Title
SPECIFICATIONS FOR VERSION 1.0 OF THE ARMY DATA ENCYCLOPEDIA

Authors
Gey, F.
Holmes, H.

Publication Date
1988-09-01

LBL-25746
UC-405

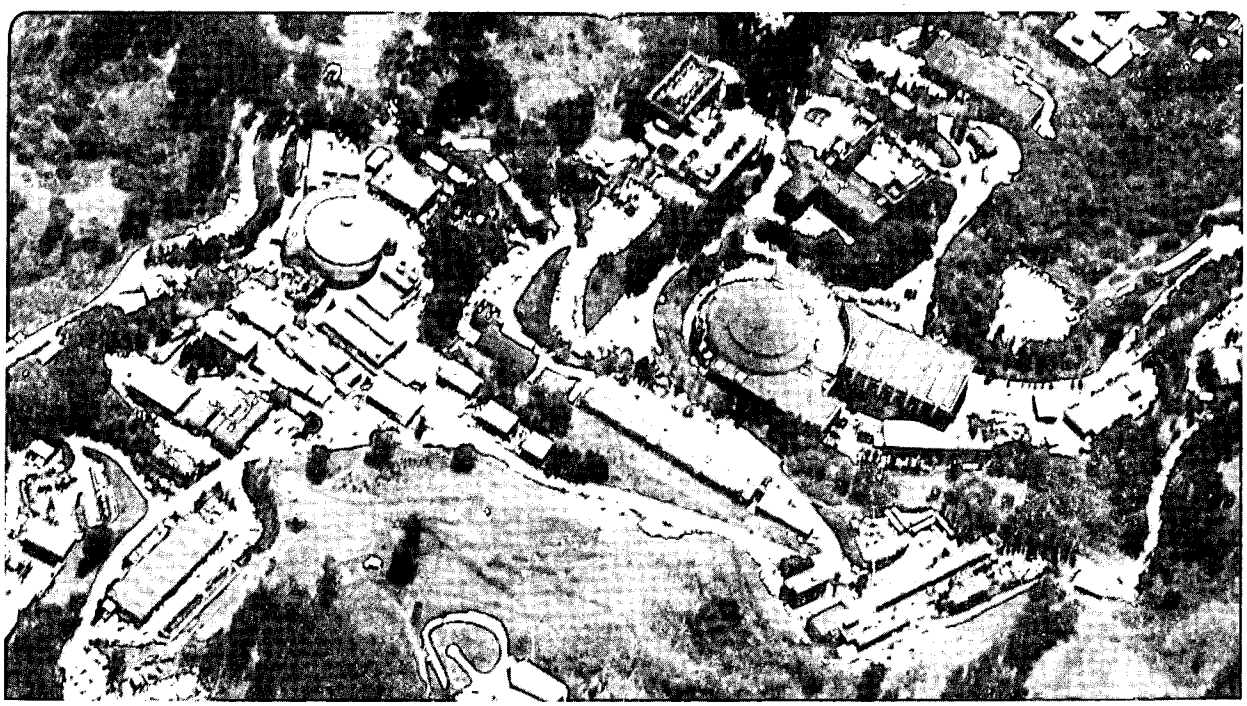
Lawrence Berkeley Laboratory
UNIVERSITY OF CALIFORNIA, BERKELEY

Information and Computing Sciences Division

Specifications for Version 1.0 of the Army Data Encyclopedia

F. Gey and H. Holmes

September 1988



Prepared for the U.S. Department of Energy under Contract Number DE-AC03-76SF00098.

DISCLAIMER

LBL-25746

SPECIFICATIONS FOR VERSION 1.0 of the ARMY DATA ENCYCLOPEDIA

Fredric Gey & Harvard Holmes

Computing Sciences Research & Development
Information & Computing Sciences Division
Lawrence Berkeley Laboratory
1 Cyclotron Road
Berkeley, California 94720

Prepared for the

U.S. Army Information Systems Engineering Command
Data Management Directorate

September 1988

This work was performed under the auspices of the Department of Energy under Contract No. DE-AC03-76SF00098 with interagency funding from the U.S. Army Information Systems Engineering Command. Views and conclusions contained in this paper are those of the authors and should not be interpreted as representing the official opinion or policy of the USAISEC or the Department of Energy.

