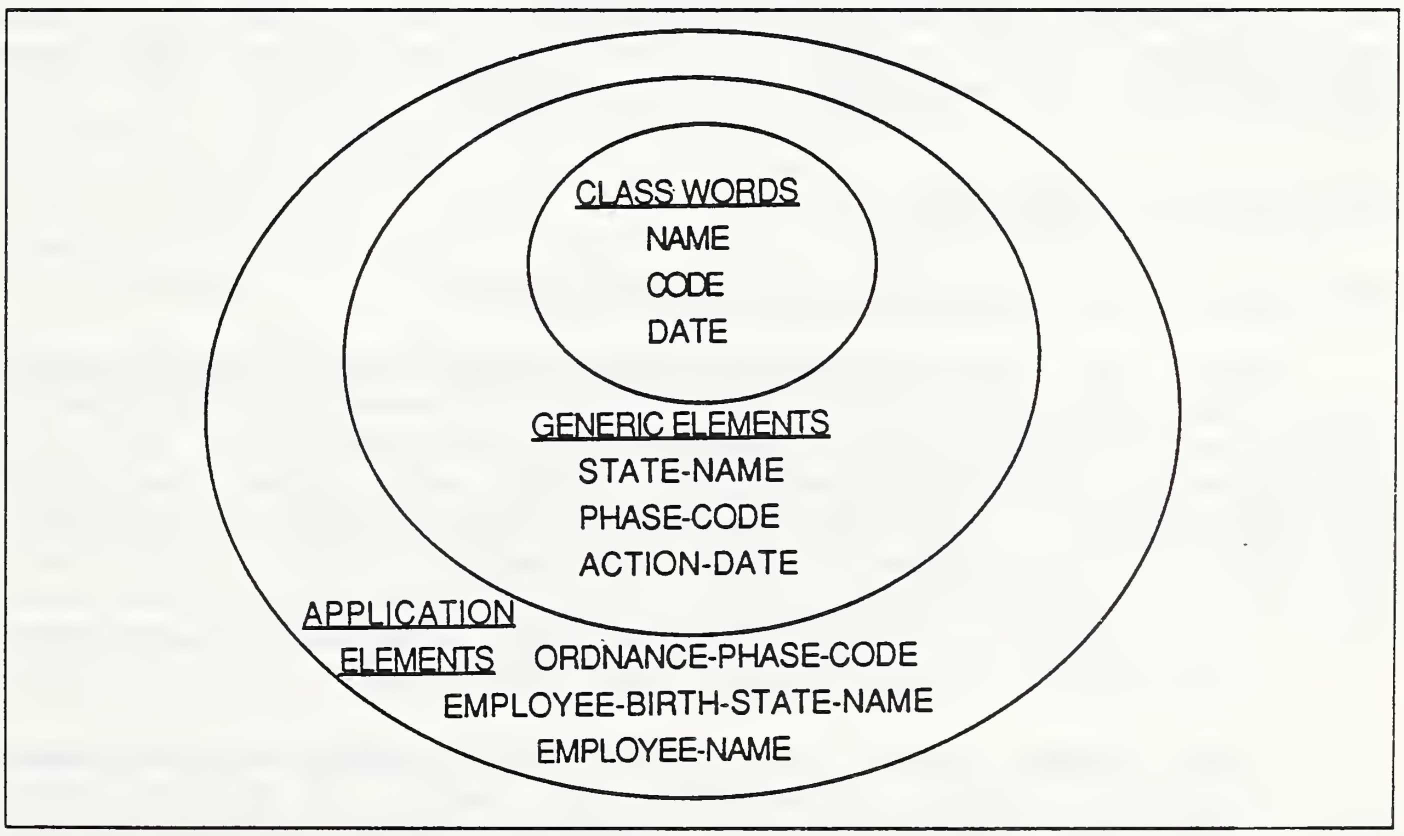


Figure 2

GUIDE TO IRDS APPLICATIONS: GENERAL CONCEPTS & STRATEGIC SYSTEMS PLANNING

Speaker
Margaret H. Law
National Bureau of Standards
Institute for Computer Sciences & Technology

WHAT IS THE IRDS?

An Information Resource Dictionary System (IRDS) is a data dictionary system used to design, monitor, protect, and control information systems. The IRDS standard represents Federal and national efforts to provide quality data dictionary system support for information engineering and management.

The extensible schema capabilities of the IRDS permit the representation of a wide variety of CASE, Data Administration, and other system life cycle information in an Information Resource Dictionary (IRD), an application of the IRDS.

The Guide to IRDS Applications [3] provides the user with guidance and examples of how to use an IRD for system development applications. The use of the extensible schema capabilities of the IRDS are illustrated.

FEATURES OF THE IRDS

Entity-Relationship-Attribute Modeling

Using the Entity-Relationship-Attribute (E-R-A) model, the IRDS supports the representation of semantic information in terms of entities, relationships between two entities, attributes describing entities, and attributes describing relationships. The IRDS uses the E-R-A model to support a three layered structure, which includes:

- o **IRD Schema Description Layer** provides Meta-Entity descriptors that support the IRD Schema

- o **IRD Schema Layer** supports both predefined and user-defined schema structures, in terms of Entity-Types, Relationship-Types, Relationship-Class-Types, Attribute-Types, and Attribute-Group-Types

- o **IRD Metadata Layer** supports user-defined metadata, in terms of Entities, Relationships, and Attributes.

The **metadata** stored in an IRD (or other data dictionary system) is different from **data** stored in a database. Metadata describes the format and meaning of data structures that are to be stored in a database. **Metadata** is defined during information system development, operations, and redesign to **describe** the operational system that supports **data**. Since system design metadata is often very complex, a structure is required to support this information that can represent a high degree of complexity. The IRDS provides support for the representation of metadata.

The E-R-A model provides the basis for representing metadata in an IRD. Metadata is information describing the characteristics of an organization's data, activities, systems, and holdings. **Entities** correspond to **nouns** (i.e., either subjects or objects), **relationships** correspond to **verbs, attributes that describe entities** correspond to **adjectives**, and **attributes that describe relationships** correspond to **adverbs**.

The **types** assigned in the IRD Schema Layer permit the user to define, and the IRD to recognize, categories of entities, relationships, and attributes that are defined in the IRD Metadata Layer. For example, the Entity-Type **ELEMENT**, predefined in the IRD Schema Layer, provides the basis for any number of data elements that can be defined in the IRD Metadata Layer, such as EMP-NO and SOC-SEC-NO. An example of an E-R-A model for a user defined schema is illustrated in Figure 1.

<u>Extensible Schema Definition Capability</u>

The IRDS supports an extensible schema that permits users to define and modify IRD schema structures. Predefined and user-defined schema structures are integrated in the IRDS. The extensible schema capability provides the user with the flexibility to design an IRD schema to fit the particular metadata requirements of an organization or life cycle phase.

<u>Predefined Schema Structures</u>

The **IRDS Core** provides the framework for all other predefined and user-defined schema structures. Among the schema descriptors defined in the Core are a number of Meta-Attribute-Types that can be used for IRD metadata control and validation. Metadata validation can be supported by the predefined schema structures of FORMAT and ATTRIBUTE-TYPE-VALIDATION-PROCEDURE.

The IRDS **Minimal Schema** provides critical schema descriptors needed to structure every IRD, such as the Entity-Types, Relationship-Types, Attribute-Types, and Attribute-Group-Types used to capture information about users, views, partitions, time, date, user permissions, etc. The **Basic Functional Schema** is specified as a required module for the FIPS and as an optional module for the ANSI standards. The Basic Functional Schema provides an initial set of schema structures that can be used as an example schema, and can be built upon through schema extensions. For example, the Attribute-Group-Type ALLOWABLE-RANGE is predefined here with its associated Attribute-Types, LOW-OF-RANGE and HIGH-OF-RANGE.

Command Language and Panel Interfaces

A conforming implementation of the IRDS standard must contain either the Command Language Interface, or the Panel Interface, or both interfaces.

The **Command Language Interface**, based on the E-R-A model, uses one command language to support both IRD schema and metadata definition. For ease of use, the IRD schema commands and metadata commands have similar structures.

The **Panel Interface** provides sets of panels through which the user can access and manipulate an IRD. The Panel Interface is specified in terms of functional characteristics without definition of screen or window design. Each IRDS panel provides six information areas for the user, including: State Area, Data Area, IRD Schema Area, Action Area, Message Area, Help Area.

Sets of panels, called Panel Trees, are included in the IRDS standard to assist IRD users in performing: metadata maintenance, metadata output, entity-lists, schema maintenance, schema output, and schema and metadata interchange between IRDs.

Extensible Life Cycle Phase Facility

Two types of life cycle phase facilities are specified in the IRDS standard. The **Core IRDS** has a basic life cycle phase facility that provides the user with the capability to construct partitions in an IRD corresponding to various life cycle phases. Three classes of life cycle phases are specified in the Core IRDS with corresponding life cycle phase partitions. The life cycle phase classes are Uncontrolled, Controlled, and Archived. The **Uncontrolled** class represents the system development phases, the **Controlled** class represents the system operation and maintenance phase, and the **Archived** class represents the historical records of a former phase. User-defined life cycle phase partitions belong to the Uncontrolled life cycle class. Life cycle phase partitions are accessed through **views**, specified in the IRDS Core.

Relationships across life cycle phases are supported by the IRDS Core. Since only entities are associated with a particular life cycle phase partition, relationships can be defined to span life cycle phases.

Additional life cycle facilities are provided by the IRDS **Extensible Life Cycle Phase Module**. To give the user comprehensive life cycle support, this optional module provides features for Hierarchical Phase Modeling, Relationship Sensitivity Structures, and Life Cycle Integrity Rules.

Variation Names and Revision Numbers

Variation names and revision numbers can be used to distinguish unique versions of an entity. Variation names can be useful to designate the life cycle phase partition in which an entity occurs. Revision numbers are initially system-defined at entity definition; after this, revision numbers can be either user-maintained or system-maintained.

IRD Import/Export Facility

The IRDS standard provides general specifications for an IRD Import/Export Facility supported by Abstract Syntax Notation One (ASN.1). This facility is intended to support schema and metadata interchange between separate IRDs, which may be located in one or more IRDSs. An IRDS schema checking capability will assist in supporting this facility when users want to exchange information between IRDs that have differing schemas. Additional specifications for this interchange facility are being developed.

Security Facilities

The Security Facilities Module supports the restriction of user access permissions to an IRD, IRD-SCHEMA-VIEW, IRD-VIEW, entity-type, individual commands, and individual entities. Access permissions can restrict the user's ability to read, add, modify, and delete schema and metadata definitions. IRDS Security Facilities offer two levels of IRD access control:

o **Global Security** provides user access restrictions according to entity-type, meta-entity-type, and partition

o **Entity-Level Security** provides user access restrictions to specific entities.

IRDS Security Facilities can be used to protect metadata stored in an IRD. An IRD, in turn, can be used to support security restrictions to protect data stored in an application database.

Procedure Facility

The Procedure Facility optional module provides the user with the capability of defining and executing new IRDS procedures, or macros, for IRDS commands. Various statement types are permitted in procedures, such as Assignment Statements (i.e., to assign values to variables), Do Statements (i.e., to group instructions together and execute iteratively), If Statements (i.e., to specify conditions for executing procedures), etc.

Application Program Interface

The Application Program Interface permits standard programming languages to interface with the command language of the IRDS. The Call feature of a standard language, such as COBOL, PL/1, or FORTRAN, can be used to access the metadata in an IRD. The IRDS standard does not specify particular language bindings. With this module, users can write programs to collect metadata from, and pass metadata to, an IRD.

STRATEGIC SYSTEMS PLANNING

The Guide to IRDS Applications [3] illustrates the design of an IRD application to support the Strategic Systems Planning phase of the system life cycle. Directions and examples are given illustrating the IRD user's development of:

- o Problem statements representing the systems analysis to be accomplished
- o A comprehensive E-R-A model sufficient to represent the stated problem
- o IRD schema definition commands, metadata definition commands, and IRD output results.

The Guide also provides users with advice on the preparation of an organization's standards and conventions document, which defines the procedures necessary to regulate the use of the IRDS within an organization.

Question: The Command Language Interface and Panel Interface that you described sound like they will not provide users the full functionality of a Graphics Interface, such as those available on most CASE tools. Is a Graphics Interface being planned for the IRDS?

Answer: At this time, there are no plans in X3H4 to add graphics functionality to the IRDS standard. The Panel Interface is intended to provide a user-friendly interface, although no graphics specifications are now included. NBS, however, is interested in seeing a Graphics Facility added to the IRDS Panel Interface, and may be able to develop generic specifications for such a facility.

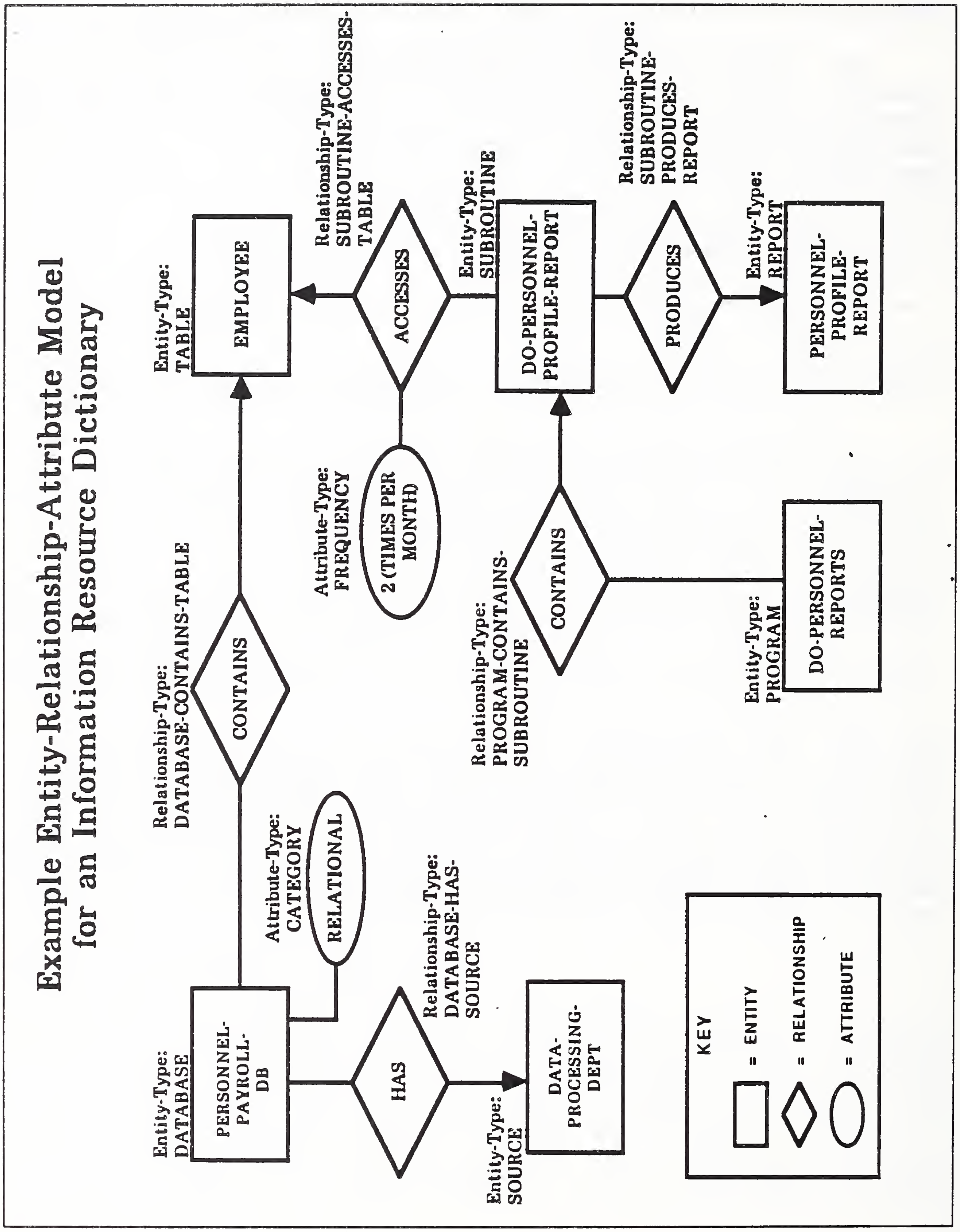

Example Entity-Relationship-Attribute Model
for an Information Resource Dictionary
Entity-Type:
DATABASE
PERSONNEL-
PAYROLL-
DB
Attribute-Type:
CATEGORY
RELATIONAL
Relationship-Type:
DATABASE-CONTAINS-TABLE
CONTAINS
Entity-Type:
TABLE
EMPLOYEE
HAS
Relationship-Type:
DATABASE-HAS-
SOURCE
Entity-Type:
SOURCE
DATA-
PROCESSING-
DEPT
Attribute-Type:
FREQUENCY
2 (TIMES PER
MONTH)
ACCESSES
Relationship-Type:
SUBROUTINE-ACCESSES-
TABLE
Relationship-Type:
PROGRAM-CONTAINS-
SUBROUTINE
CONTAINS
Entity-Type:
SUBROUTINE
DO-PERSONNEL-
PROFILE-REPORT
PRODUCES
Relationship-Type:
SUBROUTINE-
PRODUCES-
REPORT
Entity-Type:
PROGRAM
DO-PERSONNEL-
REPORTS
Entity-Type:
REPORT
PERSONNEL-
PROFILE-
REPORT
KEY
= ENTITY
= RELATIONSHIP
= ATTRIBUTE

Figure 1

FARM CREDIT ADMINISTRATION

Speaker
Marlene A. Palmer
Records and Projects Division
Office of Administration

INTRODUCTION

I'm going to speak to you not about our application or implementation of the IRDS, because we don't presently have one. I shall speak instead about a report[1] which I was asked to write for an agency task force involved in upgrading a major financial reporting system. This system, the Consolidated Reporting System (CRS), is a database of quarterly financial statements submitted by Farm Credit System lending institutions for access by the Farm Credit Administration (FCA) Office of Examination, Office of Analysis and Supervision, and Office of the Board. I should first like to say a few words about the agency, its mission, and its automated data resources, in order to explain our interest in the IRDS. I should then like to describe the contents and share some of the conclusions drawn in the report.

FARM CREDIT ADMINISTRATION

Mission

The Farm Credit Administration is an independent financial regulatory agency in the Executive Branch of the Federal Government, with regulatory, examination, and supervisory responsibilities for the Farm Credit System banks, associations, and related institutions chartered under the Farm Credit Act of 1971, as amended. Its mission is to assure the safety and soundness of Farm Credit System institutions and to protect the interests of borrowers, stockholders, investors, and the public. It is funded not by appropriated funds, but through assessments upon the regulated financial institutions.

---

1 Palmer, Marlene A., Data Dictionary/Directory Systems, Records and Projects Division, Office of Administration, Farm Credit Administration, October, 1987.

Farm Credit System

The Farm Credit System consists of 37 banks and 382 associations and related institutions located throughout the country. They include Federal Land Banks which make long-term farm mortgage loans through local Federal Land Bank Associations; Federal Intermediate Credit Banks, which provide discounted loan funds to Production Credit Associations for short-term and intermediate-term loans; and Banks for Cooperatives, which make loans to agricultural cooperatives.

Historical Perspective

FCA was founded as an independent agency in 1933. It became a part of the Department of Agriculture in 1939, where it remained until 1953. It has been an independent agency since that time. FCA was physically located in USDA office space until 1969 and occupied office space in l'Enfant Plaza from 1969 until 1984, when it moved to a new building in McLean, VA, owned by the Farm Credit System banks. With the move to the new building, the agency was completely automated from the top down. It was at this time that the FCA IRM program, of which I am a part, was implemented, within the Records and Projects Division.

When FCA was reorganized in 1985, one of the functions previously delegated to the Farm Credit System Banks (association examination) became an FCA responsibility once more. This resulted in the creation of nine new field offices and doubling the agency staff. Most of the new employees are field-office examination and supervision personnel who require access to both financial data and full-text precedential documents for end user manipulation and decision support.

IRM

The FCA IRM function is located in the Records and Projects Division, whose responsibilities also include the contractor-staffed library and the records management function. Among the accomplishments of the two-member IRM team during its first three years are implementation of the contractor-developed Farm Credit Retrieval System (FRS) for full-text precedential documents, development of the Farm Credit Thesaurus, a controlled indexing and retrieval vocabulary, and implementation of a commercial thesaurus-management package used for validating FRS indexing and search terms and for producing camera-ready copy for the hardcopy thesaurus.

My responsibilities have included thesaurus development for both form (software implementation) and content (lexicography). I have also managed the database population function, including document selection, remote data entry via contracted optical scanning, and quality control.

EDP Environment

The FCA EDP function is located in the Information Processing Division. The hardware configuration includes a cluster of VAX super mini-computers (11/750, 11/785, 8650), running under a VMS 4.7 operating system. Each headquarters and field-office employee has a microcomputer or word processor which can be used independently or as a terminal for accessing the VAX system via a United Technologies LEXAR digital telecommunications system. A high-speed leased line network to the district offices is presently being installed. Field examiners are equipped with portable laptop computers with built-in modems. The most widely-used agency software tools are All-in-One (electronic mail/office automation), WPS+ (word processing), LOTUS (spreadsheet), Polycom (communications), and ORACLE (relational database management system).

The Information Processing Division was reorganized in 1987. Previously, most database work had been contracted out, resulting in several systems characterized by suboptimal performance and divergent data elements. Among the significant changes in the reorganized division were elimination of contracted systems development, installation of ORACLE, and conversion of major systems to ORACLE.

Consolidated Reporting System (CRS) Conversion

The most important ORACLE conversion was the Consolidated Reporting System (CRS) mentioned earlier, a contractor-produced COBOL/FORTRAN system. Farm Credit System lending institutions submitted the requisite financial reports to FCA in various electronic and hardcopy media which had to be converted prior to entry into the CRS. This resulted in unacceptable delays in providing accurate and timely electronic data, frequently imposing the use of unmanipulatible hardcopy. The recent farm credit crisis exacerbated this situation. New farm credit legislation in 1987 enabled FCA to strengthen reporting requirements, and to prescribe standardized electronic and scannable hardcopy formats in order to expedite data entry into the new ORACLE-based system.

Corporate Data Concept

Conversion of the Consolidated Reporting System and other major FCA systems to ORACLE revealed a proliferation of incompatibly-defined data elements. It underlined the need for an agency concept of corporate data ownership and control. I was asked to write a report on data dictionaries as a part of the CRS conversion task force investigation of data management tools and methodologies.

DATA DICTIONARY/DIRECTORY SYSTEMS REPORT

Methodology

I began my research with a DIALOG search for relevant book, report, and journal literature on the subject. Through professional colleagues, I identified and contacted several experts in the field. It was in this way that I learned about Dr. Alan Goldfine and his work on the IRDS. Along with the Chief of FCA's Information Systems Branch, I visited Dr. Goldfine at NBS for a demonstration of the IRDS and the IRDS Command Language. We obtained a copy of the IRDS Prototype source code and were very impressed. This was the starting point of my report.

Contents

The completed report is a distillation of recent books, reports, journal articles, and course materials describing the components, goals, uses, and benefits of data dictionary /directory systems (DD/DS), as well as methodology and standards for use in their development. At an elementary level and in non-technical language, it describes different DD/DS types, lists desirable features and discusses DD/DS and data element standards. It includes a section describing additional sources of information (indices, directories, on-line services, courses, professional associations, and experts), as well as a bibliography.

DD/DS Types and Sources

Manual Systems -- DD/DS sources are seen as a continuum of manually-maintained through highly automated systems. They include coded-card systems such as edge-notched and "peekaboo" cards, and a simulated data dictionary on 3x5 cards, some of which are preliminary stages for automated systems.

Custom-Developed DD/DS Software -- Custom-developed software was not recommended as a DD/DS source. Arguments against any custom-developed software apply: cost, maintenance, documentation, upgradability, etc.

Generalized vs. DBMS-Specific Packages -- Among the advantages of generalized or independent DD/DS packages are extensibility and portability. However, in environments where most databases utilize the same DBMS, the tight integration possible with a dependent or DBMS-specific package offers more active or enforcement capabilities.

Thesaurus Software -- Because lexicographic analysis is very similar to data analysis, lexicographic tools and techniques used to build and maintain thesauri or controlled indexing and retrieval languages could be employed in the data-analysis phase of DD/DS construction. For example, semantic factoring, the process of analyzing and decomposing complex concepts into elementary concepts; facet analysis, the process of dividing concepts into facets (subject categories); and hierarchy building, the process of arranging the components of facets into part-whole hierarchical structures, are lexicographic techniques used in thesaurus construction and maintenance which could also be used for data analysis and DD/DS construction. Other similarities are the need in both thesauri and data dictionaries for synonym-homonym control and for indices of the subject terms/data elements, such as a key-word-out-of-context (KWOC) index.

Thesaurus software, which performs similar functions and processes, could therefore be used, with minor modifications, to build and maintain a DD/DS. However, it would not be a viable alternative unless it were already installed on the site and multiple usage presented no licensing agreement problems.

It is to be hoped that vendors will become interested in an integrated thesaurus product offering control of both subject and descriptive (author, title, document type, etc.) elements.

IRDS Prototype Software -- The report discusses the IRDS prototype in detail. Addresses of vendors who have announced plans to develop IRDS-based products are included. Advantages of the NBS IRDS prototype software as a DD/DS source are: free source-code, modularity, hardware independence, DBMS independence, and the fact that it is the basis of the forthcoming ANSI and FIPS data dictionary standards. Disadvantages arise from the fact that as a prototype, it may not be thoroughly debugged. Software support, documentation, and upgrading are also potential problems.

Compromise Solution -- A compromise or interim solution exists for organizations who need a dependent or integrated DD/DS, but whose DBMS vendor does not as yet offer DD/DS software or for those who need a generalized package, but wish to wait until one is available which incorporates the forthcoming ANSI and FIPS standards. Construction of the DD/DS could commence, utilizing manual, semi-automated, or automated methodology, while complying with forthcoming ANSI and FIPS standards. This is essentially the course taken at FCA, while waiting for ORACLE IRDS-compatible data dictionary software.

Conclusions

The following are among conclusions or recommendations:

- o Begin with a pilot project, with a selected set of data.

- o Consider the IRDS, since it will shortly become a standard. Arrange to meet with Dr. Goldfine, if possible.

- o Obtain NBS data-naming guidelines before establishing data-naming conventions.

- o Avoid custom development.

- o Because it requires management commitment, implementation of a DD/DS is a consciousness-raising activity. It could be the first step toward implementation of the data administration function, strategic information planning, an organizational information policy, etc.

A final conclusion is based upon the unexpectedly-large number of requests for this very elementary report from both Federal and private-sector organizations. This, I believe, indicates that there is a lot of semantic confusion as to what data dictionaries really are, and/or that many organizations have recently recognized a need for data dictionary control in their systems environments and are presently considering the acquisition of appropriate software.

Additional copies of the report are available upon telephone or mail request.[2]

Question: I assume from what you said that you would strongly support the development of extensions or modules of the IRDS that would support, even more explicitly, such things as thesaurus capability.

Response: Definitely. Something that I forgot to mention is that I think that a good IRDS should be able to control manually-maintained systems. It should be able to control, for example, an organization's document-processing unit, validating organizational names, etc.

[2] Marlene Palmer, Information Resources Specialist, Records and Projects Division, Office of Administration, McLean, VA 22102-5090, (703)883-4120.

U.S. DEPARTMENT OF EDUCATION

<u>Speaker</u>
Alexis Poliakoff

What I'm going to present is what I call a "data dictionary retrospective" at the Department of Education. It concerns a major project and what we learned from that project. We learned some valuable lessons that you may be able to take advantage of. In 1985, the Department of Education started an aggressive data element dictionary project within the Model 204 database environment. The project had the objective of capturing information concerning about 33 application systems comprised of about 600 files and over 5000 data elements in the Department's data element inventory. During this project we learned about and participated in some of the early discussions concerning the IRDS, and one of the goals of the project was to include, as far as it was feasible, the IRDS data model. We found the model very helpful at a practical level, in terms of obtaining definitions of relatively simple terms, such as the "controlled" and "uncontrolled" life-cycle-phases of a system. It made it a lot easier in defining terms concerning systems and software at the Department of Education. We also included a major effort to standardize the data elements of the student financial aid community in the Department--we have a pretty large student financial aid program, both the student loan program and the Pell Grant program which provides direct grants to students and institutions of higher education. This was all done, using the IRDS model to the extent possible, with the Model 204 data dictionary. The major product we produced was a dictionary of data elements in the Department--it contained over 5000 definitions. We provided automated database documentation, extensive documentation of the Model 204 system, and tools to manage the database environment in the Department.

Now, what did we find out? Because of the lack of overall corporate data planning, the systems that existed in the Department of Education were not all true DBMS applications. We found that, both in terms of the size of the database and in the number of users, the applications that were developed in the late 1970s and early 1980s more appropriately should have been PC based applications, since they existed at the direction of one or only a few users and had very changeable characteristics. Possibly threequarters of the applications would not, these days, be mainframe DBMS applications, due to either size or lack of multiple-users. We found that there were provable, demonstrable cost-benefits of the database activity in the areas of database documentation and life cycle standards required for database documentation, and that we could produce this very handily. There was a significant tension between the active role of the Model 204 dictionary that we used, and the IRDS role. There are products for downloading or interfacing between SAS, Lotus, and other products that tended to make the developer of the active dictionary unconcerned about the IRDS model. Changes to the new releases of the product and the data dictionary could be very traumatic for IRD definition. Major changes in this model could be swept away to accommodate practical needs for the developer of the DBMS.

Now what happened after we developed the data dictionary, populated it, and produced the reports? Well, first of all, there was a constraint on resources due to changes in management--something that most of you in this room probably have experienced. There was a shift in the software platforms for developing systems away from mainframe DBMSs towards minicomputers and small computers in particular, which made more difficult the immediate access we had to the Model 204 system. There was limited user input in standardizing of data elements. There was a limited role for the IRDS in developing new systems, which I consider a nearly fatal error in a development of this kind. Basically, the project collapsed of its own weight, or half-life. I consider that data has a half-life just like radioactive material; left alone, half of it just disintegrates over the half-life. In fact, without an elaborate, expensive maintenance effort for small systems, which probably wouldn't have been worth it at that level, the project was dropped.

What did we learn from this particular project? First, in spite of the data dictionary effort being at first a descriptive effort, where one allocates resources is a very important issue and should not be dismissed or easily dealt with. The major systems that are bona fide database applications are precisely where the IRDS resources should be allocated. Maintaining management commitment is very important. The distinction between doing the database project and selling it can lead to an easy trap to fall into for people who are enthusiastic about development. What I learned is that you have to continue this selling effort.

It's not sufficient to sell it to the point of funding, but throughout its development. Perhaps even more important is selling the maintenance. I don't believe the cost of maintaining a data dictionary system is well documented, but clearly, at the Department of Education, the maintenance of the data dictionary might be as much as half of what the original development costs were.

Second, as Marlene Palmer said, pick projects that are early in their life-cycles, pick pilot projects, pick projects that have management tension, that have the likelihood of continued interest even if there are changes in management.

Overall, I think that the primary message of this retrospective is to make the argument for the IRDS, which is difficult, in spite of demonstrable figures of cost-benefit for system and program documentation. The marketing effort can't be stopped, and is a continuous problem in this whole process.

<u>Question</u>: When you mentioned "implementing the IRD model," were you referring to the IRD Schema and the Basic Functional Schema, rather than the functionality of the IRDS?

<u>Answer</u>: That's right. We did not implement the functionality, we implemented the basic concepts. We used the entity-attribute model, we used the constructs of the IRDS--the one that comes to mind first is the system life-cycle-phase structure of archived, controlled, and uncontrolled. We used the definitions in the IRDS documents, and felt that it was a very valuable use of our time.

<u>Question</u>: Since this was a mainframe application, was there ever any thought given to putting the application in a PC environment, so that users could see their data showing up right in front of them on their own machine?

<u>Answer</u>: Back in 1985 and 1986 when this project was underway at the Department of Education, the PC population was relatively small. What I proposed, and I had some success, was to produce a document that was similar in every sense to "Webster's Dictionary." I did this for two reasons. One was that PCs were not available to any significant extent, and two, we did not have good software tools to maintain interfaces to them, and there still is a large population of users and even programmers who are not completely comfortable with using terminals. I thought that to have a

physical, printed dictionary would be the way to get the widest support.

Question: So your end product was the printed dictionary, rather than a browse capability...?

Answer: We had a browse capability, we had all the normal capabilities one would have in a dictionary system--the names and definitions of the data elements, what systems they were in--for about 5000 data elements.

Question: Did you do any reconciliation or purification. In other words, did you try to find data element redundancies throughout the system--throughout multiple systems?

Answer: Yes, but we were not terribly successful. There were clear and obvious redundancies, and great management enthusiasm for eliminating these redundancies. I never did know how to implement the elimination of redundancies when we had small systems, perhaps even tiny systems, operating very independently. The redundancies would not be cheap to eliminate. I believe that the greatest benefit of a dictionary is to show alternate sources of information that already exist. The assumption was that information is very expensive to collect, that the collection of information, not the storage of information, is the expensive element in all this, and that people who could see where the information already existed would not go to the expense of collecting it themselves.

Question: Could you expand on your statement that the cost of maintaining the dictionary was half the cost of developing it in the first place?

Answer: Well, the cost of loading the dictionary was about $100,000. To maintain all of those systems would have cost about $75,000 a year.

Question: Was this maintenance in connection with using the dictionary instead of using an alternate method to support such things as life-cycle-phases, so that in a sense it would be a justified cost?

Answer: I think it was a justified cost. In environments like the Department of Defense perhaps, and some others, the value of documentation is clearly seen by all levels of management. In our environment, the user is always more eager to get the next report than to have a system documented. In the competition for resources between the next report and the system being documented, the documentation always loses. Under that environment, the documentation isn't done, so you're not saving real dollars by producing documentation, you're saving hypothetical dollars.

U.S. DEPARTMENT OF AGRICULTURE

Speaker
Robert C. Kling
Soil Conservation Service

As announced, I work for the Department of Agriculture. I've been with the Department for about 20 years. However, I've been with the Soil Conservation Service (SCS) for only the last four. My background and interest in data dictionaries goes back to the late 70s when I went to an ACM seminar where vendors of database management systems were trying to market their products.

The Soil Conservation Service has been, historically, a controlled, mainframe environment, and with the advent of relatively inexpensive minis, we soon had a proliferation of equipment, software, data--quite a bit of an uncontrolled environment. It was decided by upper-level management that something had to be done.

The solution, of course, was a data dictionary. Contractors were brought in to sell, to those members of management beyond those few people who already understood, why the agency would benefit from a dictionary. We developed a pilot dictionary system that basically went into our mainframe environment and extracted the data definitions out of our COBOL library. That was very nice, but, as most of us know, hardly useful when we were really trying to develop a data dictionary to work from. There were no naming conventions prior to that, and we really had our work cut out for us. Most people advised that we start from scratch.

The Department's response to the Paperwork Reduction Act also was to show an interest in a data dictionary. Because each agency within the Department was required to use Department centers, a standard dictionary across the Department made a lot of sense. A number of agencies were looking at data dictionary software, and so a task force was put together. The Department chose Datamanager from MSP. Unfortunately, the Department's requirements were a little different from ours. One of the Department's key requirements was a stand alone dictionary system; we would have much preferred an integrated system. However, since the package was purchased and made available, we began to populate the dictionary.

We got a contractor in, and he proposed a methodology, made the appropriate changes and extensions to Datamanager, and we began to collect information using a top-down process. We also had a contractor who began to work on user interfaces because the Soil Conservation Service is a very distributed agency. We have people located throughout the country, and the thinking is "distribute everything." The initial idea was that everyone and his brother were going to access this dictionary, not understanding the control aspects of a dictionary system.

In the meantime, along came reorganization, budget cuts, and RIFs. The IRM organization was abolished, the ADP group was decimated over a three year period and replaced by contract people, and the dictionary, basically, went away.

Another agency in the Department that used Datamanager had a similar experience. They had a good head start on a dictionary, populated it with the understanding that they were purchasing minis that would be placed around the country, and got all their data definitions in. When development started, it started about 2000 miles from the data dictionary. Needless to say, things soon got out of sync, the key person working with the dictionary died, and the dictionary died with her. So I've been involved with those factors. When I was in the Agricultural Research Service, we started a database project using the Cullinet IDMS product, which has its dictionary IDD, and quite often we had various conflicts.

That's my background, and you'll see it influencing the recommendations and approaches that we're taking in the Soil Conservation Service.

The Soil Conservation Service has about 10,000 employees. There are about 3,000 field offices, each of which has been blessed with its own microcomputer and has some form of an IRM requirement. The organization is very field oriented. All our priorities are driven from the field up, which is quite opposed, obviously, to a data dictionary, top-down approach. As the hardware showed up in the Agency, people began to use it, although with a lack of both software and support from headquarters.

Figure 1 was prepared for management, to try to convince them that there had to be a better way. The figure is a depiction, really, of where we were. We were basically forced into a distributed data dictionary because of the wide variety of hardware and software that we had. The tools that we were using had embedded within them some degree of a data dictionary. The dotted lines that you see were simply interface deficiencies. In other words, how in the world would you communicate between these. Every one was different, and the IRDS was not quite far enough along in terms of export-import or interface issues. The task was really monumental.

So we entered again the stage that Nolan has defined as "chaos," this time with microcomputers distributed throughout the agency. Again, there was a recognition in the agency that some control had to be brought about. One of the groups at one of these places was doing structured systems analysis, and they were collecting our information requirements. Unfortunately, no one in this group bothered to collect our data element requirements or record requirements. So we had all these groups out there using the Excelerator product, which has some form of a dictionary--it's a CASE tool. However, there was no standardization as to how to collect data, and no integration between the groups, so we had no real way of integrating or doing any data modeling as the result of it.

So after one of the groups had spent considerable time and had thought that they were ready to develop the system, they were informed that they were far from ready because they had not defined their information needs. A request was made to the IRM division that presented this issue: "The current method of priority setting for software development and overall coordination of IRM activities appear to be cumbersome, overlapping, and inconsistent with overall agency objectives. The goal is develop a structure to ensure that there is a process that addresses the issue."

What the IRM division decided to do with that request was to create what we call a "vision statement." This statement was basically to give the idea of the many independent activities that were going on where the direction was headed. If you're going to proceed, at least proceed within these boundaries until we can introduce some control and provide some standardization and input from the top down.

There's a very good article by Daniel Appleton in the March 1, 1987 issue of <u>Datamation</u> that talks about this concept. It's good because it fits what I believe is really where an IRDS fits in.

The data dictionary has come a long way since the days when it was simply a listing of data elements. The modern data dictionary is the key to integrating and planning software development, workbenches with programming environments, and application systems in the MIS organization. It is the ultimate repository for data about the information resource environment. It is also the key to controlling information products and keeping their production costs and lead-times down.

But before you can implement this important support software, you must first re-evaluate your control architecture and make any necessary changes. Only then can the productivity pluses promised by data dictionaries be realized.

The idea that the dictionary is, in fact, for control architecture is really the point that I'm making. If it becomes a control architecture, it will be fed on time, it will be fed accurately, and it will be fed as a by-product of the policies and procedures of your agency. The systems that I mentioned before were separate entities, separate projects. They were initiated with the idea of achieving the benefits of a data dictionary, but in no way did that occur, simply because there was no integration into the organization and the way the organization operated.

Here is a draft statement that I put together to give an overview to the people who requested the vision statement: "The objective of an SCS IRD information system would be to facilitate SCS compliance with Federal and agency IRM reporting requirements. It would also support SCS information resources management throughout the systems development life cycle, including planning, feasibility, cost-benefit, design, development, testing, and maintenance phases when fully implemented. It would provide a controlled repository of information about SCS information and information resources. By having up-to-date and accurate information on information and related resources, much of the current duplication of effort would be eliminated. A properly implemented IRD system would facilitate structured systems analysis and data modeling activities planned in SCS. Information on structured systems analysis would be maintained in the IRD, and would be related to current and planned SCS information systems software and databases. The IRD would be used to assist SCS in establishing and maintaining standard data definitions and usage. A fully implemented SCS IRD would consist of many components developed over a considerable period of time. If SCS is to benefit from the implementation of an IRD system, we must first clearly define our requirements. Implementations should proceed through the same system development life cycle that any information system undergoes. At present, there is no coordinated effort in SCS to collect, analyze, and disseminate information about SCS information resources, such as data, data definitions, files, databases, software, IRM skills, processing facilities, systems requirements, information flow, reports, forms, systems planned, under development, or in production, expenditures, budgets, etc. With the potential of over 3,000 processing facilities and associated software, data, and personnel, it is obvious that the volume of data alone implies that some automation is necessary if the information is to be useful to managers in planning, developing, operating, and maintaining information systems to support the mission and programs of SCS in the future."

Figure 2 is a pictorial view of the proposed vision. The whole thing is the IRDS--the information resources management system. The data dictionary, the IRD, is just a piece of that. The way I view it, the system has underneath it many of the already maintained information systems in our organization, and the idea is really to feed on them. If we don't have to create a new organization to collect data or to feed the dictionary, if all we have to do is take the information that's available and then solve the maintenance issues, the opportunity of success is much greater because the IRD will be missing the information it needs only if the organization stops performing one of its functions. Essentially, you're taking a two-pronged approach: a long-term and a short-term.

The idea of approaching this as you would any information system is that we need to do a requirements analysis in the context of the mission and resources of SCS. What we've done is to contract with the General Services Administration, which in turn has under contract a number of vendors, to help us put together a strategic plan for information resource management. The two primary deliverables of the contract are the strategic plan and the schema for an IRD that will support and integrate with that plan. The agency has gone to a standard methodology for systems development life cycle, and within that is the structured systems analysis, which is being used to collect information related to actual, proposed systems to be developed.

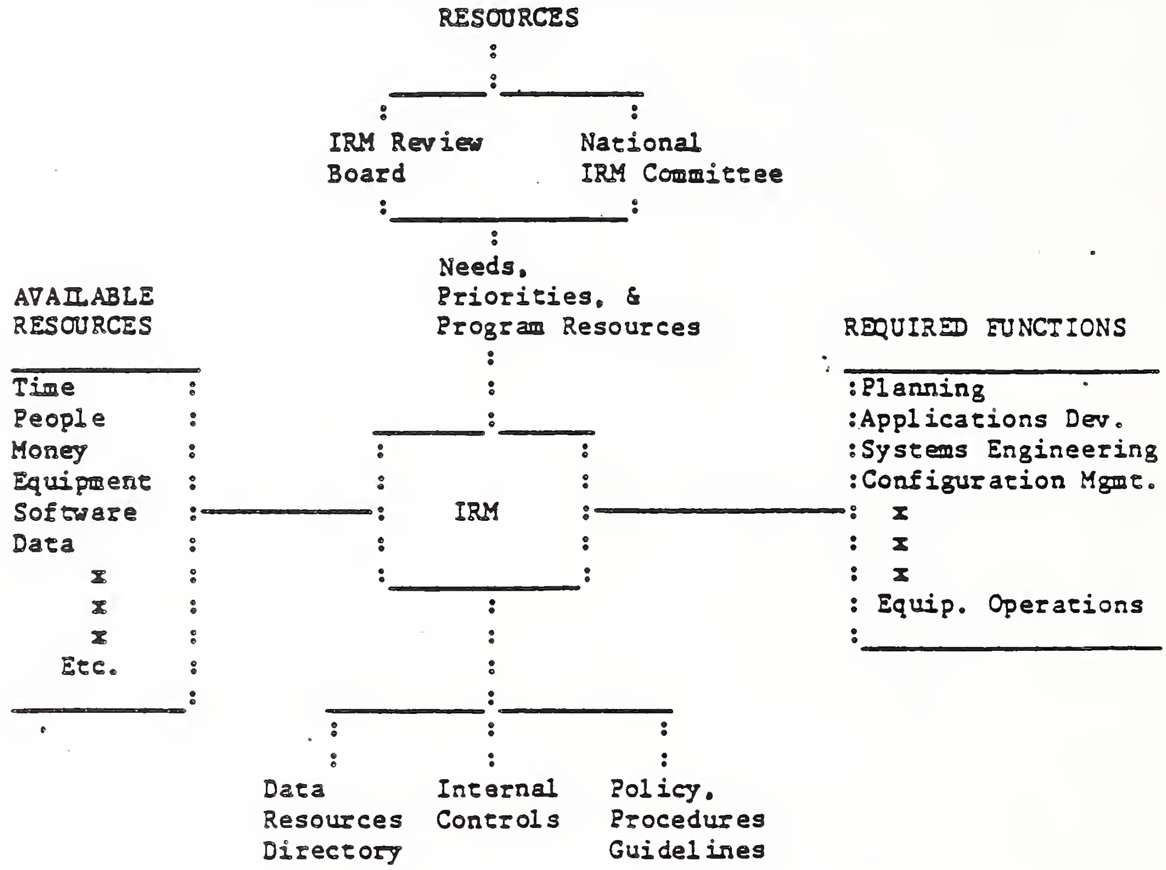
Since we have many of these projects in motion, I have been working half-time on the long-term side of it, and half-time on the short-term. The long-term project will create, over a ten year period, some 50 million dollars worth of software to support our engineering functions in the field offices. We're starting with a top-down structured systems analysis. We've inherited the Excelerator software for the short-term. It's unfortunate, but using that package for three years the agency never did come up with a data entry form or a standard or a naming convention or anything else, so if anybody has done that for Excelerator, I would be glad to hear about it. I've got two weeks to put something together!

To just explain Figure 2 and what I view as the total IRM model here, essentially, all these types of information are what's necessary to do the functions of IRM, the planning functions, and the priority settings. You have management that gives us money, personnel, and what they believe we need in the form of support for the mission. We have certain resources available that we have to use in the most efficient way we can to meet as many of those needs as we can. We have the various planning functions, priorities that need to be consistent with the budgets, the hardware, and the software that we have. We need to identify where we need to get more people and skills, and whether we need to change our skills.

All that type of information exists in an organization. The policy out there requires software inventories and we're supposed to know the hardware we have. Personnel, in theory, knows the people we have, Security tells us that we have to classify all our data, so we must have that. We have to classify our facilities, because that's how we determine our security plans. All that information is being collected, but unfortunately it's being collected by an individual solely for his purpose, and nobody else really gets access to it.

By implementing this within the strategic plan of the agency and within the system development life cycle structure and methodology that we've picked, we will not allow a project to get approved without certain information and we will not allocate resources to it without additional information. In that manner, we hope to feed this environment. As the figure shows, that information is all over the country. With communications it is actually possible to access the information. The Department is working on and does have in place a network, so it is feasible. It's just that it will not fit into one software package for a long time to come.

GENERIC SCS INFORMATION RESOURCES MANAGEMENT MODEL



Information Resources Management Functions

In order to establish the required resources and to incorporate policies, procedures, standards and guidelines into IRM-related functions, the following information is needed for each function. Ultimately, these functions and related information relate back to functions within position descriptions, vacancy announcements, branch functions, etc.

Function: (A description of the function to be performed)

Roles & Responsibilities: (Who has ultimate responsibility and authority? Identification of tasks and technical & management responsibilities)

Resources Required: (Levels of expertise needed to perform this function. Detailed knowledges, skills, and abilities (KSAs).) Time and money required. Special resources required (Equipment, Data, Software).

Policy & Procedures: (References to existing Federal, USDA, or SCS policy and procedures)

Standards & Guidelines: (References to existing Federal, USDA, or SCS standards and guidelines)

Figure 1

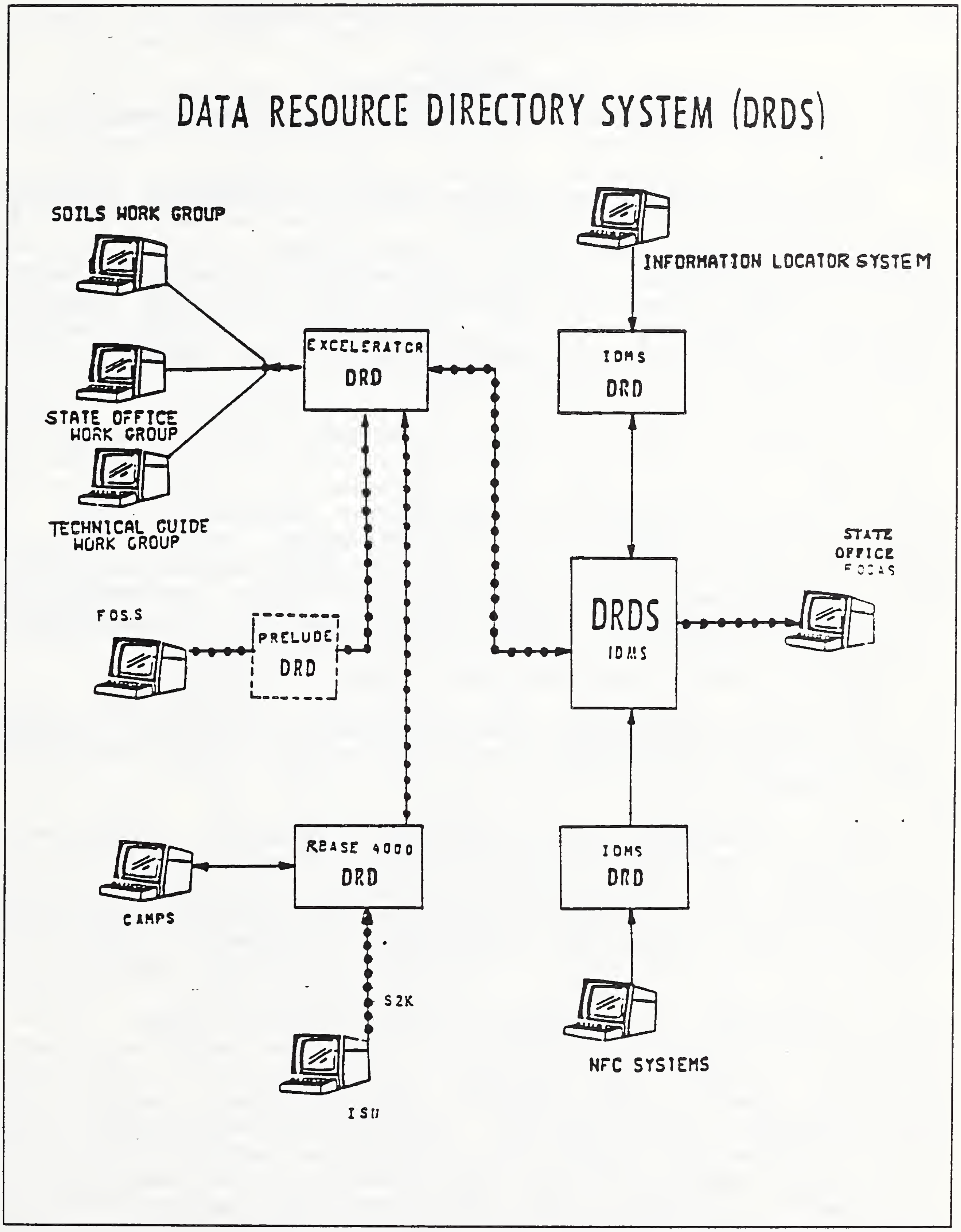

DATA RESOURCE DIRECTORY SYSTEM (DRDS)
SOILS WORK GROUP
STATE OFFICE WORK GROUP
TECHNICAL GUIDE WORK GROUP
EXCELERATOR DRD
INFORMATION LOCATOR SYSTEM
IDMS DRD
FOS.S
PRELUDE DRD
DRDS IDMS
STATE OFFICE
RBASE 4000 DRD
CAMPS
IDMS DRD
S2K
ISU
NFC SYSTEMS

Figure 2

ENVIRONMENTAL PROTECTION AGENCY

Speaker
Mary Lou Melley
Office of Solid Waste and Emergency Response

When I received the letter from Alan Goldfine verifying the invitation to this workshop, I called and asked him what the specific topic should be. He used the word, "perspective", a perspective on the Information Resource Directory System (IRDS) from EPA and the Office of Solid Waste and Emergency Response. I then visualized a telescope, from EPA, through my office, all the way to the National Bureau of Standards and the IRDS (Figure 1).

I'm going to cover in my presentation the view from EPA, from my office and our data activities through to NBS and the IRDS. I shall cover the functions and activities of the Office of Solid Waste and Emergency Response in EPA, the specific functions of my organization, the Information Management Staff and, in particular, the Information Management objectives in the areas of life cycle management and data administration. Then I'll describe our project to implement a data resource directory and compare it to NBS's Information Resource Dictionary System (Figure 2).

The Office of Solid Waste and Emergency Response (OSWER) supports the implementation of a number of environmental acts (Figure 3).

The Comprehensive Environmental Response, Compensation and Liability Act, or Superfund, addresses the nation's abandoned hazardous waste sites. Activities include identifying such sites through a preliminary assessment, followed by a site inspection. The site may or may not be placed on the national priorities list (NPL) according to its hazard ranking score. A feasibility study is carried out to develop and analyze cleanup alternatives. Then a record of decision is produced which documents the remedy selected. The cleanup process continues, which may involve remedial design and remedial action, then operation and maintenance. The remedial process is longer-term action taken to prevent, minimize, or mitigate exposure and damage to human health or the environment. Another type of response to the discovery of a hazardous waste site is a short term action taken to prevent, minimize, or mitigate damage to human health, welfare, or the environment. This is the emergency, short term situation.

The Resource Conservation and Recovery Act (RCRA) regulates facilities which generate, transport, treat, store or dispose of hazardous wastes. Activities include inventorying hazardous waste sites, managing the permitting process, monitoring the sites, and enforcing standards.

Title III, Emergency Planning and Community Right to Know, requires industries which manufacture, store, or use certain chemicals to cooperate by identifying facilities where chemical emergencies may occur, providing information on the quantities and characteristics of these chemicals, and planning emergency response efforts with local committees. It incorporates State and local governments into what is now a national emergency response structure for hazardous substances.

The legislation regarding underground storage tanks directs EPA to support the technology to identify leaking underground storage tanks in every locality, and to identify measures to correct leaks and also to prevent them. Implementation is the responsibility of the localities.

As a result of these programs' activities (Figure 4), the information collected and maintained by OSWER is extensive, both in volume and in the range of information. Databases have been defined and produced concerning inventories of sites, permitting facilities, enforcement actions, and status of cleanup activities. Databases have been developed to support scientific and technical standards; laboratory analyses results have been recorded. Information is available concerning the status of hazardous waste facilities nationwide, and computers are used to implement environmental models, geographic information systems, expert systems and dispersion models.

In OSWER (Figure 5), the Assistant Administrator has recognized the need for a strong information management function. The Information Management Staff, in the Immediate Office, is responsible for developing the policy and procedures for managing the information activities within the Office. The program offices are asked to submit their information resource management plans at the beginning of the fiscal year, and to update them at midyear. These plans describe the activities which may include system development projects, surveys, or development of procedures and tools.

Following a review of the programs' plans the Office's Senior Information Resource Management Officer (SIRMO), who is the Director of the IM Staff, approves the plans. The SIRMO has the authority to review and approve all procurements which relate to information activities and, in that way, he ensures that the procurements reflect the approved plans.

The SIRMO has defined the charter of the Information Management Steering Committee, a senior management group which reviews and approves the accomplishments of major system development projects at frequent intervals. To support the system development process, the IM Staff is completing its System Life Cycle Management Guidance, which details the activities and products from the initial description of the information need through the archive stage of a system (Figure 7). Figure 8 illustrates our life cycle phases and stages. They correspond, in general, to the NBS standards and also our agency-wide standards.

In addition, the IM Staff is developing a Data Administration program (Figure 9). The data administration policy states that data is an important resource. We have defined the relationships of organizations to data: data stewards have the responsibility for defining the data and ensuring its overall quality; data custodians have physical custody of the data. The data administration program has both short and long-term objectives. One of the long term objectives is an approach toward an enterprise model for OSWER which will form the basis, or the logical, over-arching design for future information systems.

We have incorporated data management principles into our Life Cycle Management Guidance, requiring data element dictionaries and logical data models. We have also produced a Practice Paper to supplement the Guidance; it explains the concepts such as data stewardship and the products, in greater detail. We shall be defining what our approach toward the enforcement of data management practices will be.

As a short-term project for the data administration program, we are defining and developing a Data Resource Directory (DRD) (Figure 10). The high level requirements for the DRD involve two levels of information management. One is for data documentation during the system and data life cycle. Another is a data resource directory for all of the systems, databases, and models which originate in OSWER. To be useful, the inventory will incorporate keywords which relate to the programs' functions. The data documentation capability, or dictionary, will be guided by naming conventions and well defined data element (and record, file, group, etc.) characteristics. The DRD will be the repository of data standards and standard definitions of program terms. The information about data elements and systems in OSWER may be analyzed to identify unnecessary redundancies and to move toward integrating similar databases.

In the search for a short-term solution to our need for a DRD, we are evaluating the software already available in EPA. We are examining PREDICT, an active dictionary for ADABAS, Data Catalogue 2, and FOCUS. We plan to develop a user-friendly interface to one of those packages and proceed to establish an inventory and encourage its use as a data element dictionary.

Upon analyzing our approach to the guidance for developing individual dictionaries, with respect to the NBS's IRDS, there are some variations on the same theme (Figure 11). In our LCM Guidance we require the development of dictionaries to document the detailed data requirements, the physical database design, and then the production database design. Each of the dictionaries is produced and accepted as documentation to "freeze" the requirements, or progress, at those points in the life cycle. The requirements dictionary becomes a permanent documentation item. This contrasts with the NBS approach which implies that the dictionary, in a developing system, is generally in the "uncontrolled" stage. For a production system, the dictionary is "controlled archived." There is no contradiction between the two approaches; one is more specific than the other.

In Figure 12 we note the similarities between the DRD and the IRDS in terms of definitions, functions, and procedures. The DRD entities and attributes will correspond to those of the IRDS. The data dictionary for OSWER systems corresponds to concepts in the proposed Data Management Module of the IRDS. OSWER's programmatic and technical metadata will reside in the DRD, and this corresponds, again, to the IRDS entities and attributes. The naming conventions OSWER will define will be based on the NBS naming convention guidance. Additional DRD concepts, such as configuration management and, in the future, some type of data quality indicators, are also described in the IRDS proposed Data Management Module.

In summary (Figure 13), the IRDS entity, relationships, attributes, and functional capabilities will be reflected in the OSWER DRD. NBS naming conventions will be used as guidelines. Our Life Cycle Management Guidance emphasizes dictionary requirements during the system life. The key words that we are developing will support data sharing and analysis.

At EPA, within OSWER, we coordinate all of our information management activities with the Office of Information Resource Management.

The perspective of NBS/IRDS from this end of the telescope is that OSWER looks to NBS for the concepts and specifications which will help to make our products useful, not only to us, but to others.

ENVIRONMENTAL PROTECTION AGENCY

OFFICE OF SOLID WASTE AND EMERGENCY RESPONSE

PERSPECTIVE OF

THE INFORMATION RESOURCE DICTIONARY SYSTEM

PRESENTED TO:
NATIONAL BUREAU OF STANDARDS

MARY LOU MELLEY
INFORMATION MANAGEMENT STAFF
EPA/OSWER
3/24/88

Figure 1

OFFICE OF SOLID WASTE AND EMERGENCY RESPONSE

INFORMATION MANAGEMENT STAFF

INFORMATION MANAGEMENT

LIFE CYCLE MANAGEMENT

DATA ADMINISTRATION

OSWER'S DATA RESOURCE DIRECTORY AND NBS'S INFORMATION RESOURCE DICTIONARY SYSTEM

Figure 2

OFFICE OF SOLID WASTE AND EMERGENCY RESPONSE

- COMPREHENSIVE ENVIRONMENTAL RESPONSE, COMPENSATION AND LIABILITY ACT (CERCLA) OR SUPERFUND
- RESOURCE CONSERVATION AND RECOVERY ACT (RCRA)
- EMERGENCY PLANNING AND COMMUNITY RIGHT-TO-KNOW ACT (TITLE III)
- UNDERGROUND STORAGE TANKS

Figure 3

INFORMATION MAINTAINED IN OSWER

- SITE INVENTORIES, PERMITTING, ENFORCEMENT ACTIONS, CLEANUP STATUS
- SCIENTIFIC AND TECHNICAL STANDARDS
- LABORATORY ANALYSES RESULTS
- STATUS OF HAZARDOUS WASTE FACILITIES NATIONWIDE
- ENVIRONMENTAL MODELS, GEOGRAPHIC INFORMATION SYSTEMS, EXPERT SYSTEMS, DISPERSION MODELS

Figure 4

INFORMATION MANAGEMENT STAFF - RESPONSIBILITIES

- POLICY DEVELOPMENT AND IMPLEMENTATION
- IRM PLANNING
- APPROVAL OF PROCUREMENTS
- GUIDANCE AND PROCEDURES
- REVIEW OF ACTIVITIES AND PRODUCTS

Figure 5

GUIDANCE AND PROCEDURES

- LIFE CYCLE MANAGEMENT GUIDANCE
- DATA ADMINISTRATION PROGRAM

Figure 6

LIFE CYCLE MANAGEMENT GUIDANCE

- NINE STAGES
- OBJECTIVES, ACTIVITIES, DECISIONS, PRODUCTS
- REVIEW OF ACTIVITIES AND PRODUCTS
- MANAGEMENT REVIEW AND APPROVAL AT END OF EACH STAGE

Figure 7

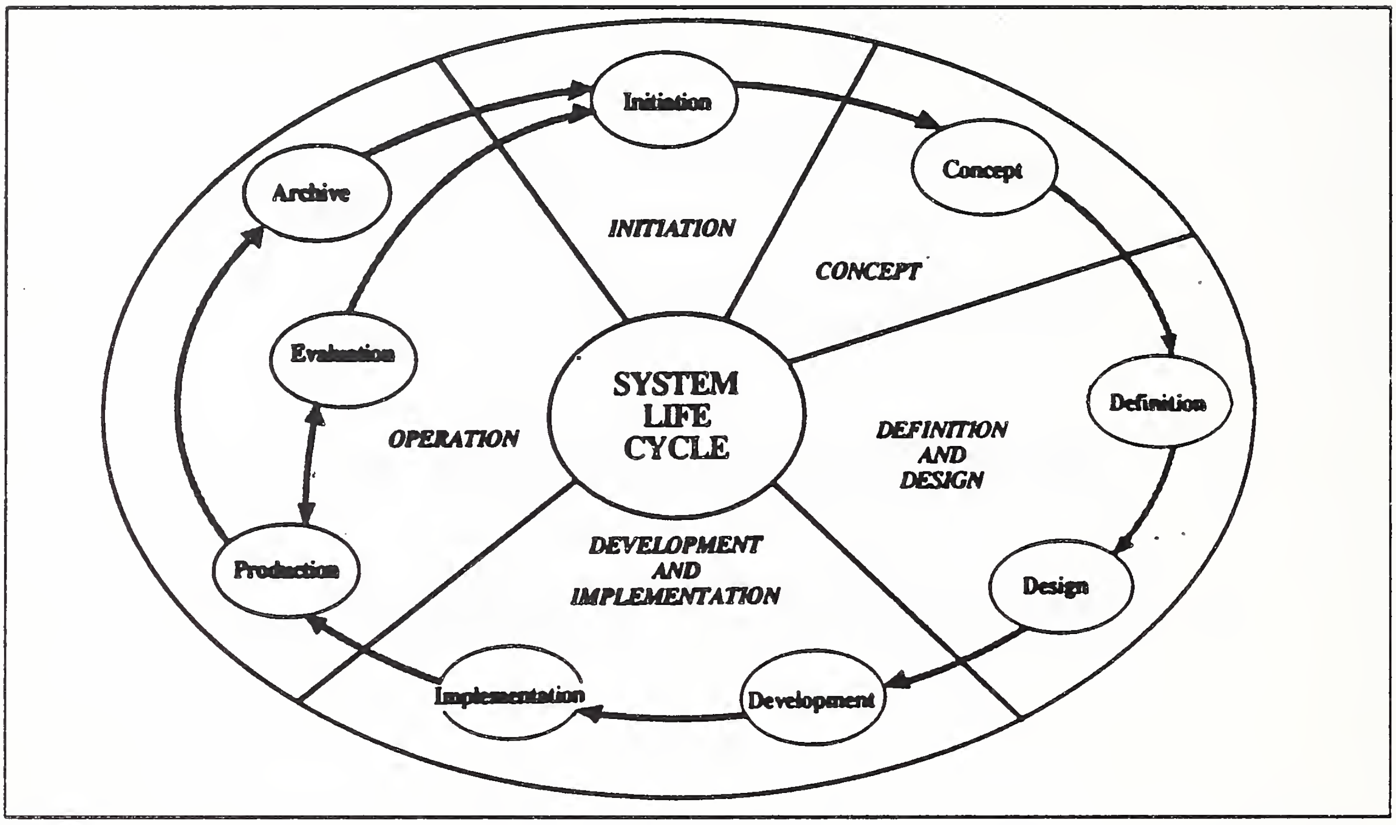


Figure 8

DATA ADMINISTRATION PROGRAM

- **POLICY**

  RECOGNIZE VALUE, STEWARDSHIP, AND CUSTODY OF DATA

- **PLAN**

  SHORT AND LONG TERM/STRATEGIC SYSTEMS PLANNING, CORPORATE DATABASES

- **GUIDANCE**

  INCORPORATED INTO LIFE CYCLE MANAGEMENT GUIDANCE

- **ENFORCEMENT**

Figure 9

DATA RESOURCE DIRECTORY SYSTEM

- DATA DOCUMENTATION DURING SYSTEM AND DATA LIFE CYCLE
- DATA RESOURCE DIRECTORY FOR OSWER SYSTEMS (INCLUDES KEY WORDS)
- PROGRAMMATIC AND TECHNICAL INFORMATION ABOUT OSWER DATA
- STANDARD METHOD FOR DATA NAMING
- DATA INTEGRATION

Figure 10

OSWER'S DRD AND NBS'S IRDS

DRD	IRDS
LIFE CYCLE SUPPORT	**UNCONTROLLED, CONTROLLED ARCHIVED**
REQUIREMENTS DATA DICTIONARY DETAILED DATA REQUIREMENTS	
DESIGN DICTIONARY PHYSICAL DATA BASE DESIGN	
PRODUCTION DICTIONARY PRODUCTION DATA BASE DESIGN	

Figure 11

OSWER'S DRD AND NBS'S IRDS

DRD	IRDS
DRD ENTITIES AND ATTRIBUTES	RDS ENTITIES AND ATTRIBUTES
DATA DICTIONARY FOR OSWER SYSTEMS	DATA MANAGEMENT MODULE
PROGRAMMATIC, TECHNICAL METADATA	IRDS ENTITIES, ATTRIBUTES
DRD NAMING CONVENTIONS	NBS NAMING CONVENTIONS
FUTURE DATA QUALITY INDICATORS	DATA MANAGEMENT MODULE
CONFIGURATION MANAGEMENT	DATA MANAGEMENT MODULE

Figure 12

SUMMARY OF IRDS AS SEEN BY OSWER IM

- IRDS ENTITY, RELATIONSHIP, ATTRIBUTES AND FUNCTIONAL CAPABILITIES WILL BE FOLLOWED
- NBS NAMING CONVENTIONS WILL BE USED AS GUIDELINES
- OSWER DICTIONARIES WILL REPRESENT AN EXPANSION OF LIFE CYCLE-RELATED DICTIONARIES' CONCEPT
- KEYWORDS FOR ENTITIES WILL SUPPORT DATA SHARING, ANALYSIS AND PREVENT DUPLICATION AND CORRESPONDS TO FUTURE DATA MANAGEMENT MODULE
- POSSIBLE FUTURE DATA QUALITY INDICATORS WILL RELY ON DATA MANAGEMENT MODULE

Figure 13

OFFICE OF THE ARMY CHIEF OF STAFF

Speaker
Bruce Haberkamp

[Editor's Note: This talk is a general discussion of data standardization activities within the U.S. Army. One of the major tools being developed in this area is the Army Data Encyclopedia, which is an extension of the IRDS. The Army Data Encyclopedia is described in the next talk, from Lawrence Berkeley Laboratory.]

DATA STANDARDIZATION IN THE ARMY

The Army has formed an Information Mission Area, since it considers information to be a vital asset which must be managed as a resource. The Information Requirements and Data Management Division, Architecture Directorate (ADR) is where I work. ADR is part of the Office of the Director of Information Systems for Command, Control, Communications and Computers (ODISC4), which is located in the Office of the Secretary of the Army. The ODISC4 is the Army's senior policy official for information management and includes the disciplines of records management, printing and publishing, audio-visual, automation, and communications. These disciplines together form the Information Mission Area (IMA). The resources and activities of the IMA are employed in the acquisition, development, transmission, use, integration, retention, retrieval, and management of information.

As Figure 1 indicates, I will talk about data standardization in the Army. I will present the goal of the Army's Information Mission Area; consider the problem of inconsistent data; and describe the approach the Army is taking to deal with this and other data management problems.

IMA Goal

The goal of the Army's Information Mission Area, according to recent guidance noted in Figure 2, is to meet the information requirements of the Army. DoD Directive 7750.5, Management and Control of Information Requirements, defines information requirement as "the functional area expression of need for data or information to carry out specified and authorized functions or management purposes that require the establishment or maintenance of forms or formats, or reporting or recordkeeping systems, whether manual or automated." The IMA goal can be achieved by managing information resources, including data elements, Army wide. It cannot be achieved by allowing separate organizations and systems to collect and handle information independently.

Inconsistent Data

Because information has not been managed as a resource Army wide, Figure 3 displays the current situation: user information requirements are not being effectively and efficiently met. The problem, which created the need for ODISC4 and the IMA, is that data shared across organizational boundaries in the Army did not meet user requirements for consistency. The relatively uncontrolled and unguided development of databases has led to inconsistent data attribute definitions. When data is not consistently defined, it may mean different things to different people. When managers are not able to obtain consistent views of various operations in their organization because of inconsistent data, they are not able to rely on that data for decision making.

Stovepipe Systems

Figure 4 provides us with a prime cause for the problem of inconsistent data: Information systems are not integrated and interoperable within the Army. Interoperability, according to the Joint Chiefs of Staff (JCS Pub 1), is "the ability of systems, units or forces to provide services to and accept services from other systems, units or forces and to use the services so exchanged to enable them to operate effectively together."

For years, information systems were not conceptually or logically designed to be integrated and interoperable Army wide in terms of a common information architecture. The result is a current environment of stovepipe systems. As users with differing needs use data originating in information systems outside their organization, they frequently experience problems with the timeliness and accuracy of the data they receive.

Generally, systems designers within the Army have been left alone to develop data attribute definitions with little or no guidance or coordination from headquarters. When the data is eventually shared across systems, data from one database is transferred to another database with the help of hardware and software support. However, where data attributes are inconsistent, data sharing is not possible even if the hardware and software for information exchange is available. The result is that data which is shared among different databases often is inconsistent in at least four ways:

(1) The same data in different databases has different names. In part this may be due to a lack of guidance and guidelines to ensure consistent naming of data elements.

(2) Different data in separate databases has the same name. This may result because a common set of attributes to manage data is not used in all databases.

(3) The same data is updated in different databases at different times (uncoordinated update cycles).

(4) Information about errors discovered and corrected in one database is not communicated to those in charge of other databases containing the same data.

Complex Organization

Another major cause for the problem of inconsistent data is that the Army is a large, complex organization. As we can ascertain from Figure 5, the Army is indeed a large and complex organization. It consists of numerous units with over two million people serving in one capacity or another. There are 1,254 installations of which 60 are major installations. Over $78 billion was authorized this fiscal year, of which $5.1 billion was allocated to the Information Mission Area. To meet user information needs, there are over 4,000 information systems, not to mention the accompanying databases.

In such a large organization as the Army, information systems and their associated databases have been developed in relative isolation. In the haste to get an information system up and operating, few if any centralized policies and controls have been placed on the development and maintenance of the databases. A rapidly changing environment coupled with organizational complexity increases management's need for information from other parts of the organization and from sources external to the organization. However, information must be available, timely, accurate, and consistently defined in order to be useful.

Need for Data Management

The need to share information across organizational boundaries has increased. Although Figure 6 shows organizational relationships, it could also represent data sharing patterns. This increased need to share data is in part due to the following: 1) new information technologies which allow data sharing by uploading, downloading, networking, etc; 2) an effort by Congress to reduce the number of information collections; and 3) budget cuts.

However, the potential for information sharing is greater than the actual sharing which now occurs. Why? As was mentioned, in the past different information systems have been developed in isolation from each other. Data scrubbing efforts in the Army reveal that data element definitions are vague and poorly documented. Given such conditions, there is a relatively low level of trust that one will receive data from another organization at the level of accuracy and timeliness required.

This current situation of inconsistent data in derivative databases frustrates the achievement of the IMA goal. This situation cannot be tolerated much longer. There is too much data redundancy and waste. The mission and the information requirement still exist. According to DoD Directive 7740.1, DoD Information Resources Management Program, a major policy objective of the DoD IRM Program is to support "decisionmaking with information that sufficiently meets the need in terms of availability, accuracy, timeliness, and general quality."

The need for a data management program which manages information about information resources has never been more critical. Because there is no shared source of information about information resources and no program which matches information requirements with data elements Army wide, such incidents as the following may frequently occur: Two agencies were leasing the same data from an organization in the private sector. Both were paying a lot of money to receive the information each month. By chance it was learned that they had common needs and were receiving the same data. The solution and happy ending? One agency stopped its subscription and instead worked out an arrangement whereby it received the data from the other agency three days later than normal.

So far in this presentation we've seen that the Army's IMA goal is to meet the information needs of the Army. We have looked at the problem of inconsistent data and examined several underlying causes. Let us now examine the Army's approach to solve the problem and satisfy the information requirements of users (Figure 7). The Army's approach includes the following four interrelated components: 1) the Army Data Management Program; 2) the Standard Elements Life Cycle; 3) Data Element Naming Conventions; and 4) the Army Data Architecture.

Army Data Management Program

Figure 8 shows the Army Data Management Program (ADMP), which includes the following: data management policy; a data encyclopedia; data element standards; a data standards life cycle; and quality control and enforcement. The ADMP is one approach the Army is taking to solve the problem of inconsistent data and achieve the IMA goal of meeting user information needs.

Policy and guidance for the Army Data Management Program is contained in Army Regulation (AR) 25-9. AR 25-9 addresses the management of data (whether processed manually or with the use of automation) down to the data element level and concentrates on the identification, definition, and processing of Army corporate data. Army corporate data is that data or information about personnel, equipment, organizations, facilities, services, or dollars that is passed between the organizational blocks depicted in Figure 6 or between different blocks at the same level (e.g., between two installations).

In order to achieve data standardization, the Army Data Encyclopedia (ADE) will provide an automated, on-line repository of information about existing standard elements. This encyclopedia may be queried for standard elements which meet the user's information needs. If a standard element cannot be located which meets the need, an encyclopedia facility will assist the user to create the attributes or documentation for obtaining approval for a new standard element which others also may want to use. Another facility of the encyclopedia will allow a standard element developer to submit a candidate element to those who must review and approve the documentation before the element is allowed to be installed in information systems. This electronic staffing facility will greatly speed up the data standardization process in the Army.

Every standard element has a functional proponent assigned to be responsible for it. According to AR 25-9 (draft), standard elements will be used by new information systems and existing information systems undergoing major redesign. Standardization procedures and standard element attributes provide for the documentation needed to coordinate data sharing across the Army.

The Standard Element Life Cycle contains the phases necessary to define, approve, implement, assess and review standard elements. Each standard element has a standardization status: candidate element, approved element, installed element, or archived element. Each status marks the progress of the element through the Standard Element Life Cycle. For each status there is an additional amount of attribute information available in the Army Data Encyclopedia.

In regards to quality control and enforcement, AR 25-9 (draft) states that compliance with established standard elements will be ensured through formal data management reviews conducted immediately after the data requirements are defined, during the System Design Test (SDT), during the Software Qualifications Test (SWQT) , and during the System Acceptance Test (SAT). Additionally, compliance with standard elements will be monitored during structured walkthroughs and reviews of technical documentation. The goal is to identify problems as soon as possible so that they can be resolved with a minimal amount of redesign. Violations may highlight areas of confusion or disagreement that should be investigated.

Standard Element Life Cycle

Figure 9 shows data standardization in the Army in terms of the Standard Element Life Cycle (SELC). The basic rationale for standardizing data is that data elements developed in one organization should be available to other organizations if those organizations have a need to know that data in order to carry out their functions effectively.

The phases of the SELC relate to the life cycle phases of an information system. According to AR 25-9, data elements required to support a proposed application should be identified in the earliest phases of the information system's life cycle. The system's developer and/or the system's functional proponent should compare these proposed elements to existing standard elements in the Army Data Encyclopedia to determine if existing standard elements can satisfy the information requirement of the application.

The Standard Element Life Cycle (SELC) consists of four phases. Phase 1 of the SELC begins when no standard element is found or a change is required to an existing standard element. Phase 2 begins when the appropriate attribute information is documented and the element is submitted as a candidate for review and approval. Candidate elements which have passed functional and technical reviews are upgraded to approved elements, which begins phase 3. On the assigned installation date, an approved element is upgraded to an installed element, which begins phase 4. Databases will have been modified to accommodate the actual data values based on the newly installed element, and all affected information systems must then operate using updated versions of the databases. During the assessment phase, data will be tracked to determine whether it is still considered essential and is being used. Installed elements become archived elements when they no longer support an information requirement.

Data Element Naming Conventions

Data elements are the smallest units of data that are meaningful and about which characteristics (attributes) are defined. Rules for naming data elements in the Army generally follow guidance found in NBS Special Publication 500-149, Data Entity Naming Conventions [2]. According to AR 25-9 (draft), when information sharing and compatibility are required across two or more information systems, data element names must adhere to the same 15 rules (i.e., Rule 4: The prime term will precede the reference element name in a data element name). In this way, common data elements can be identified.

Figure 10 shows the structured name of a data element. The name is composed of two components: a prime term and a reference element name. The prime term consists of a prime word which may be further modified by qualifiers. A prime term identifies and represents the object or relationship between objects about which the Army wishes to maintain information. An object is represented by a prime word and optional qualifiers which further defines its functional role. A reference element name consists of a class word and optional qualifier(s). The reference element name identifies the domain of values or type of information which can be attached to an object(s).

Army Data Architecture

For several years the Army has been systematically identifying its information needs through carrying out information requirements studies. One of the products of this effort is the Army Data Architecture. Figure 11 portrays the Army Data Architecture.

The Army Data Architecture can be thought of as an entity-relationship model that depicts the fundamental data relationships among 28 subject areas. These subject areas are logical groupings of data concerning those persons, places, things, concepts, events or activities about which the Army wants to keep information.

The Army's Data Architecture was published on October 28, 1987, as the framework for all Army data management and development projects. The Army Data Architecture will be used as the capstone for all data architecture development. Each Army organizational level will develop a data architecture that is appropriate to that level and that relates to the other data architectures. Organizational data architectures will be developed which relate to the data architectures of the next higher echelon.

Since architectures are developed with top leadership commitment and perspective and reflect the information needs of the organization, there will be a significant improvement in future information systems development over the parochial (stovepipe) systems development approach of the past. Properly developed architectures will assist in eliminating the "stovepipe" problems by providing an integrated development approach.

Figures 12 and 13 show the application of the data element naming conventions to personnel-related data elements. Note that the prime word "personnel" relates to one of the 28 subject areas of the Army Data Architecture. The prime word qualifiers specify more precisely the object which the data element is describing. Class words identify the type of information used to describe the object. Besides enabling the user to locate data elements which meet his information requirements, the 15 naming rules result in a data compression and structuring which reduces the number of data elements formulated and submitted for approval as standard elements.

In conclusion, I believe that the approach the Army is taking will change in large measure the current situation of user information requirements not being effectively and efficiently met. The Army Data Management Program, the Standard Element Life Cycle, the data element naming conventions, and the Army Data Architecture are interrelated components to the approach which is aimed at solving the problem of inconsistent data. This approach will aid significantly in the achievement of the IMA's goal of meeting the information needs of the Army.

DATA STANDARDIZATION
IN THE ARMY

BRUCE HABERKAMP
DISC4 (SAIS-ADR)
694-0754

Figure 1

ARMY'S INFORMATION MISSION AREA (IMA) GOAL

"MEET THE INFORMATION NEEDS OF THE ARMY"

Annex C, Information Management Planning (IMP) Guidance 29 Jan 88

Figure 2

PROBLEM STATEMENT

USER INFORMATION REQUIREMENTS ARE NOT BEING EFFECTIVELY AND EFFICIENTLY MET

Figure 3

WHY?

INFORMATION SYSTEMS ARE NOT INTEGRATED AND INTEROPERABLE WITHIN THE ARMY

Figure 4

THE FY88 ARMY IS A LARGE ORGANIZATION

UNITS

- 28 COMBAT DIVISIONS (18 ACTIVE, 10 NATIONAL GUARD)
- NUMEROUS ARMY RESERVE UNITS

PEOPLE

- 781,000 ACTIVE ARMY PERSONNEL
- 469,000 ARMY NATIONAL GUARD PERSONNEL
- 679,500 ARMY RESERVE PERSONNEL
- 445,000 CIVILIAN PERSONNEL

2,374,500 TOTAL

Figure 5

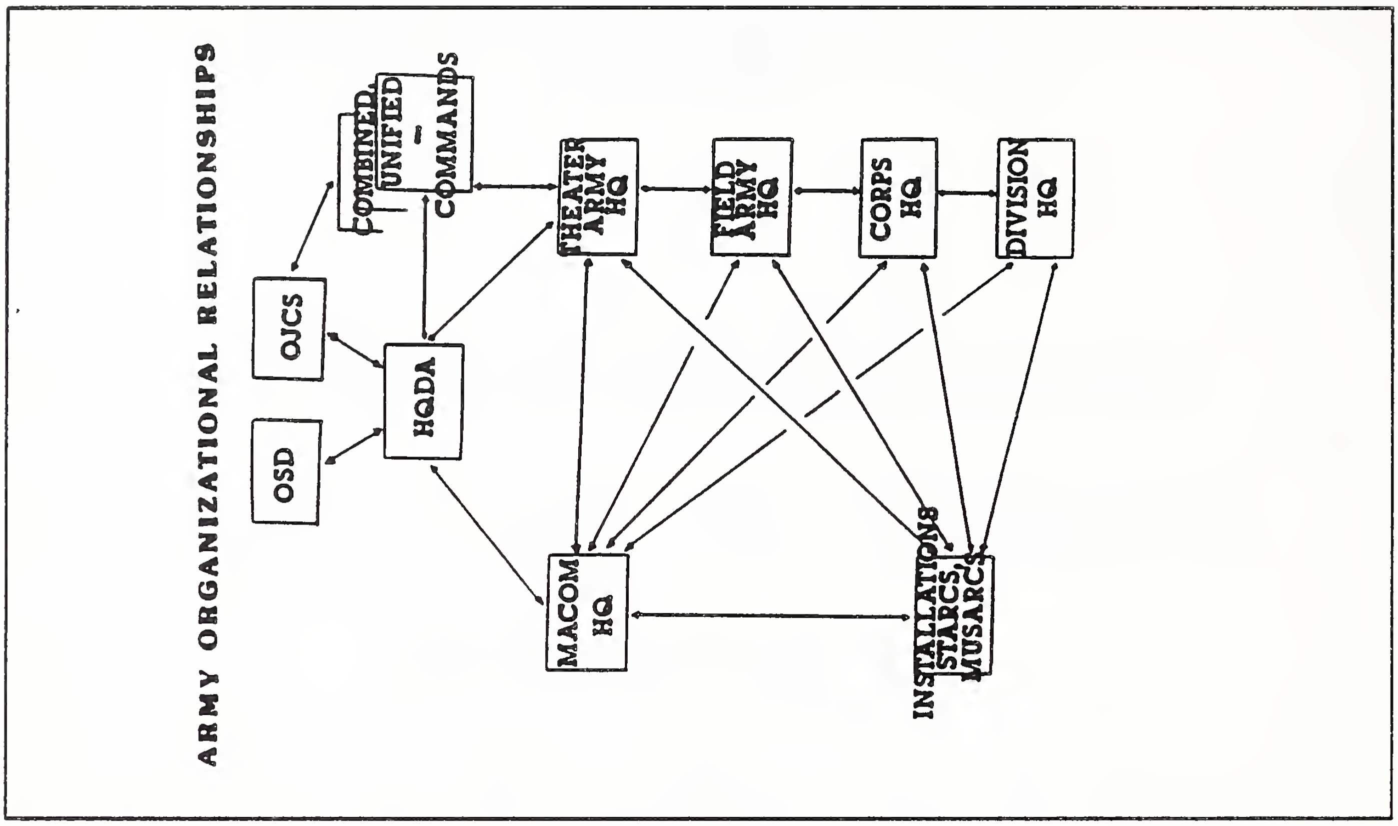


Figure 6

ARMY'S APPROACH TO

SOLVE THE PROBLEM

Figure 7

ARMY DATA MANAGEMENT PROGRAM

- **DATA MANAGEMENT POLICY**
- **DATA ENCYCLOPEDIA**
- **DATA ELEMENT STANDARDS**
- **DATA STANDARDS LIFE CYCLE**
- **QUALITY CONTROL AND ENFORCEMENT**

Draft Army Regulation 25-12
17 Nov 87

Figure 8

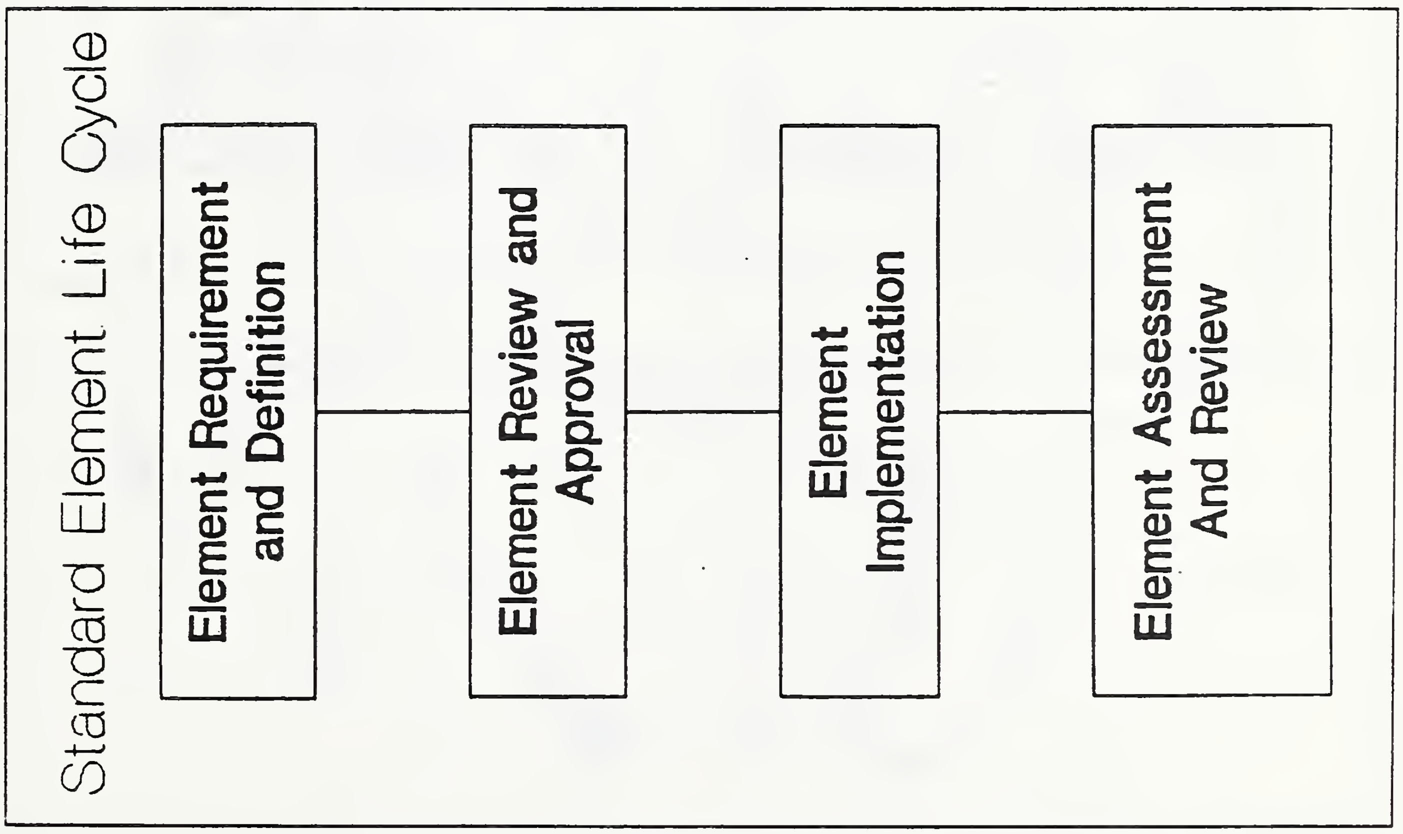


Figure 9

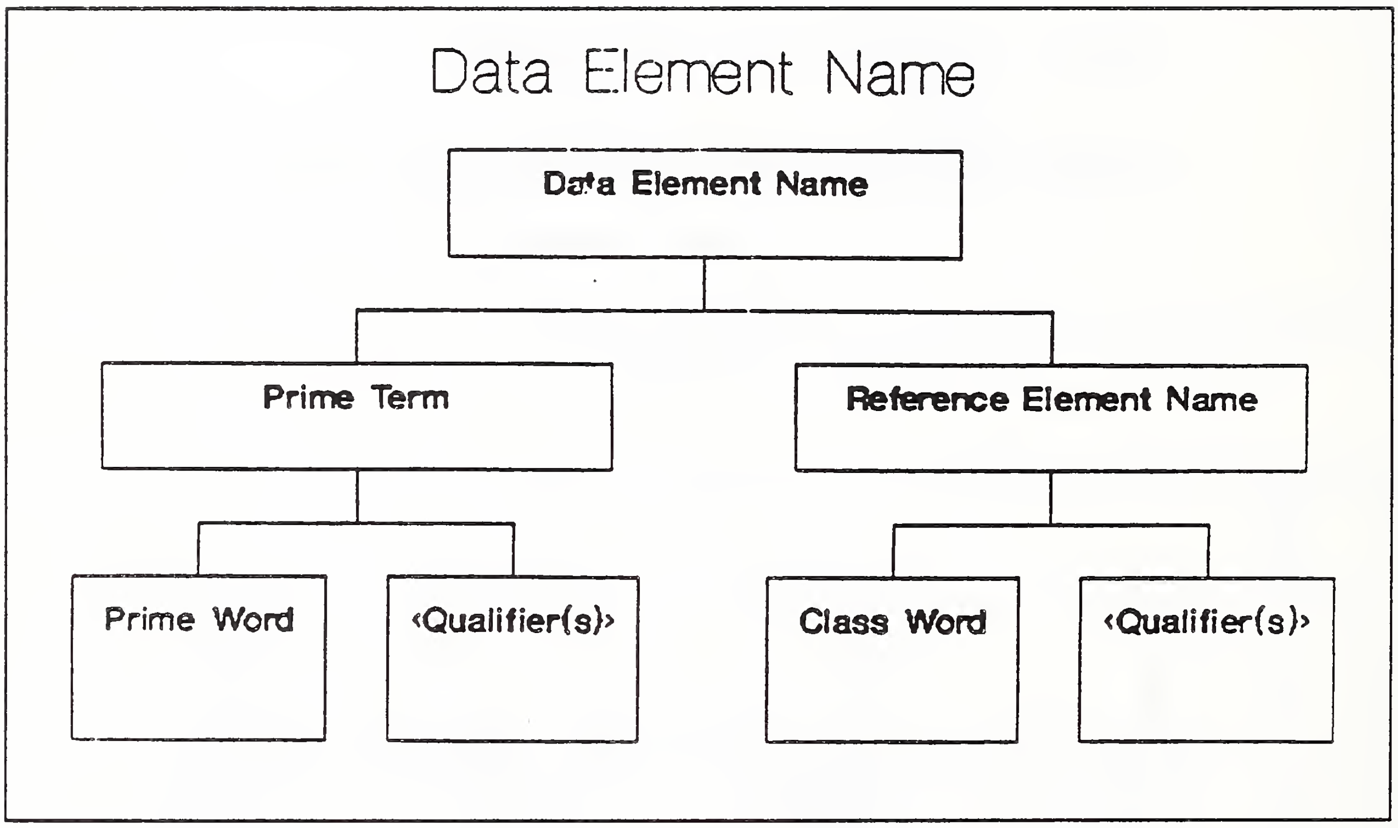

Data Element Name
Data Element Name
Prime Term
Reference Element Name
Prime Word
<Qualifier(s)>
Class Word
<Qualifier(s)>

Figure 10

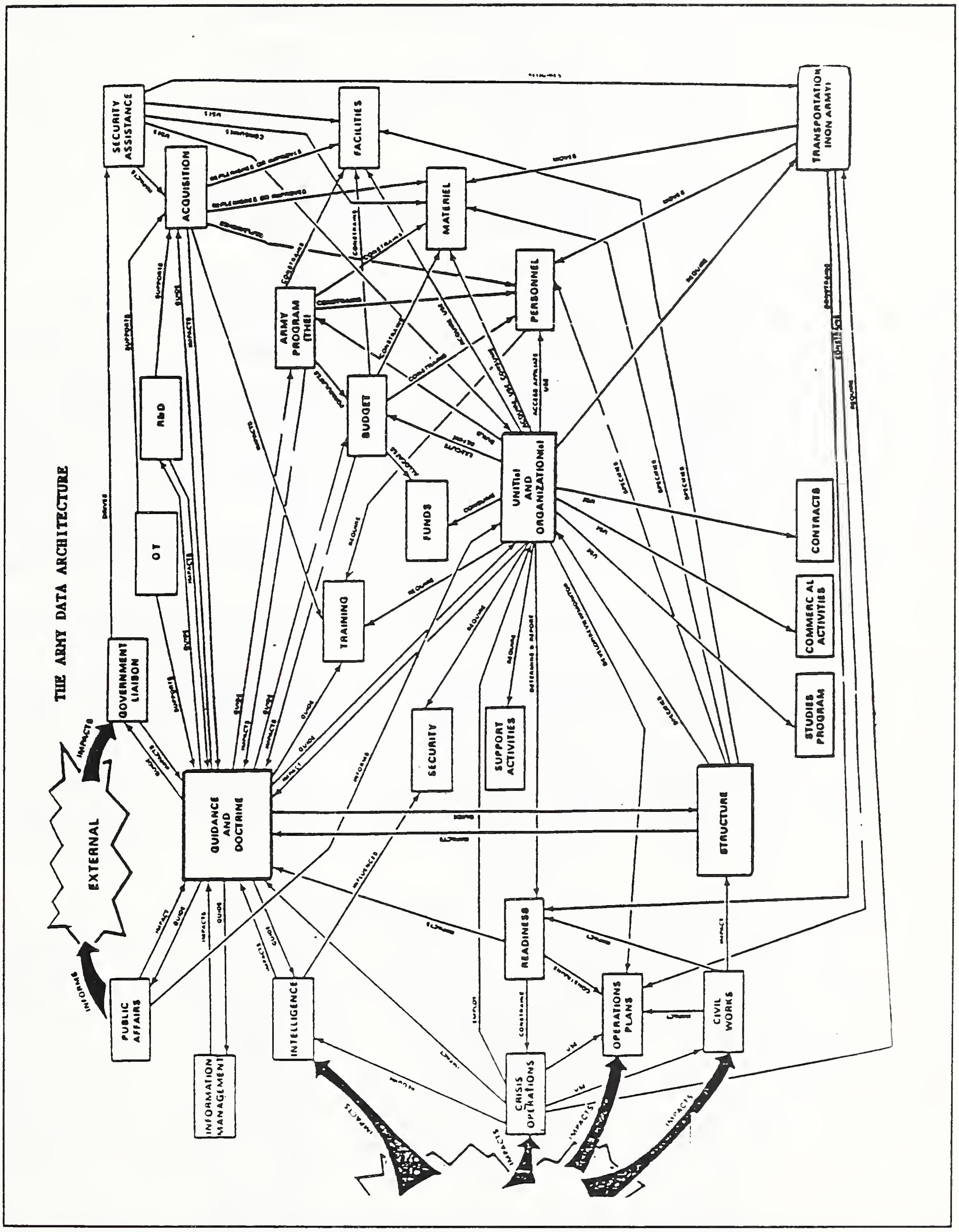

THE ARMY DATA ARCHITECTURE
SECURITY ASSISTANCE
ACQUISITION
FACILITIES
MATERIEL
PERSONNEL
TRANSPORTATION (NON ARMY)
ARMY PROGRAM (THE)
BUDGET
R&D
OT
FUNDS
UNIT(s) AND ORGANIZATION(s)
CONTRACTS
COMMERCIAL ACTIVITIES
STUDIES PROGRAM
TRAINING
GOVERNMENT LIAISON
SECURITY
SUPPORT ACTIVITIES
EXTERNAL
GUIDANCE AND DOCTRINE
STRUCTURE
READINESS
PUBLIC AFFAIRS
INFORMATION MANAGEMENT
INTELLIGENCE
CRISIS OPERATIONS
OPERATIONS PLANS
CIVIL WORKS
IMPACTS
INFORMS

Figure 11

	PRIME WORD	QUALIFIER(S)	CLASS WORD
1.	Personnel	Absence Reason	Identifier
2.	Personnel	Service Term	Identifier
3.	Personnel	Absence Start	Date
4.	Personnel	Absence Stop	Date
5.	Personnel	Absence	Duration
6.	Personnel	Weight Control Profile	Year-Month
7.	Personnel	Weight Control Profile	Identifier
8.	Personnel	Color Vision Defect	Identifier
9.	Personnel	Visual Acuity Correctability	Identifier
10.	Personnel	Physical Readiness Test	Indicator
11.	Personnel	Physical Readiness Test	Score
12.	Personnel	Physical Readiness Test	Year-Month
13.	Personnel	Skill Qualification Test	Score
14.	Personnel	Skill Qualification Test	Percent
15.	Personnel	Dental Visit	Date
16.	Personnel	Marksmanship Class	Identifier

Figure 12

17.	Personnel	Marksmanship Class	Year-Month
18.	Personnel	Life Insurance Value	Amount
19.	Personnel	Pay Grade	Identifier
20.	Personnel	Pay Grade	Date
21.	Personnel	Promotion List	Serial Number
22.	Personnel	Promotion List	Year-Month
23.	Personnel	Training Program	Identifier
24.	Personnel	Award	Identifier
25.	Personnel	Award	Count
26.	Personnel	Award	Class
27.	Personnel	Assignment Area	Identifier
28.	Personnel	Assignment Preference	Identifier
29.	Personnel	Assignment Command	Identifier
30.	Personnel	Assignment Start	Date
31.	Personnel	Assignment	Duration
32.	Personnel	Assignment Position	Name
33.	Personnel	Assigned Service	Identifier
34.	Personnel	Skill	Identifier

Figure 13

LAWRENCE BERKELEY LABORATORY

Speaker
Frederic Gey
Computer Science Research Department

A DATA ENCYCLOPEDIA ARCHITECTURE WITHIN AN EXTENDED IRDS FRAMEWORK

Let me begin with just a little bit of background about Lawrence Berkeley Laboratory (LBL). We're a national laboratory of the Department of Energy, operated by the University of California, and we do work for the Department of Energy and also for other agencies under inter-agency agreements with the Department of Energy. The work that I'm going to describe is funded by the Army's Information Systems Engineering Command Data Management Directorate in Fort Belvoir. The Laboratory's mission for the Energy Department is to do high quality general science and advanced technology research.