1. Introduction and Organization

This document provides a more detailed description of the Army Data Encyclopedia (ADE) Version 1.0, in accordance with the Data Encyclopedia Architecture for Army Information Management [CSR 88]. The software and contents of the ADE are key mechanisms that are necessary to achieve interoperability, integration and synchronization of Army information systems. The ADE architecture is intended to provide a global view, long-term direction, and the conceptual foundation for further development. In accord with the architecture, the ADE will develop through a sequence of increasingly powerful versions, supporting a series of Army-wide Information Mission Area (IMA) efforts, such as data element standardization.

The major components of the ADE Version 1.0 are an ANSI-FIPS Standard Information Resource Dictionary System (IRDS) framework [ANSI 88], which is implemented using a relational DBMS, a user interface for schema and data maintenance, a user interface for the Data Element search/retrieval/approval process, and the loading of the actual data of the ADE.

Consistent with recommendations in the Data Encyclopedia Architecture, this paper assumes that the Army will develop the ADE Version 1.0 internally (with LBL assistance), and it is structured accordingly; however, software, schema and data organizations given here can be used to guide the procurement of commercial systems if that is desirable. Much of the work, e.g., documentation, data loading, and distributed access mechanisms, would not be covered in commercial systems in any event.

This paper assumes the reader is familiar with the Data Encyclopedia Architecture document. It provides background for ADE Version 1.0 and a brief status report of ADE related activities in progress (Section 2). It outlines a software structure for the ADE (Section 3), functionality to be implemented within the ADE 1.0 (section 4), structure of the data element approval process and user interface (Section 5), initial data content of the ADE (Section 6), documentation needs of the ADE (Section 7), and remote user access strategies for the ADE (Section 8).

2. Background for ADE Version 1.0

2.1 Goals, Objectives and Benefits

The most immediate goal of the ADE (and its predecessors) is to store the metadata which is now being collected in a variety of formats, locations, and degrees of consistency, accuracy and standardization, and to make the ADE available to personnel involved in the Army Data Management program and IMA life cycle management. It is critical that this metadata be stored in a flexible, extensible structure as embodied in an IRDS. The next most immediate goal for the ADE is to improve the user interface to the ADE, and to stabilize the ADE development process so that its evolution can be controlled and managed.

Development of the ADE is part of a broad effort to restructure the Army's information systems so that they serve a higher organizational purpose. The intent is to structure information technology to gain competitive advantage both through efficiency-lower cost-as well as through effectiveness-better outcomes. While savings in the Army are expected in cost displacement and cost avoidance within the Army's information mission area, the greater part of the savings will be from increased efficiency in the use of personnel and materiel throughout the total Army. An important goal is the effective use of information that serves strategic, tactical, and sustaining purposes. This enables the Army to sharpen its response by superior organization of physical resources to gain maximum effect with minimum loss.

2.2 Functionality

The Army Data Encyclopedia (ADE) will contain many different types of information. For the most part, it will not contain "data" *per se*, but rather data *about* data, i.e., meta-data. For example, the encyclopedia will contain information *about* the data element "social security number" (e.g., the fact that it contains nine single digit numbers and which systems use a social security data element), but it will not contain any actual social security numbers. Data itself will reside in data management systems, files, etc. throughout the Army, while metadata in the encyclopedia will be used to access and manage such data. Some special kinds of information about data-related items such as computer systems, programs, subject areas, and certain data models may also be stored directly in the ADE because they are closely linked to metadata.

The IRDS standard provides a good foundation for organizing the Army's metadata. This new standard is based on the Entity- Relationship-Attribute (ERA) model, which is already used extensively throughout the Army for systems planning and design. An IRDS entity represents or describes a "real world" concept, person, event, or quantity (e.g., the data element "social-security-number"). A relationship is an association between two IRDS entities (e.g., Personnel Record "CONTAINS" Social Security Number). Attributes represent properties of an entity or relationship. For example, one attribute of the entity "Social Security Number" is its LENGTH (e.g., 9 characters).

The IRDS standard will help organize the ADE in three ways. First, the IRDS model itself specifies standards not only for how an IRDS data dictionary will describe application-specific entities, attributes, and relationships, but also for how the dictionary objects themselves (e.g., data element, system, etc.) will be specified. This multi-level specification is illustrated by the schematic diagram and examples in Figure 1.