Encyclopedia Configurations (Passive/Active)

What I'd like to do is generically address the architectural components of a data encyclopedia. You may ask what a data encyclopedia is, and is it different from an information resource dictionary system? I'll try to describe that as I go along. An encyclopedia, or an information resource dictionary extended, provides the beginnings of data integration and data and metadata integrity.

In a passive configuration (Figure 2), we have an encyclopedia that has software and data components which support data administrators, operational systems, and software design systems through an export-import capability. Therefore, design data in external design systems can be imported and exported, and standards can be applied to maintain authoritative design data and authoritative operational metadata for operational systems. The latter include production systems, reports, forms, queries, decision support systems, and tools to operate the organization.

In an active configuration (Figure 3), the encyclopedia becomes a layer between these various systems, and so all the organizations go in-and-out of the encyclopedia to obtain their data. Operational systems may or may not go directly through the encyclopedia for operational data because, for efficiency reasons, it may be important to have the operational system apply directly to the operational data without a layer that would introduce inefficiencies. This is done by allowing the metadata to be compiled into the operational system. That is, an operational system doesn't go on-line unless its metadata organization is compiled into its operation which thus controls the methods by which it accesses the operational data.

Encyclopedia Metadata

In order to do this, you need metadata layers of control. Data at a particular level of abstraction becomes metadata, i.e., structure and control information for the next lower level of abstraction. You need tools to design, model, and manage database applications using these layers of abstraction to organize and understand systems. The layers of metadata and data form a continuum; however, they're generally represented by a few major layers.

The benefits of viewing this metadata as data provides you with interoperability, synchronization, system integration, the advantages of simulation and, certainly, impact analysis. The layers of abstraction of metadata in an encyclopedia form a multidimensional prism (Figure 5). Everything, from the general to the specific, supports the operation of the business: business description, business operations, requirements analysis, software (in the sense of all things which fall within the software development life cycle), and the technology and networks that implement actual production information systems. One might say that it's not clear where requirements analysis falls. I'd say that global information requirements of an organization drive the development of systems, yet there are also specialized requirements analyses needed for a particular system.

We see the vertical dimension as three layers from the more abstract to the more concrete: a conceptual understanding of operations, a logical description of these operations, and then a physical or system description of these operations. This organization is facilitated by building on the existing IRDS functionality to create an information resource thesaurus, and adding those functions for those entities and relationships that address specific business needs. This has a lot in common with John Zachman's framework for information systems architecture. If any of you are not familiar with it, it was in the IBM Systems Journal in March, 1987. A very fine article.

Encyclopedia Structure and Operation

The Encyclopedia consists of a data repository interfaced with a set of tools (Figure 6). These tools may be decision support systems, query (assistance) tools, systems modeling tools, CASE tools, automated data standards tools--that is, tools for manipulation and utilization of the metadata stored in the Encyclopedia.

The Encyclopedia stores IRDS entities, relationships, and attributes, as well as extensions for subject search and navigation of the database through subject terms--narrower, broader, etc.--and encyclopedia domain and integrity information. Domain and integrity information is central to the reliability of the data of the organization. In this sense, we want to capture not only data integrity in operating systems, but we want to capture data integrity in the operation of the business itself.

Take, for example the acquisition process in a government organization: you want to have stored into the Encyclopedia rules that say "approval is required by a certain organization if the cost is greater than a certain amount" and "a request for proposal is required for costs greater than some other amount." We wish to capture all these business rules in addition to local software system integrity.

The operation of the Encyclopedia (Figure 7) has data administrators operating through the panel or command languages, and system services which support data administration standards and validation and approval authorization--that is, the approval of information belonging to the actual functional areas of the organization. You have to have import and export for data exchange to outside users. This operation is layered over an IRDS which calls a DBMS which accesses the metadata.

National and International standards should guide the development of the Encyclopedia (Figure 8). Entities, attributes, and relationships are the purview of the X3H4 committee IRDS. The import/export--the IRDS interchange--follows the standards given by ISO 8824/25, ASN.1. Attributes of data elements fall in ANSI X3L8. Conceptually, there are possible mappings and transformations to the ANSI-SPARC 3-schema model, and the implementation is done within NDL or SQL, which is the ANSI X3H2 committee. LBL standards activities for the Encyclopedia have included (Figure 9):

- o Participation in the X3H4 IRDS standards committee. In particular, we are specifying the IRD-IRD export/import format using ASN.1.
- o A research associate activity with NBS where we're doing data encyclopedia experiments with IRDS prototype software. We hope to adapt the IRDS prototype to the IBM-PC. We hope to adapt the prototype software to call embedded SQL instead of using the existing CALL interface--if this is done, the code will be potentially portable to other database management systems and, over a network, to SQL servers.
- o Some participation, just beginning, in X3L8, the subtask on the classification and attribution of data elements.
- o Participation, also just beginning, on the IEEE CASE interchange task force.

IRDS Export-Import Specifications

Figures 10 through 14 begin specifying the IRD-IRD export/import format. As you can see from Figure 10, what we're doing is building a file that contains the schema--the "S" schema--that defines the entities and then the relationships. These will expand into the entity definitions and each of the attribute definitions, followed by the data instances of each one of those entities.

For example, if the entity is a SYSTEM, then there would be a data instance of say, a personnel system, followed by all the attributes of that personnel system, then possibly a payroll system followed by all the attributes of that system, and then a relationship between these systems. The way this gets specified at the meta-schema level is through the language of ASN.1.

More specifically, for IRDS entities (Figure 12), a RECORD can be defined as the record name and all the attributes associated with the RECORD, an ELEMENT can be defined as the element name and its attributes, a CONTAINS relationship can be defined as the two entities, the RECORD, the ELEMENT it contains, the attribute giving the relative position, and possibly other attributes.

Now, in Figure 13, we get down to the actual data. PAYROLL contains certain ELEMENTs. The PAYROLL RECORD contains the name and the social security number (SSN), the PERSONNEL RECORD also contains the name and the social security number, the PAYROLL RECORD also contains the hours worked. This encodes, in ASN.1, as the kind of string shown in Figure 14. It finally compiles into a binary representation that's very compact.

Organization of the Encyclopedia

In the organization of the Encyclopedia (Figure 15), we see the Core Schema: ELEMENT, SYSTEM, PROGRAM, FILE, RECORD, USER; relationship-types; and then extensions to support the business needs: Mission and Policy and, at the hardware level, the Network_Node and the Protocol. Also there are certain new relationships, such as a certain mission is carried_out_by a certain suborganization, a certain file may be replicated_at multiple nodes across a network. You then get some systematic description of data fragmentation of a distributed database system. Encyclopedia extensions also support the Thesaurus, with such definitions as DOMAINS, SUBJECT TERMS, and relationships such as BROADER_THAN, NARROWER_THAN, and RELATED_TO.

Now, what we see stored in the IRDS (Figure 16) are the familiar layers of abstraction: The meta-schema, the IRDS's description of itself, which is implementation dependent, and the Schema, which is extensible and contains the various "types." You can only add entity-types, attribute-types, and relationship-types. Then, within the IRDS data layer, you have all those things that support the layers of abstraction of the Encyclopedia. In other words, all Encyclopedia abstractions must compress down to this one layer of the IRDS. Therefore, in order to define and store a real multi-dimensional view of an organization, you have to expand this layer into something that's more than one-dimensional.

The Thesaurus View of the Encyclopedia

For Thesaurus extensions to the IRDS Functional Schema (Figure 17), we would add SUBJECT_TERM, and have as an attribute the definition of the SUBJECT_TERM (basically the "vocabulary"). We would add the relationship-types: BROADER_THAN, NARROWER_THAN, RELATED_TO (which is "see also"), and USED_FOR. The benefit of this kind of extensibility is that it provides for navigation of the objects of the Encyclopedia, and it also provides a capability of data purification. That is, you can index on all the descriptions, and then find related objects described by the same SUBJECT_TERMS.

The functionality of the Thesaurus (Figure 18) gives you: hierarchical groupings of objects in the Encyclopedia, a subject oriented search, access via multiple rather than one hierarchy, a glossary and definition, and resolution for synonyms, homonyms, and aliases. At the System level you have Integrity Constraints--table lookup of valid values. For data manipulation, there is Units Conversion Information. This goes beyond the usual idea of a thesaurus, so we've coined the term "Data Thesaurus."

Domain Integrity in the Encyclopedia

Domain Information (Figure 19), sometimes referred to in the DoD community as "data items," has lists of codes and valid values. At the Thesaurus level it would have Controlled Vocabularies--lists of valid entries, class words, prime words, modifiers, and other definitions. In our mind, both domains and elements follow a hierarchy as well. As shown in Figure 20, a conceptual level is called the "reference element." The logical level is the data element, and there is a physical or a system level--what actually is utilized in an information system. Similarly for domains: we have domain classes, domains, and actual lists of values and codes. As an example of that (Figure 21) we have, for a particular attribute, the various domain classes attached to it. In other words, a whole group of data elements could be in the time area, i.e., their units are of the time category, but specifically what they are is unspecified. Similarly, we have the length category, meters, feet, and miles, and the temperature category. Attributes can be Minimum, Maximum, a Standard attribute or let us say a standard measurement unit, and various Conversion factors. Units have an important place in information systems. An example given to me by a Canadian data administrator is that of an Air Canada plane that was refueled in Toronto and had to emergency land in Manitoba, halfway to its destination of Vancouver, because it got filled up in liters instead of the requested gallons.

Limitations of the Existing IRDS Standard

We have some questions and concerns about the limitations of the IRDS (Figure 22). It allows only binary relationships between entities, although it's quite possible to have relationships in which multiple entities participate. The family relationship is a good example, where parents, grandparents, aunts, uncles--whole classes of entities could participate in a single relationship. Attributes can't be entities. Data flow diagrams contain "operations," and real time command and control is more complicated.

As an example (Figure 23), consider systems like personnel and payroll which have a number of entities in common. If you have information that's shipped from one system to another, which can be defined by a list of entities, the list of entities should really hang off of that relationship. This would denote the information transfer from one system to another system. It can be implemented by an attribute which is a pointer to an entity list, but that's kind of faking it. To actually implement this within an IRDS, you must take the relationship and decompose it into two additional relationships R_2 and R_3, where the entities actually exist within the relationship definition.

Capturing and Implementing CASE Constructs

Figure 24 illustrates a simple purchase order process. We're interested in this diagram because we've talked to the developer of it and they've implemented data flow capture within their system, which is quite nice. You can have entity-relationship diagrams for the purchase order process (Figures 25 and 26), but to really capture the process, you have to have a data flow diagram that shows the operation. The data administrator of Atomic Energy of Canada recently visited us. He has extended the entity-relationship model to create what he calls the "entity-relationship-operation model" where in hanging off particular entities you can have operations, and the operations can be triggered by actual triggers within your database management system when you invoke that entity.

Question: Isn't that really the notion of message passing and object oriented programming?

Answer: It probably is. My understanding of an object oriented program is that you attach some rules to objects.

Trigger mechanisms are right on the borderline of what needs to be done with the IRDS to make it more general in the direction of object oriented programming.

Unresolved Issues and Research Areas

Some other unresolved issues (Figure 27) include temporal data semantics, real time decision support, mappings and transformations--that is how do you actually do these mappings and transformations from entity-relationships to SQL and NDL, and advanced text retrieval. If we consider real-time requirements analysis, we have a situation where events arrive either synchronously or asynchronously to a place where a decision has to be made and responses have to be sent out. I visualize this event as a triple of an entity, a desired action, and a desired result. I don't see, off the top of my head, how to model the synchronicity or asynchronicity of these events using the existing IRDS.

In the area of advanced text retrieval applied to Encyclopedia development, we have Figure 28. Most text retrieval systems operate on a boolean model in which the objects are described by text terms. In the Figure, three terms are used to describe eleven objects, which classifies them into eight sets. If you want to retrieve the objects, you intersect the terms through combinations of ANDs, ORs, and NOTs. It is well known in the information retrieval community that this can also be represented as a vector space in the minimal number of descriptive terms, and you can then introduce similarity measures. You can calculate Euclidean distances between these various objects and get better retrieval. You can also weight the terms in a particular request so that you can get a Retrieval Status Value (RSV) attached to every object in a collection. Then you can order objects by highest retrieval status value, and this procedure will yield higher relevance of the object to a particular request. It's been done in experimental text systems for about 20 years now, but no production system uses the method.

Question: Doesn't that mean anything?

Answer: My opinion is that the large information services form an oligopoly, and they don't want anything interfering with their bottom line!

Question: You mentioned quite a bit in the latter part of your talk about potential or desirable extensions or enhancements of the IRDS. Of course, many or most of the things that you discussed are things that have been kicked around from time to time, such as n-ary relationships.

<u>Answer</u>: I must say, we're kind of new guys on the block.

<u>Question</u>: Actually, my question refers to a point you brought up a little earlier in your talk, where you were talking about the Encyclopedia and certain extensions to the IRDS either in the direction of a thesaurus or otherwise. Do you think that any or all of those are candidates for standardization, or was it merely that for this particular type of application, you need to add more to the IRDS, without the thought that it might become a "standard"?

<u>Answer</u>: I think that those generic extensions which increase the functionality of the Encyclopedia might be considered for standardization. One would have to be very careful about extensions specific to business missions.

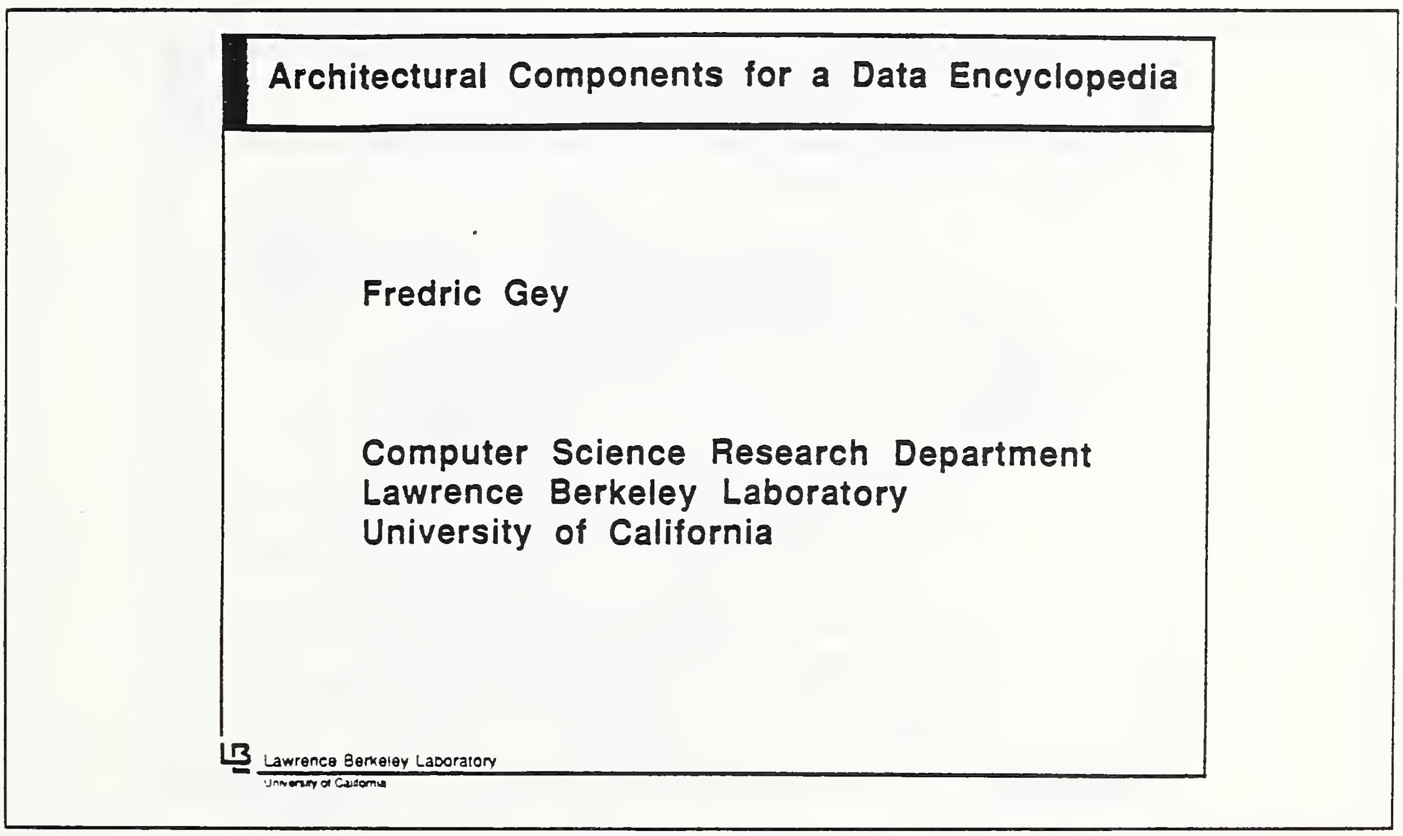


Figure 1

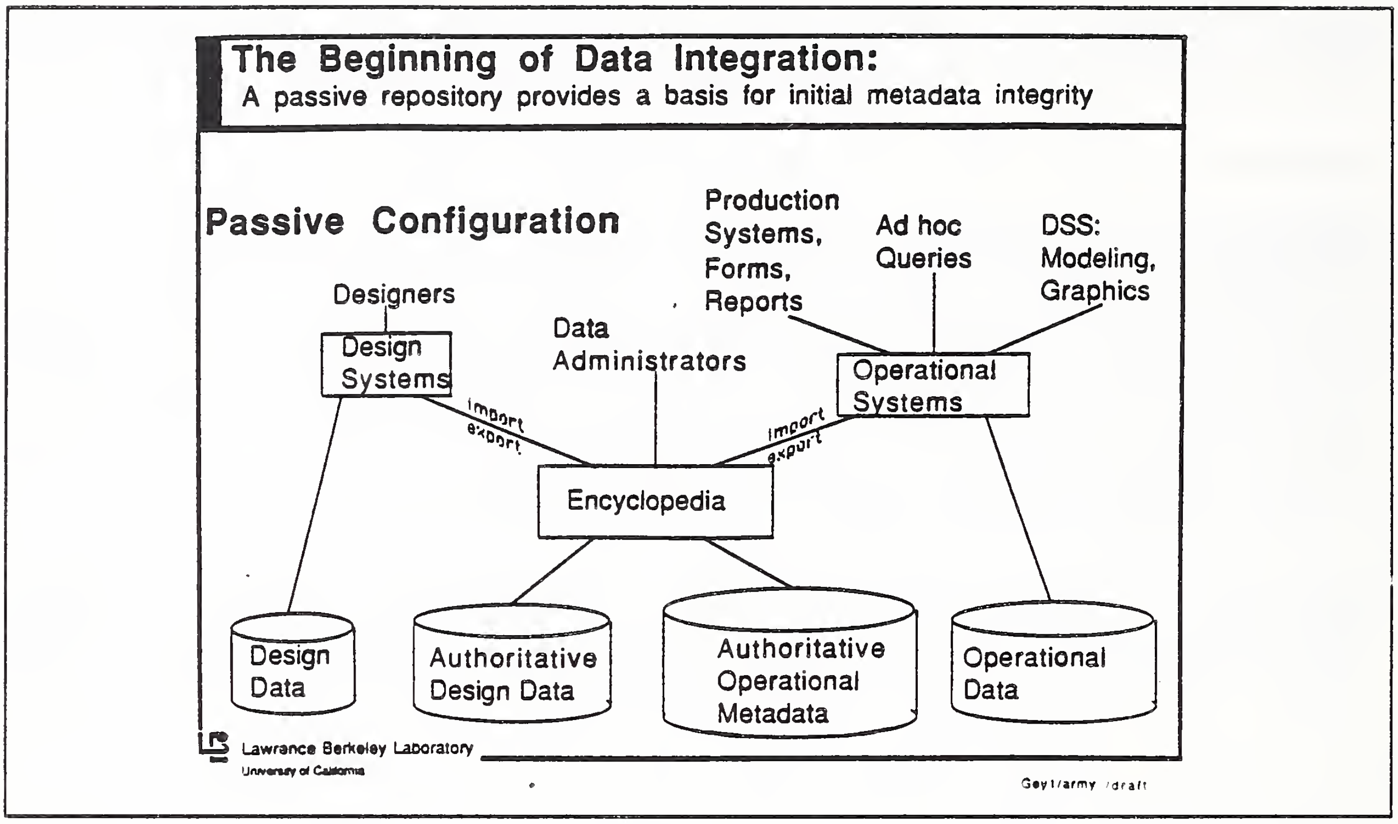


Figure 2

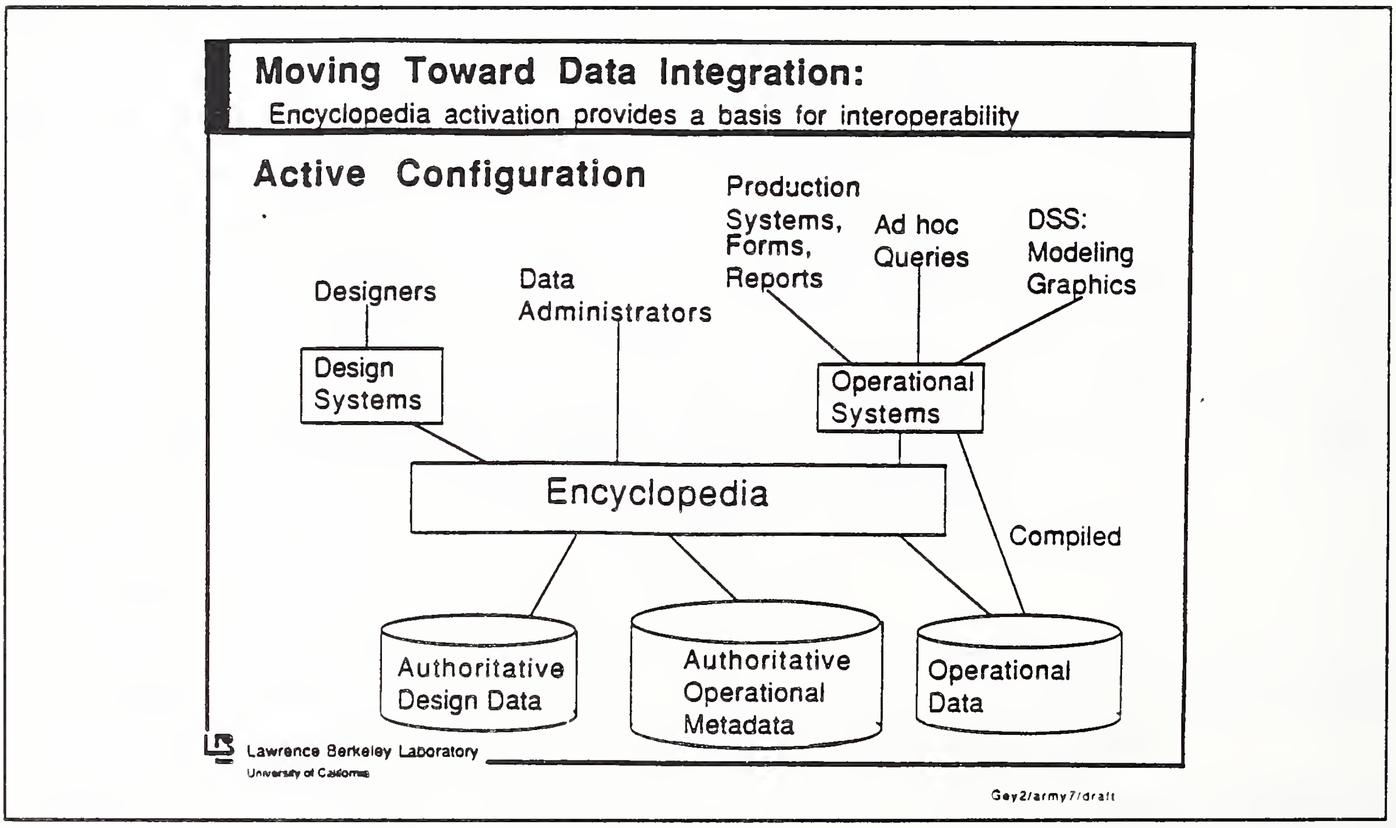


Figure 3

Metadata-Data Layers of Control

- Data at a particular level of abstraction becomes metadata (structure and control information) for the next lower level of abstraction.
- Tools to design, model, and manage (database) applications use layers of abstraction to organize and understand systems.
- The layers of metadata/data form a continuum.
- What are the benefits of viewing metadata as data?
  - Interoperability
  - Synchronization
  - System Integration
  - Simulation
  - Impact analysis

Lawrence Berkeley Laboratory
University of California

Gey3/army11/draft

Figure 4

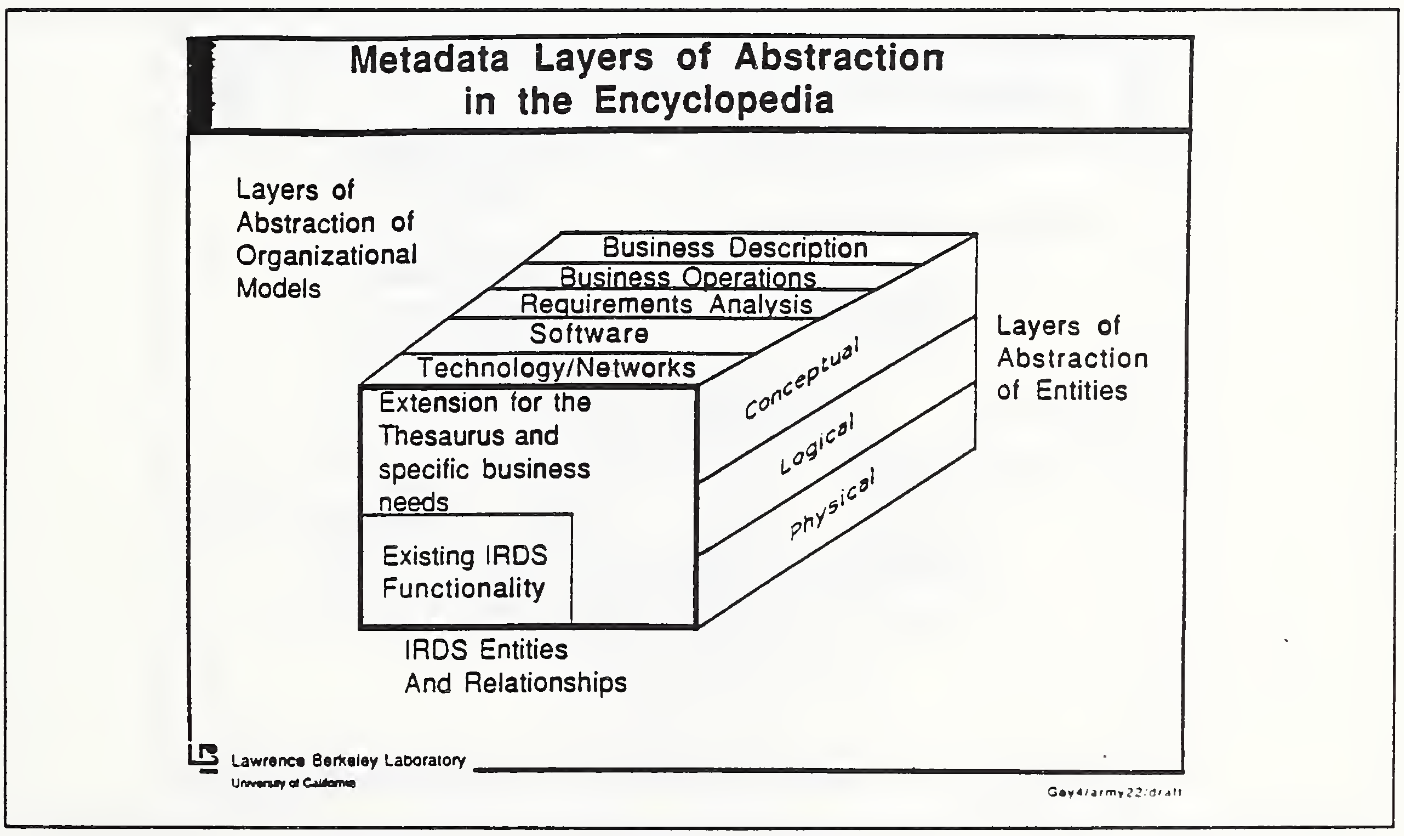


Figure 5

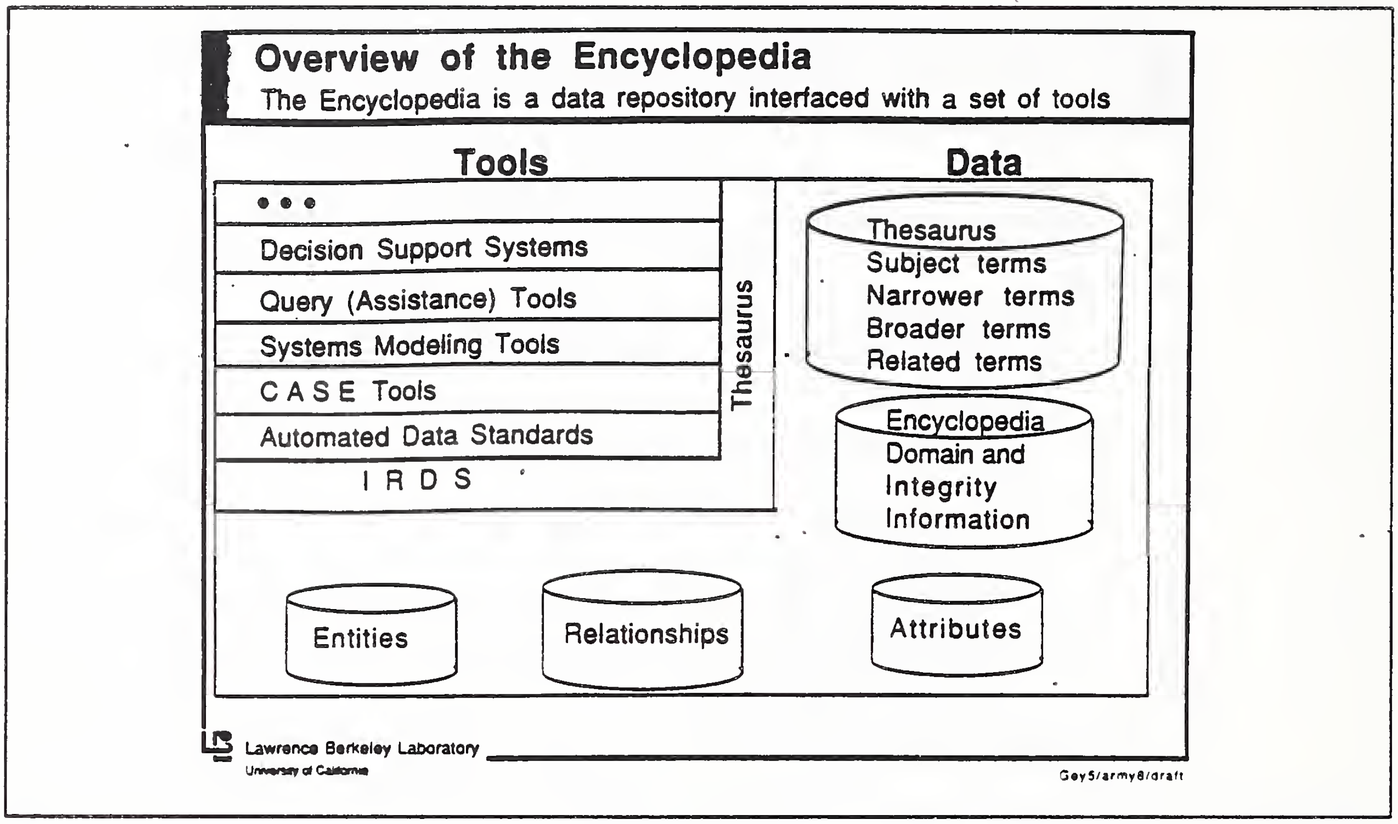


Figure 6

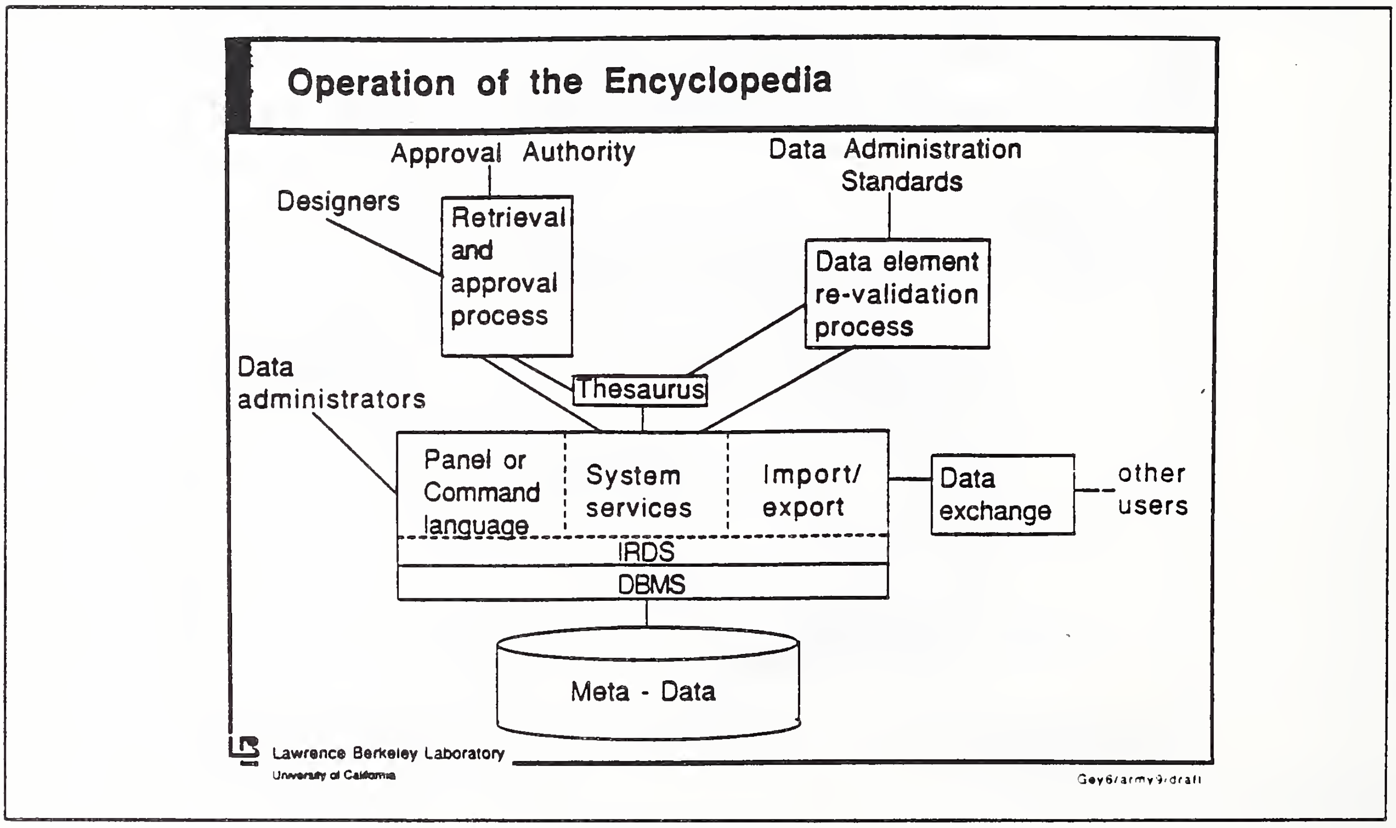


Figure 7

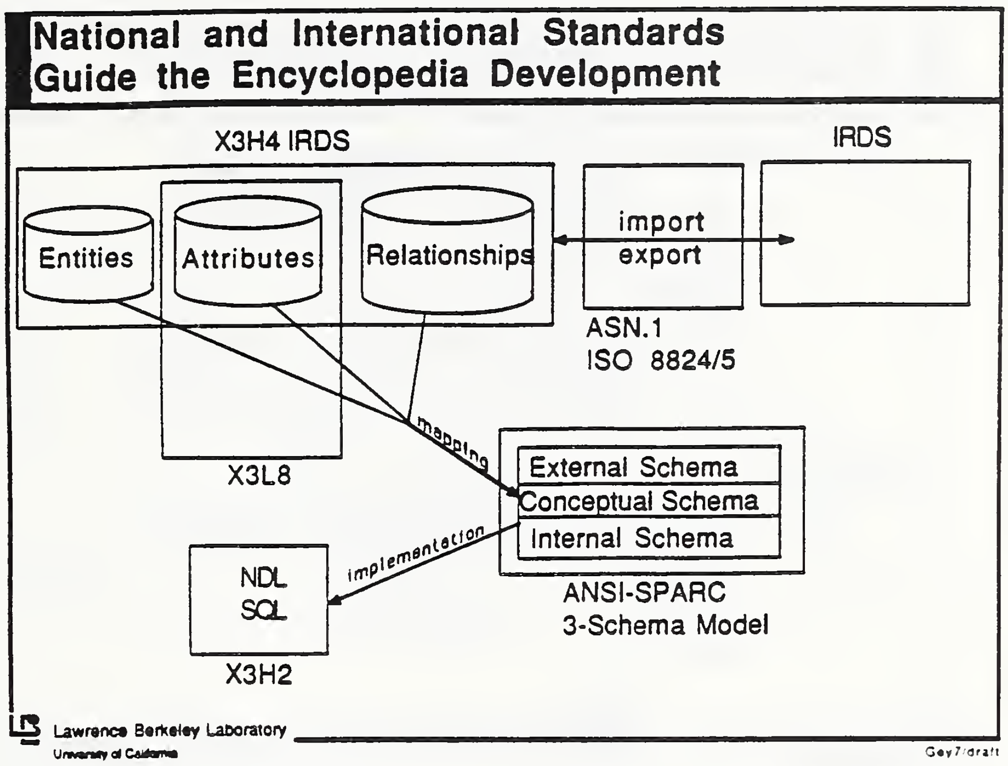


Figure 8

LBL Standards Activities for the Encyclopedia

- Participation in X3H4 IRDS standards committee
  - Specification of IRD-IRD import/export format using ASN.1
- Research Associate Activity with NBS
  - ADE experiments with IRDS prototype software
  - Adapt IRDS prototype to IBM-PC
  - Adapt IRDS prototype to embedded SLQ
  - Experiment with parser generation for IRDS command language
- Participation on X3L8
  - Classification and attribution of data elements
- Participation on IEEE CASE interchange task force

Lawrence Berkeley Laboratory
University of California

Figure 9

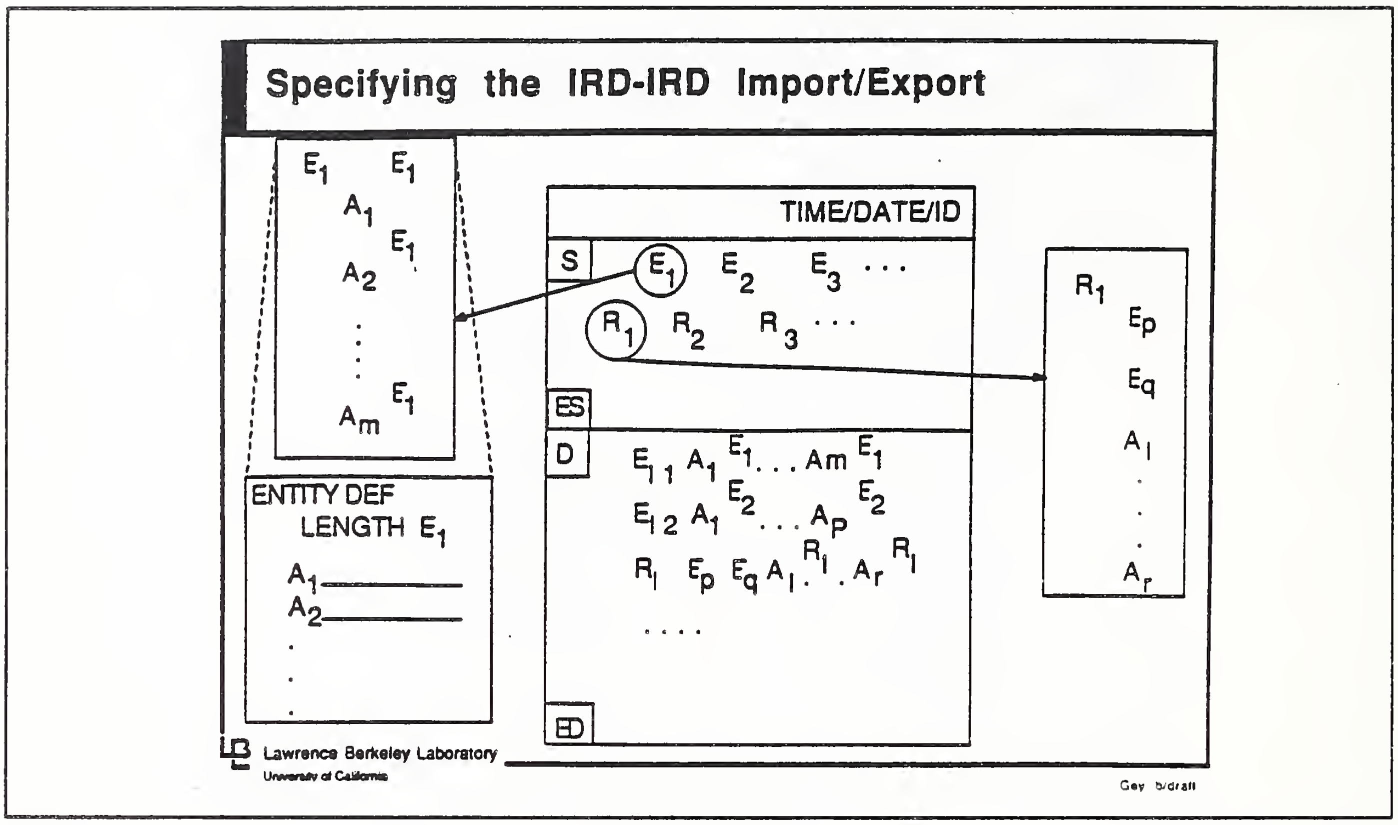


Figure 10

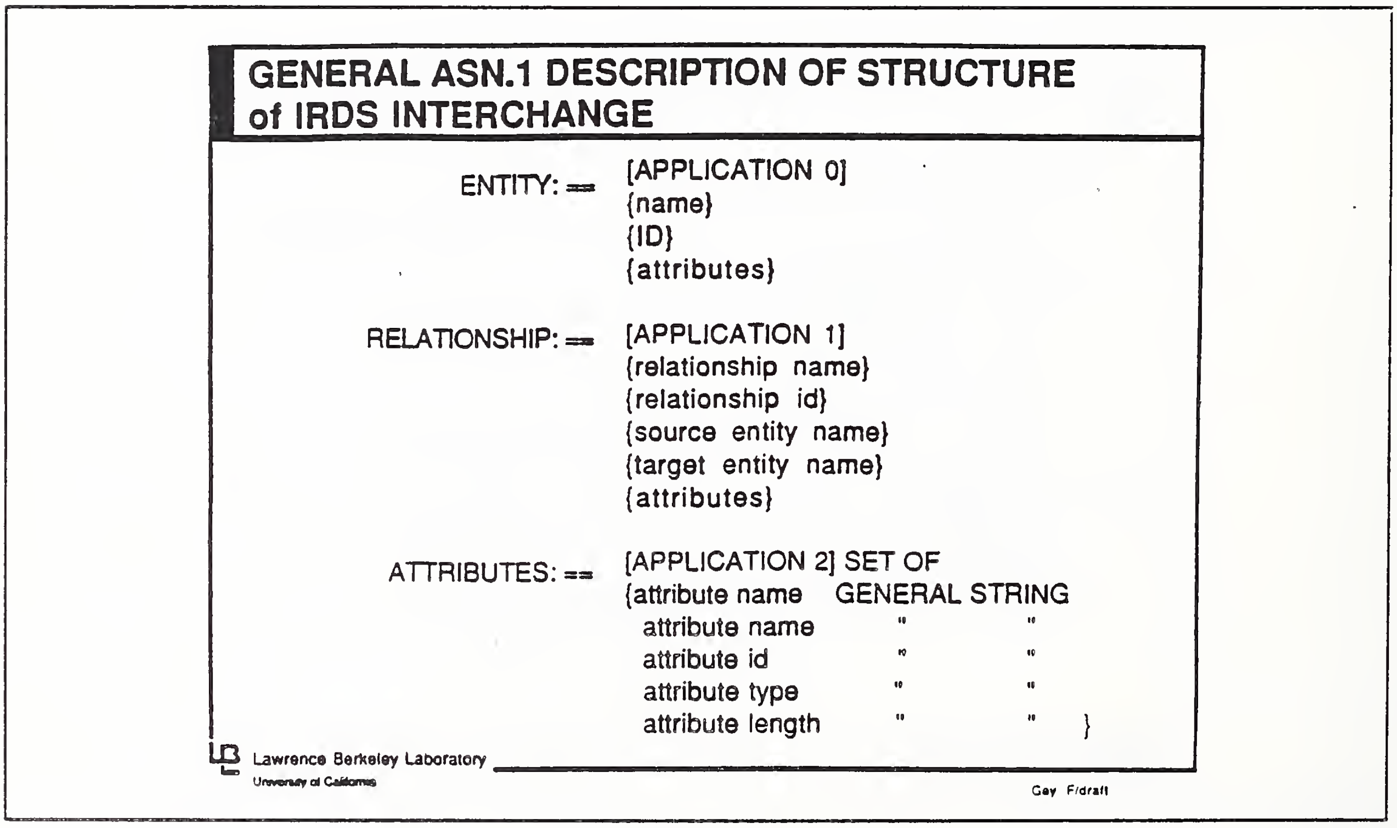


Figure 11

Specific ASN.1 Description of Structure of IRDS Interchange

```
record: ==    [Application 0] IMPLICIT SEQUENCE Type D(E)
                rec_name [0] general string
                A        [1] general string type 2
                .         .      "      "   type 2
                .         .      "      "   type 2
                A        [n]     "      "   type 2
element: ==   [Application 1] IMPLICIT SEQUENCE Type
                e name   [0] general string
                A              "      "   type 2
                .                          .
                .                          .
                A        [m]   "      "    .
contains: ==  [Application 2] IMPLICIT SEQUENCE Type 1
                record   [0] general string Type 0
                element  [2]  "      "   Type 0
                rel_position [2] "   "   Type 2
                Ar       [3]  "      "   Type 2
```

Lawrence Berkeley Laboratory
University of California

Gey C/draft

Figure 12

Some Irds Data to be Exported

CONTAINS

RECORD	ELEMENT	relative position
PAYROLL	name	1
payroll	SSN	2
personnel	name	1
personnel	SSN	3
payroll	hours_worked	3

Lawrence Berkeley Laboratory
University of California

GeyA/draft

Figure 13

ASN.1 Encoding of the IRD data

```
{record  {{rec name "payroll", a "...",..., An "..."},
          {rec name "Finance", . . . .              }},

element  {{ele name "name", . . . .                 },
          {ele name "SSN", . . . .                  }}

contains  {record "payroll", element "name", relpos "1", Ar "a"},
          {record "payroll", element "SSN",   relpos "2", Ar "b"},
          {record "personnel" element "name", relpos "1", Ar "c"},
          {record "personnel" element "SSN", relpos "3", Ar "d"},
          {record "payroll", element "hrs.worked", relpos "3", Ar "e"}}
```

Lawrence Berkeley Laboratory
University of California

Gey D/draft

Figure 14

Organization of the Encyclopedia

IRDS: Entity and Relationship Types with Encyclopedia Extensions

	Entity Types	Relationship Types
Core Schema	Element System/program File Record User	goes_to contains responsible_for
Extensions to Support Business needs	Mission Policy Node Protocol • • •	carried_out_by replicated_at
Extensions to support the Thesaurus and other manipulations	Domain Subject term • • •	broader_than narrower_than related_to

Lawrence Berkeley Laboratory
University of California

Gey9/army23.draft

Figure 15

Organization of the Encyclopedia Thesaurus Functionality

- User (Information Retrieval)
  - Hierarchial groupings (Narrower, Broader & Related terms)
    Subject Oriented search
    Access via multiple hierarchies
  - Glossary & Definitions
  - Synonym, homonym & alias resolution
  - Grouping of related information
  - Keyword Searches (this is not in IRDS)
- System (Integrity Constraints)
  - Table lookup for valid values
- Data Manipulation
  - Units Conversion Information
  - Labeling & Edit Masks for I/O (forms, reports, etc.)

Lawrence Berkeley Laboratory
University of California
Gev12/army25/draw

Figure 18

Domain Information - Components of the Encyclopedia

Qualitative Domains (Discrete Valued)

- Users, Organizations, locations and other entity list
  - List of valid names
  - Location of descriptive information
- Controlled Vocabularies
  - List of valid entrys
  - Class words, Prime words, modifiers, others definitions

Lawrence Berkeley Laboratory
University of California
Gev13/army18/draw

Figure 19

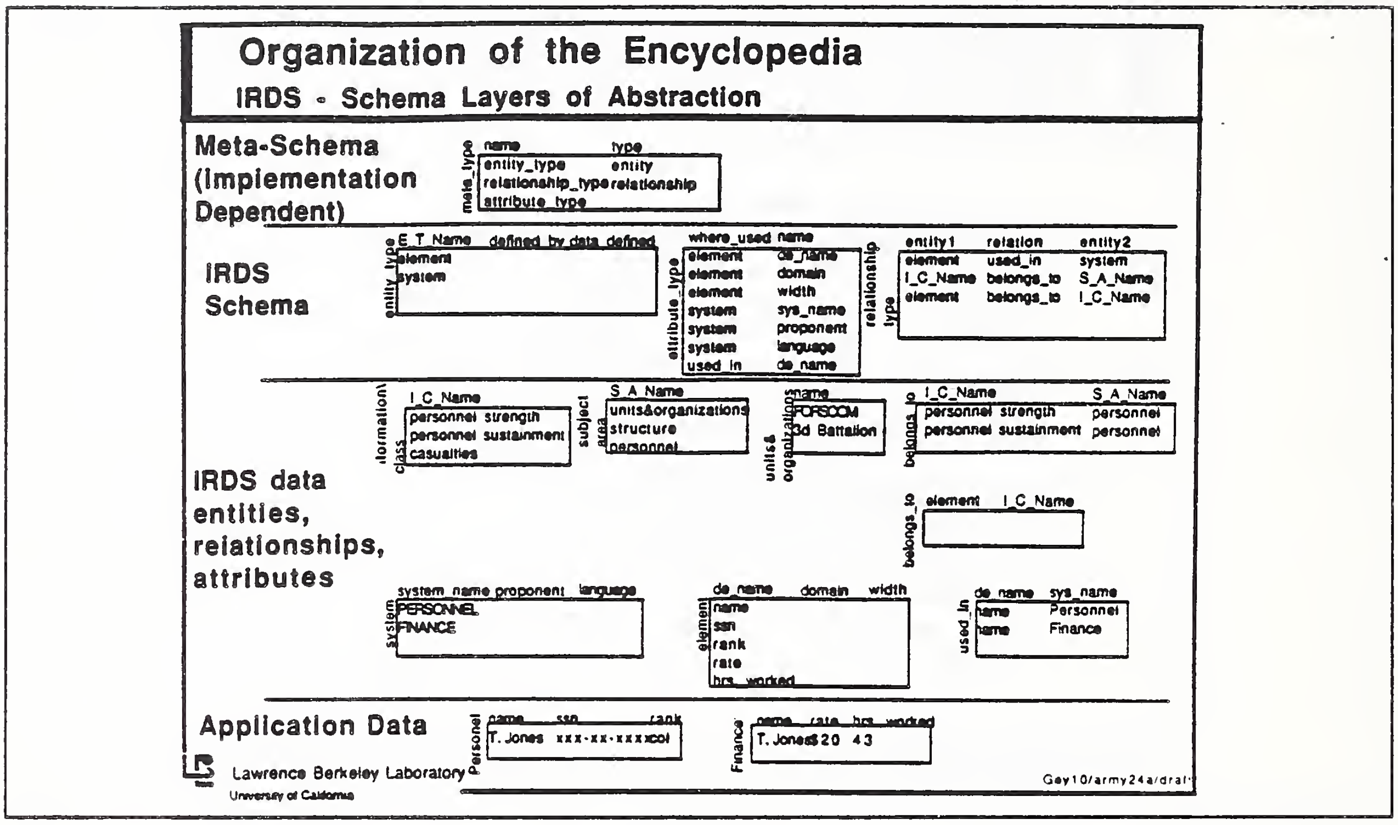


Figure 16

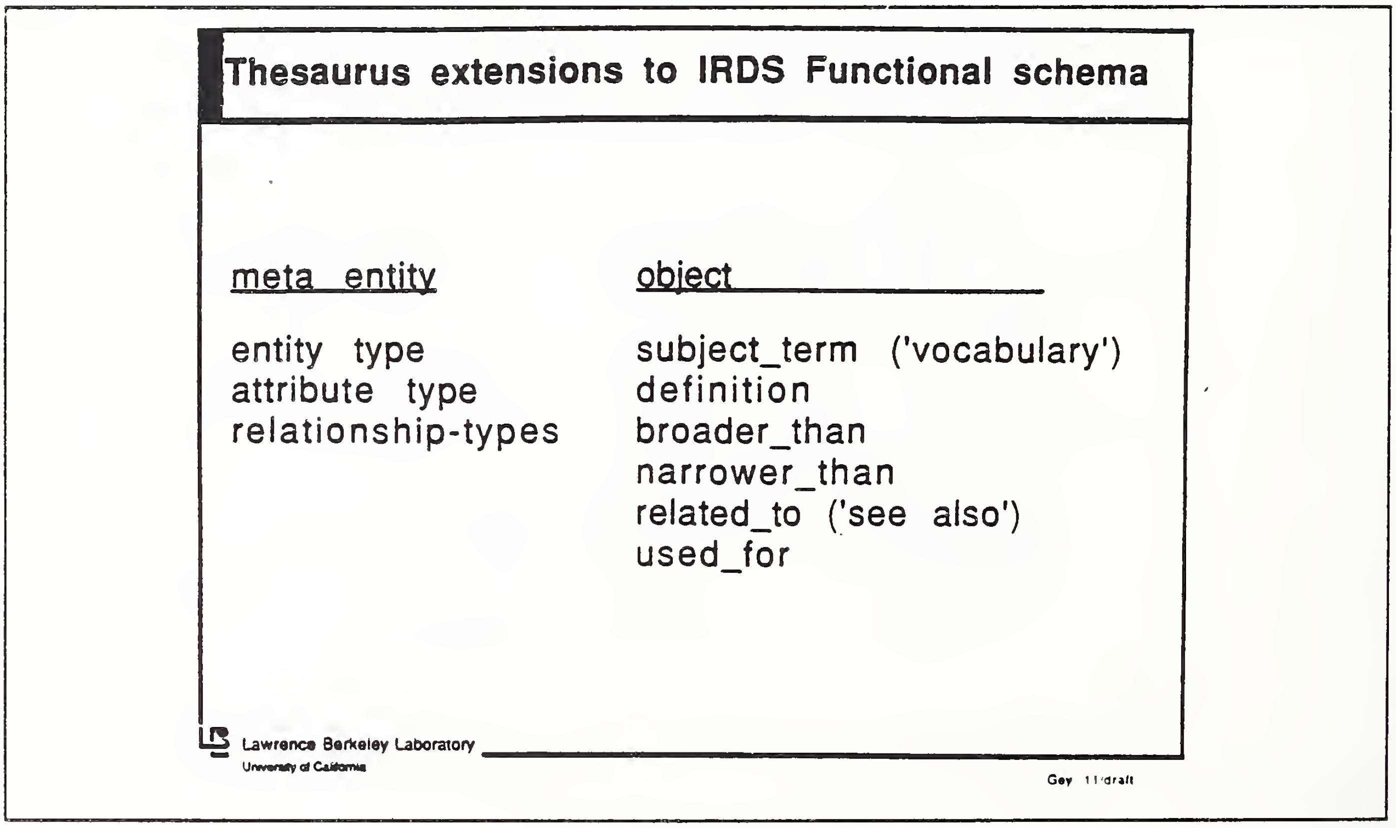


Figure 17

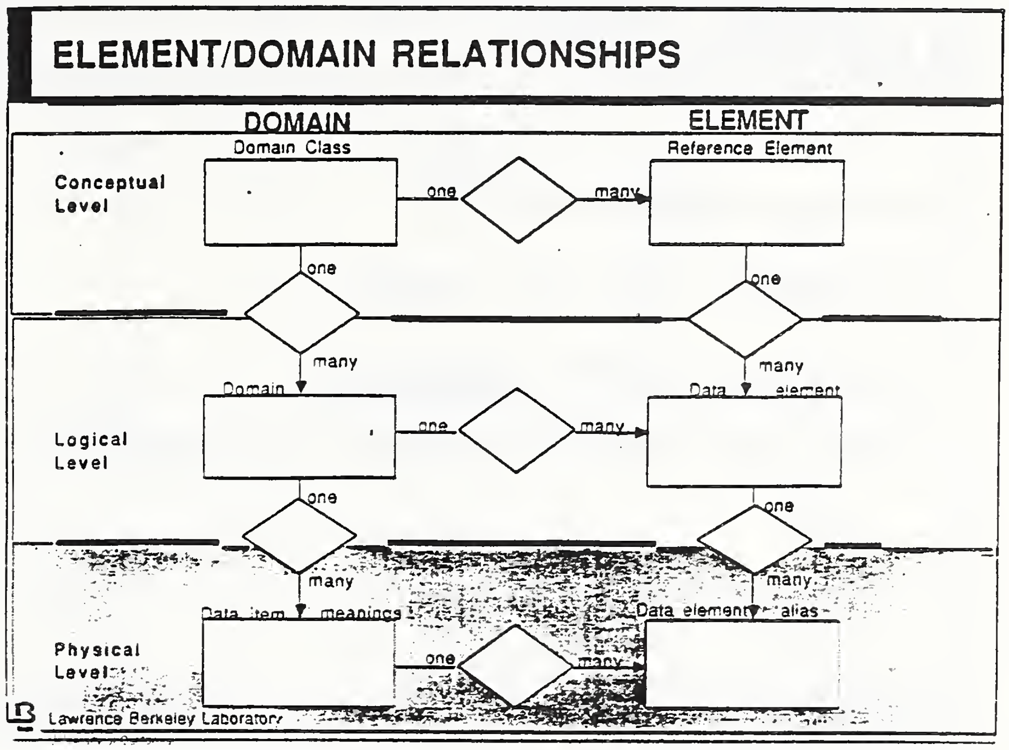


Figure 20

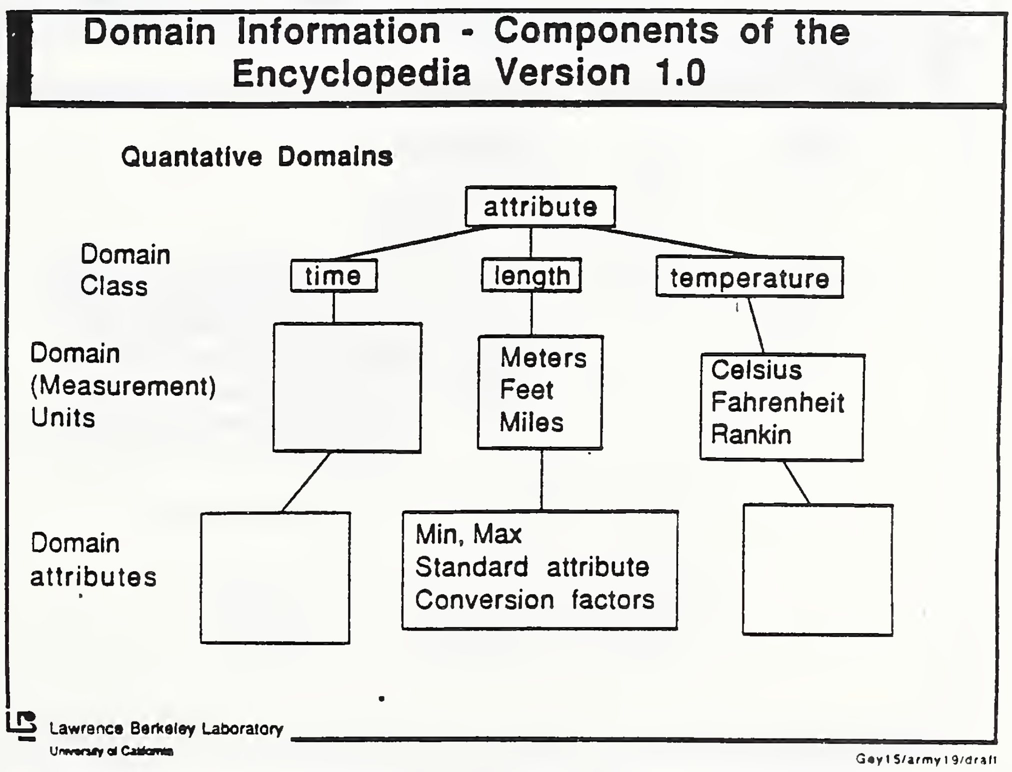


Figure 21

Limitations of IRDS

- Binary relationships only
- Attributes cannot be entities
- Data flow diagrams 'operations'
- Real time command & control more complicated

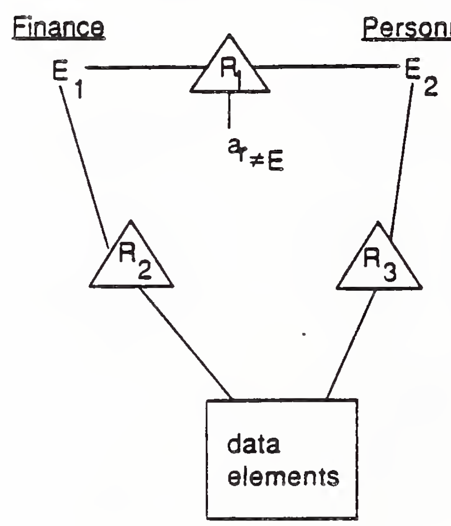
Lawrence Berkeley Laboratory
University of California

Gey16/draft

Figure 22

Limitations of the IRDS

IRDS Attributes cannot be Entities



Since in IRDS, a relationship, e.g. R_1 cannot have an entity, e.g. a data element as an attribute, the data elements common to two applications must be shown through separate relationships, e.g. R_2 and R_3

Lawrence Berkeley Laboratory
University of California

Gey 16a/army?b/draw

Figure 23

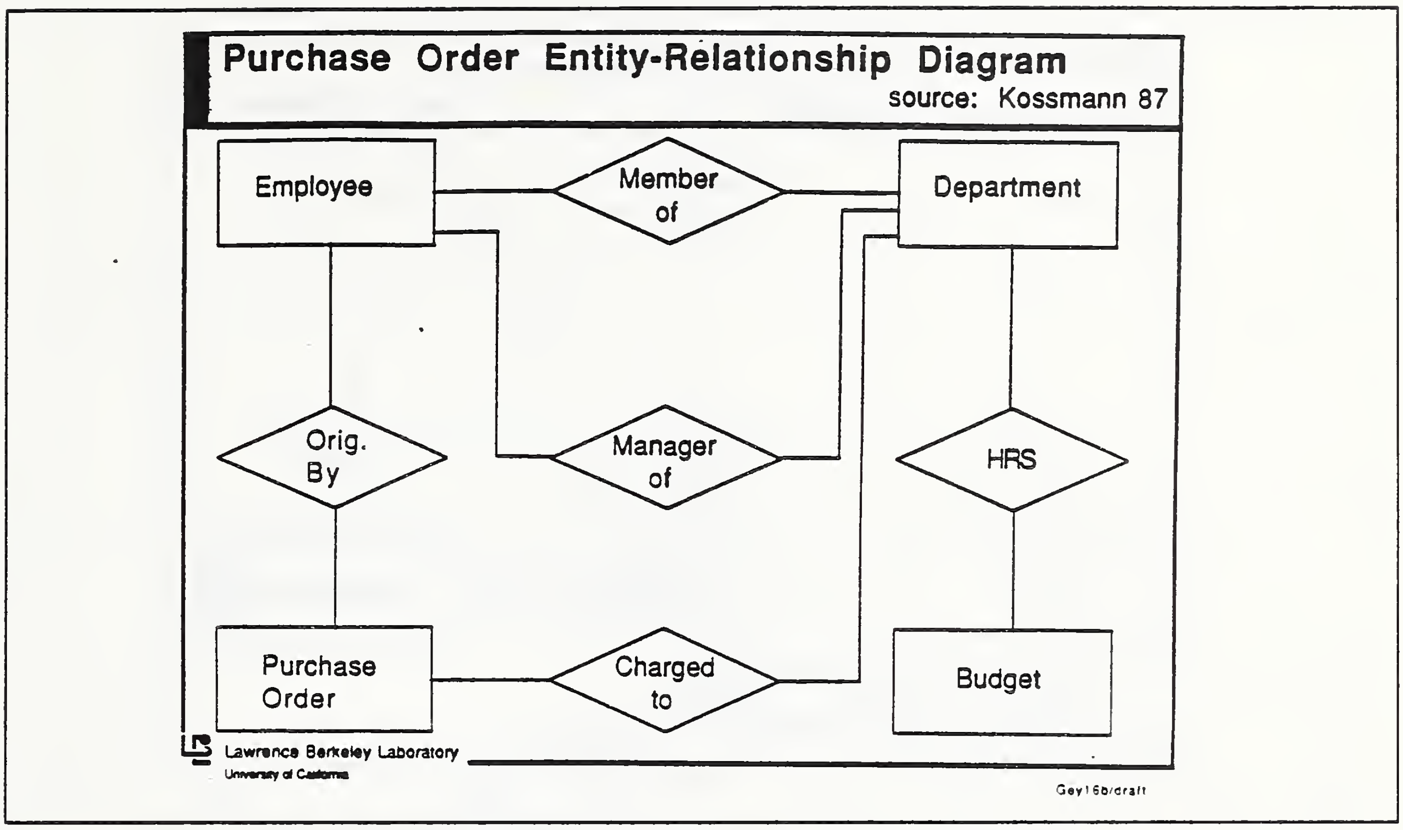


Figure 24

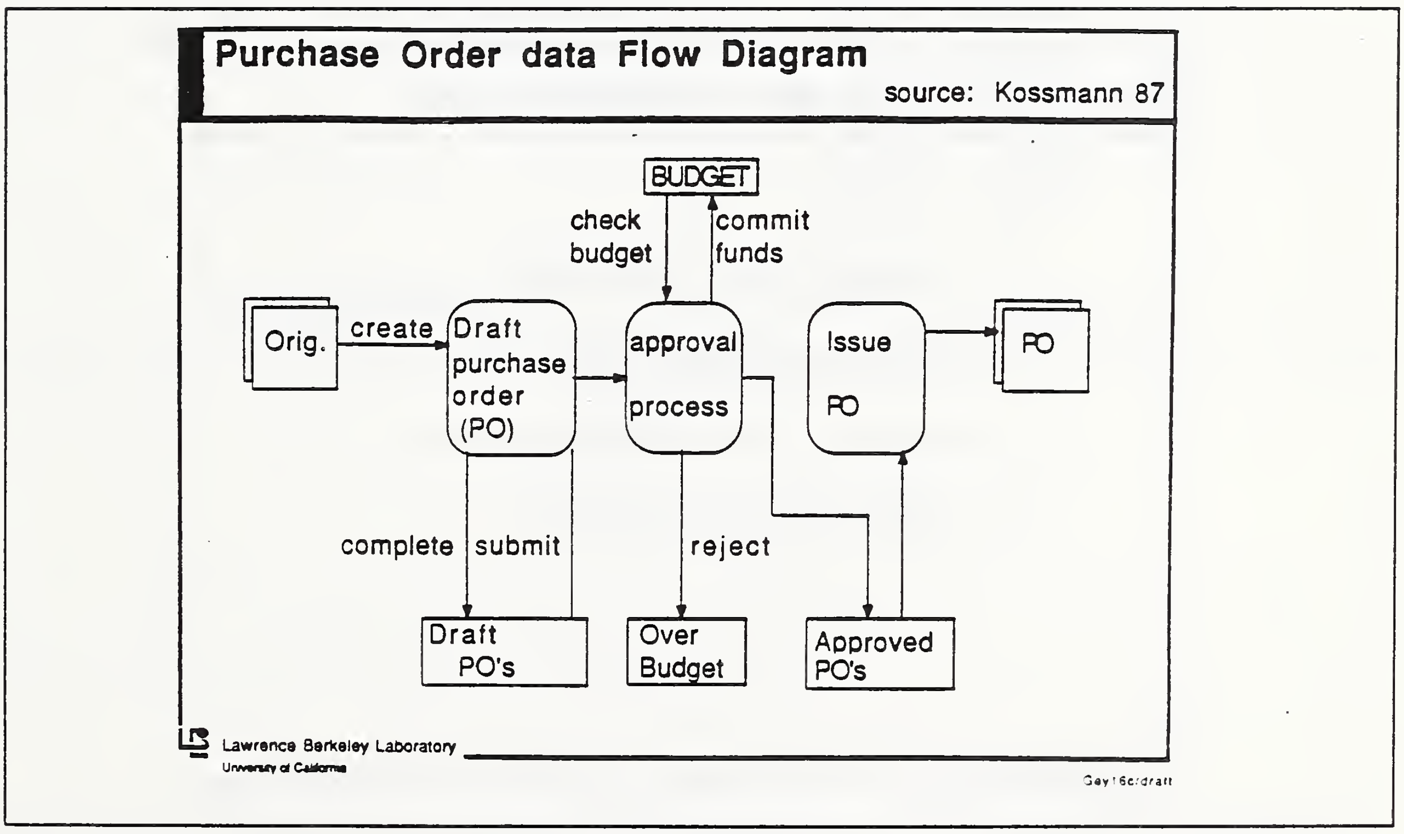


Figure 25

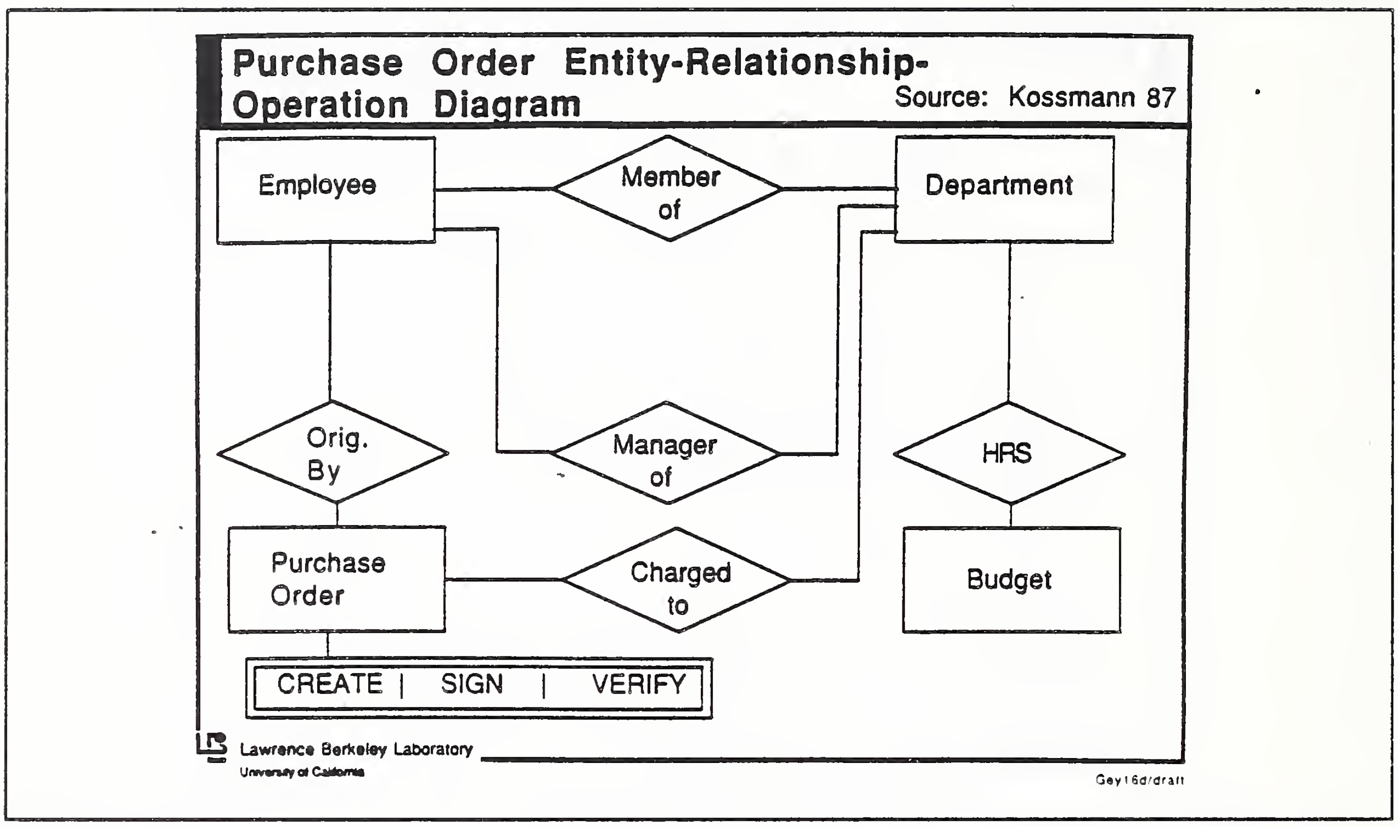


Figure 26

Research Issues and Unresolved Areas

- Temporal Data Semantics
- Real time decision support
- Mappings and Transformations
- Advanced text retrieval

Lawrence Berkeley Laboratory
University of California
Gey17/draft

Figure 27

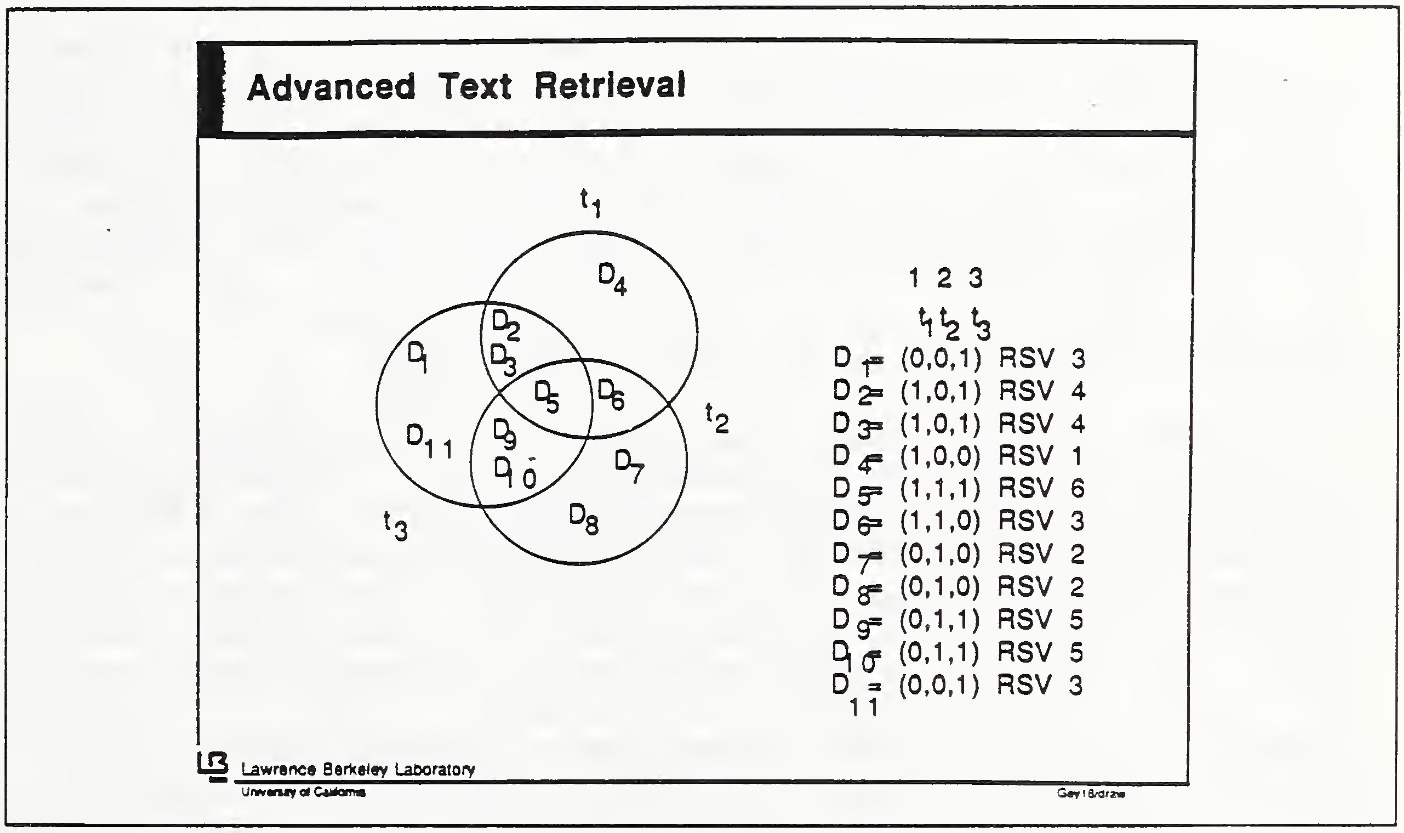

Advanced Text Retrieval
t1
t2
t3
D4
D2
D3
D1
D5
D6
D11
D9
D10
D7
D8
t1 t2 t3
D1 = (0,0,1) RSV 3
D2 = (1,0,1) RSV 4
D3 = (1,0,1) RSV 4
D4 = (1,0,0) RSV 1
D5 = (1,1,1) RSV 6
D6 = (1,1,0) RSV 3
D7 = (0,1,0) RSV 2
D8 = (0,1,0) RSV 2
D9 = (0,1,1) RSV 5
D10 = (0,1,1) RSV 5
D11 = (0,0,1) RSV 3
Lawrence Berkeley Laboratory
University of California

Figure 28

ORGANIZATION OF THE JOINT CHIEFS OF STAFF

Speaker
Major Reed Borman
U.S Army

I'm going to talk today about the data administration program within the Joint Chiefs of Staff. I'm from the office of J7. What that stands for is the Directorate for Operational Plans and Interoperability. That's a long title, but the interoperability question is the one at hand in the area of data administration.

Probably all of you have heard about the Goldwater-Nichols Act. Our tasking to look at interoperability came from that Act. Instead of all the services working individually, they need to come together, with joint doctrine, procedures, and how they're supposed to work together as a joint operational community. From the Act, the restructuring required better advice to the Chairman. The area that we are looking at is to consolidate operational planning, especially in the area of interoperability.

Figure 8 indicates the scope of our problem. That's where our information system operates. We're trying to organize the data elements associated with the information system, the Worldwide Military Command and Control System (WWMCCS) Information System (WIS). You might get the impression that things are scattered everywhere, and you might even ask: Is there any way to get all that done? As you look at the organization itself, there are assorted software programs that consolidate the information for the joint community. There are other application programs that have to do with specific requirements of each particular command. It's further broken down--for every site, every place where we have a computer, there are software packages that work against those data elements also.

Now, what are the problems concerning interoperability that we are finding with all of this? The way we do our war planning today, there are two major programs: the Joint Deployment System and the Joint Operation Planning System. What happens is that when we start to do our force planning, we use the Joint Operation Planning System to build what we call five-day force deployment data. That information is then applied against the Joint Deployment System database and manipulated there. The same type of data elements, called differently, structured differently--the mnemonics are different--go back and forth between these systems. As you follow the arrows back and forth in Figure 11, you can see the interoperability problem continue to build and build, because the data elements are not structured the same, the naming conventions are not quite the same, there's no consistency. If you think that doesn't cause problems in our planning, you're mistaken. We have a major headache here. What we're trying to do with our planning and execution system is to integrate the two programs, the Joint Operation Planning System and the Joint Deployment System into what we call JOPES, the Joint Operation and Planning Execution System. We're hoping to improve the ability and the agility for rapid reaction for missions. One of the things that we've started within the last year or so, and which has grown out of the interoperability issue, is the data administration program for WWMCCS. This effort is aimed at improving the accuracy and timeliness of war planning. This is what we're concerned with--to support our National Command Authorities in the making of their decisions.

Right now, some things that we're working on include the focus on the JOPES program and the building of a WWMCCS Information System Dictionary for Information Management, called WISDIM. We are looking at data administration to provide a way of managing and controlling information to provide consistent, timely, and accurate information to the decision maker. If we provide "poor quality" information to the decision maker, he will make a good decision using bad data. That gives us a bad decision.

The program that we're working on is that of implementing procedures for data standardization, data integrity, data access and security. Since we can't tackle all the issues at the same time, our focus is on data standardization and the building of a dictionary to build interoperable computer systems. The data administration program is performed by J7, as the WIS Data Administrator. This is put into regulation in JCS Publication 19, Annex M. We are supported by the WIS Database Administrator, Hedrick Mitchell at the Joint Data Systems Support Center at the Defense Communications Agency, who will be speaking next. In addition to that, all the services, all the sites, the data and database administrators in the field are supporting us in our efforts to try to put this together. The benefits of doing this are that we're facilitating interoperability, supporting the accuracy of the data, enhancing the ability to share data between computer systems, controlling proliferation of data, and aiding the interpretation of data to make sure that when we refer to a particular type of data element, we are always talking about the same thing.

One of our tasks in building the dictionary system is that today, our dictionary looks like that in Figure 15. You see a file cabinet with a bunch of hard-bound volumes. We're trying to change that, and implement a standardization process for the JOPES/WIS data elements. This is a four step process: analysis, system engineering, the approval process, and the publication of the data elements.

The first step is the analysis of the data. This is where we look at all the data elements that exist. The initial review of the data requirements document gave us over 7,000 data elements and 4 or 5 million lines of code that we were dealing with in the Joint community. After we did the analysis, we had to take into account other existing data elements, other standards within the Department of Defense, the Defense Intelligence Agency, Defense Logistics Agency, Defense Communications Agency, and other areas where we could go to try to find some standards.

The next step in our efforts was to apply a systems engineering approach to our data administration role. This meant establishing data element naming conventions, and these we adopted from the National Bureau of Standards guidelines. The next thing we did was to put into place a data dictionary. We used the IRDS implementation as we built that. There are two data dictionaries that we are working on now. One is WISDIM (the WWMCCS Information System Dictionary for Information Management), implemented using a Model 204 DBMS, which is operational although not fully complete yet. The other dictionary is implemented on a PC using the ORACLE DBMS.

We've been working on this system for just about a year using the concepts of the IRDS, and are trying to fully include the NBS naming conventions as well. The next step in our systems engineering process is to recommend the best standard we can for our data elements. These standards include not only the element definitions, but other attributes, the data structures, the size formats, the mnemonics, the standard date, and so on. We then conduct further discussion with the community to make certain that we have satisfied all the requirements for the standardized data elements that they're going to use with the associated metadata. This is the search for the "best" solution. The final step is the approval by the WIS Data Administrator.

We're working now on making the formal document an automated dictionary--not a publication that sits on a shelf, but a formal, on-line dictionary that is accessible in two different ways, either through the defense data network or through the PC version on the workstation.

WISDIM is the central tool that we're using to support our data administration program and to build the required interoperability. We look at it as the data administrator's automated tool, and we're going to make it available to the JOPES-WIS community. WISDIM on the Model 204 is the central repository of metadata. In that central repository, we are looking at adding all the currently existing dictionaries, including the DoD dictionary, DIA, DLA, our own dictionary for the Joint community, and the dictionaries of all four of the services. This will enable us to do analysis of the data elements and standards existing outside of the JCS view. This should help us come up with some new recommended standards for DoD. We've asked the Office of the Secretary of Defense to look at establishing an Assigned Responsible Agency for C3I data elements. That's still in staffing, and there's no definite outcome of that yet.

Other functions within WISDIM track the standardization process for each data element. One of the attributes we have corresponds to the standardization itself. Is it an approved standard, is it going to be an approved standard, is it a submitted data element that will become a standard, or is it a proposed standard. This answers the question of where does this data element fit in the standardization process.

One of the tools we're working on now supports software prototype development and software development by allowing the user to go to the data analysis section and look at the current systems as they exist today, and ask which systems exist, or to ask for a particular system. Let's say we wanted to look at a civil engineering support program. All the data elements that exist within that program will be deployed, along with their preferred mnemonics. We will then list the standardized mnemonic, the new standard data structure associated with that data element, and all the information for the programmer listed by system as he works at rewriting, recoding, or rebuilding the particular software application.

As I mentioned before, a minicomputer will provide the central repository for metadata. This will be the one location where we can go to "get the truth" about the data standardization program for the Joint community. The PC version will provide local copies. What we envision is something similar to a Bernoulli box where you can put a cartridge in. Our current PC-WISDIM dictionary is about 20 megabytes of data, and we would put copies of that on cartridges and mail them to all the sites, where they would put them up on their PCs, and they would have their updated dictionaries. We're looking at using the mainframe WISDIM to perform a data call. We would go to the community and say "look at PC-WISDIM, what information do you find there, what additional information (data elements) are you using for your site-unique programs, what additional data elements are you using," and then have the central repository serve as the gathering point for the additional data elements that need to be standardized at the Joint level. This will start to build interoperable computer systems that will support the National Commands Authorities and improve our war fighting capabilities.

The status of PC-WISDIM is that we have completed an alpha test of this product at nine sites. We're now incorporating changes and enhancements to the product, and we should be ready for beta testing around July, 1988.

My database administrator, Hedrick Mitchell will follow me. Our discussion is a little different from the others today in that we brought both our data administrator and our database administrator!

<u>Question</u>: I was at a meeting yesterday where we were struggling and arguing about the term "interoperability." Have you guys wrestled with that same problem and come up with a formal definition related to data management?

<u>Answer</u>: Well, let me go ahead and give J7's view of interoperability. I suppose everyone's got their own definition of the term. Interoperability has to do with making things work together. Whether it deals with Joint warfighting doctrine, operational procedures, or with insuring that the same type of bullet comes out of a French, German, or American rifle, or that the same type of artillery shell is used, or that the same type of communications is used with the Army, Air Force, or Navy radios. That's interoperability. The same is true for computer systems. If we build

computer systems that can take data elements from one particular software application and move that information to another application, that's interoperability.

Figure 1

J-7

OPERATIONAL PLANS AND INTEROPERABILITY
DIRECTORATE

PLANNING SYSTEMS DIVISION

Figure 2

WIS

DATA ADMINISTRATION

OJCS J-7

Major Borman

Figure 3

JCS REORGANIZATION

GOLDWATER-NICHOLS REORGANIZATION ACT

ENACTED
30 September 1986

Figure 4

PURPOSE OF THE ACT

- IMPROVE MILITARY ADVICE TO THE PRESIDENT
- PLACE CLEAR RESPONSIBILITY WITH THE UNIFIED AND SPECIFIED COMMANDS FOR THEIR MISSION
- IMPROVE FORMULATION OF STRATEGY AND CONTINGENCY PLANNING
- PROVIDE ADVICE TO THE SECRETARY OF DEFENSE, "ON PRIORITIES OF THE REQUIREMENTS IDENTIFIED BY THE COMMANDERS OF THE UNIFIED AND SPECIFIED COMBATANT COMMANDS"

Figure 5

OJCS RESTRUCTURING

- BETTER ADVICE TO CJCS/MORE RESPONSIVE
- ASSESS FORCE STRUCTURE, STRATEGY, RESOURCES
- CREATE FOCUS FOR INTEROPERABILITY
- CONSOLIDATE OPERATIONAL PLANNING
- FOCUS RESPONSIBILITIES OF DIRECTORATES (JS)
- ENHANCE STAFF PROCEDURES, EFFICIENCY

Figure 6

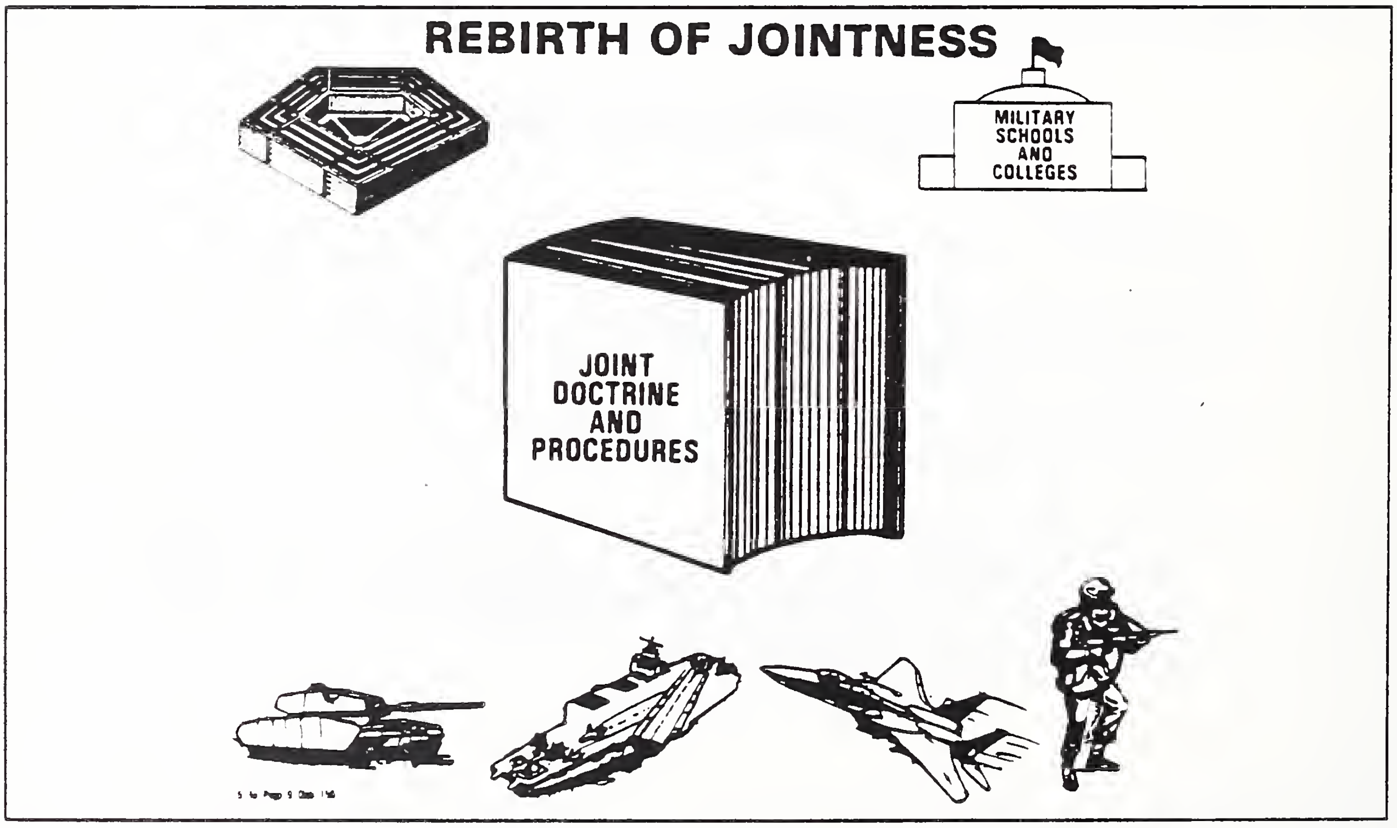


Figure 7

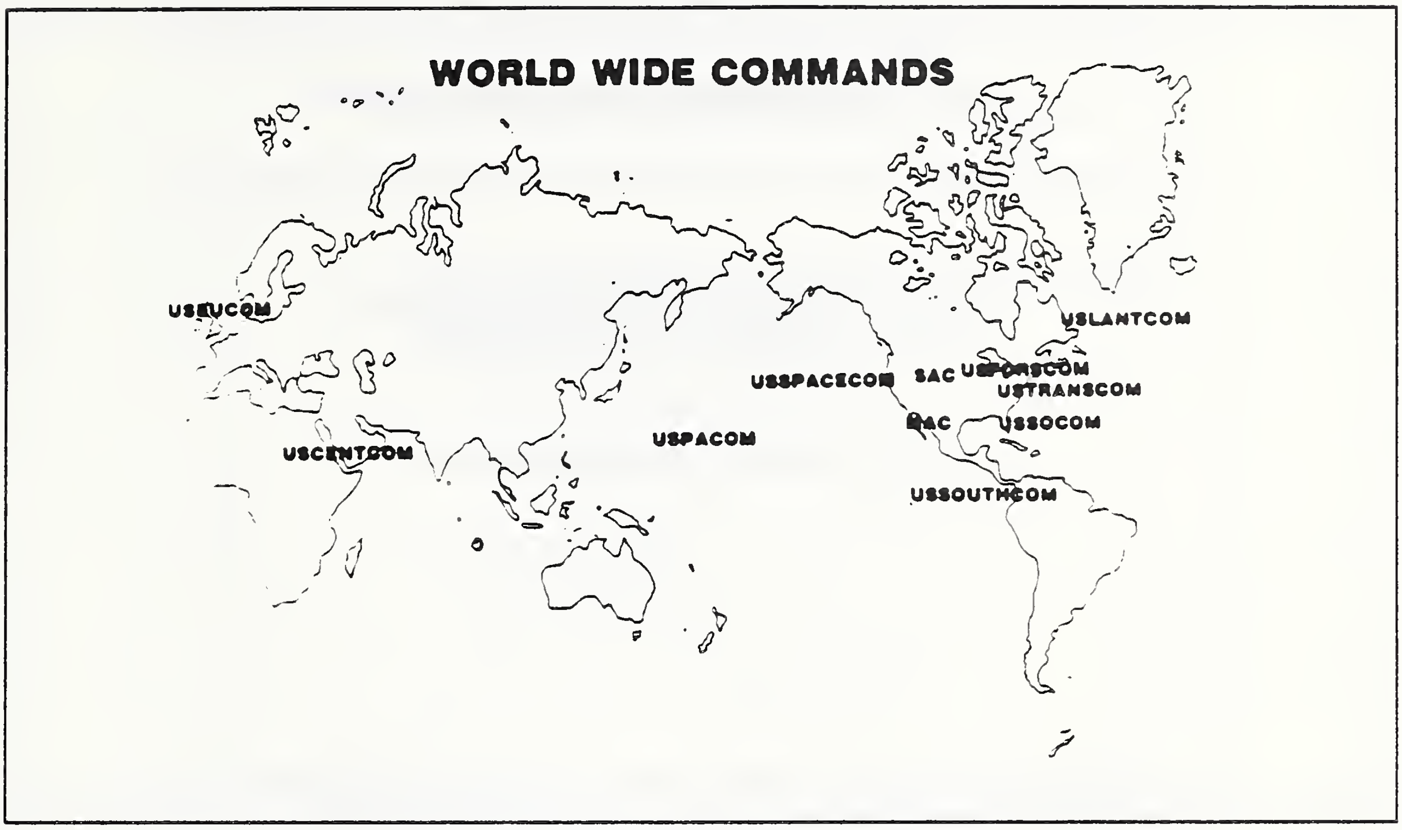


Figure 8

DATA ADMINISTRATION

MANAGEMENT AND CONTROL OF INFORMATION TO PROVIDE:

- CONSISTENT
- TIMELY
- ACCURATE INFORMATION

TO THE DECISIONMAKER

Figure 9

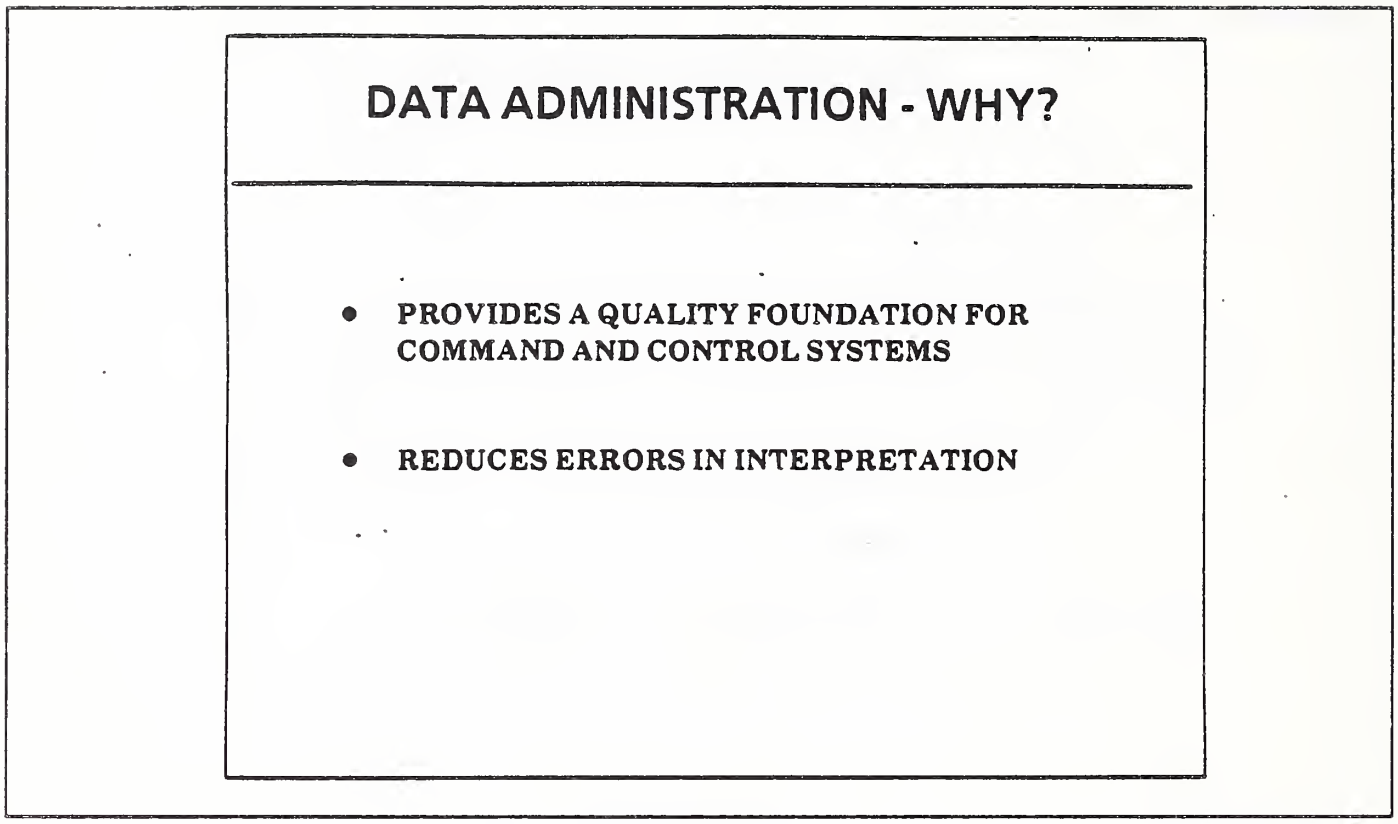


Figure 10

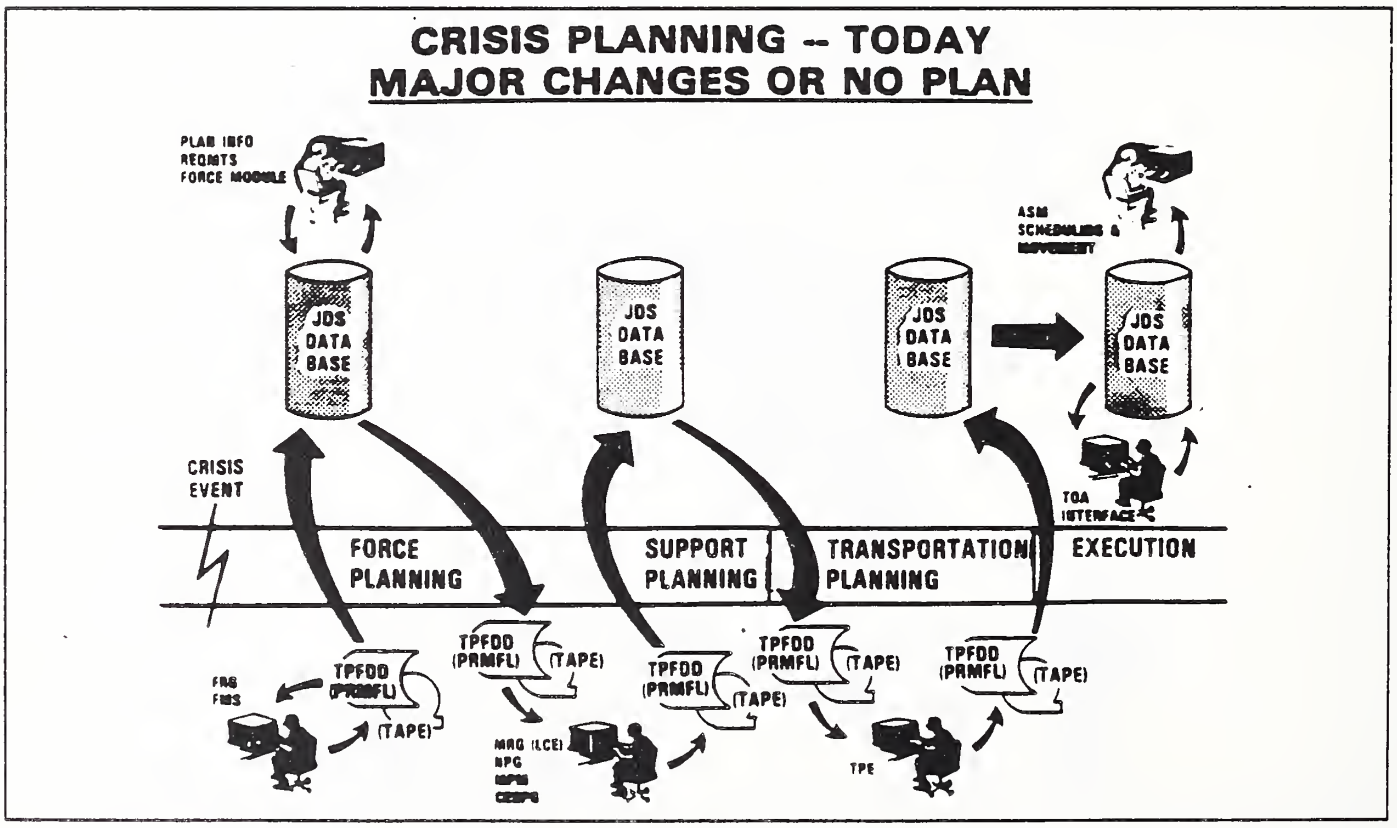


Figure 11

DATA ADMINISTRATION - HOW?

- ENACT, IMPLEMENT AND ENFORCE POLICY AND PROCEDURES TO GOVERN:
  - -- DATA STANDARDIZATION
  - -- DATA INTEGRITY
  - -- DATA ACCESS
  - -- DATA SECURITY

Figure 12

DATA ADMINISTRATION - WHO?

OJCS J-7 IS THE WIS DATA ADMINISTRATOR (WIS DA)

IN ACCORDANCE WITH JCS PUB 19, ANNEX M

SUPPORTED BY:

WIS DATA BASE ADMINISTRATOR (WIS DBA)

-- OJCS J-6 AND JDSSC

SERVICEs, SITEs, CINCs, SYSTEM DAs, AND DBAs

Figure 13

DATA ADMINISTRATION

WHAT ARE THE BENEFITS?

--STRONG CENTRAL MANAGEMENT TO:

- **-- FACILITATE INTEROPERABILITY**
- **-- SUPPORT ACCURACY**
- **-- ENHANCE DATA SHAREABILITY**
- **-- CONTROL PROLIFERATION**
- **-- AID INTERPRETATION OF DATA**

Figure 14

DATA STANDARDIZATION - TODAY

JCS PUB 6

**TODAY'S
DATA
DICTIONARY**

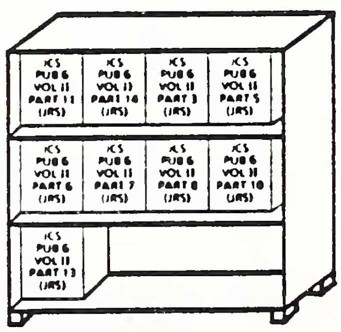


Figure 15

STANDARDIZATION PROCESS - JOPES/WIS

REQUIRES:

- ANALYSIS
- SYSTEMS ENGINEERING
- APPROVAL
- PUBLICATION

Figure 16

STANDARDIZATION PROCESS - JOPES WIS

ANALYSIS:

-- COLLECTS SIMILAR DATA ELEMENTS

-- COMPARES THEM WITH:

  -- EACH OTHER

  -- EXISTING (DOD, DIA, USMTF) STANDARDS

Figure 17

STANDARDIZATION PROCESS - JOPES/WIS

SYSTEMS ENGINEERING PROCESS:

-- NAMING CONVENTIONS

-- DATA DICTIONARY

-- RECOMMENDED STANDARD

Figure 18

STANDARDIZATION PROCESS - JOPES/WIS

APPROVAL:

-- DISCUSS WITH COMMUNITY

-- SEARCH FOR "BEST" SOLUTION

-- APPROVAL BY WIS DA

Figure 19

STANDARDIZATION PROCESS - JOPES/WIS

PUBLICATION:

- ENTRY INTO WIS DICTIONARY FOR INFORMATION MANGEMENT
- PROVIDE INFORMATION PACKETS
- ENTRY INTO PERTINENT JCS PUBS

Figure 20

WISDIM

THE CENTRAL TOOL TO PROMOTE AND SUPPORT COMMAND AND CONTROL SYSTEMS INTEROPERABILITY

- DA's AUTOMATED TOOL
- AVAILABLE TO JOPES/WIS COMMUNITY
- CENTRAL REPOSITORY OF INFORMATION ABOUT DATA

Figure 21

WISDIM

THE CENTRAL TOOL TO PROMOTE AND SUPPORT COMMAND AND CONTROL SYSTEMS INTEROPERABILITY

- TRACKS STANDARDIZATION PROCESS
- SUPPORT SOFTWARE PROTOTYPE DEVELOPMENT
- MINI COMPUTER VERSION PROVIDES GLOBAL DIAL-UP ACCESSIBILITY
- PC VERSION PERMITS LOCAL COPIES

Figure 22

DEFENSE COMMUNICATIONS AGENCY

Speaker
Hedrick Mitchell

I work for the Defense Communications Agency (DCA), Joint Data Systems Support Center (JDSSC), General Applications Division (Figure 1). I will discuss how we at DCA are implementing the IRDS standard. As shown in Figure 2, the National Military Command System environment has four major areas of interest. The first two are being incorporated into a joint services planning and execution capability. The other two have their own data administration capabilities planned, but we in JDSSC recognize the need to develop a capability for information management that would bring into a coordinated effort the management of all data from each of these major initiatives. In order to accomplish this, we are endeavoring to develop an Information Resource Management (IRM) capability.

We have started with Missions and Functions (Figure 3) and use them as guidelines to modify the Information Requirements that are presented to JDSSC. We have either developed or modified IRM tools and applications, which we apply to the logical databases that we control.

Figure 4 illustrates some of the information resource management uses that we address. The significant idea here is that we intend to use IRM to support both manual and automated systems. Figure 5 lists some of the benefits we anticipate deriving from IRM. We intend to refine the uses of IRM and document them for specific users. For example, our functional users will be provided with procedures and documents that will allow them to do the kind of work shown, and our applications developers and requirements analysts will be supported through actions as listed in Figures 7 and 8.

As far as the actual implementation of the IRDS is concerned, we generally conform with the standard. The first three columns of Figure 10 show the entity-types provided by the Basic Functional Schema of the standard. However, we have added the three additional entity-types on the right: Function, Process and Procedure. We found the need to do that because we anticipate a requirement to incorporate entities that relate to more than one Procedure or System at one time. This is exemplified in the JOPES organization by "JOPES support elements." As previously stated, there is a need to address suppliers and one standard type of supply unit. Also to be taken into account is, where are those things obtained and how are they transported. That entire process includes a set of processes JOPES has identified as Procedures.

We are also working on the Security Module from the standard. The things that pertain directly to the standard are in the left column of Figure 11. However, because of our particular usage of hardware and software, we are able to incorporate other security measures and checks and balances in the data dictionary system. Those capabilities are shown in the right column.

The software that we use for the dictionary system is the Model 204 database management system (Figure 12), which runs on an IBM 4361. Our Computer Services Directorate supports the computer system.

We have developed a partitioning mechanism because we have more than one application and more than one initiative or organization to address. For example, in Figure 13, CMS refers to "Configuration Management System." We are looking to improve configuration management of both hardware and software within our organization. We feel that we can easily describe the entities to support this application within the capabilities of the IRDS, since we can define any kind of entity-type for any kind of application. WISDIM has already been described. DA2 and DA3 are references to other kinds of extensions. We have started work on DA2, a second OJCS application which addresses those entities that handle office automation functions within the Joint Chiefs of Staff. DA3 refers to an application describing the entry of metadata about systems that are developed within JDSSC. As can be seen from the diagram, each extension incorporates part of the Core. Additionally, there are other entity-types and other extensions that we will have to add.

To handle the system, we decided that we needed a particular set of management procedures, and these can be separated by organizational components (Figure 15). The Technical Organizations include the M204 System Manager, the File Manager, and the Computer Operations and Support Personnel. The Management Organizational components consist of the IRM Manager, the IRDS Manager, and the Configuration Control Board. The Functional Organizational components are the CM OPR, the DA/DBA OPR, and other OPRs as necessary.

When we have these offices completely staffed, we will be able to provide sufficient configuration management for the Directorate. In figure 16, CI refers to Configuration Item. We will start with new CIs and later process modified CI constructions as received. The CIs will go through a technical review process within JDSSC. One of the things that we can easily handle within our organization is an independent system test capability for all the CIs that are addressed in the configuration management process. So we have an unbiased branch take a look at CIs in order to provide some independent control and input to the configuration management process. We will perform similar processes on the metadata that is entered into the IRDS (Figure 17). There will be configuration management done on the maintenance activity, the notices, and the user reviews. There will have to be people assigned responsibility for this kind of review and approval. In fact, OJCS J7 would be a prime example of the office that would be responsible for the approval of the CIs dealing with the JOPES environment.

Our plans for software include things that will allow us to interface with the PC environment. It has been pointed out that there is already a WISDIM capability developed for the PC. We also want to be able to download to other 3270-compatible PCs, hence the reference to PC/204 in Figure 18. What we are looking for is similar to the thoughts that have gone into the development of the IRDS standard. In that regard, we want other applications, especially those for the PC, to be accessible to the subsets or the entire database contained on the host DBMS. Another significant point is that we are going to contain, as one of the extensions, the DoD data element standards, so users will be able to access these standards directly, subject to proper approval.

Our hardware plans (Figure 19) include the completion of a DDN connection to our operating support facility. IOC here refers to Initial Operating Capability. The DBMS is operational. However, its development is not frozen. We intend to have a series of extensions where we have a number of databases online, depending on the requirements of future applications.

Question: Who came up with the acronym "WISDIM"? That's a good name?

Answer: Major Borman (OJCS J7). I don't bear any responsibility for that. A little heat maybe, but no responsibility!

Question: We have been looking at configuration management, the handling of which we feel is too rudimentary in the present standard. Have you put in any extensions in this area?

Answer: Yes, we have. We are familiar with Military Standard 483A (USAF), and we have borrowed heavily from the configuration management that has been used in the WWMCCS program.

Question: Within the IRDS, what additional control features have you put in to make sure this is followed?

Answer: We have developed a schema for entering those entities that we feel would be applicable in a configuration management environment. That is what comprises the Configuration Management extension to the system. What we will do to complement that application is to develop a set of procedures to implement a configuration control board, of knowledgeable people, to whom these questions will be referred.

Question: Have you devised reporting procedures, that would catch configuration management errors. I think the previous question was really asking what, in addition to the schema, have you come up with to check or validate the data?

Answer: We are working on applicable plans. We have drafts of a Database Administration and a Configuration Management Plan that set policy for specific procedures. Right now we have configuration management control built on the IRDS itself. We have a series of forms that control either changes to the software and hardware, or modifications of how the data will be sent to the configuration control board.

Question: Just for clarification, the way we have been looking at this, is that it's really a meta-schema problem. I mean, we have to define some additional semantics that the schema will understand, so that what the IRDS will do is not let any mistakes occur.

Answer: That's true. We are already using such things as version control that are already built into the IRDS. I get the impression that the rest of the configuration management implementation is not quite complete. When it is, we would certainly conform to it.

<u>Question</u>: In your IRDS, which is supporting a very large system, what did you have in mind for the development of a new significant system?

<u>Answer</u>: We are working on entering all of the other OJCS systems into this database. We will have a series of extensions. We are developing an entire information architecture so that one can go through the definition of entities, relationships, data-flow diagrams, and so on. We hope to be able to handle all those capabilities, but we may not do them all on the host. We are looking to offload some of that work to a PC, so that one can apply any kind of application to the information that we have on the host.

JOINT DATA SYSTEMS SUPPORT CENTER

INFORMATION RESOURCE DICTIONARY SYSTEM

Figure 1

DEFENSE COMMUNICATIONS AGENCY

NMCS ENVIRONMENT

RESOURCE & UNIT MONITORING
CONVENTIONAL PLANNING & EXECUTION
NUCLEAR PLANNING & EXECUTION
TACTICAL WARNING & SPACE DEFENSE

Figure 2

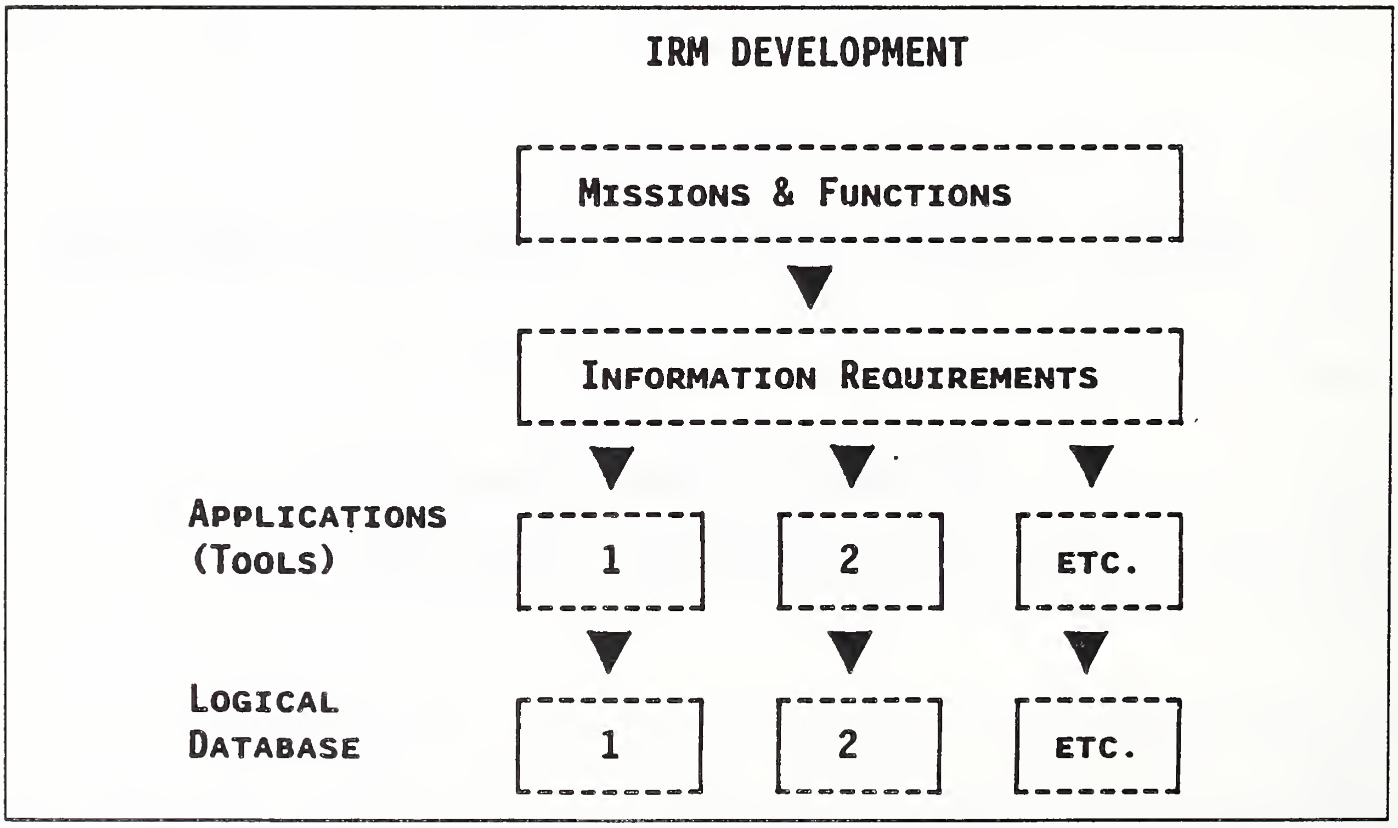


Figure 3

IRM USES

- AID IN DEVELOPMENT, MODIFICATION, AND MAINTENANCE OF MANUAL AND AUTOMATED SYSTEMS.
- SUPPORT DATA ADMINISTRATION AND STANDARDIZATION PROGRAMS.
- SUPPORT RECORDS, REPORTS, AND FORMS MANAGEMENT IN MANUAL AND AUTOMATED ENVIRONMENTS.
- SUPPORT INFORMATION RESOURCE MANAGEMENT ACTIVITIES.

Figure 4

BENEFITS OF AN IRM

- SHARE EXISTING INFORMATION RESOURCES.
- REDUCE UNNECESSARY DEVELOPMENT OF APPLICATIONS WHEN SUITABLE ONES EXIST.
- SIMPLIFY SOFTWARE AND DATA CONVERSION THROUGH CONSISTENT DOCUMENTATION.
- TRANSPORT SKILLS AND REDUCE TRAINING COSTS.

Figure 5

FUNCTIONAL USERS

IDENTIFICATION OF THE LOCATION OF DATA
DATA DEFINITIONS
PREFERRED USES
REPORTING REQUIREMENTS
REQUIRED FREQUENCY OF UPDATE

Figure 6

APPLICATION DEVELOPERS

PROVIDING TECHNICAL ATTRIBUTES FOR DATA ELEMENTS
DETERMINING A DATA ELEMENT'S INTERNAL REPRESENTATION
DETERMINING AUTHORIZED USE OF DATA
RECOMMENDING NEW DATA ELEMENTS
RECOMMENDING CHANGES TO EXISTING DATA ELEMENTS

Figure 7

REQUIREMENTS ANALYSTS

CORRELATING DATA ELEMENTS WITH PROCESSES
PERFORMING ANALYSES AMONG DATA ELEMENTS
TRACING DATA REQUIREMENTS TO SOURCES
TRACING DATA REQUIREMENTS TO REPORTING SYSTEMS

Figure 8

IRDS IMPLEMENTATION

Figure 9

IRDS ENTITIES

USER	SYSTEM	DOCUMENT	FUNCTION
	PROGRAM	FILE	PROCESS
	MODULE	RECORD	PROCEDURE
		ELEMENT	

Figure 10

IRDS SECURITY

ANSI/FIPS	JDSSC
	SYSTEM
	DBMS
	IRDS
GLOBAL	GLOBAL
	FILE
ENTITY	ENTITY

Figure 11

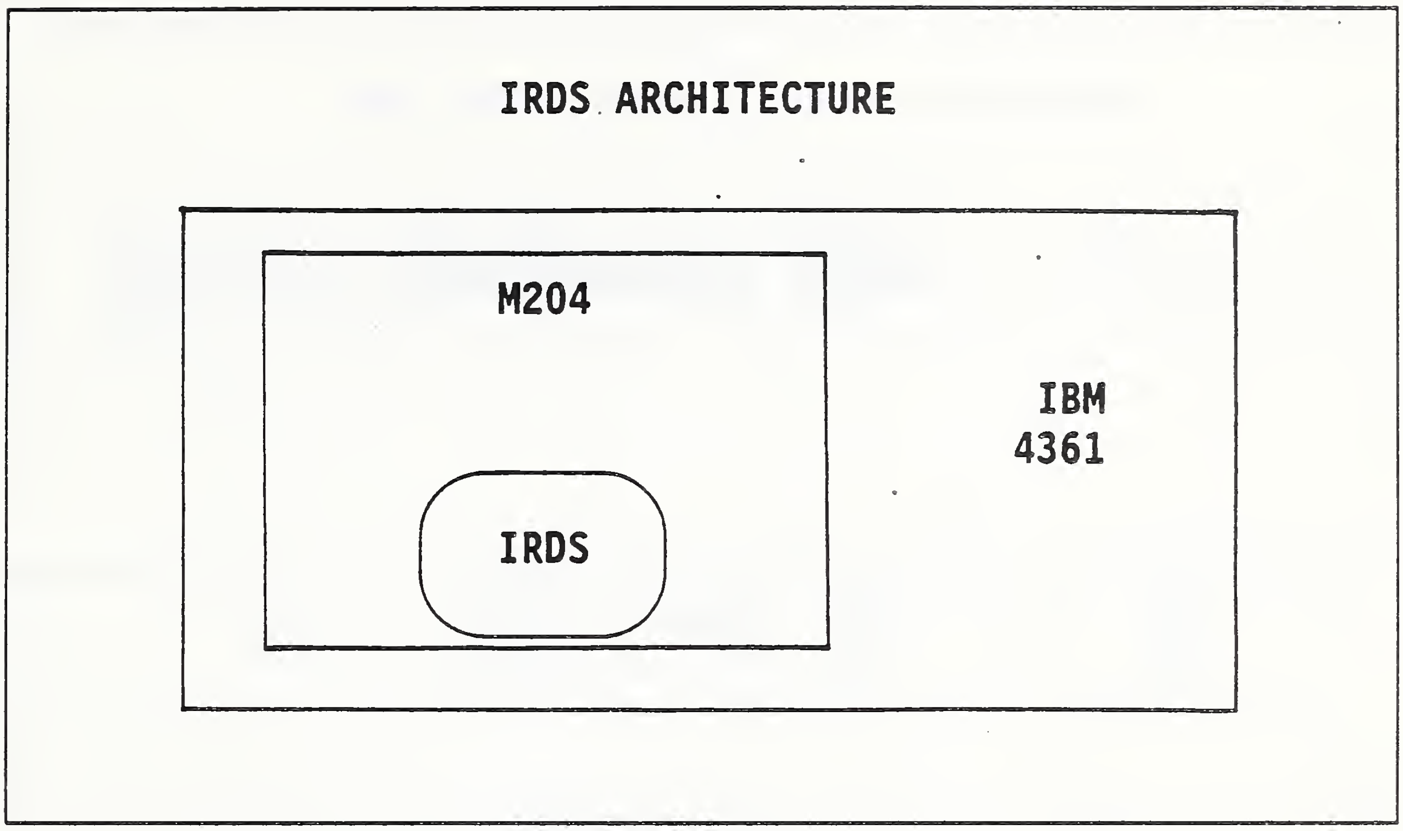


Figure 12

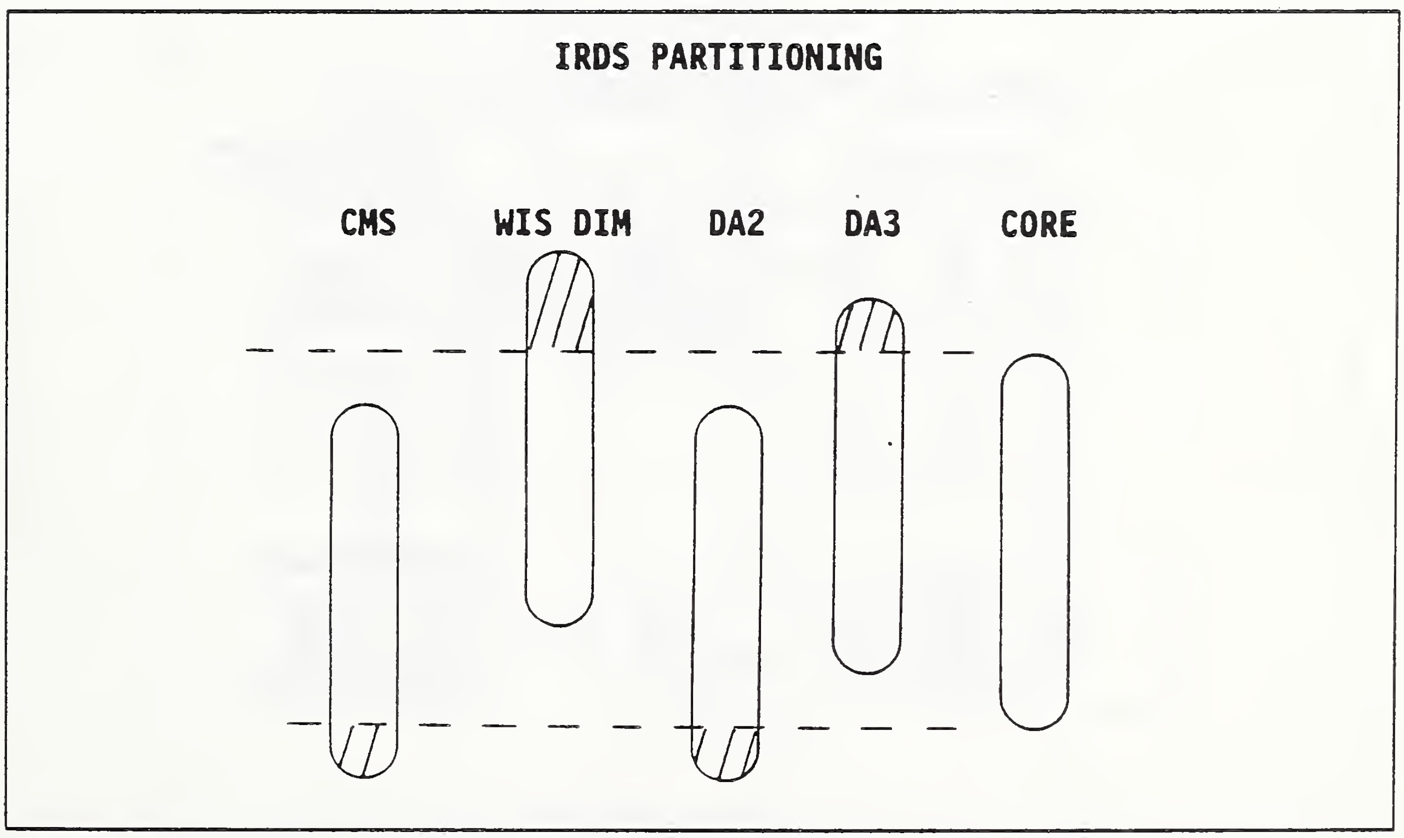


Figure 13

IRDS MANAGEMENT

Figure 14

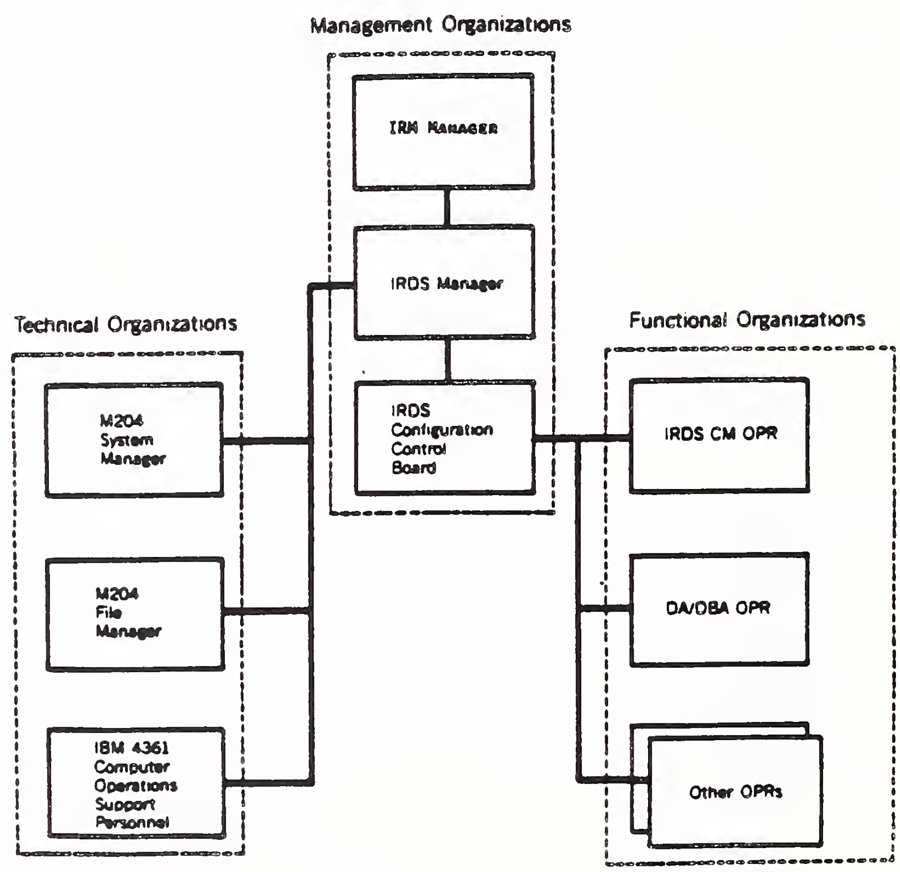


Figure 15

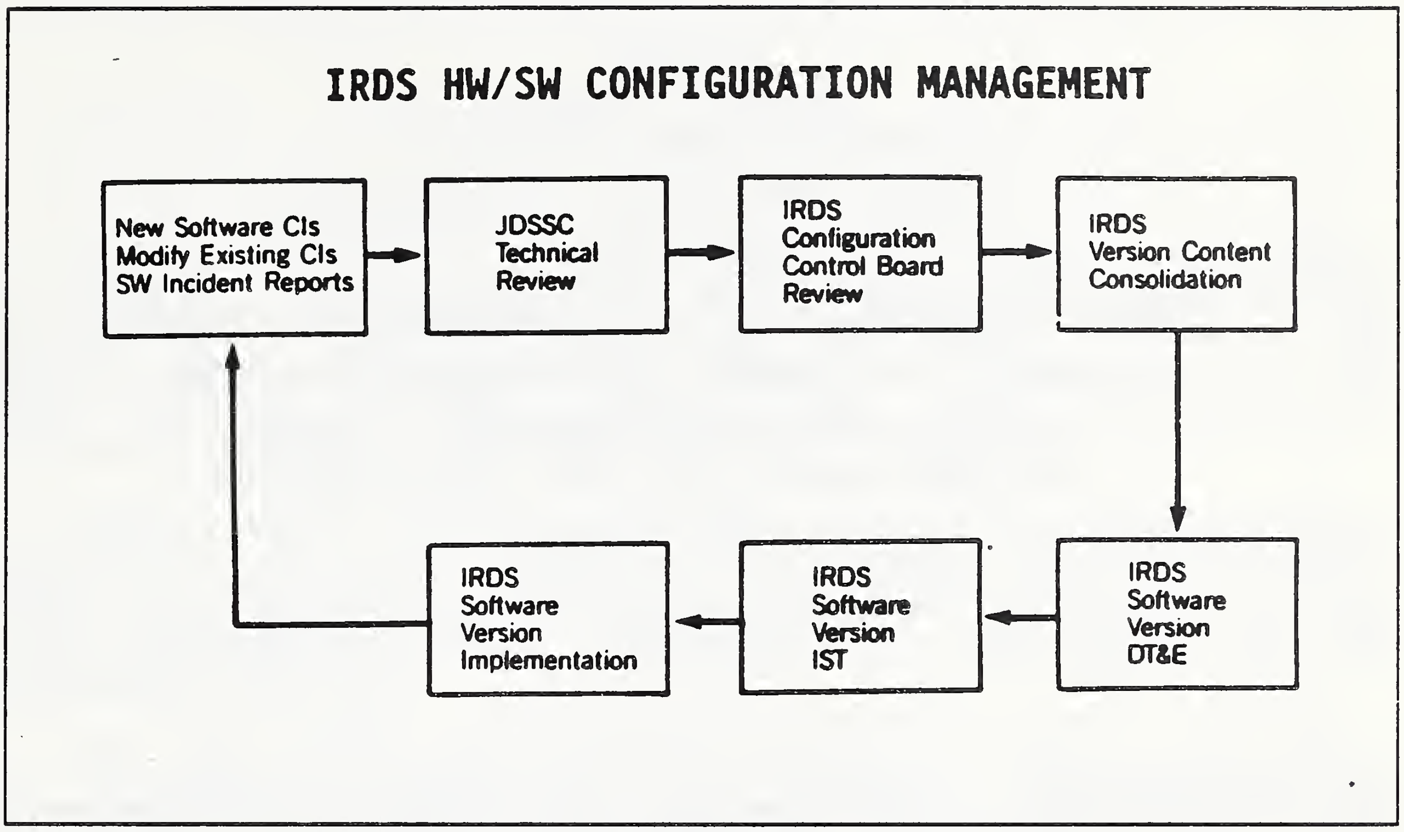


Figure 16

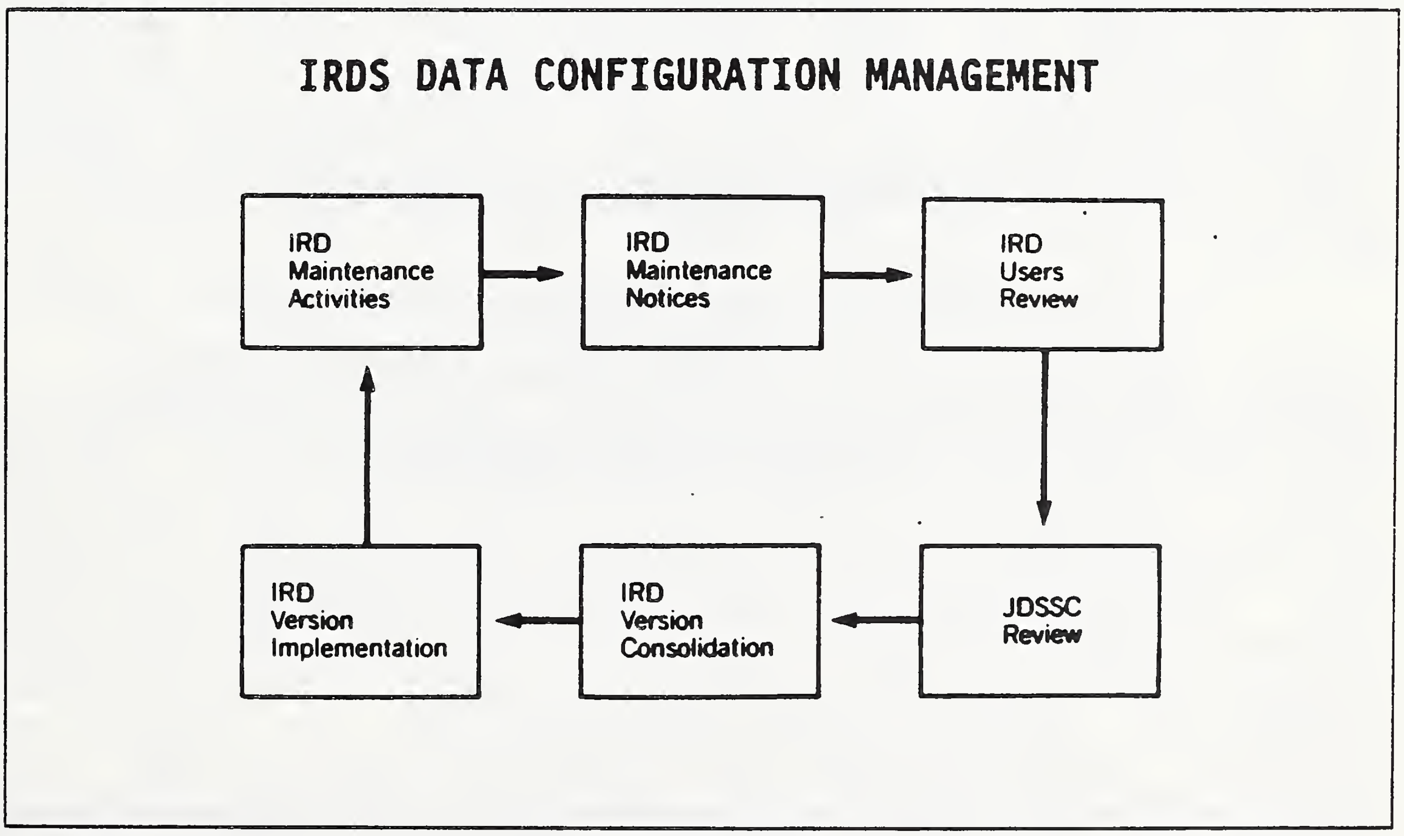


Figure 17

SOFTWARE PLANS

DEVELOP DOWNLOADS TO PC/WISDIM AND PC/204
DEVELOP PC-BASED COMPLEMENTS TO IRDS
INSTALL DOD DE STANDARDS IN IRDS
PROVIDE LINKS TO OTHER DICTIONARIES

Figure 18

HARDWARE PLANS

COMPLETE DDN CONNECTION AT OSF
TEST IOC AT PENTAGON FROM JSSIS NETWORK
COPY IRDS DATA BASE FROM OSF TO PENTAGON
UPGRADE OSF FACILITIES

Figure 19

U.S. DEPARTMENT OF COMMERCE

Speaker
James E. Squier
Office of the Secretary

Good morning. My name and organization are on Figure 1. It's sort of a carryover from when I was young and my mother used to pin a little tag on me that said: "This is Jim Squier and he belongs at 65 M Street." As you can see from this figure, I'm the Acting Chief, Data Administration Division, Office of the Secretary of Commerce.

What I want to do this morning is give you background on what we're doing at Commerce with respect to Data Administration, and how we plan to use the IRDS Specifications and ICST recommended naming conventions as the basis for our standards program.

The Department of Commerce, like many Federal agencies, is involved in Reform 88 initiatives. Departments and agencies throughout the Federal Government have been directed by OMB Circular A-127 to implement departmental financial systems within five years, and to consolidate, department-wide, all their administrative and management systems. The current priority within the Department of Commerce is to consolidate application systems with the objective of shortening the length of time required for systems implementation. Therefore, we intend to acquire off-the-shelf systems, and use the systems of other Federal departments and agencies to meet this objective. In other words, we won't, to the extent possible, build our own systems from scratch, we won't reinvent the wheel.

Without a data administration function, the next step or subsequent move towards integration of the current in-house systems, the new package systems, and the systems from other agencies will lead to a chaotic data environment. One of the most important things, not only in Commerce but elsewhere, is to establish, formalize, and institutionalize the data administration function. We are doing that now. It's important to identify the data administration function organizationally, staff it properly, and continue to promote and defend it. As resources become tighter, IS managers will have to make difficult decisions regarding functions and staffing. It's difficult to justify the Data Administration function for the long term, so we are establishing the DA function, and hope to institutionalize it as soon as our reorganization goes through. In addition, we plan to recruit a data administrator who will serve as the supervisory focal point for this organization. This position will be established at the GM-14 level. I estimate that, if things proceed as we anticipate during the next five years, the DA function will become increasingly important within the organization, and the position will justify a higher grade level.

We're also in the process of developing our Commerce enterprise functional and data models. Our business models will describe administration in Commerce. That is, the centralized, corporate functions that are found in the Office of Administration at the Secretary level, as well as administrative functions that reside in Commerce bureaus. We've found this task to be much more difficult than the examples one finds in textbooks. Their models have six or seven little blocks. We're finding four or five hundred entities, and a complex network between them. We will need the side of a wall to put up our entity diagrams. We've made some progress, and will continue this work, which will be the logical foundation of our enterprise model and subsequent decomposition for the structure of our data dictionary.

Another initiative that we have in progress is the development of a data standards manual for administrative and management systems in Commerce. We're not addressing program area data. That's too diverse and complex, but we can focus on the administrative and management areas. We intend to implement standards and enforce them.

When we say that we're going to issue data standards that conform to IRDS specifications, we mean that our data standards will be operative in an IRDS based dictionary environment. Therefore, we want to ensure that our standards conform to the adopted IRDS specifications. I understand that within the last year there has been some change, representational change, to the proposed IRDS Core structure. Consequently, we don't want to publish a standards manual that doesn't fit. We're very conscious of a potential conflict with the IRDS specifications and are eager for their final adoption as an ANSI standard.

We've also adopted the data naming conventions recommended by the Institute for Computer Sciences and Technology. Not only are they logical suggestions from our perspective, but we've incorporated the recommended approaches, and are comfortable with them.

Because it may be one to two years before there are data dictionaries on the market that use the IRDS specifications, we plan to begin collecting information for a dictionary now, using our data standards, and possibly develop our own dictionary of very limited functionality. This interim step would be taken in anticipation of using the collected information as input to a dictionary system based on IRDS specifications when such a system is marketed.

JAMES E SQUIER

CHIEF, TECHNICAL SUPPORT DIVISION

(ACTING CHIEF, DATA ADMINISTRATION DIVISION)

OFFICE OF THE SECRETARY

U. S. DEPARTMENT OF COMMERCE

TEL: 202-377-2855

Figure 1

U. S. DEPARTMENT OF COMMERCE
DATA ADMINISTRATION INITIATIVES
* * * * *
● ESTABLISH DATA ADMINISTRATION FUNCTION AND RECRUIT DATA ADMINISTRATOR
● DEVELOP ENTERPRISE, FUNCTIONAL AND DATA MODELS

Figure 2

U. S. DEPARTMENT OF COMMERCE
DATA ADMINISTRATION INITIATIVES (Cont'd)
* * * * *
● ISSUE DATA STANDARDS THAT CONFORM TO IRDS SPECIFICATIONS
● ADOPT DATA NAMING CONVENTIONS RECOMMENDED BY ICST
● DEVELOP INTERIM DATA DICTIONARY

Figure 3

NATIONAL BUREAU OF STANDARDS

Speaker
Joan E. Tyler
Center for Manufacturing Engineering
Factory Automation Systems Division

This talk will focus on the IRDS in the context of support for another emerging standard, the Product Data Definition Exchange Specification (PDES). I will also describe how our factory prototype here at NBS is linked to PDES and IRDS.

As you can see from Figure 3, the AMRF stands for the Automated Manufacturing Research Facility. Our major source of funding for our research comes from the Navy's Manufacturing Technology program. We also work closely with industry and universities through the NBS Research Associates program. Our facility is intentionally composed of manufacturing and computing equipment from many vendors which provides a real "testbed" for interface standards. For the factory of the future, the ability for a company to start with a numerically controlled machine, add a robot, and add a PC or other equipment as the company grows and has capital to invest in flexible manufacturing, is extremely important. Dealing with complex data problems with dissimilar computing systems and developing data driven automation concepts are examples of the kind of work I have been involved with.

I am holding a piece of raw material used in our machining process in the factory. This (see Figure 4) pipeclamp is the result after the drilling and milling process. As we look at it we can see that it has holes or circles, it has a shape, it has a surface, etc. All this information is important data and must be captured and distributed to other systems. As the part shape changes, data from robots, sensors, and time-related information becomes important and needs to be integrated into the data system. Defining, integrating, distributing, and building a common data system to handle a number of dissimilar computer systems, data systems, and databases are some of the complexities we are dealing with here at the AMRF.

Now, I would like to talk about PDES, and how our work here is helping this specification evolve. The PDES objective is to develop and apply the technology necessary to communicate digital product definitions within a heterogeneous computing system environment involved in industry automation. I would like to explain how PDES evolved. In 1984, the Initial Graphics Exchange Specification (IGES) community found the data format that allows geometric data to be exchanged between two different types of computer-aided design systems was inadequate for the broader goals of PDES. They realized the need to pass information about features which use geometry, and to pass other complex data types needed in flexible manufacturing. They began to look at work sponsored by the Air Force Computer Integrated Manufacturing (CIM) office at Wright-Patterson Air Force Base. This large initiative was called Product Data Definition Interface, PDDI. They started by describing and defining the data about just four aircraft parts. You would not think that would be so difficult, but when you see the complexity of the kind of data you need to describe just this pipeclamp I have in my hand, you begin to see the level of difficulty in describing the data semantics about large objects, an aircraft wing, for instance. Then, after capturing the data, you must solve the problem of integrating it into a factory environment. That's analogous to what Jim Squier talked about earlier--integrating data across any boundary. That's what we're all involved in--this sharing of information, and sharing the meanings for this information.

As the Air Force PDDI initiative was ending, the PDES community started its own effort to drive PDDI work toward standardization. The three-schema ANSI/SPARC architecture was used in the PDDI work.

What we call CIM is Computer Integrated Manufacturing. We are trying to address all the requirements to keep our manufacturing base competitive. We need an edge, and we believe that technology is the edge that's going to keep us competitive. So what we've really been involved in is helping the small manufacturer integrate.

I want to tell you what the status of PDES is. As you can see from Figure 7, it's still developing. Brad Smith, a local person, is the Chairman. We have a joint development with ISO. We have a strong voluntary effort with 260 companies represented by 655 individuals organized into 19 technical committees. I'm on several of the committees.

AMRF started with a proof of concept. This proof of concept was looking at all application areas and putting together a conceptual model for each one of these. We did this in something called the NIAM information analysis methodology. Some really good results came out of that about how to integrate data, and how to drive an enterprise through a conceptual model. I won't go into all of these different testing drafts. I want to say that PDES Version 1 will be cut at the end of this year.

Now, I'd like to marry the two things together. As Alan mentioned, the AMRF is very closely allied with PDES because of the work that we're doing. These are all things that industry is involved in--they have a need for data to drive automation. So they have to have conceptual models that identify the data. When I talk about conceptual models I'm talking about information models. I differentiate between logical and conceptual. I put conceptual and information models at the very top. There's a lot of information that never gets down to the logical model, company proprietary data for instance.

For integration, you identify your connection points, your interaction points with other departments, for instance, or your other pieces of data that you need to interact with. You need the connect points for identification and integration purposes. Also, we've discovered that we want to share databases across departments, and we want to share between companies. But there is security information or proprietary information that can never be shared. We need to inform other people that we know it's there, but that we can't share it. The conceptual model allows that visibility; it allows us to get our arms around all the things we have about the enterprise.

What I'm trying to show is the marriage between an application and our world of the AMRF. I've given you several examples. We're dealing with computer integrated manufacturing, we're dealing with factory automation. CALS requirements involve standards. We're closely aligned here. As you can see from Figure 8 we have integrated models, we have physical files and databases, glossaries, and dictionaries. That's where we feel that the IRDS can help us. We are an evolving standard. We're in our infancy. We have nineteen conceptual models, and you can imagine all the versioning we've gone through. The companies and people who have contributed come from different contexts, they see the world differently, and we allow them the ability to put together different generic models and share across these companies. So, you can see we've gone through a lot of versions that we need to manage. So the IRDS can really help us in managing these versions. We're looking at some of these issues: How we can use the IRDS standard to help us with our life cycle management, our versioning?

I'd like to talk a little bit more about the idea of a conceptual model. Yesterday we heard a little bit about conceptual models. Two years ago, when we stated modeling, there were no tools to help us except pencils and paper. As Jim Squier said, these models evolve to the point where they take up whole walls. We had to do this by hand because there were no tools to help us automate. Now, I can happily say that there are several tools available. DACOM has a tool, CDC has a tool. CASE tools are emerging to help us as analysts, as generalists, as DBAs, to help put together these conceptual models. I believe that we now know from the PDES world what we need from conceptual models. These 655 people and 260 companies can say that they will eventually be able to come up with a consensus on what is a conceptual model.

In a product life cycle, one of the first things you do is define the activities and functionality of the project. Now, these we can decompose, and decompose, and decompose, but at least we've got our arms around the activities up front. In the middle of Figure 11, where one activity flows into another, we need to start tracking. These activities represent the information which the IRDS can help track the cycle.

Figure 12 lists some of the many information model standards issues. We need guidelines for assessing completeness and conceptuality. We need review mechanisms during model development. We need a means of controlling computer files related to models. We need a related dictionary suitable for DBMS use. We may require some "soft standards" concepts that are evolving that we don't even know about. In the world that I'm working in, we really believe that the thing that will hold all these things together and shed light on the whole thing is the conceptual data model.

Question: Are you getting any information and input from the CAD/CAM world?

Answer: Yes, constantly.

Question: Are they developing a set of standards in conjunction with you, or are you joining their activities? The thing I'm getting at is that numerically controlled machines have been around for a long time.

Answer: That's part of the whole scenario. We're getting a lot of information from them. For example, on our factory floor we have a lot of information coming in from NC machines in the CAD/CAM area. They feed us information and we use it in our data driven factory.

AMRF ROLE
IN
STANDARDS

JOAN E. TYLER
CENTER FOR MANUFACTURING ENGINEERING
FACTORY AUTOMATION SYSTEMS DIVISION

Figure 1

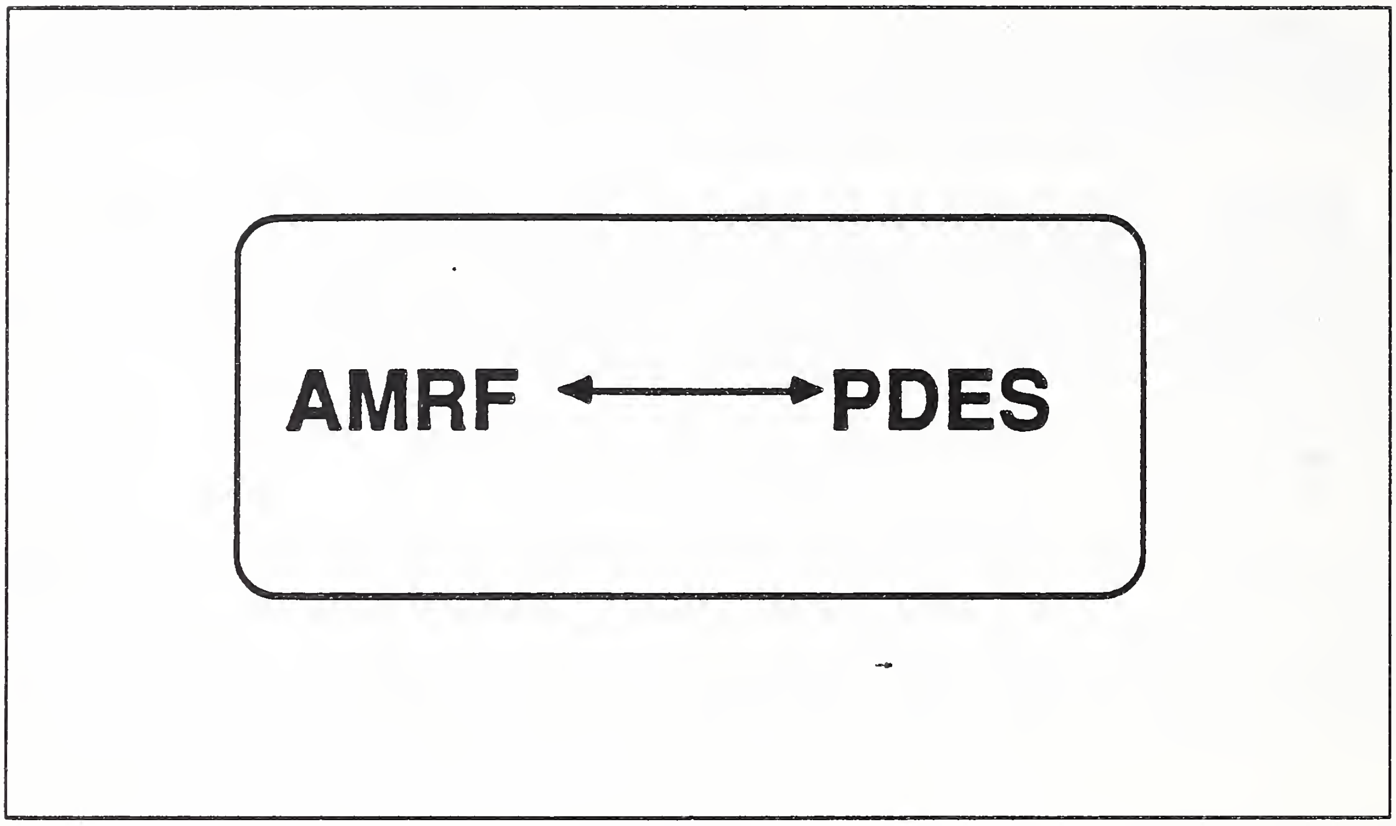
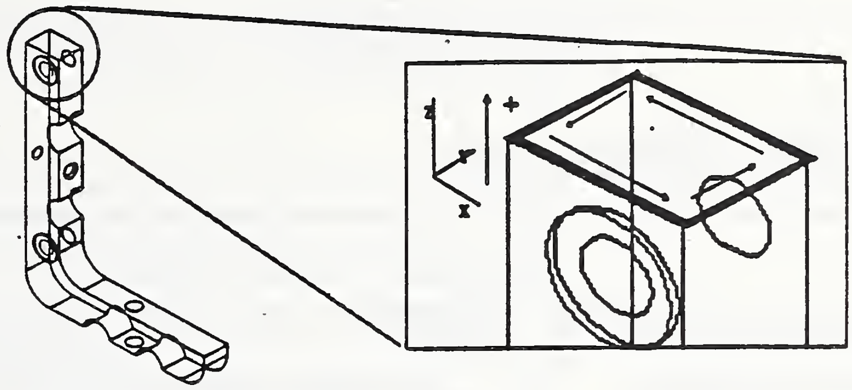


Figure 2

AUTOMATED MANUFACTURING RESEARCH FACILITY (AMRF)

- o Provide a Testbed for Factory Automation
- o Develop Data Driven Automation Concepts
- o Aid and Advise Projects and Standards (CALS, PDES)
- o Technology Transfer

Figure 3



```
/PART_MODEL
/HEADER
   PART_NAME = 'PIPECLAMP_FV' .
/END_HEADER
/TOPOLOGY
/SHELLS
   SHL001;
   FAC001,FAC002, FAC003, FAC004, FAC005,
   FAC006, FAC007, FAC008, FAC009, FAC010,
   FAC011,FAC012, FAC013, FAC014, FAC015,
   FAC016, FAC017, FAC018, FAC019, FAC020,
   FAC021,FAC022, FAC023, FAC024, FAC025,
   FAC026, FAC027, FAC028 .
/END_SHELLS
```

```
/FACES
   FAC001; LOP001; SUR001 - .
   FAC002; LOP002; SUR002 + .
   FAC003; LOP003; SUR003 + .
   FAC004; LOP004; SUR004 - .
   FAC005; LOP005; SUR005 + .
   FAC006;LOP006,LOP030;SUR006 +.
   FAC007; LOP007; SUR007 + .
 FAC011; LOP011; SUR011 +.
   FAC011; LOP011; SUR011 + .
   FAC012; LOP012, LOP032; SUR012 - .
   FAC013; LOP013; SUR013 - .
```

Figure 4

What Is Product Data and Product Definition Data?

Product Data

- Product Definition Data and Product Support Over Life Cycle

Product Definition Data

- A Subset of Product Data Relating to Geometry, Topology, Tolerances, and Features of Components or Assemblies and Materials

Figure 5

PDES Objective

Develop and Apply the Technology Necessary To Communicate Digital Product Definitions Within a Heterogeneous Computing System Environment

Figure 6

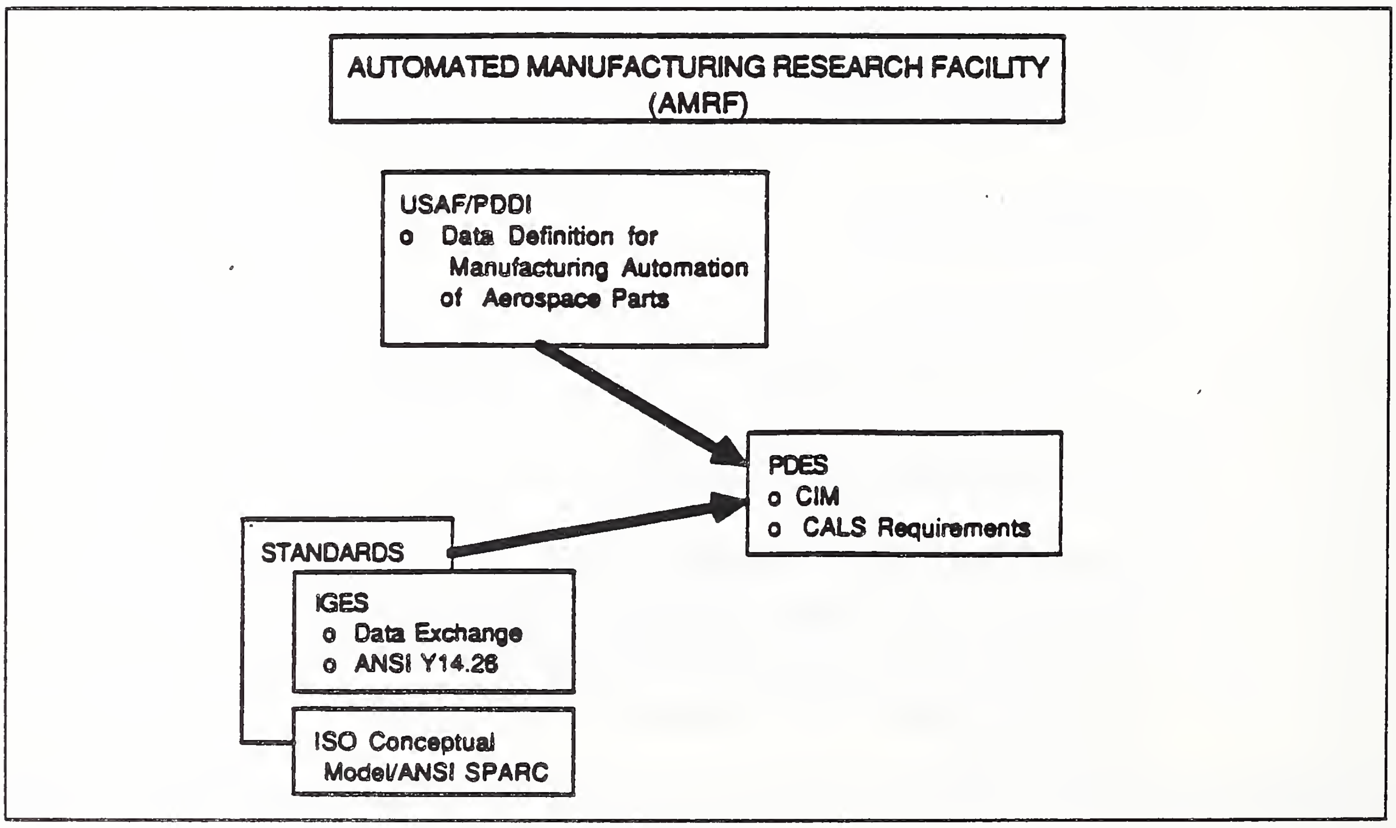


Figure 7

PDES Status — October 1987

National Bureau of Standards (NBS)

IGES-PDES National Chairman - Brad Smith -
Automated Production Technology Division

- Began as IGES Initiative in 1984
- Now Under Joint Development With ISO as World Standard
- Voluntary and Intermittent Effort Contributed by About
  - 260 Companies
  - 655 Individuals
  - 19 Technical Committees

Deliverables

• Proof of Concept and Initiating Activity	→ Completed 1986
• Version 1.0 Conceptual Data Models	→ Under Development
• Initial Testing Draft	→ Released March 1987
• Second Testing Draft	→ Released September 1987
• Model Validation and Integration	→ Begun 2nd Quarter '87 and Continuous
• Third Testing Draft	→ Scheduled January '88
• Fourth Testing Draft	→ Scheduled April '88
• Draft of Version 1.0	→ Scheduled End of '88

Figure 8

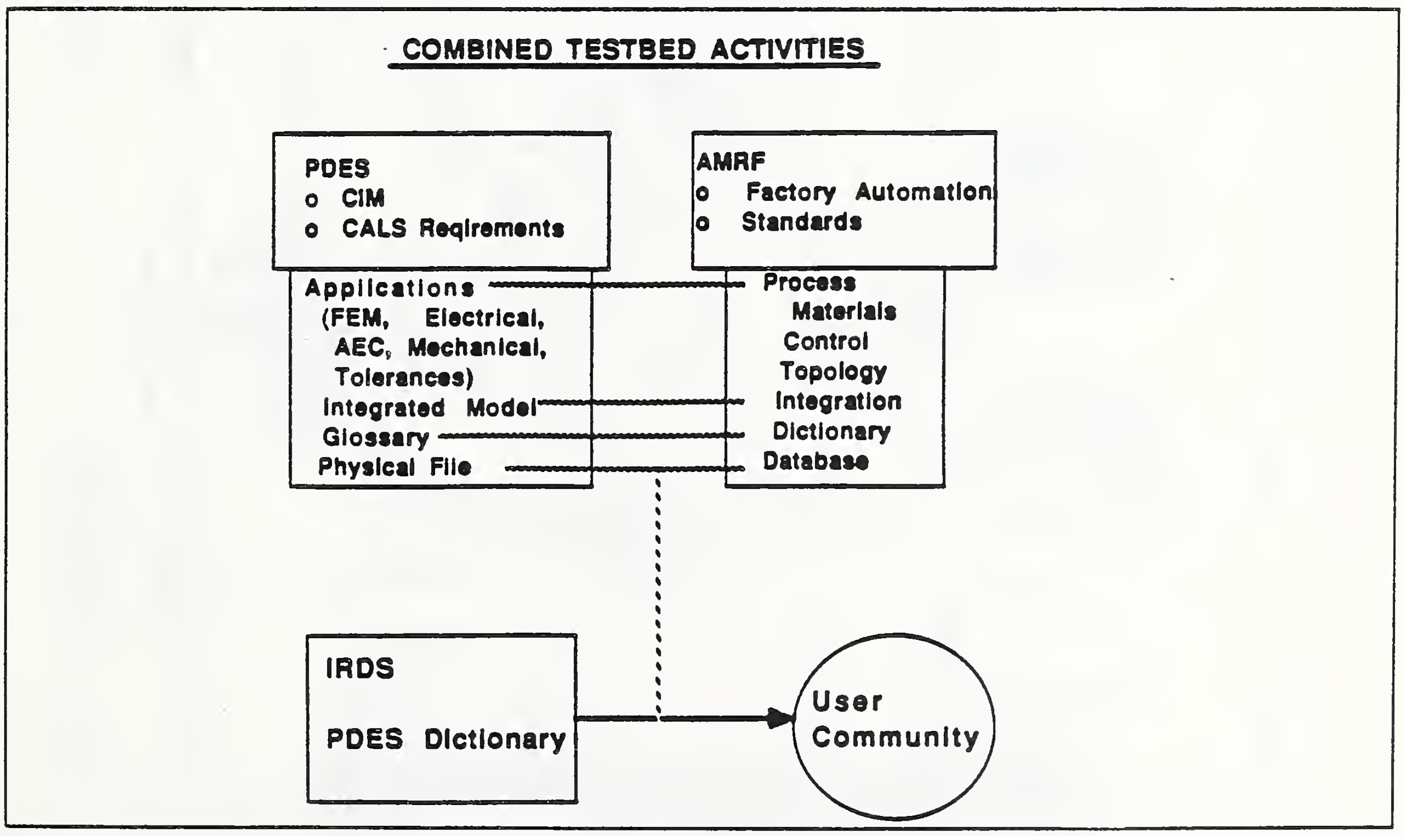


Figure 9

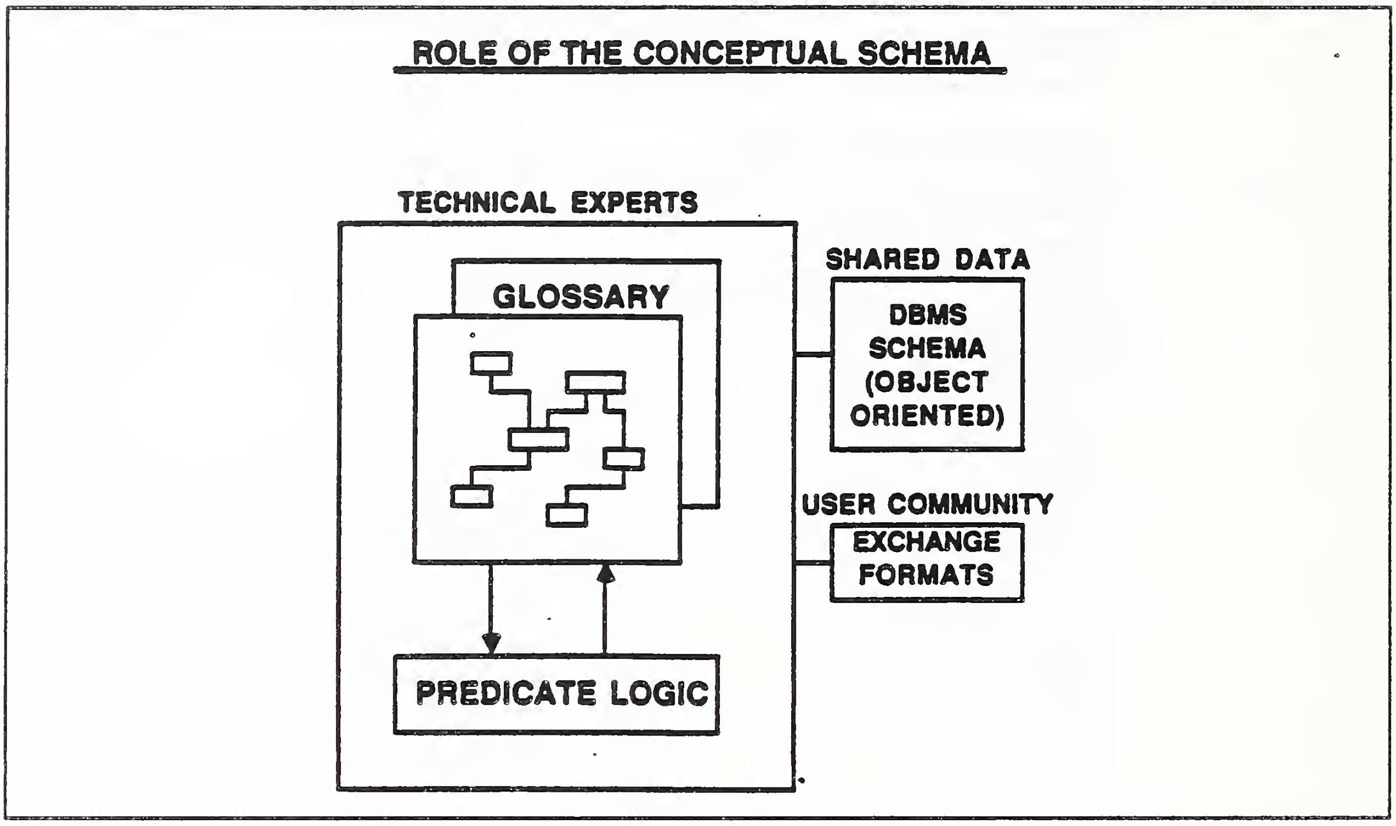


Figure 10

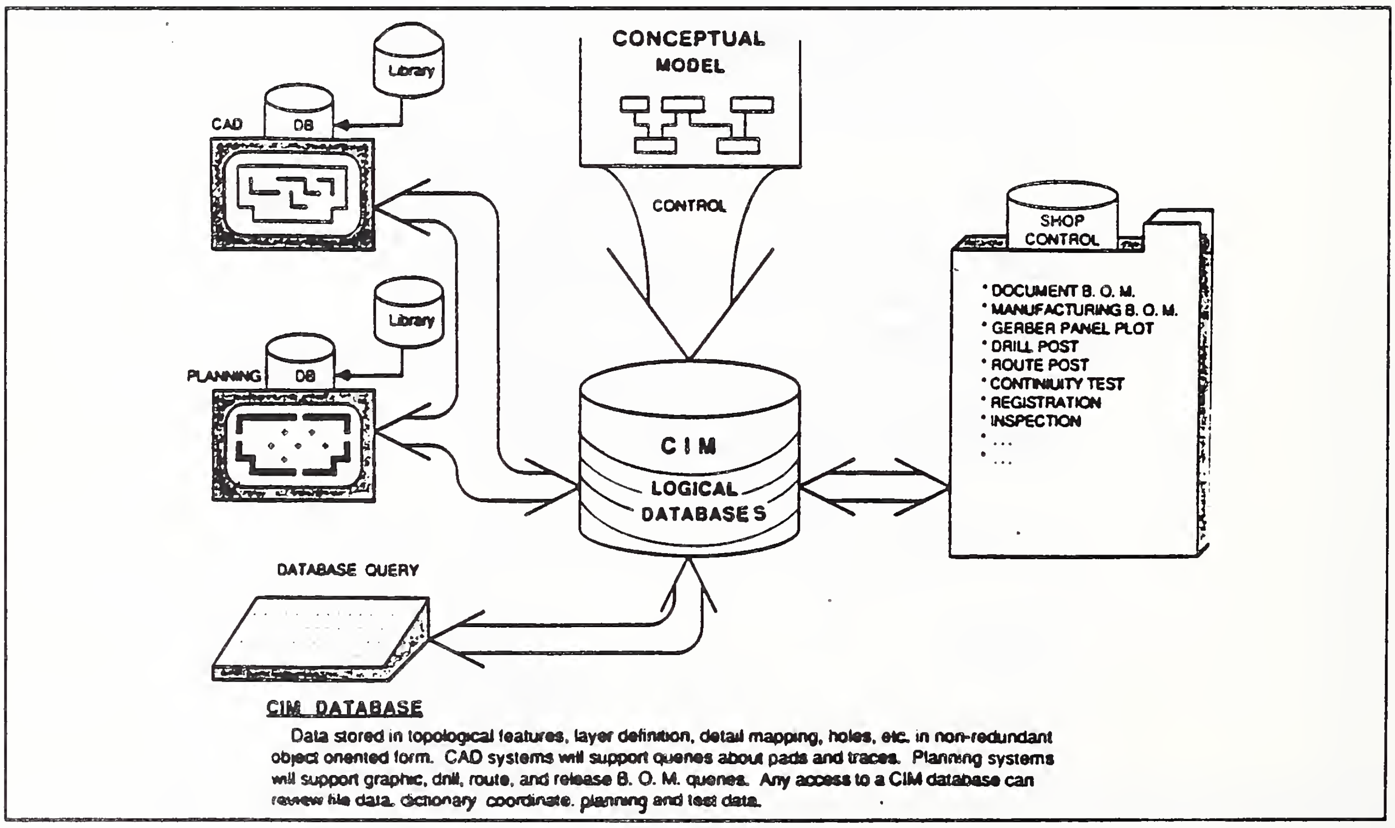


Figure 11

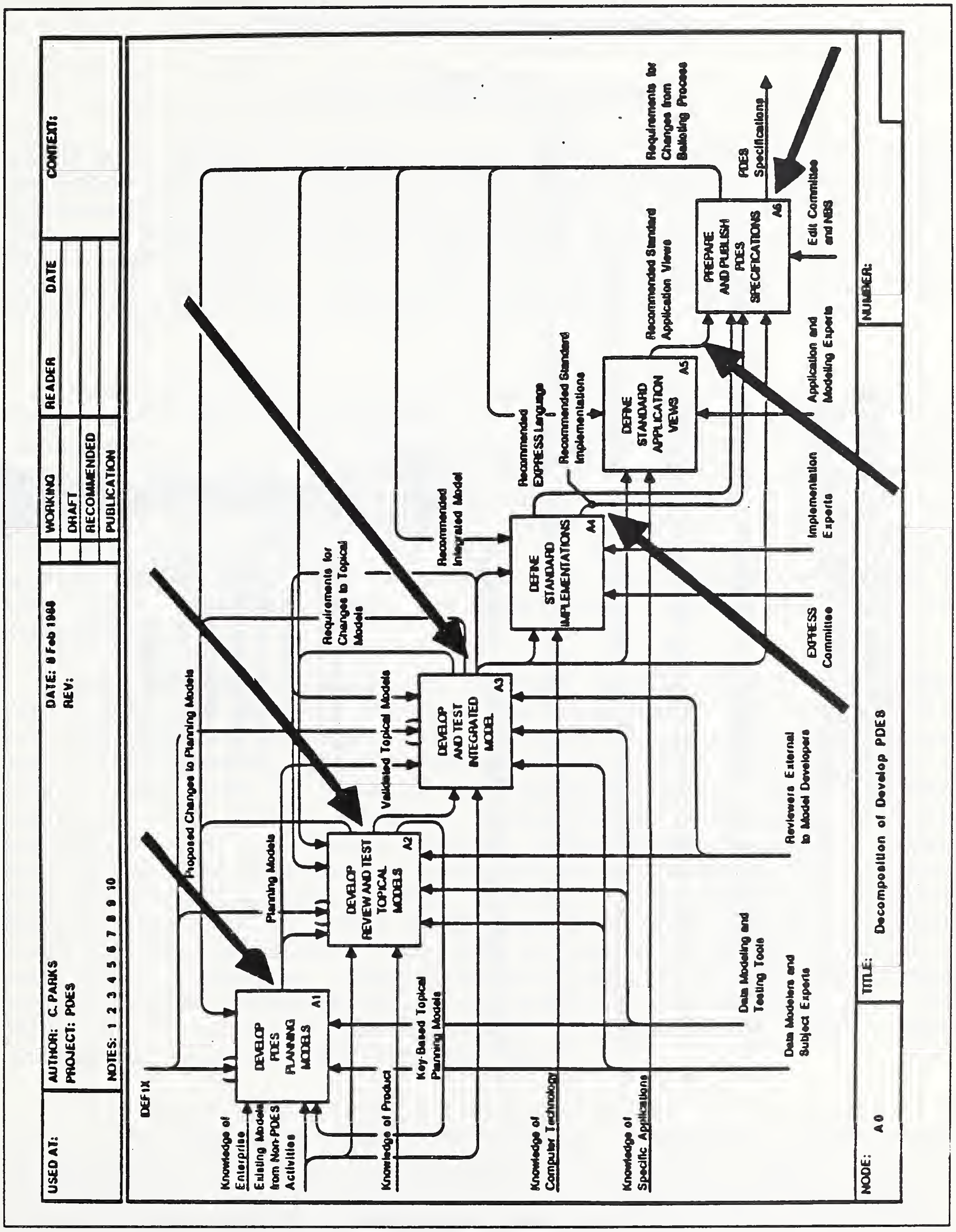

USED AT:
AUTHOR: C. PARKS
PROJECT: PDES
NOTES: 1 2 3 4 5 6 7 8 9 10
DATE: 8 Feb 1988
REV:
WORKING
DRAFT
RECOMMENDED
PUBLICATION
READER
DATE
CONTEXT:
DEVELOP PDES PLANNING MODELS
A1
DEVELOP REVIEW AND TEST TOPICAL MODELS
A2
DEVELOP AND TEST INTEGRATED MODEL
A3
DEFINE STANDARD IMPLEMENTATIONS
A4
DEFINE STANDARD APPLICATION VIEWS
A5
PREPARE AND PUBLISH PDES SPECIFICATIONS
A6
IDEF1X
Knowledge of Enterprise
Existing Models from Non-PDES Activities
Knowledge of Product
Knowledge of Computer Technology
Knowledge of Specific Applications
Proposed Changes to Planning Models
Planning Models
Key-Based Topical Planning Models
Validated Topical Models
Requirements for Changes to Topical Models
Recommended Integrated Model
Recommended EXPRESS Language
Recommended Standard Implementations
Recommended Standard Application Views
Requirements for Changes from Balloting Process
PDES Specifications
Edit Committee and NBS
Application and Modeling Experts
Implementation Experts
EXPRESS Committee
Reviewers External to Model Developers
Data Modeling and Testing Tools
Data Modelers and Subject Experts
NODE:
A0
TITLE:
Decomposition of Develop PDES
NUMBER:

Figure 12

INFORMATION MODEL STANDARDS ISSUES

- Need guidelines for assessing completeness, conceptuality
- Need review mechanism during model development
- Need a means of controlling computer files related to models
- Need a related dictionary suitable for DBMs use (CS 610.?, ANSI X3K5?)
- May require a "soft standards" concept

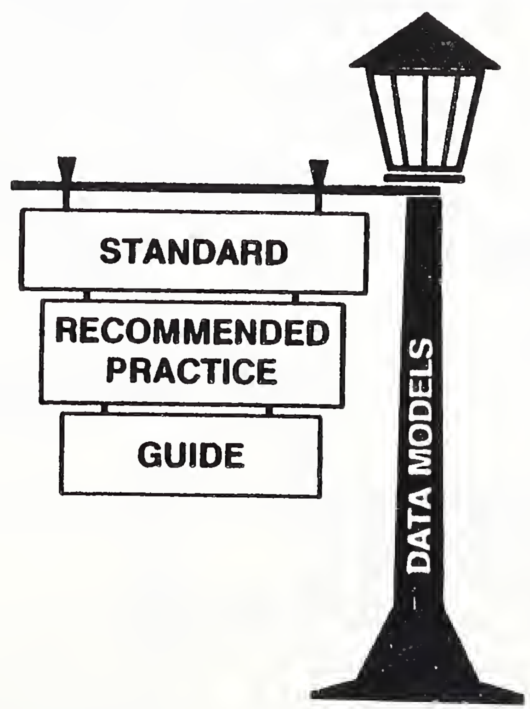


Figure 13

NASA--JOHNSON SPACE CENTER

<u>Speaker</u>
Sandra Anderson (MITRE)

[Editor's Note: Sandra Anderson was asked to attend the Workshop by Davis Howes, the data administrator of the Engineering Directorate at NASA-Johnson Space Center (JSC), who was unable to attend himself. She focused her talk on the general use of the IRDS at the Johnson Space Center. The next speaker, Steve Ritzman, also representing NASA, concentrated on the IRDS and the Space Station program.]

The Information Resource Dictionary System is becoming more and more recognized as an integral part of the management of data at the Johnson Space Center. The initial impetus for using a global IRDS came from the Space Station Program (SSP). The IRDS standard was specified in the Technical and Management Information System (TMIS) Request for Proposal (RFP). The TMIS is an SSP-wide information system supporting the technical and administrative needs of the NASA centers, the international partners, and the customers.

The Data Administration Working Group (DAWG) was formed in October of 1986 to provide the management and integration of data across the SSP. The working group name was changed to the Database Integration Working Group in November of 1987. The DIWG is made up of representatives from the NASA centers, the Space Station Contractors, the international partners, and the Space Station information systems. There are three Space Station information systems: TMIS, Space Station Information System (SSIS), and Software Support Environment (SSE). The SSIS supports the real-time operation of the station. The SSE supports the development of real-time software. The DIWG provides policies and standards for the integration of information across the SSP. The DIWG is to develop and maintain a global data model, data standards, and an IRDS for the SSP.

Data Administration is a recognized function at JSC. It is a new function which is gaining momentum, having been introduced into JSC in the past two years. There are two directorates and one project office that have data administration functions. They are the Engineering Directorate, the Mission Operations Directorate (MOD), and the Space Station Project Office (SSPO). The SSPO is responsible for the administration of JSC's Space Station data, regardless of where the data resides. A good deal of JSC's Space Station data will reside with the Work Package 2 contractor. The Engineering Directorate is responsible for the management of the engineering data at JSC, for both the Shuttle and the Space Station. MOD is responsible for the management of the operations data for both the Shuttle and the Space Station. There is a lot of overlap of data requirements between the directorates, the contractors, and the project office. There needs to be a way of minimizing data redundancy and providing data integration. The IRDS is seen as an important tool that, in conjunction with data standards and a global data model, can provide a framework for the integration of data at JSC. In this environment, where data and data dictionaries are located on information systems distributed across JSC and at the contractors, there is a need for an IRDS which will provide a common standard for interfacing these dictionaries and for the capture of the global data model.

Integration is one of the primary objectives of data administration at JSC. The three schema architecture is seen as one possible approach for defining an environment for that integration. The three schema architecture was defined by ANSI as a means to provide data independence and integration. The three schemas are the internal, the conceptual, and the external. The internal is the structure of the data as it is implemented, the conceptual is the global data model, and the external is the user views of the data.

The conceptual schema is seen as the integrating factor for the data at JSC. JSC is presently looking at the Product Data Control Model (PDCM) being developed by Rockwell for the Air Force. The PDCM is a generic data model representing the information needed in the development of a product. JSC is looking at the applicability of this model in the definition of the information used to develop the Space Station and the possibility of extending this model to address operations data.

We are starting a pilot project using the PDCM to provide integration from the operations area and the engineering area. We have a prototype data dictionary which supports the three schema architecture that we will use to store the PDCM and capture the physical structure of the data as it is implemented.

I don't want to be negative--I think that the IRDS Standard is very good and needed for interfacing dictionaries, but we need to have a more comprehensive standard for the content of the IRDS. This is where I think that Module 2 of the IRDS Specifications falls short. Figure 8 describes the meta-data to reside in the IRDS proposed for JSC. The content of the IRDS can be grouped from two perspectives. One is the three schema and the other is the three categories: "data," "process," and "other." "Data" pertains to the information about the data, such as data model, database, etc. "Process" pertains to those process-oriented items such as the system that processes the data. "Other" is a catch-all for anything other than data or process, such as sponsor or project. The meta-entities that are highlighted are those specified in Module 2 of the IRDS standard.

Module 2 addresses only meta-data describing the internal schema. These meta-entities define what exists on the system. I see this as leading to a bottom-up approach, ignoring the three schema approach and the conceptual model. Our fear is that we will have these complex tools, but instead of providing an integrated environment with a conceptual model, there will be a tendency for people to take the IRDS and implement it in the old way of doing business by taking and documenting data as it exists rather than attempting to integrate.

Data integration is a significant objective at JSC for both Shuttle and Space Station data. The recognition of data administration in the SSPO, MOD, and Engineering shows management's emphasis in this direction. The conceptual data model and the three schema architecture are seen as the foundation of this integration. The IRDS can play an important role if it captures the conceptual aspects of the data. This is where we see the IRDS standard lacking. The standard does not address the three schema architecture and it does not specify that the data model be included in the starter-set. Module 2 includes those meta-entities that document the physical aspects of the system without addressing the conceptual. It is the conceptual that can best provide the foundation for data integration.

Question: I'd like to pick up on that last point. Keeping in mind that the Basic Functional Schema is extensible, are you arguing for the development of additional modules that would support, for example, the three schema architecture, or a global data model?

Answer: I'd like to see an extension of Module 2 so that it would be more comprehensive. I realize that you can extend the IRD definition to define your E-R model or your data flows, but in the environment that we're working in there is a cultural gap. At first review of the IRDS standard documentation, it was perceived that the entity-relationship diagram was part of the standard. We realize now that it's not; it's just part of the approach to define the IRDS. If the data model were part of the Standard, it would provide the impetus needed to gain acceptance of the conceptual data model into the NASA environment.

Question: So you're saying that there should be a standard extension of the Basic Functional Schema, and that, in the same sense that the Basic Functional Schema was defined in the first place, there should be an official, standard content module to support some of these other things?

Answer: Yes, definitely.

Question: In order for something to become a standard, there pretty well has to be consensus. When it comes to a global data model, do you think you can ever achieve consensus?

Answer: I'm not looking for a standard of the actual data model of NASA's data requirements, but of the types of objects that you could store in the IRDS, such as "entity," "relationship," and "attribute."

Question: But I still ask the same question. Do you think there ever is a chance for consensus?

Answer: We are looking at the data model developed by Rockwell, the PDCM, to see if it is applicable to the JSC environment, because we would be miles ahead by starting with something that has been under development for years, rather than starting from scratch. A consensus would be needed for any global data model developed, regardless of whether it was based on the PDCM or not. The DIWG has as its agenda to develop and maintain a global data model for the Space Station.

IRDS/NASA

Sandra V. Anderson

24 March 1988

Figure 1

Background of IRDS/Space Station

- Space Station (SS) Technical and Management Information System (TMIS)
  - Space Station Program-wide information system supporting
    - NASA Centers
    - Headquarters
    - International Partners
    - Customers
  - IRDS standard specified in TMIS RFP

Figure 2

Background of IRDS/Space Station (continued)

- SS Database Integration Working Group (DIWG)
  - Formed October 1986 as Data Administration Working Group (DAWG)
  - Changed name to Database Integration Working Group (DIWG) November 1987
  - Representatives from the NASA centers and support contractors

Figure 3

Background of IRDS/Space Station (concluded)

- SS Database Integration Working Group (DIWG) is tasked to:
  - Provide policies and standards for the integration of information across the SSP
    - -- Technical and Management Information System (TMIS)
    - -- Space Station Information System (SSIS)
    - -- Software Support Environment (SSE)
  - Maintain global data model for Space Station
  - Provide an interim Data Dictionary (DD) for SSP
  - Develop requirements for Information Resource Dictionary System (IRDS)

Figure 4

Background of IRDS/JSC

- Data Administration
  - Space Station Project Office (SSPO) at JSC
    - Data Administrator for Space Station Data at JSC
  - Engineering Directorate
    - Data Administrator for Engineering Data both Shuttle and Space Station at JSC
  - Mission Operations Directorate
    - Data Administrator for operations data both Shuttle and Space Station at JSC

Figure 5

Background of IRDS/JSC (concluded)

- Need for IRDS across JSC for Shuttle and Space Station Data
- Work Package 2 (WP-2) contractor
  - Interface WP-2 DD to SSPO IRDS

Figure 6

IRDS Standard

- Need DD to support the integration of shared information at JSC
- Three Schema Architecture is needed to define this environment
- Integration of data based on conceptual schema (global data model)
- Only meta-data defining the Internal Schema specified by IRDS Standard

Figure 7

Metadata to Reside in IRDS Proposed for JSC

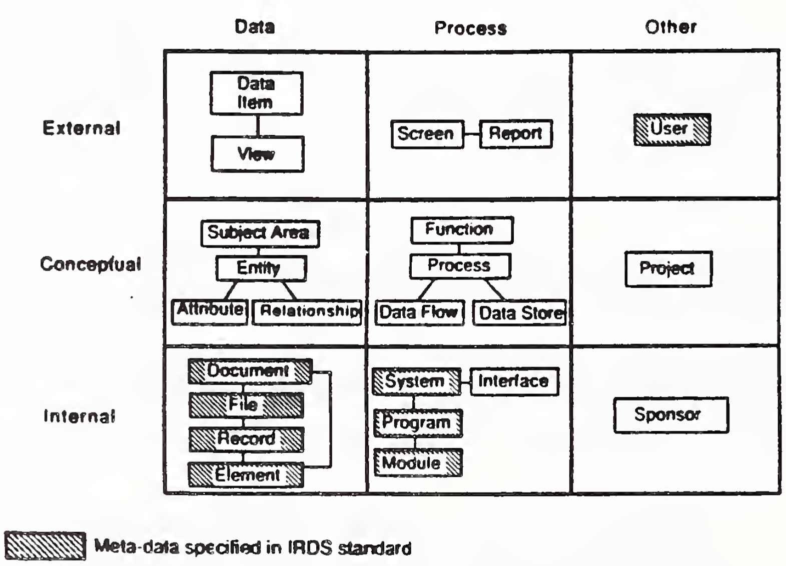


Figure 8

Summary

- IRDS standard is being introduced across NASA with its beginning in the Space Station Program (SSP)
- Data Administration function has been initiated at JSC in the

  - SSPO
  - Engineering Directorate
  - Mission Operations Directorate
- However, the IRDS standard is lacking in that:

  - Does not cover all three schemas
  - Needs to capture the global data model

Figure 9

NASA--SPACE STATION PROGRAM

Speaker
Stephen J. Ritzman (Booz, Allen & Hamilton, Inc,)

Let me begin by giving you an idea of why I'm here, who I represent, and what my role is. I'm here as a representative of NASA's space station program managed out of Reston, Virginia. My talk concerns the Technical and Management Information System (TMIS) project, but actually is a little bit broader in that I'd like to talk about issues related to other systems within the space station program and the impact of the IRDS on them as well. The way I'd like to do this is, first of all, to borrow your telescope, Mary Lou, and look up at the IRDS, but maybe a little bit further towards where the space station is supposed to be, and get an idea of how the IRDS can work in the environment that we're envisioning, in the information environment surrounding the Space Station Program.

First of all, I'd like to talk about the information system goals in the space station program, then mention three specific information systems that are currently under development. I'll then look at information resources that the space station will have to deal with. A lot of this will just be defining the breadth, scope, and diversity of the information resource that we're going to have to harness with the IRDS. Then I'll identify some working groups that were put together to solve the "ility" problems. "Ility" stands for "Interoperability, Data Transportability, Commonality." I asked the gentleman yesterday if he had a definition for interoperability because I was at a meeting and somebody came forward with, I think, eleven or thirteen levels of interoperability in the space station program. This is an important topic, and I think that there are a lot of good opinions about it. Finally, I'll be talking about the IRDS goal and the plans we have of getting to that goal in the space station program.

We have really three basic goals or challenges relating to the space station program information systems. The first is information management. What I mean by this is that we have a very diverse environment. We have international partners, NASA organizations that are not part of the space station program, different levels of organizations within the space station program, and four major NASA centers located around the country, each working with separate work package contractors. We have the potential of people, say, here at NBS or at different institutions, using the space station by setting up an experiment and running that experiment from their own environment. So we are dealing with the management of the information that's going to be passed from sources to destinations, whether they're between ground-based destinations, say a scientist to someone in the space station program, or from ground-based to the space-based elements, or even between space-based elements. I'll talk a little later about a figure I have that defines some of those areas. So we're concerned with managing or transferring that information and providing some common interfaces to that information, to make our work a little bit easier and the use of our work in future cycles of the program a little more useful.

We're also concerned with program management. In that same environment, you've got a real problem with respect to program management. When you've got multiple organizations, multiple levels of the same organization, all with very strong, good opinions on how things should be done, you need a centralized--I don't want to say a centralized database of opinions and issues--but you need some way to centralize the diversity of issues and opinions about how the program should be developed, and you want to get that program information to the right people at the right time so that they can do their portion of the project effectively.

Finally, you have software development. Currently, we're having software developed not just by one source but by a variety of sources. It's difficult enough when you have just one organization developing software, but when you have four or five organizations developing software, and you're also integrating software from previous projects for which you did not set up standards and that you may have to make some changes to, you need a very good software development and support environment to carry on that activity.

So with these three goals in mind, and with these three challenges with respect to the information in the space station program, there were three systems, or types of systems, that were established to solve those problems. The first is the Space Station Information System (SSIS), which deals with getting information from sources to destinations in an environment that looks something like Figure 5. What we're looking at is the information flow between some user facilities on the ground, or possibly some experiments on the space platform, or from someone who mans the space station. It's a complex environment, and the space station information systems are being defined with common interfaces, where possible, and the use of standards. In the area of communications, we're going to use the OSI standards where applicable. We're looking at the idea of using the IRDS as the standard data dictionary. There are a lot of other national, NASA, and space station program standards that we'll be using, to try to bring the project to a manageable and, hopefully, workable place where we can replace and add components and people, and the project would still go on.

Question: Are you trying to use the IRDS to characterize the information that's being passed around, or to characterize the components of the communications, or both?

Answer: At the current time, we're establishing requirements for the use of the IRDS. My feeling is that we will probably do both, in some fashion, within the space station program. I have a feeling that it's like a lot of things--once you get the snowball rolling, it builds up a lot of momentum. I certainly feel that it will add a lot of value to our program.

The second information system is the Technical and Management Information System (TMIS). This again is to address the problems of distributed management with a variety of perspectives on how things should be done. Figure 6 is a picture of the TMIS environment. What I really want to point out is that TMIS is being defined with commercial, off-the-shelf products. The goal is to do little or no development in the areas of electronic mail, project management packages, DBMSs, document management, scheduling, etc. We want to buy everything we can and plug it together as much as possible. To expand on TMIS a little bit, there's also going to be a variety of things that are called information systems, which to me are applications that have large amounts of technical and management data associated with them, very little algorithmic processing, and a screen interface with users. These systems will also pop up in this environment, so that an experimenter who is trying to determine the feasibility of his experiment with things that are currently going on, or could be done, can access the databases in the TMIS world to verify or validate whether or not his goals and projects are useful. So there's going to be a lot of data that the TMIS world has that's going to be useful to a lot of people. And then of course, the SSIS can be used to transfer that information, for someone who doesn't have direct access to a TMIS station.

So we have information transfer systems in control of the space station and technical and project management systems. Unfortunately, I don't have a real nice picture of the Software Support Environment (SSE) world, but I will say that this system is currently being developed down at the Johnson Space Center by a single contractor. It's a very sophisticated software support environment. It's got strict and stringent configuration management control based on the system life cycle process. I've been here the last couple of days listening to what people are saying, and I'm wondering, gee, I know the IRDS is going to fit in there, but I wonder how easy it's going to be to put it in, because they're already in development of the SSE and they have future goals with it, but with their goals in configuration management and life cycle development or control, the IRDS seems to be a very good candidate for solving their problems. In fact it'll be interesting to see how they plan to solve their problems, because they've got to have some kind of data dictionary somewhere that does a lot of the same kind of stuff.

In summary, the types of information systems in the Space Station Program are systems for the transfer of information with common services to that information, technical and project management information systems, and software support environment or software development information systems are the types of things we'll be dealing with within the space station program information area.

Now, let's look at three characteristics of the information resources that are going to exist in these different information system domains (Figure 7). The first is diverse information sources. We're going to have databases, file systems, commercial off-the-shelf products data, and a variety of data that I'm probably not even aware of. Some will be developed within the space station program. Some will be developed by other organizations within NASA, some will come from the international arena, and who knows where the rest might come from. So it's a variety of diverse sources and types of data or information that has to be captured, and I believe the IRDS can do that. Secondly, the resources are widely distributed. It's going to be global information, in the sense that it's world-wide information. We will want to know about information relating to the space station that sits in Europe, Japan, and Canada. We've got to capture information that sits on the space station itself. We've also got to capture information that sits at the different institutions where scientists are working on experiments that are based on the space station. Finally, we have information resources governed by distributed management control, some with different views on how the resources should be used within project management or execution of systems. I hope this gives you an idea of the variety and diversity of information resources that have to be captured in some kind of a dictionary system.

Finally, to help solve these problems, and to deal with some of the "ilities," transportability, interoperability, commonality of environment, a number of different working groups were established by the space station information system office (Figure 8). I've been involved with the working groups on Database Integration, Operating System Services, and Networking Services, but I'm here primarily in connection with database integration. Our goal is to provide database commonality throughout the space station program. As Sandra Anderson mentioned, the group started out as the Data Administration Working Group, but at that time it was related only to the TMIS environment. Now it's been given the charter of establishing database commonality throughout the space station program, so it's no longer just data administration issues related to TMIS, but database integration across the entire program. A definition, that is not necessarily endorsed but hasn't been scoffed at yet, is my concept of database commonality. I know that this was a definition that was used for some of the high level requirements for database commonality in the space station program. Database commonality is just the agreed use of standards, policies, procedures, and guidelines for the definition, development, access, and administration of data and databases. Again, we see the IRDS coming into play in the area of enforcement of standards, and in the administration of data and databases. We're currently looking at the naming conventions as being a standard for our data and databases.

The Database Integration Working Group leads me to my final slide, Figure 9, and that is our singular goal with respect to the IRDS. We hope to have a space station wide ANSI/FIPS IRDS. We are endorsing the ANSI/ISO/FIPS standards for our work on the space station. As far as our plans, we are looking at and actually are playing around with several systems to serve as an interim data dictionary, to solve some of our early problems and to gain some understanding and insight in the data dictionary arena. We're also forming a team to work on formal requirements for the data dictionary within the space station program.

Question: If you're endorsing the ANSI/FIPS IRDS, what are you doing formulating requirements?

Answer: I guess I should ask what you mean, but I won't. I had a similar question. If we've got a standard out there, and it's a specification of what the thing looks like, why are we coming up with "requirements"? An answer that I have in my own mind for that is first, that we need to determine which components of the standard we need to bring into our environment, and at what time. In other words, which optional modules would we use, and when. The SSE is, from what I can see, a very good system in the area of life cycle and product management. Well, if we view life cycle as being important with respect to the data dictionary, how are we going to do that merging, or pushing forward, say? That, I think, is the driving force behind the requirements. Also, knowledge. I think that with the IRDS, as with other tools, there isn't as much perception of need as conception of need. So the feeling is "let's use it," rather than the question "why do we need to use it."

Question: I think it may also have to do with content, that Module 2, the Basic Functional Schema needs to be expanded with respect to the space station program.

Answer: Right.

IRDS

WITHIN THE

SPACE STATION PROGRAM

Figure 1

- SSP INFORMATION SYSTEM GOALS
- SSP INFORMATION SYSTEMS
- SSP INFORMATION RESOURCES
- SSP WORKING GROUPS
- IRDS GOAL & PLANS

Figure 2

SSP INFORMATION SYSTEM GOALS:

- INFORMATION MANAGEMENT - provide automated information management across the SSP over the full SSP life-cycle.

- PROGRAM MANAGEMENT - provide automated tools to facilitate the management of the program development process.

- SOFTWARE DEVELOPMENT - provide automated tools to minimize the cost and risk of program software development.

Figure 3

SSP INFORMATION SYSTEMS:

- Space Station Information System (SSIS)

- Technical and Management Information System (TMIS)

- Software Support Environment (SSE)

Figure 4

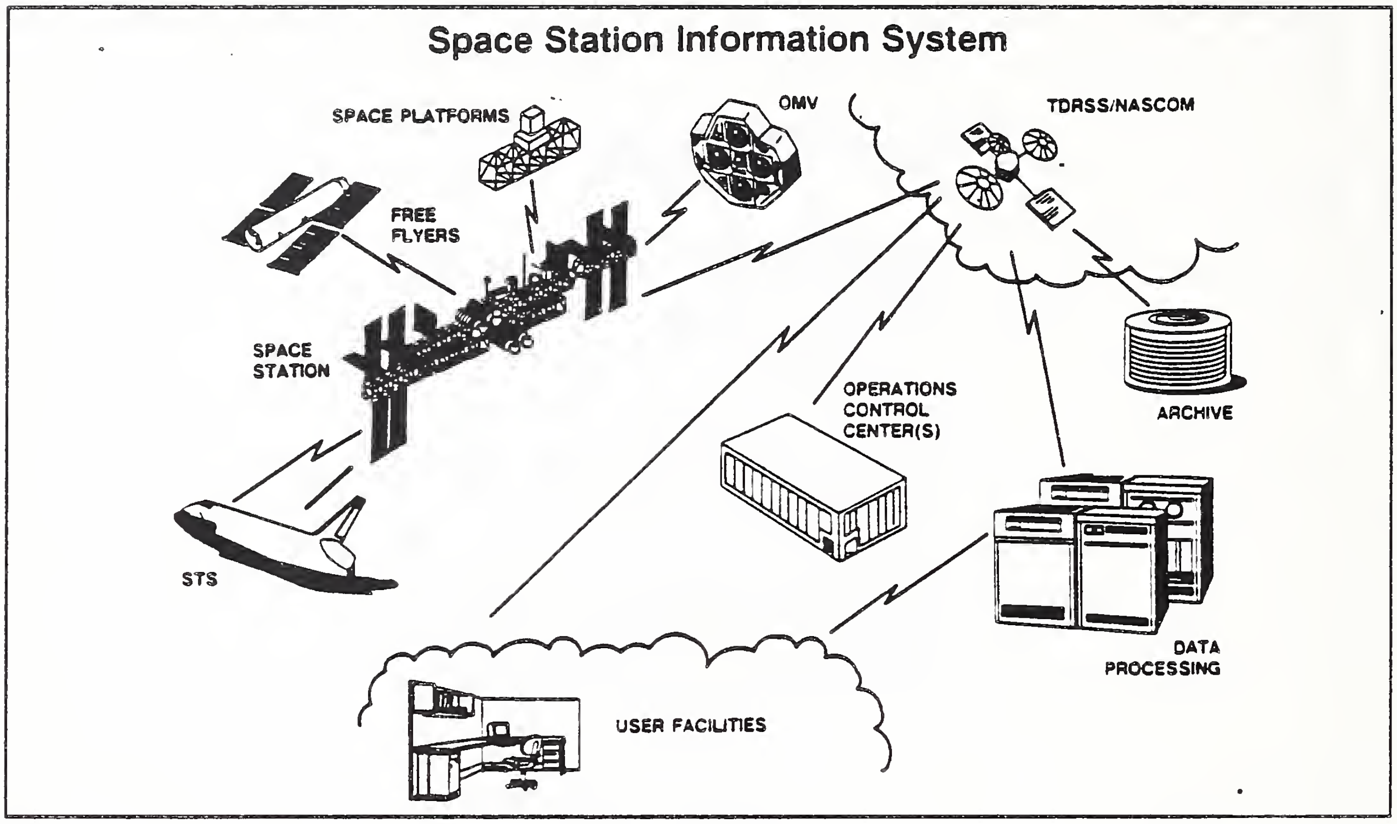


Figure 5

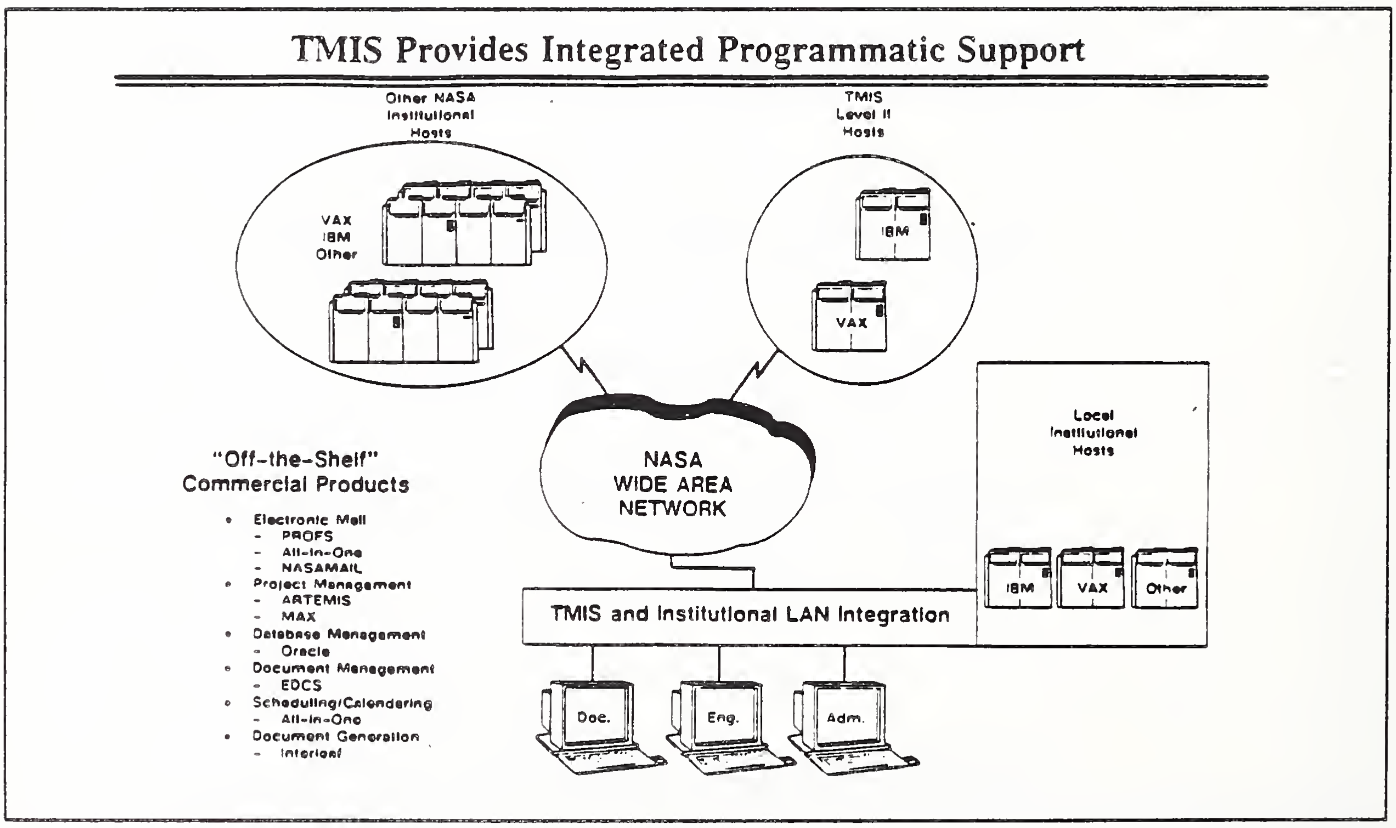


Figure 6

SSP INFORMATION RESOURCES:

- Diverse Information Sources
- Widely Distributed Data
- Distributed Management Control

Figure 7

SSP WORKING GROUPS:

- Database Integration WG
- Operating System Services WG
- Networking Services WG
- User Support Environment WG
- AI and Advanced Technology WG
- Security and Privacy WG

Figure 8

IRDS GOAL:

- SSP-WIDE ANSI/FIPS IRDS

IRDS PLANS:

- Interim Data Dictionary

- IRDS Requirements (RDD)

Figure 9

U.S. AIR FORCE--ROME AIR DEVELOPMENT CENTER

Speaker
Patrick McCabe

Good afternoon. My name is Pat McCabe and I work for the Rome Air Development Center (RADC). I'm involved with the Center's research into the area of command, control, communications, and intelligence database techniques. I thought I'd start with a little bit of a discussion on what exactly RADC does.

The role of RADC (Figure 2): We don't really build database management systems, and don't necessarily build databases directly. We are not in a position to set standards. What we do is try to identify emerging standards and technologies and to monitor existing standards and technologies for use in statements of work, specifications, and technology forecasts which we can then fold into the development of advanced systems.

Figure 3 is an overview of the information system architecture that we've been working with for some time. It's relatively generic; it goes across quite a few commands and quite a few different application areas. We have these kinds of inputs on the left, and those are the outputs on the right, with many of our users being essentially knowledge bases. We have a communications shell around the system, with the database generation shell and what we call "correlation and fusion" capability on the other side. What we focus on, in my area of interest, is the database itself.

Figure 4 shows the model that we use for a generic database. We have a rather unique view of the Information Resource Dictionary System, in that we view it actually as an active control mechanism, rather than being a database itself, or enveloping a traditional database. As such, it presents an interface for the communications, applications, and other capabilities within the outer circle to interact with the database itself, and then we come back out.

Our initial interest in the IRDS came from our need to communicate between different and very heterogeneous databases and database management systems (Figure 5). We saw the IRDS as a mechanism that we could use to try and present a network view of the information that is available for users, so a user can, in a transparent way, go out and request different kinds of data that might meet his needs. We also see similar kinds of applications in what we call "diverse object databases" where you have a database consisting of things like photographs, bit-map graphics, natural language text, or formatted kinds of record oriented information that we traditionally associate with a database.

So we view the IRDS really as a control mechanism. We're building on top of it some research capabilities, looking at some alternate solutions to what we see as the interoperability problem across different database management systems and within different types of database management systems.

Question: So you're using the IRDS as a way to solve problems connected with an active data dictionary, and you're doing research into that?

Answer: Yes.

Question: Good. It's a real problem.

Question: Could you describe your hardware and software environment, or have you gotten that specific?

Answer: My group works with long term, down the road techniques rather than currently operational systems. What we have in the field today, running together, is IBM mainframes, Honeywell mainframes, a couple of Amdahls, a lot of minicomputers running different kinds of applications--DEC, Data General, Hewlett-Packard. We've also got a lot of very strange micros that are starting to pop up in the guise of workstations--home grown workstations, commercial off-the-shelf workstations, and they're all starting to have their own databases built on top of them. They all have completely different hardware suites, and even some of the systems with the same hardware have different software suites. We use Cullinet, M204, a couple of home grown ones, one called SARP which was developed by Eaton Corporation back when it was Bunker-Ramo, and UNIFY in the micro area. It runs quite a gamut, and we're hoping to come up with a way to start to have a higher level of abstraction to unify this mess.

Question: I'm familiar with the Air Force personnel system that logically and conceptually is very powerful and useful. Are you going to take that into consideration? Are you going to build a pilot implementation once you figure out what the solution is?

<u>Answer</u>: The people who sponsor us come out of the command, control, communications, and intelligence side, so what we'd do is probably focus on an information system from one of the operational commands, get their support, and build a pilot project around that, and try to go from there.

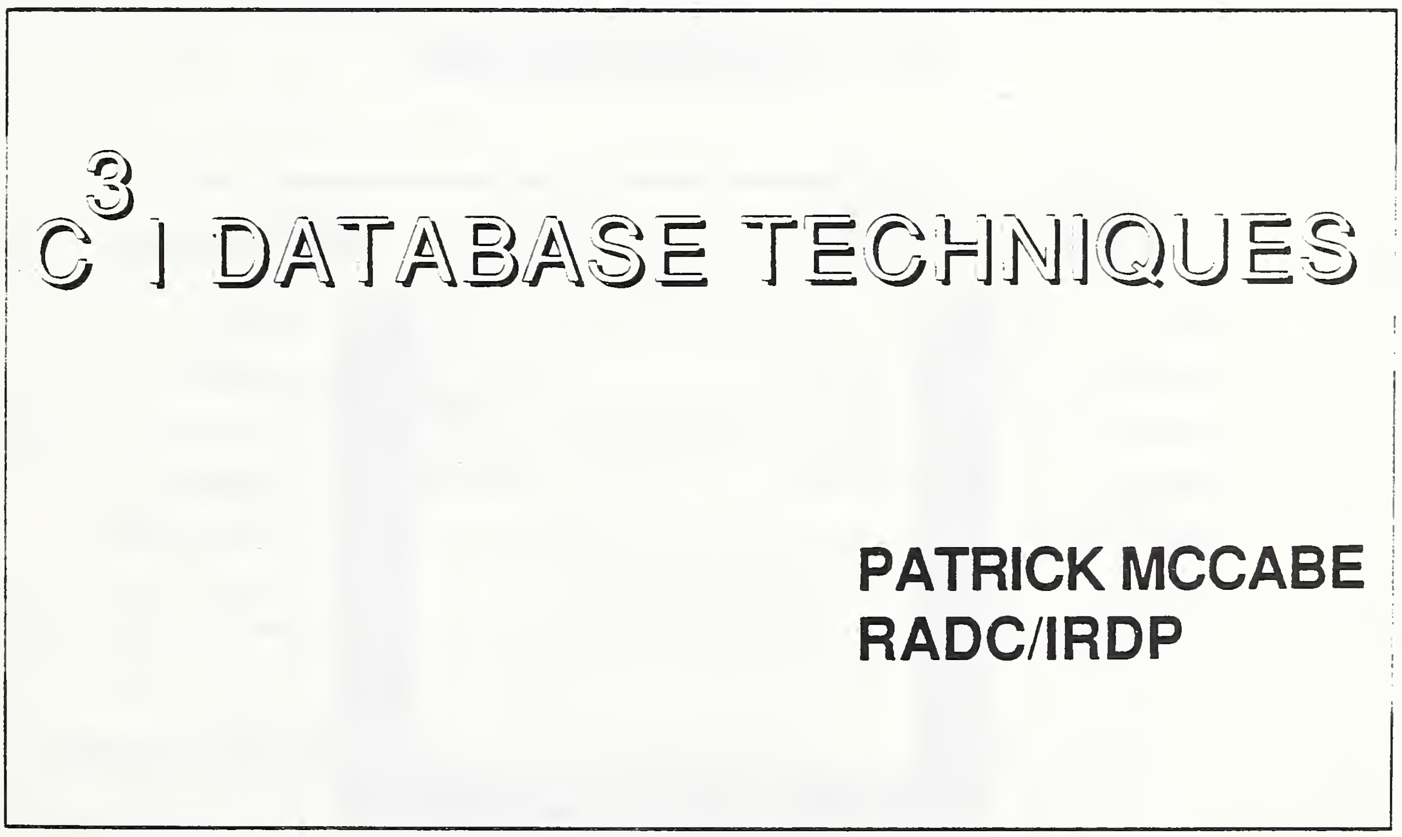


Figure 1

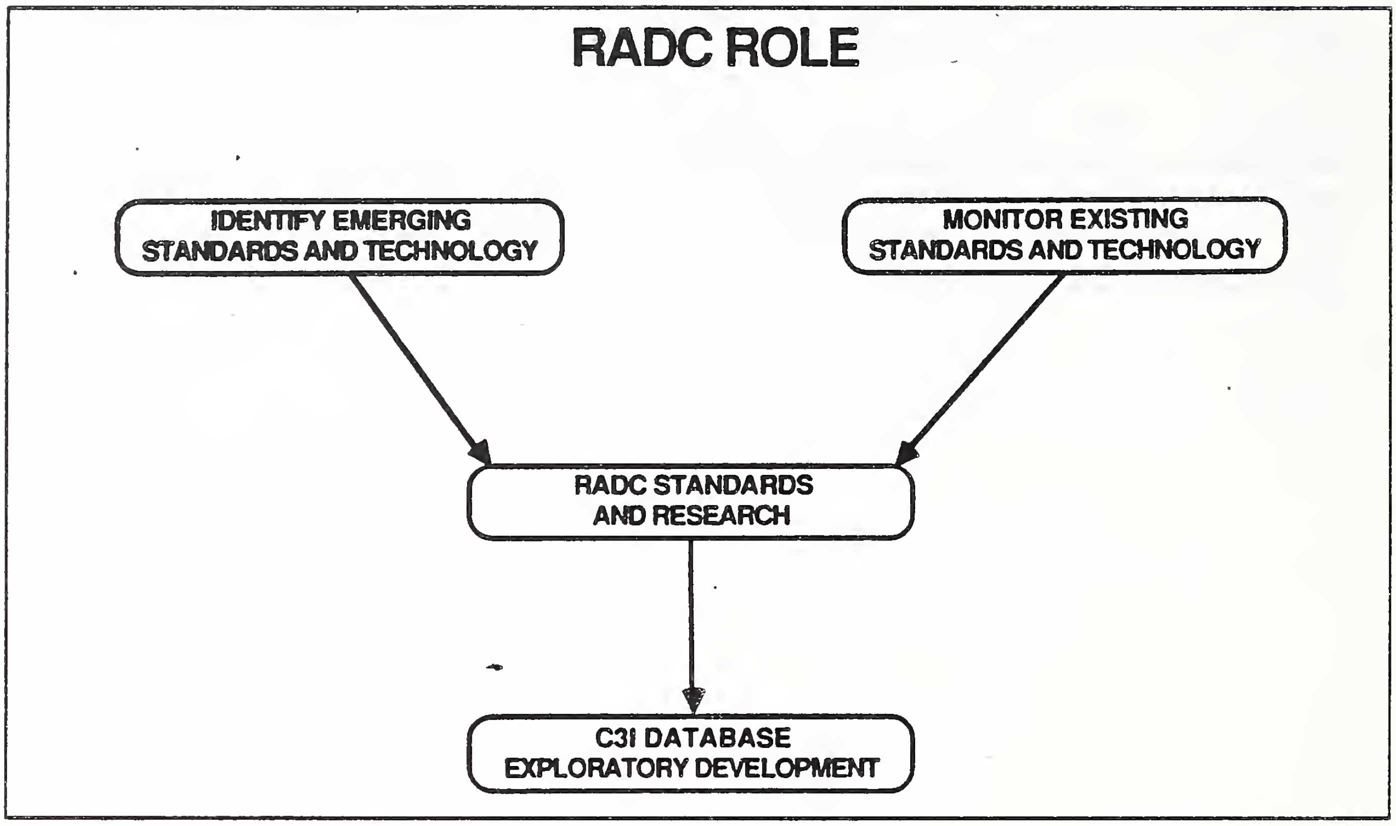


Figure 2

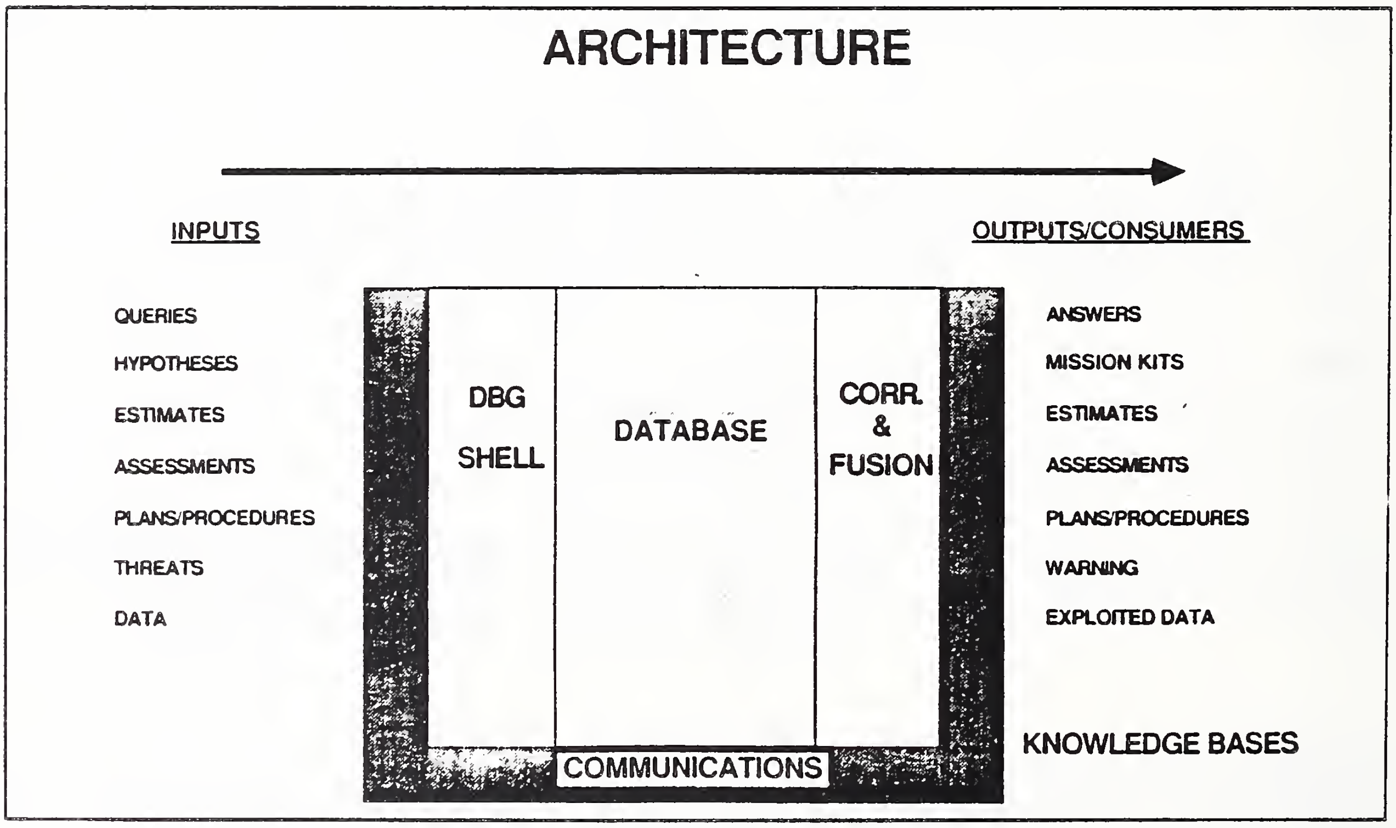


Figure 3

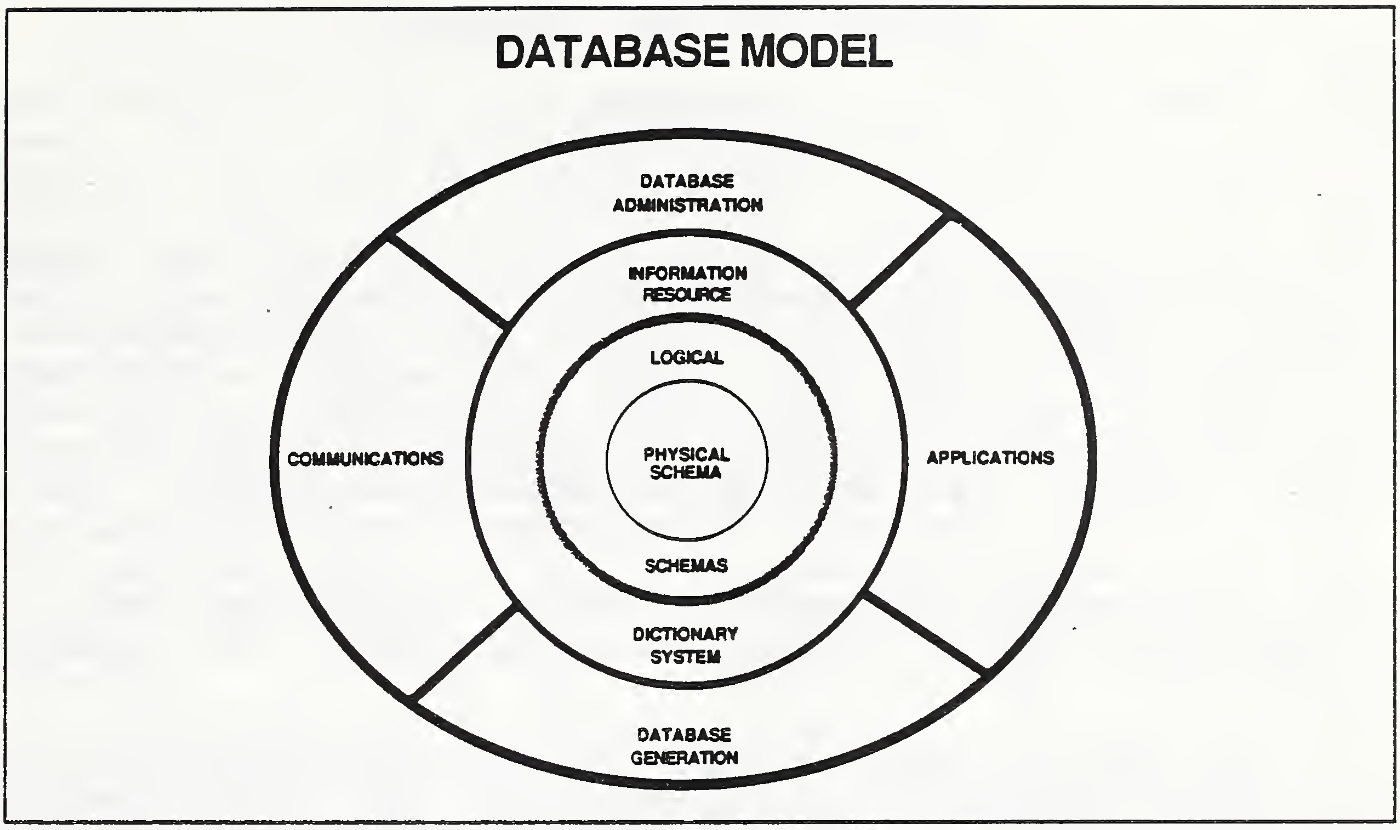


Figure 4

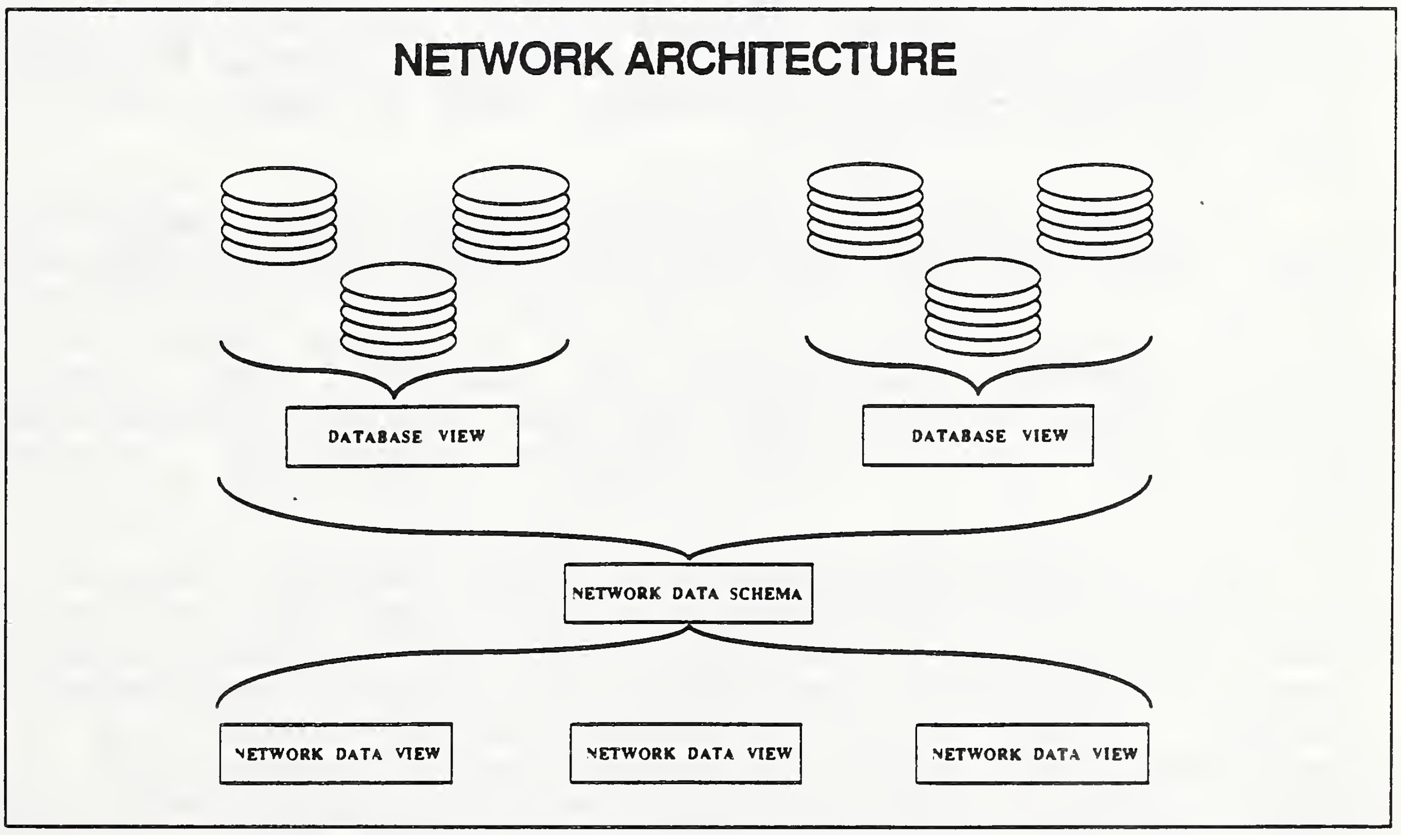


Figure 5

ARGONNE NATIONAL LABORATORY

Speaker
Greg Robinson

My name is Greg Robinson. I work for the Argonne National Laboratory, specifically the Energy and Environmental Systems Division. One of our responsibilities is information systems for energy management, environmental management, and modeling. The Organization of the Joint Chiefs of Staff is an agency to support the operation of the Joint Chiefs. It's made up of eight directorates. The one that we're working with is the Force Structure, Resource, and Assessment Directorate--J8. J8 is in the position right now of trying to modernize their operations concerning computer hardware, software, data management, and organization.

The specific part of this that I'm working on is related to data management in association with the simulation models. J8 runs a large variety of simulation models, ranging from 65,000 lines of code to over 350,000 lines of code. Almost all of these are designed in an old-fashioned, FORTRAN, flat-file input structure, with very little coordination attempted between them. What we're doing in our support program for the modernization effort is identifying various operational needs. One of these is for an IRDS capability (Figure 1).

J8 uses multiple simulation models with common data sets. They are derived out of a common set of algorithms produced in the late 1960s and early 1970s. But each model is run independently, and data is collected independently. Often, officers responsible for various models request similar types of data from the same external source almost at the same time. We're trying to help them coordinate that.

So in the coordination of all the simulation models, and in the need for standardization, is where we see that the IRDS can help significantly. The IRDS is also needed to improve coordination and timeliness of data. Sometimes, the results of model runs are compared between models and it is found that the results are significantly different. After doing some checking, we find that one model's data is two weeks old and the other's is two months old. We need improved coordination of that type of data use.

A further use of the IRDS is the support of new model development. They have new models under development right now. One is called JAWS, the Joint Analytical Warfare System. We like that name! JAWS is being developed in a database environment already, but, as is the case with other model development or modification work, they tell their contractor "this is what we want the model to do," and the contractor comes back, gives them a model, and says "okay, this model will do that, but you need these pieces of data." Then they realize that they don't have that data, and they're not sure how to get it. Part of what we want to do with the IRDS is to establish what the data sources are that can be used in model development.

J8 sponsored an INGRES version prototype of an IRDS (Figure 2). The first half of this was completed in November, 1987 and was presented at the INGRES users conference that month in Tampa, Florida. There are significant problems with this initial prototype. It was just a demonstration, not to be put in as an operational system anywhere. It's buggy, and by no means completely functional in many particular areas. We are currently reviewing proposals to expand this prototype to make it more usable and to develop it more to our specific needs (Figure 3). We've received all the responses to the RFP, and an award is expected around mid April. Once the expanded prototype is finished, Argonne itself will port it to a SUN system.

<u>Question</u>: A few of us at NBS did see a demo of the original prototype. That prototype was of the Panel Interface. Are you specifically going to continue the development of the Panel Interface?

<u>Answer</u>: Our prototype was of the Panel Interface, not of the Command Language. We are going to continue work on the panels. Eventually, and this will come up later in this talk, we are looking at the possibility, not in the current proposals but further down the road, of taking the National Bureau of Standards prototype and converting it to our use. Our first aim is a Panel Interface--this interface is more graphically oriented, and we're trying to bring that type of approach to the users that we're dealing with.

Along the lines of standardization, Major Borman has already talked about the standardization of data elements within J7 for use in interoperability across the various services. The J8 directorate will be using the standardized data elements as determined via the WISDIM system to help with the standardization models within the Joint Chiefs.

My presentation is kind of a mixture between my own slides and some from the J8 Directorate that I just received Wednesday. I'm not fully familiar with Figures 4 through 11, so I may not be able to answer all your questions. These figures are primarily oriented towards what we feel we need to do and expect to do, long-term, with the IRDS.

We look at the IRDS to provide a central, coordinating data repository for data definitions, a unified data administrator domain, and enforcement of naming and usage conventions. We have been required by J8 to use INGRES as our relational database management system. They have a significant investment in INGRES at this time, including the development of the JAWS system.

Around the repository, our first aim is a Panel Interface--a menu or panel operational interface that will control both data and meta-data definitions. Theoretically, the coding for an INGRES version prototype is complete, but, as I've indicated, it's quite buggy and not at all operational. Hopefully, after the next section of work that Argonne is sponsoring, we will have a more usable beta version, perhaps in July or August. Since the work is being developed by the Government, it will become available as a piece of public domain software.

The Command Language interface is also of interest to us. We're very familiar with it, we have a copy of the NBS prototype that uses ORACLE which we have put up, for test purposes, on a MICROVAX at the Laboratory. We are considering porting it to INGRES, but that would not occur until next fiscal year at the earliest.

An area that we're interested in that isn't really dealt with by the standard is the use of entity-relationship diagram graphics tools. We want to be able to show, in our simulation operations, data structure operations and entity-relationship structures. How this is to be done is a subject of research for us. At this time we don't have a good idea how we're going to do this. Simplify, which does take advantage of E-R operations, is being looked on as a possible tool, and we would need to interface that with the IRDS. Obviously, this would be a valuable tool for documenting our various systems.

Another need for us is in the area of the overall documentation of the various simulation models that we deal with. We're looking at the concept of what we refer to as backloaders. These are systems to take existing documentation and information about systems and convert this into an IRDS structure. There would be modules for INGRES, DEC's RDB and CDD, FORTRAN and C. FORTRAN is a major concern for us, in that most of these modules were developed in that language to begin with. In connection with the IRDS Services Interface, we're very interested in looking at the FORTRAN impact here. We're also looking at pattern matching languages to help build backloader modules.

Question: Do you have any specifics yet on pattern matching languages?

Answer: The AWK language is one that is being considered. I'm not as familiar with the concerns for this as some other people are. When Dr. Goldfine contacted me about this workshop, I was really not directly involved any longer with the IRDS operation. It had been assigned to a staff member who decided, a few weeks ago, to leave us. So I ended up back in it, coordinating our IRDS effort after being out of it for several months, so I'm kind of in the middle here.

The IRD-IRD Interface is also going to be very important for us. We need to be able to interface data to and pull data from the WISDIM system, as it's been implemented using ORACLE and Model 204. We're looking forward to seeing the Abstract Syntax Notation for the interface.

We certainly see the need for the IRDS Services Interface in allowing us to have an active IRDS. We view that as critical for our operations and control. We also need, immediately, an active link with CASE tools and database design tools, and network control for the overall system.

The environment that we're aiming at is a workstation oriented environment. The hardware for this is part of our research task. The types of systems under consideration are systems like SUN workstations, which take advantage of high-level graphics capabilities, parallel processing systems like Sequent Balance or Symmetry machines, or the Alliant FX8 systems. We will be working in a distributed database environment using INGRES/STAR. So we'll end up with distributed IRDs. Part of it is to aim our usage of systems to the question of what systems run what models and what type of hardware configuration is best for that model. If a model can be particularly aimed at a parallel processing machine, and you can get much faster response out of it, you would want to run it there. If it won't work, if it's better on a serial processor like a SUN, we'll be able to move it to whatever system it needs to run on.

Figure 12 gives you just one idea of how the IRDS might work, in the area of database driven models. We see the IRDS as a repository for information for a central reference database or central storehouse of common data elements that are used across the various networks. The IRDS would contain information about the reference database and about each specific model. It would show what data in the reference database maps to what data in the model, in terms of changes of formats, names, everything. We would use the IRDS to help coordinate the reference database, our knowledge database which is in preparation, and model specific data to build a specific model scenario database to run a particular model operation. In the short-term, we're not allowed to interface the models themselves directly with the database, so we have to stick with flat file structures, although in the future the model will be modified to read the database directly and drop out the set of input file preparation routines, so we would go from flat files to a much more modern operating system.

Another use of the IRDS that we see focuses on new model development that will be oriented towards a uniform development environment. When J8 contracts with companies for the development of new models, they tell them what they want, but don't give them that many guidelines on how to develop it. What we want to do is to help in standardizing how things are developed, how models are developed for J8. Of course, the standardization efforts for J8's data elements will be provided as an IRDS to the contractors to define fully the type of environment where everything comes back in a consistent, documented, complete, and standardized manner, so that J8 can simplify its operations.

<u>Question</u>: What was the portion of the Standard that was incorporated into the prototype of the Panel Interface that you mentioned had been developed?

<u>Answer</u>: We implemented the basic core of the IRDS Standard based on the Panel Interface and the Basic Functional Schema. Beyond that, I'm really not sure, since I wasn't involved in writing the original SOW that gave the extent of the work--essentially, Module 1 and Module 2.

<u>Question</u>: So you're saying that you included all the schema commands?

<u>Answer</u>: Yes.

Argonne National Laboratory
J-8 Support Program - J-8 Modernization Effort
<u>Need for IRDS type of Capabilities</u>
Multiple Simulation Models/Common Data
Need for Standardization
Improve Coordination and Timeliness of Data
Support of New Model Development

Figure 1

Argonne National Laboratory

Project Associated Work

INGRES Prototype IRDS - Nov 1987

Further expansion of Prototype

RFP released

Responses Received

Award around Mid April

Expanded Prototype will be ported to SUN/Unix - INGRES System by ANL

Figure 2

Argonne National Laboratory

Outline of RFP Tasks

Submit Prototype for Review of NBS

Analysis of Compilance with proposed IRDS Standard

Analysis of Features for implementation under INGRES 6.0

Test Prototype for bugs coordinate with ANL for fixing

Implement a Panel Interface according to the Proposed ANSI Standard

Installation at J-8

Develop Documentation

Conceptualize interface: IRDS to designated database design tool

Figure 3

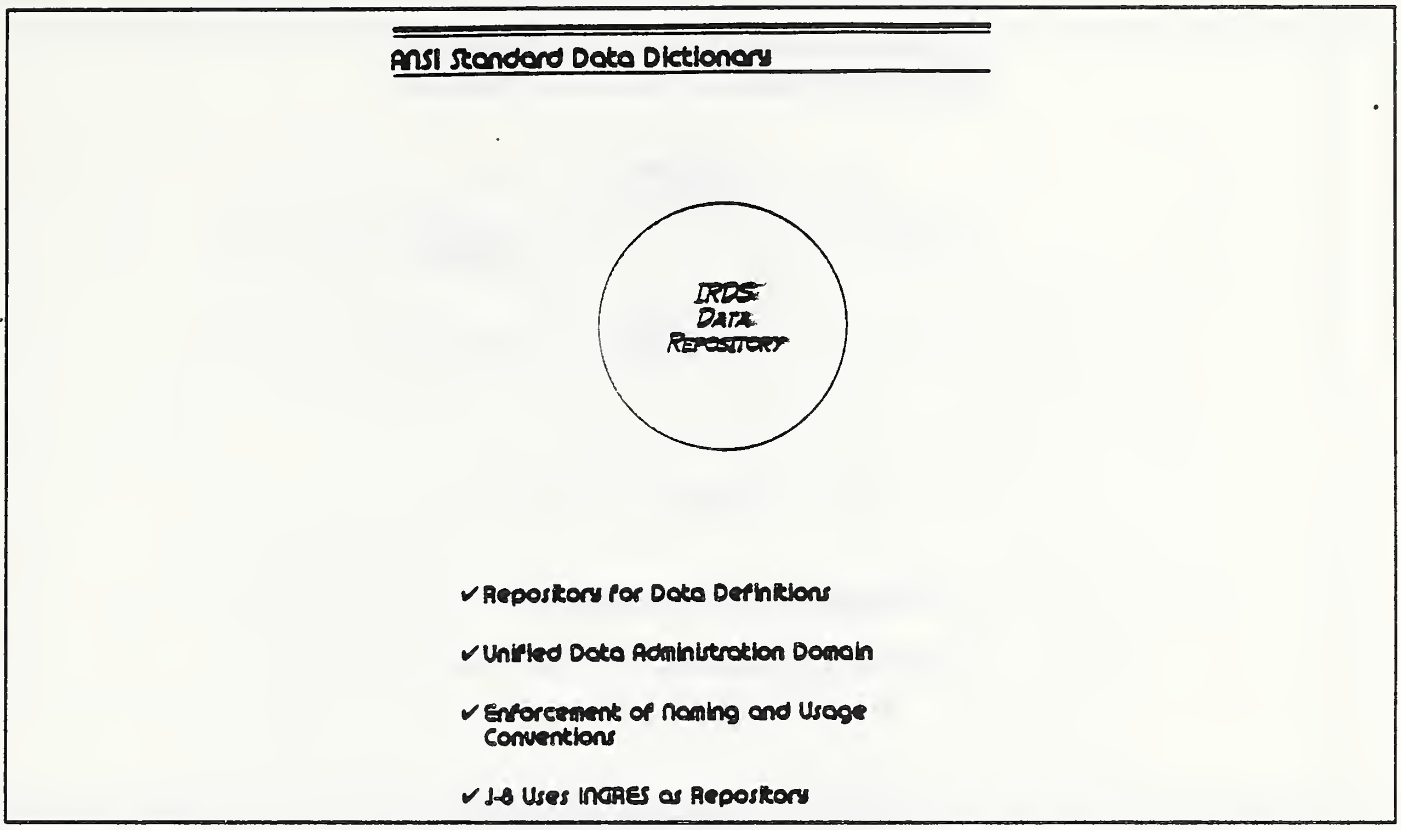


Figure 4

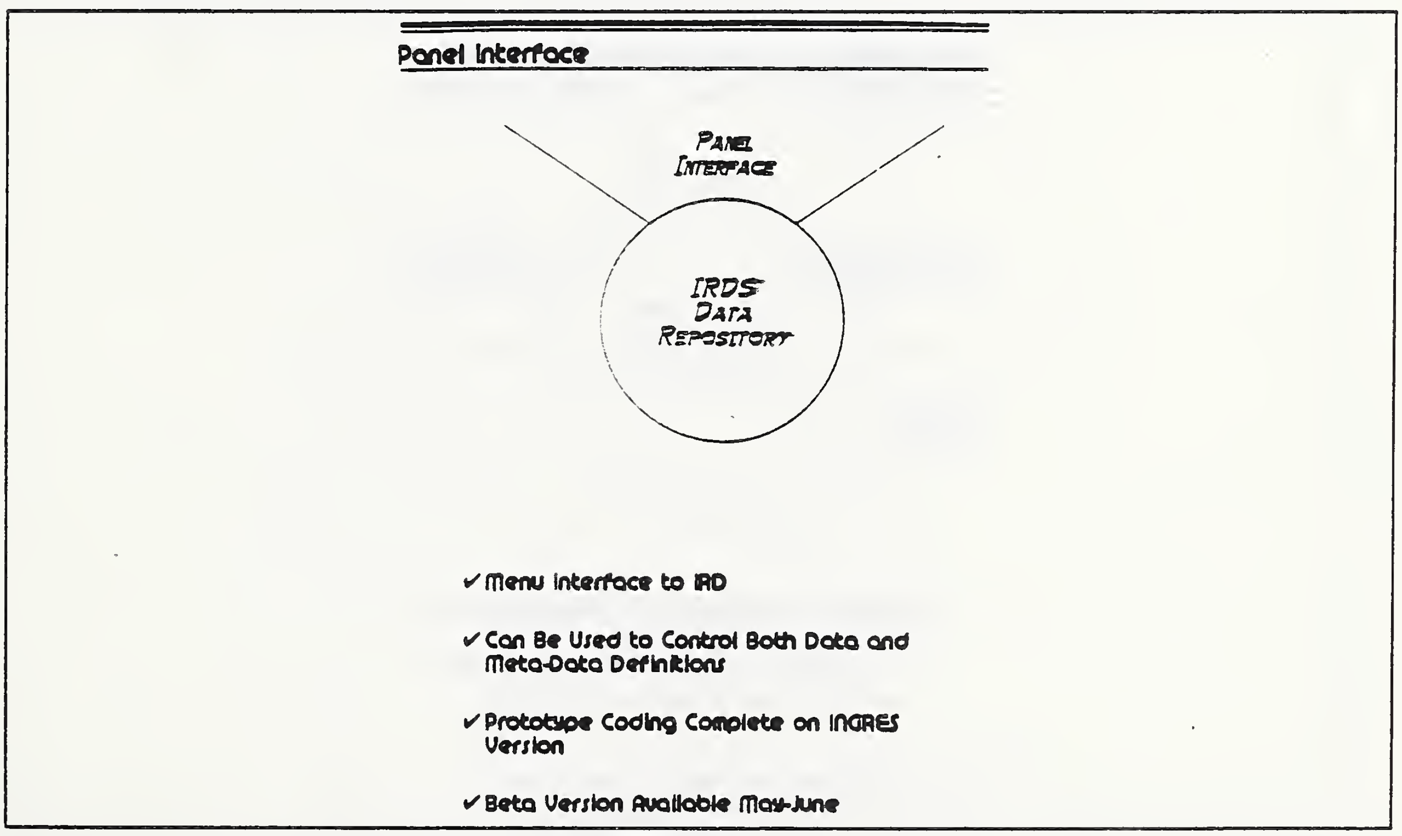


Figure 5

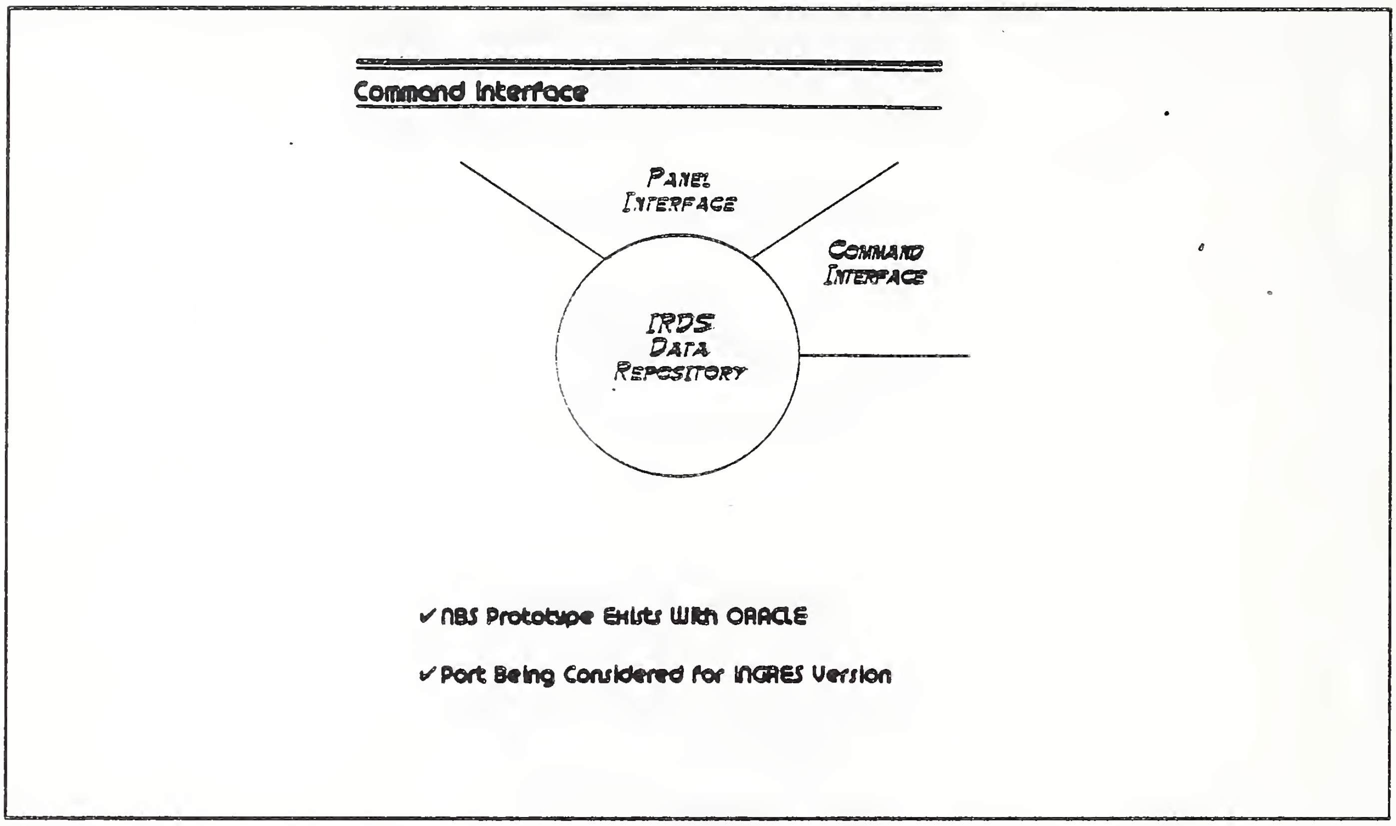


Figure 6

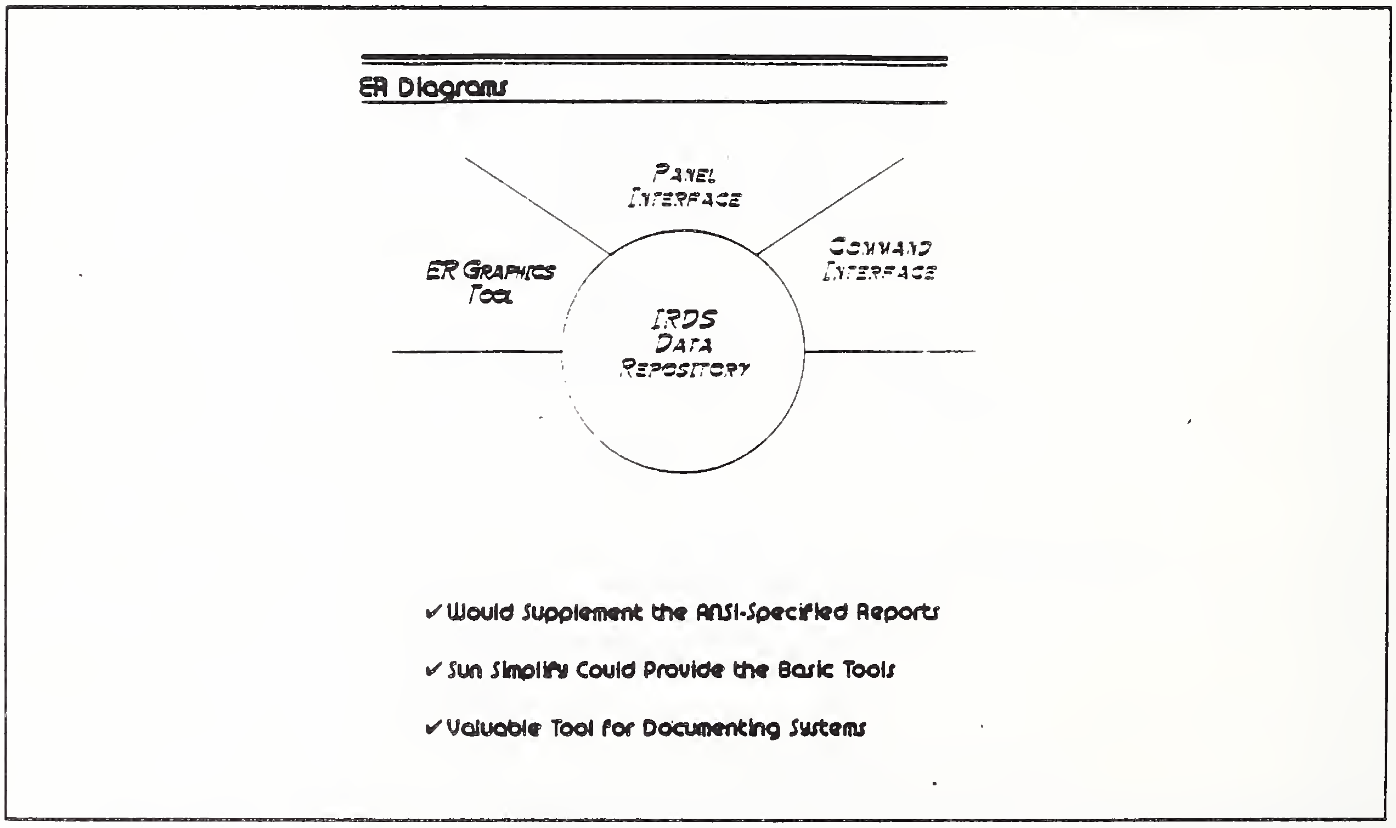
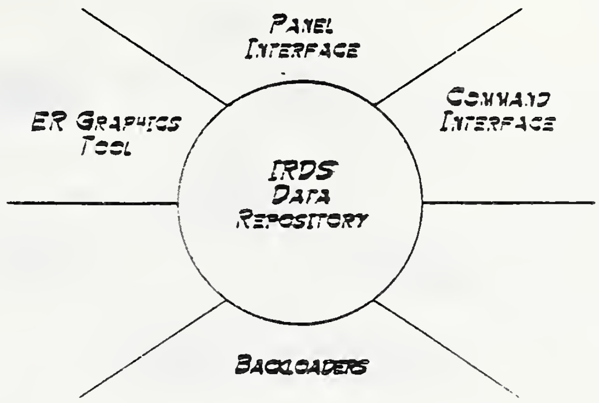


Figure 7

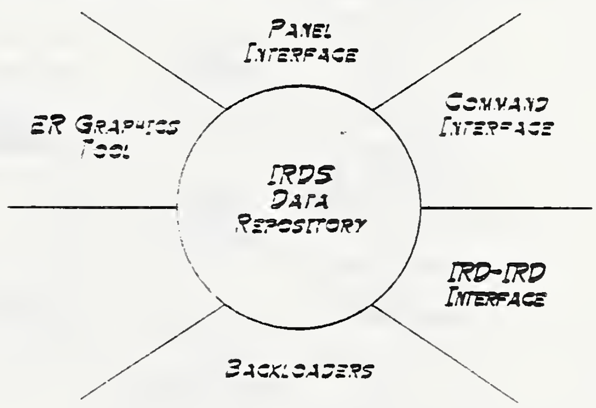
Backloaders



- ✔ Load Definitions of Existing Systems into IRD
- ✔ Greatly Simplify Documenting Existing Systems
- ✔ Modules for Ingres, DEC's Rdb and CDD, Fortran, C
- ✔ Pattern Matching Language Like AWK To Build Backloader Modules

Figure 8

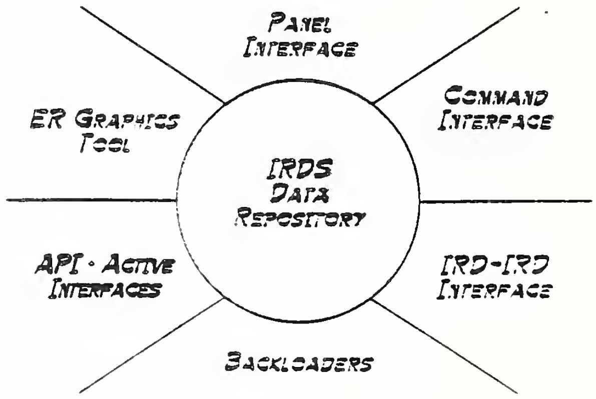
IRD-IRD Interface



- ✔ Allow Different Repositories to Exchange Data and Meta-Data
- ✔ Different Implementations of IRDS Can Coexist
- ✔ Uses Abstract Syntax Notation (ASN.1) of OSI as Language for IRD-IRD Interface

Figure 9

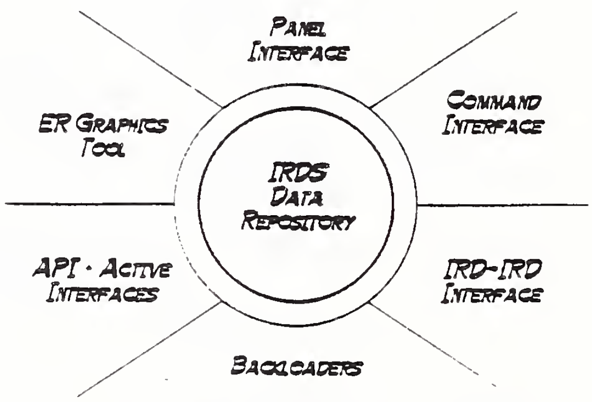
API & Active Interface



- ✓ A Applications Programming or Services Interface allows Tight Integration into Other Systems
- ✓ Active Link With CASE Tools
- ✓ Network Control Center with System Definitions Stored in an IRD

Figure 10

The Network



- ✓ Workstation-Based Environments Require High Degree of Distribution
- ✓ J-8 Using Distributed DBMS (INGRES/STAR) To Allow Distributed IRD's

Figure 11

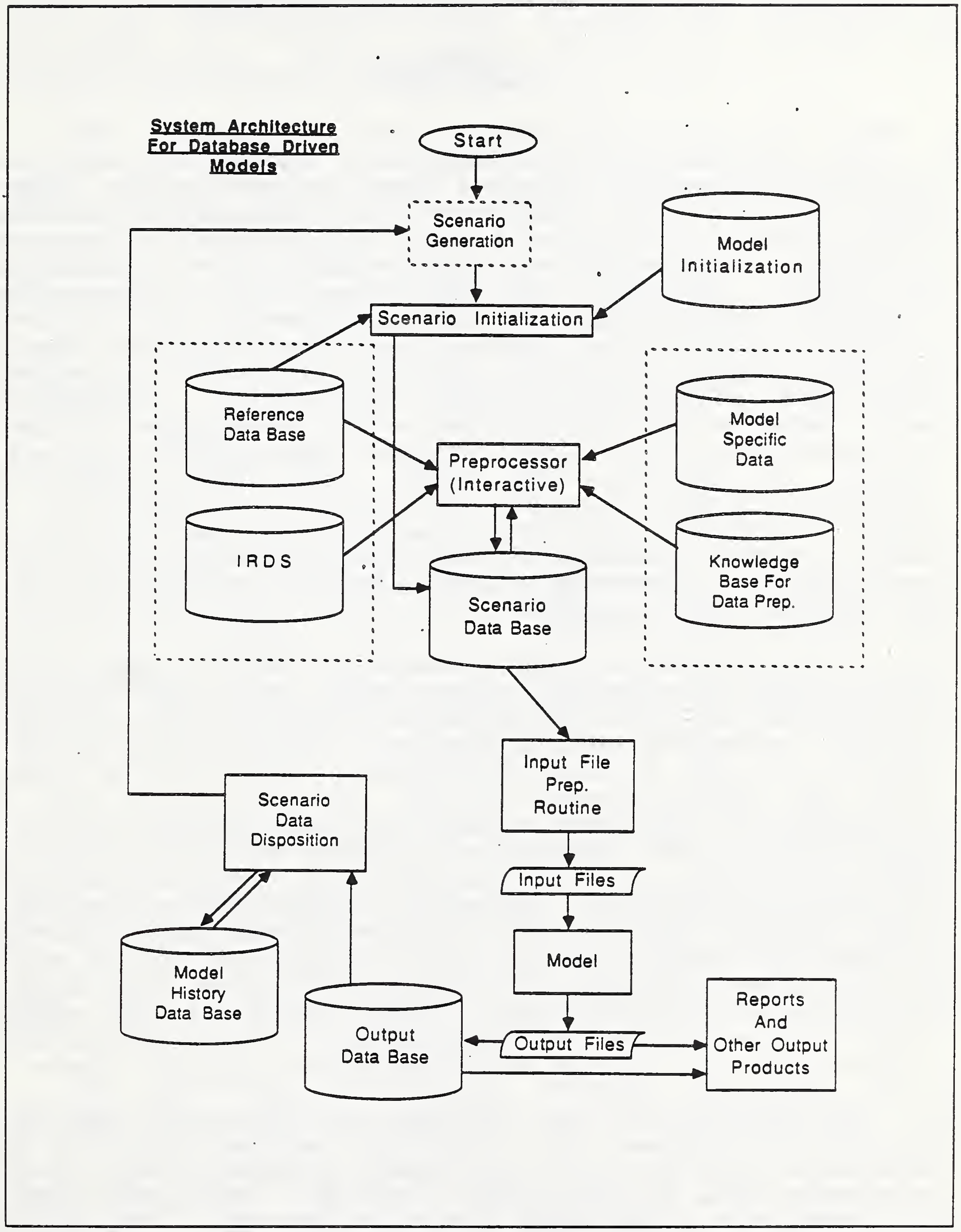

System Architecture For Database Driven Models
Start
Scenario Generation
Model Initialization
Scenario Initialization
Reference Data Base
IRDS
Preprocessor (Interactive)
Model Specific Data
Knowledge Base For Data Prep.
Scenario Data Base
Input File Prep. Routine
Scenario Data Disposition
Input Files
Model
Model History Data Base
Output Data Base
Output Files
Reports And Other Output Products

Figure 12

U.S. NAVY--DATA AUTOMATION COMMAND

<u>Speaker</u>
James Lynagh

Hi. NAVDAC is an acronym that stands for Naval Data Automation Command. We are a staff command for the Department of the Navy Directorate for Information Resources Management. Our mission is, essentially, to put together policies, standards, and plans that work towards the effective, efficient, and economical use of information resources of all kinds in the Department of the Navy. In putting together this presentation, I've taken the approach to talk to you about some of the things that we're doing that track pretty well with the IRDS effort. We feel that we're in a pretty good position to take advantage of some of the technology that's been offered here today. I have a couple of gentlemen here today, also from NAVDAC, Lt. Lubinsky and Randy Sullivan, who work in a program area for which I'm responsible. The area initially was the application software standardization sharing program, but we're taking on an increasingly large technical responsibility in the area of data management.

Within the Naval Data Automation Command, and specifically within the Software Directorate, we've got some major initiatives and concerns for which we're trying to formulate policy, procedures, and standards. The first one listed in Figure 2 is the Information Locator. It's been my experience that people across the Navy, because of the size and complexity of the organization, are not generally aware of the resources that they have available to them to help get their jobs done. This goes for all different types of organizational levels. We have a corporate need, we have a major command need, a major functional area need, and then departmental needs and end user needs. One of the things that we like about the IRDS concept is that it looks like it tracks well with that type of need. We're looking increasingly at the area of data management. I'll talk some about initiatives that we have underway in each of these areas.

Specifically, we see the need for doing more work, from a corporate perspective, on putting together effective data management policy, guidelines, and, in some measure, a data dictionary/data element locator system. Information system interface is a big area for us right now. We see an increasing demand on us to look at ways to interface our major information systems. Finally, we have responsibility in terms of helping to put together the policies and standards that increase the productivity of our major software development activities.

The Information Locator function that I alluded to earlier (figure 3) is one that tracks very well with the IRDS initiative here, and one for which we'll attempt to take advantage of that technology. The function is being pursued through the life cycle management process. We put together a mission element needs statement, which has yet to be approved, although I anticipate approval of that document. This document sets out the deficiencies that we have right now, in terms of our ability to identify, locate, and access accurate and up-to-date information about information resources. As a way to define and generate requirements, and to support any automated solution that we would develop to address those deficiencies, we have now a prototype Navy information directory.

I've found it interesting that I've heard today the terms "information resource dictionary," "data resource dictionary," and "data resource directory." One of the things that we're beginning to see now is perhaps some consensus of what it is we're talking about. I think that as we move further into this area, and become more proficient in what we're doing, we need to agree on some terminology, and get a good, solid definition down as to what it is we're talking about. That, of course, helps us in briefing more senior levels of management.

The Information Clearinghouse alludes to a concept that we've worked on for some time where we try to make available to organizations throughout the Navy, across functional areas and across major commands, such things as good, off-the-shelf application software packages, particularly in the area of micro computer systems. We've found a tremendous demand for good, core functional applications to get a job done, and we've been able to share applications across many activities, saving significant resources.

The prototype information resource directory that I alluded to a little earlier, and I emphasize the word "prototype" because it's in its very early stages, is being used in an iterative process to help us define the requirements for our long term implementation of a system to consolidate information about information resources. The more I hear about the IRDS, the more I see the role the IRDS will play in this effort. The prototype right now is based on a functional model, which is really the output of strategic planning within our command. It identifies the major categories of information on information resources. For example, my model right now has major categories for technology management, information management, there is a program management module, an AIS or major systems module, and an information resource management module. So each of the major categories of information that IR managers would be concerned with appears in that model. The prototype runs on a PC, and it provides pointers to other sources of useful information. We don't necessarily want to replicate information that's already been collected and that's already being maintained in some useful fashion, but we want to point people to it, initially in a manual mode, and ultimately in an automated fashion.

In the area of data management (Figure 5), the prototype that I talked about earlier appears to me to be the foundation for a data dictionary for IR managers. It'll be the data dictionary for the IR data administrator. We also have a project underway to identify existing data dictionaries in the Department of the Navy, as well as subject databases and the information systems that the dictionaries support. We have to keep in mind that the first thing we need to do for IR managers is to let them know where these resources are, so that location becomes important. Subsequently, we can factor in the management of the data resources, and then access to those databases.

There is, within the Navy, a subcommittee for data administration of the Information Systems Standards Committee. This is a group of people who have been brought together to get input from all the major commands of the Navy, and to identify the area where the greatest need is for some corporate level of resourcing, management attention, and focus. What they say is that we need some policy, we need some guidelines, and we need a way to locate these databases and gain access to the information.

In terms of the major software development activities that are part of the Navy, Figure 6 shows some of the things for which we're attempting to put policies and procedures in place. Obviously, we want to identify opportunities for sharing. This started out as an application software sharing initiative, but we're turning an increasing level of attention to the data itself. We're doing everything to increase productivity for these activities. We're beginning to gain some intelligence about corporate databases that are being maintained in the major functional areas such as supply, payroll, personnel, etc. We want to put into place an infrastructure that allows people access to this data. This is where the IRDS, as I'm beginning to understand it, is going to be very important.

I read, again, the IRDS Overview over the weekend, and Figure 7 shows the features and Modules that seemed to jump out at me as being most useful at this time. In particular, the IRD-IRD interface could be very beneficial in our environment for the reasons that I've given. We have an organization that is split in different ways by function and by major command, and the major commands have suborganizations. So I can see the need for providing some mechanism at each of those organizational levels to move data across them.

Figure 8 shows some of the other IRDS services that look very attractive, and the Modules that, certainly from my perspective, we'd be interested in investigating and factoring into our management structure.

<u>Question</u>: How far are you along on the prototype that you're working on?

<u>Answer</u>: It's in a preliminary stage. I've got the first copy that's been given to me as a backbone, as something to build on. The backbone allows you to load up on a micro and to look at the different categories of information, and to begin to do such things as data fill to let people, through an information process, tell you how valuable an information category is to them at their particular workstation. I anticipate that this thing will develop in a phased implementation, and at some point in the life cycle process, the requirements will have to be developed more definitively in the concept development stage. That's where I see ourselves looking at the IRDS in a great level of detail. This thing will continue to build over time. I'm going to try to phase it so that I will have useful products at specific intervals, because of budget constraints and that sort of thing. I want to be sure I can deliver something in a phased way that will have some value to our managers in the field, but at the same time aiming towards a longer term goal.

<u>Question</u>: What software are you implementing it in?

<u>Answer</u>: dBASE III. We did an analysis going into the prototype to determine what would be the most effective environment to do it in, and we felt that we ought to put it on something that people could be able to load up on their own workstations, and get them involved in a dialogue and an iterative process. At the same time, I alluded to a mission on the needs statement where we're developing the requirements that the prototype will help drive.

NAVY
IMPLEMENTATION

OF THE

IRDS

Figure 1

GOALS

- INFORMATION LOCATOR
- DATA MANAGEMENT
- INFORMATION SYSTEM INTERFACE
- S/W DEVELOPMENT ACTIVITY MANAGEMENT

Figure 2

INFORMATION LOCATOR

- MENS
- PROTOTYPE NAVY INFORMATION RESOURCE DIRECTORY
- INFORMATION CLEARINGHOUSE

Figure 3

PROTOTYPE NAVY IRD

(INFORMATION RESOURCE DIRECTORY)

- DIRECTORY OF INFORMATION RESOURCES ORGANIZED INTO MEANINGFUL CATEGORIES
- RUNS ON PC UNDER MS/DOS, WRITTEN IN DBASE III, COMPILED
- PROVIDES POINTERS TO USEFUL SOURCES OF INFORMATION. ("ELECTRONIC YELLOW PAGES")

Figure 4

DATA MANAGEMENT

- DATA DICTIONARY LOCATOR
- DONISS COMMITTEE PROJECT (DEPARTMENT OF THE NAVY INFORMATION SYSTEMS STANDARDS COMMITTEE)

Figure 5

S/W DEVELOPMENT ACTIVITY MANAGEMENT

- IDENTIFY OPPORTUNITIES FOR SHARING (DATA AND APPLICATIONS)
- IMPROVE PRODUCTIVITY THROUGH IMPLEMENTATION OF NEW TECHNOLOGY
- ESTABLISH CORPORATE DATABASES

Figure 6

RELATIONSHIP TO IRDS MODULES

MOST CURRENT REQUIREMENTS

- CORE IRDS
- DD STRUCTURES
- IRD—IRD INTERFACE
- APPLICATION PROGRAM INTERFACE

Figure 7

RELATIONSHIP TO IRDS MODULES

FUTURE REQUIREMENTS

- IRDS SECURITY
- LCM
- SERVICES INTERFACE
- DATA MANAGEMENT SUPPORT
- LIFE CYCLE AND CONFIGURATION MANAGEMENT SUPPORT

Figure 8

U.S. NAVY--SPACE AND NAVAL WARFARE SYSTEMS COMMAND

Speaker
Bob Moyer
SPAWAR Technical Data Center

Good afternoon. My name is Bob Moyer. I am here representing the Space and Naval Warfare Systems Command (SPAWAR) Technical Data Center (TDC) at Portsmouth, VA (Figure 1). Accompanying me is Dr. Rick Klobuchar of Advanced Technologies, Inc.

We are particularly pleased to be here to learn more about the IRDS and to present a Navy technical repository's perspective on data management and on how the IRDS may enhance our productivity. I would like to emphasize that we are here more to learn about the IRDS than to point a direction for the IRDS. We see the IRDS as potentially a very valuable "arrow" to have in our "quiver," and we want to learn how it can be applied. In the course of our discussions, we would also like to present our perspective on technical data management down where the "rubber really meets the road"--at the fleet user level.

In this presentation (Figure 2), I will briefly talk about some of the data-intensive TDC mission areas, responsibilities, and activities. This will set the stage for a discussion on why the TDC is interested in the IRDS as a potential productivity enhancer. I will then present four ongoing SPAWAR TDC initiatives. These initiatives include our development of a master library index specification, the conceptual design of a knowledge-based data delivery architecture, technical manual automation, and technical manual print-on-demand. In this discussion, I will consider elements of our design where the IRDS can potentially be a major player. Lastly, the SPAWAR TDC is open to serving as a cooperative testbed for real world implementation of the IRDS.

To understand why the SPAWAR TDC is interested in learning more about the IRDS, it is important to understand our basic mission areas, responsibilities, and activities.

Fundamentally, the SPAWAR TDC (Figure 3) is a major Navy technical data repository supporting the Space and Naval Warfare Systems Command, and is a major clearinghouse of technical data on Navy electronic equipment covering C3I systems, telecommunications, tactical data systems, and electronic warfare systems. We work closely with the other Navy systems commands and other DoD activities to help ensure that the right technical information and technical data is placed into the hands of the fleet user community.

The TDC initiatives are:

- The TDC provides SPAWAR managers with a centralized automated repository/technical data center capable of responding to requirements for technical documentation in support of acquisitions, procurements, and life cycle maintenance of SPAWAR equipment and systems.

- We are designated as one of eight primary SYSCOM repositories to be automated under the EDMICS initiative.

- The TDC develops and implements a management information control system (MICS) to support SPAWAR TDC data management requirements.

- We provide remote technical data assistance to the ISEAs for fleet support.

- The TDC develops and implements a quality control program for the acquisition of technical documentation.

- We are responsible for the SPAWAR automated technical manual initiative.

- We must stay abreast with the state-of-the-art technology for improvements in the automation process of technical documentation.

At the current time, we maintain large quantities of technical manuals and engineering drawings. Sources of the technical manuals and engineering drawings include contractors and other Navy or DoD activities. Our job is to respond quickly to requests for technical information and data. The requests can be as diverse as providing change documentation to a given ship, to scanning 100 or so pages of an existing hardcopy technical manual to provide electronic media in support of the Navy's "paperless" ship initiative.

Additionally, the SPAWAR TDC supports the Navy's BOSS program. BOSS stands for "Buy Our Spares Smartly." In BOSS, the SPAWAR TDC provides technical and engineering data to support an engineering breakout process, where frequently the part to be procured will have to be reverse engineered. In our efforts, we are also concerned with:

- o Implementation and prototyping of Computer Aided Logistic Support (CALS) specifications and standards.
- o Implementation of Standard Generalized Markup Language (SGML), Initial Graphics Exchange Standard (IGES), and raster scanning of large and small format documents.
- o Development of management information control systems to handle the technical data.
- o Implementation of local area networks.
- o Utilization and integration of page description languages.
- o Enhancement of engineering drawings.

The list of activities we support is large, and our responsibilities at the TDC are growing. If this sounds like a data intensive situation, it is. Unfortunately, both our responsibilities and the user's demand for technical information are growing at a time when budgets are being trimmed to the bone. This situation forces us to consider all possible productivity enhancing measures. This is a major reason why we are here to learn about the IRDS.

Figure 4 briefly summarizes why the SPAWAR TDC is interested in the IRDS. We have read some of the IRDS documentation and the following seems attractive to us from a real-world technical data repository point of view:

<u>(1) First, the IRDS allows data transportability over dissimilar architectures.</u>

The TDC maintains a number of databases which include:

- o SPAWAR publications master file.
- o Deficiencies database.
- o Inventory of SPAWAR publications.
- o Production reports.

- o Distribution lists.
- o Engineering drawing and technical manual contract data requirements list (CDRL) tracking systems.
- o Configuration data.

Along with these, the TDC also maintains a large engineering drawing database, and a technical manual database that includes SPAWAR publication numbers, camera-ready copy, electrically generated data, and user feedback.

As you can see, with this diversity in databases we require, within our own operation, data transportability over dissimilar architectures.

<u>(2) Portability of skills among organizations through a common panel interface.</u>

Another important factor for not only the SPAWAR TDC, but also the Navy, is that the IRDS will facilitate portability of skills between organizations through a common panel interface. Our experience is that the requirement to standardize is important. Standardization promotes productivity enhancement. The IRDS Panel Interface will permit non-technical personnel access to the database without them having to understand or use a more complex syntax of the Command Language interface.

<u>(3) A powerful tool for life-cycle management of technical data.</u>

We understand that there is an initial cost with implementing and using the IRDS. However, we can see the possibility of reduced costs over the long run.

<u>(4) Extensible to customized user environment.</u>

The IRDS can potentially offer the SPAWAR TDC an ability to customize its data delivery environment. Particularly attractive is the notion of IRDS functional modules. We see an opportunity for the data architecture to be flexible and extensible. The Core IRDS appears to have these capabilities which will enable us to customize and extend the type of data that can be stored.

<u>(5) Possibility of extending data directory to knowledge-based delivery architectures.</u>

Lastly, we appreciate that the multiple levels of IRDS functionality are extensible to knowledge-based delivery architectures. Particularly significant, from our point of view, is that level 4 is notionalized for knowledge-bases and expert systems. We don't know all of what this means yet, and that is why we are here--to learn. However, we do see and appreciate the connection of being smarter with:

- o Our generation and development of Navy technical information.
- o Our handling of the technical information.
- o Our handling of the technical information in a timely fashion to our user community.

As Figure 5 indicates, our architecture is considering four major efforts or initiatives. These initiatives are closely interrelated with the goal of delivering more timely and more accurate technical information to the fleet. These initiatives include:

- o Master Index Library Specification (draft) (SPAWAR/NAVSEA/NAVAIR/NPPSO)
- o Knowledge-based delivery architecture (draft)
- o Technical manual automation (SPAWAR/NAVSEA/NAVAIR)
- o Technical manual print-on-demand (SPAWAR/NPPSO)

I will briefly discuss each of these. First, the Master Index Library Specification.

Currently, the SPAWAR TDC is working on an existing paradigm to capture the complex interrelationship in existing technical manuals and engineering drawings. The intention of this effort, consistent with CALS direction and guidance, is to produce indexed technical information for storage, interchange, and ultimate usage by the Navy fleet user community.

Our indexing effort currently exists in draft form, soon to be promulgated for review. The indexing specifications draw heavily on MIL-STD-1840A, but extends it significantly in the area of identifying indexing elements which meet downstream user requirements for inter-activity. The thrust of the effort is currently focused on indexing of a large backlog of hardcopy technical manuals which will be scanned.

We foresee the indexing specification as being used by industry sources to produce indexed media, which would be input into our master library via an input processing routine. We are currently in the process of rapid prototyping the index specification and input processing routine as a "proof of concept" demonstration. The specification will be used by all the Navy systems commands and the Navy Publication Production Support Office.

Currently, the SPAWAR TDC is also developing a knowledge-based delivery architecture master plan. This plan, which currently exists in draft form, seeks to develop a series of "intelligent" applications programs to deliver tailored technical information to the user. Some of the functions of the knowledge-based delivery architecture would include:

- o Streamlined and intelligent directed search of distributed Navy databases.
- o Media conversion and bundling processes.
- o Development of tailored technical information packages based upon a knowledge of user needs and requirements.
- o Interaction through a natural language query in the user's vernacular.

In the development of the knowledge-based delivery architecture, we believe that data dictionary systems and the higher meta-level functionality of the IRDS can be significant players. We foresee the possibility that higher level IRDS functionality could control processes both for input of indexed technical data and for its knowledge-based delivery. We are here to explore the prospects of doing this with NBS and other interested parties.

With regard to technical manual automation, the SPAWAR TDC also has a major initiative, the Technical Manual Upgrade Program (TMUP). The goal of this program, as its name suggests, is to invoke modern technology to streamline, improve, automate, and enhance large numbers of Navy technical manuals. In this effort, we are actively working with all the Navy systems commands and with other DoD elements to bring modern electronic publishing systems to bear. Our efforts in this area include:

- o Working closely with Navy and other DoD activities in the implementation of the Standard Generalized Markup Language with a tagging set containing an output specification, MIL-M-28001.

- o Evaluation of scanners, both for raster and intelligent scanning of hardcopy technical manuals.

- o Evaluation of electronic publishing systems for cost-effective, high-volume production of technical manuals.

- o Evaluation of data dictionaries and database management systems capable of enhancing technical manual storage and interactive retrieval across distributed databases.

- o Participation in the Navy's "paperless ship" initiative as a rapid prototyping and demonstration activity.

- o Evaluation of electro-optical media such as cd-rom and worm for delivery of technical manuals.

- o Evaluation of page description languages and laser printers for delivery.

The last initiative that I will discuss is the SPAWAR TDC's participation in the print-on-demand program. Until portable electronic delivery devices are perfected, paper technical manuals will continue to be the fleet users' preferred method of receiving technical information. With the print-on-demand project, we are exploring technologies for:

- o Delivery of technical manual information on consumable electro-optical media.

- o Rapid, on-the-fly page composition of graphics and text materials.

- o Print-on-demand of specified technical manual pages.

Print-on-demand is intended to provide only the technical information necessary to meet the user's task in frequently adverse environments like the hold of a ship.

I would like to address our willingness to serve as a CALS/IRDS testbed (Figure 6). This has benefits both to the IRDS project and the Navy. We believe that we offer an opportunity to test concepts in a real-world technical data repository environment. We are a major Navy technical data repository with mission needs that the IRDS can help support. We also have rapid prototyping and evaluation talent available to support the effort. We are also the Navy's coordinator of "paperless ship" demonstration projects. Lastly, we have personnel readily available with an ideal blend of understanding of expert systems, database management systems, data dictionary systems, knowledge engineering processes, and the operational user environment.

Question: This is directed towards the Navy in general. I'm familiar with the Navy's standard system development life-cycle, because we looked at it at EPA, and continue to look at it and benefit from it. I've also been aware of the activity of incorporating data dictionary kinds of development and rules into that life-cycle process, and wondered where that fits organizationally?

Answer: It obviously fits within the systems commands. There used to be an organization called NAVMAT, the Naval Materials Command. It no longer exists, and all its functions have gotten down to the individual systems commands, who have the requirement to coordinate amongst each other. It's a very hard question to answer--where, organizationally, that responsibility lies. There is a Secretary of the Navy instruction on life-cycle management, the last iteration of which was put out several years ago. I believe that is under revision now.

Question: Having looked at it and profited from it, I know that it's very useful. We're trying to figure out what life-cycle means for expert systems, and you mentioned that. Is that reflective of what the Navy's instruction would be?

Answer: Well, part of it is a delivery architecture, because we can customize the data output to the users in the format that they are familiar with, and gradually bring the new technology to them.

SPAWAR
TECHNICAL DATA
CENTER
BRIEF

NAVAL ELECTRONIC SYSTEMS ENGINEERING CENTER
PORTSMOUTH, VIRGINIA

Figure 1

SPAWAR TECHNICAL DATA CENTER OVERVIEW

- SPAWAR TDC INITIATIVES
- WHY SPAWAR TDC IS INTERESTED IN IRDS
- SPAWAR TDC ARCHITECTURE
- PROPOSED COOPERATIVE INITIATIVE (TESTBED)

Figure 2

SPAWAR TECHNICAL DATA CENTER

INITIATIVES:

- PROVIDE SPAWAR MANAGERS WITH A CENTRALIZED AUTOMATED REPOSITORY /TECHNICAL DATA CENTER CAPABLE OF RESPONDING TO REQUIREMENTS FOR TECHNICAL DOCUMENTATION IN SUPPORT OF ACQUISITIONS, REPROCUREMENTS AND LIFE CYCLE MAINTENANCE OF SPAWAR EQUIPMENTS/SYSTEMS.
- DESIGNATED AS ONE OF EIGHT PRIMARY SYSCOM REPOSITORIES TO BE AUTOMATED UNDER EDMICS INITIATIVE.
- DEVELOP/IMPLEMENT A MANAGEMENT INFORMATION CONTROL SYSTEM (MICS) TO SUPPORT SPAWAR TDC DATA MANAGEMENT REQUIREMENTS.
- PROVIDE REMOTE TECHNICAL DATA ASSISTANCE TO THE ISEA'S FOR FLEET SUPPORT.
- DEVELOP/IMPLEMENT A QUALITY CONTROL PROGRAM FOR THE ACQUISITION OF TECHNICAL DOCUMENTATION.
- SPAWAR AUTOMATED TECHNICAL MANUAL INITIATIVE.
- STAY ABREAST WITH STATE-OF-THE-ART TECHNOLOGY FOR IMPROVEMENTS IN THE AUTOMATION PROCESS OF TECHNICAL DOCUMENTATIONS.

Figure 3

WHY SPAWAR TDC IS INTERESTED IN IRDS

- ALLOWS DATA TRANSPORTABILITY OVER DISSIMILAR ARCHITECTURES
- PORTABILITY OF SKILLS AMONG ORGANIZATIONS THROUGH COMMON PANEL INTERFACE
- POWERFUL TOOL FOR LIFE-CYCLE MANAGEMENT OF TECHNICAL DATA
- EXTENSIBLE TO CUSTOMIZED USER ENVIRONMENT
- POSSIBILITY OF EXTENDING DATA DIRECTORY TO KNOWLEDGE-BASED DELIVERY ARCHITECTURES

Figure 4

SPAWAR TDC ARCHITECTURE

- MASTER INDEX LIBRARY SPEC (DRAFT) (SPAWAR/NAVSEA/NAVAIR/NPPSO)
- KNOWLEDGE BASED DELIVERY ARCHITECTURE (DRAFT)
- TECHNICAL MANUAL AUTOMATION (SPAWAR/NAVSEA/NAVAIR)
- TECHNICAL MANUAL PRINT ON DEMAND (SPAWAR/NPPSO)

Figure 5

WHY A TESTBED AT TDC?

- MAJOR NAVY EDMICS SITE
- RAPID PROTOTYPING AND EVALUATION TALENT AVAILABLE
- COORDINATOR OF NAVY "PAPERLESS" SHIP DEMONSTRATIONS
- PERSONNEL AVAILABLE WHO UNDERSTAND:
  - EXPERT SYSTEMS
  - DATA BASE MANAGEMENT AND DATA DICTIONARY ENVIRONMENTS
  - KNOWLEDGE ENGINEERING PROCESSES
  - OPERATIONAL USER ENVIRONMENT

Figure 6

WORKSHOP ATTENDEES

Ms. Sandra Anderson
MITRE Corporation
(NASA--JSC)

Major T. Reed Borman
Organization of the Joint Chiefs
of Staff, J7, PSD

Mr. Howard Brewer
Defense Communications Agency
Joint Data Systems Support Center

Col. Wayne W. Byrd
U.S. Army
ODISC-4 (SAIS-ADR)

Mr. Roger Cooley
U.S. Department of the Interior
Office of IRM

Mr. Fred Gey
Lawrence Berkeley Laboratory

Dr. Alan Goldfine
National Bureau of Standards

Mr. Bruce Haberkamp
U.S. Army
OACSIM (SAIS-ADR)

Mr. Robert C. Kling
Department of Agriculture
Soil Conservation Service, IRM

Dr. Richard L. Klobuchar
Advanced Technology, Inc.
(U.S. Navy--SPAWAR)

Dr. Margaret H. Law
National Bureau of Standards

Mr. Joseph Leahy
UNISYS
(Defense Communications Agency)

Lt. David Lubinsky
U.S. Navy
NAVDAC

Mr. James Lynagh
U.S. Navy
NAVDAC

Mr. Patrick McCabe
U.S. Air Force
Rome Air Development Center

Ms. Donna McCloud
Defense Logistics Agency

Ms. Sarah McMahan
U.S. Department of Commerce

Ms. Mary Lou Melley
Environmental Protection Agency

Mr. Hedrick E. Mitchell
Defense Communications Agency
Joint Data Systems Support Center

Mr. Robert S. Moyer
U.S. Navy
SPAWAR Technical Data Center

Ms. Judith J. Newton
National Bureau of Standards

Ms. Marlene Palmer
Farm Credit Administration

Ms. Sahan Palmer
Veterans Administration

Mr. Curtis H. Parks
General Dynamics

Ms. Charlotte Plum
U.S. Department of Commerce

Mr. Alexis Poliakoff
U.S. Department of Education

Mr. Martin Riekse
Veterans Administration

Mr. Steve Ritzman
Booz-Allen Hamilton
(NASA--SSP/TMIS)

Mr. Greg Robinson
Argonne National Laboratories

Mr. Thomas Roche
Veterans Administration

Mr. Bruce Rosen
National Bureau of Standards

Mr. Tom Sheffler
U.S. Department of Commerce

Mr. Frank Spielman
National Bureau of Standards

Mr. Jim Squier
U.S. Department of Commerce

Mr. George Sullivan
U.S. Navy
NAVDAC

Mr. Jerry Szablowsky
Small Business Administration

Ms. Joan Tyler
National Bureau of Standards

NBS-114A (REV. 2-8C)

U.S. DEPT. OF COMM. **BIBLIOGRAPHIC DATA SHEET** *(See instructions)*	1. PUBLICATION OR REPORT NO. NISTIR 88-3896	2. Performing Organ. Report No.	3. Publication Date DECEMBER 1988

4. TITLE AND SUBTITLE

Proceedings of the Federal Information Processing Standard (FIPS) Workshop on Information Resource Dictionary System (IRDS) Applications

5. AUTHOR(S)

Alan Goldfine

6. PERFORMING ORGANIZATION *(If joint or other than NBS, see instructions)*

NATIONAL BUREAU OF STANDARDS
U.S. DEPARTMENT OF COMMERCE
GAITHERSBURG, MD 20899

7. Contract/Grant No.

8. Type of Report & Period Covered

9. SPONSORING ORGANIZATION NAME AND COMPLETE ADDRESS *(Street, City, State, ZIP)*

10. SUPPLEMENTARY NOTES

☐ Document describes a computer program; SF-185, FIPS Software Summary, is attached.

11. ABSTRACT *(A 200-word or less factual summary of most significant information. If document includes a significant bibliography or literature survey, mention it here)*

This report consists of the user presentations at a workshop on applications of the Information Resource Dictionary System (IRDS), held at the National Bureau of Standards on March 24-25, 1988. Representatives of twenty Federal Government agencies discussed current and planned applications of the IRDS at their respective organizations.

12. KEY WORDS *(Six to twelve entries; alphabetical order; capitalize only proper names; and separate key words by semicolons)*

data administration; data dictionary; data management; Information Resource Dictionary System; information resource management; IRDS

13. AVAILABILITY

☒ Unlimited
☐ For Official Distribution. Do Not Release to NTIS
☐ Order From Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402.
☒ Order From National Technical Information Service (NTIS), Springfield, VA. 22161

14. NO. OF PRINTED PAGES

196

15. Price

$18.95

USCOMM-DC 6043-P80