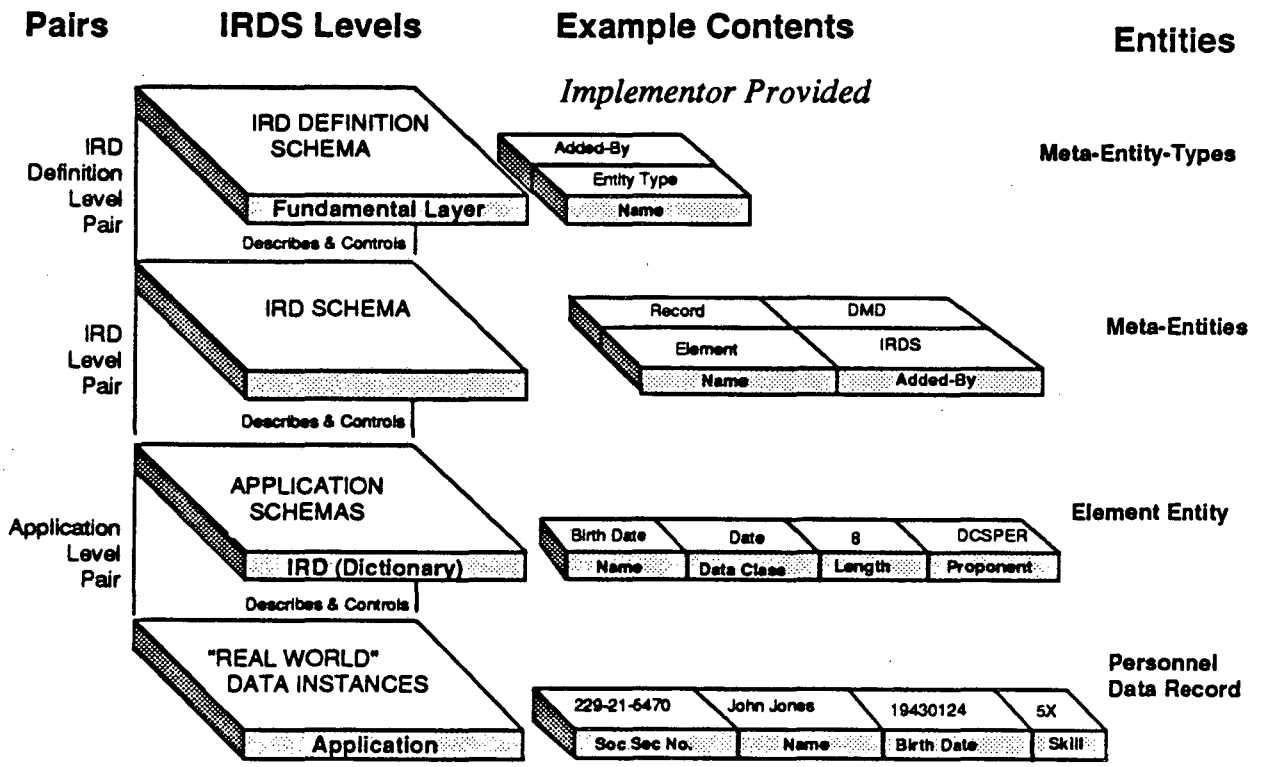


Figure 1: Overview of IRDS Architecture

Second, the IRDS basic functional schema already includes many standard entities, relationships, and attributes the Army needs to represent. Third, the proposed IRDS standard includes explicit mechanisms for extending the basic IRDS model, which can be used to tailor the IRDS to support unique Army requirements and an evolving Army Data Encyclopedia architecture.

2.3 Contents

Information Architecture -- The Army has embarked on a comprehensive Information Mission Area program to rationally describe its strategic planning and operational aspects. Central to this description is the *information model*, which is defined using the Information Requirements Study methodology. In order to characterize Army mission and organization, it is necessary to incorporate Army Information Models in the ADE.

The Army Data Encyclopedia will capture elements of information models according to the cubic structure of information class by organizational element by process,

According to AR 25-1,

> Information models will be defined at all levels, including the following tactical commands: theater Army headquarters, corps headquarters, division headquarters, and separate brigade headquarters.

This leads to the idea of the "information model cubelet" as described in Figure 2.

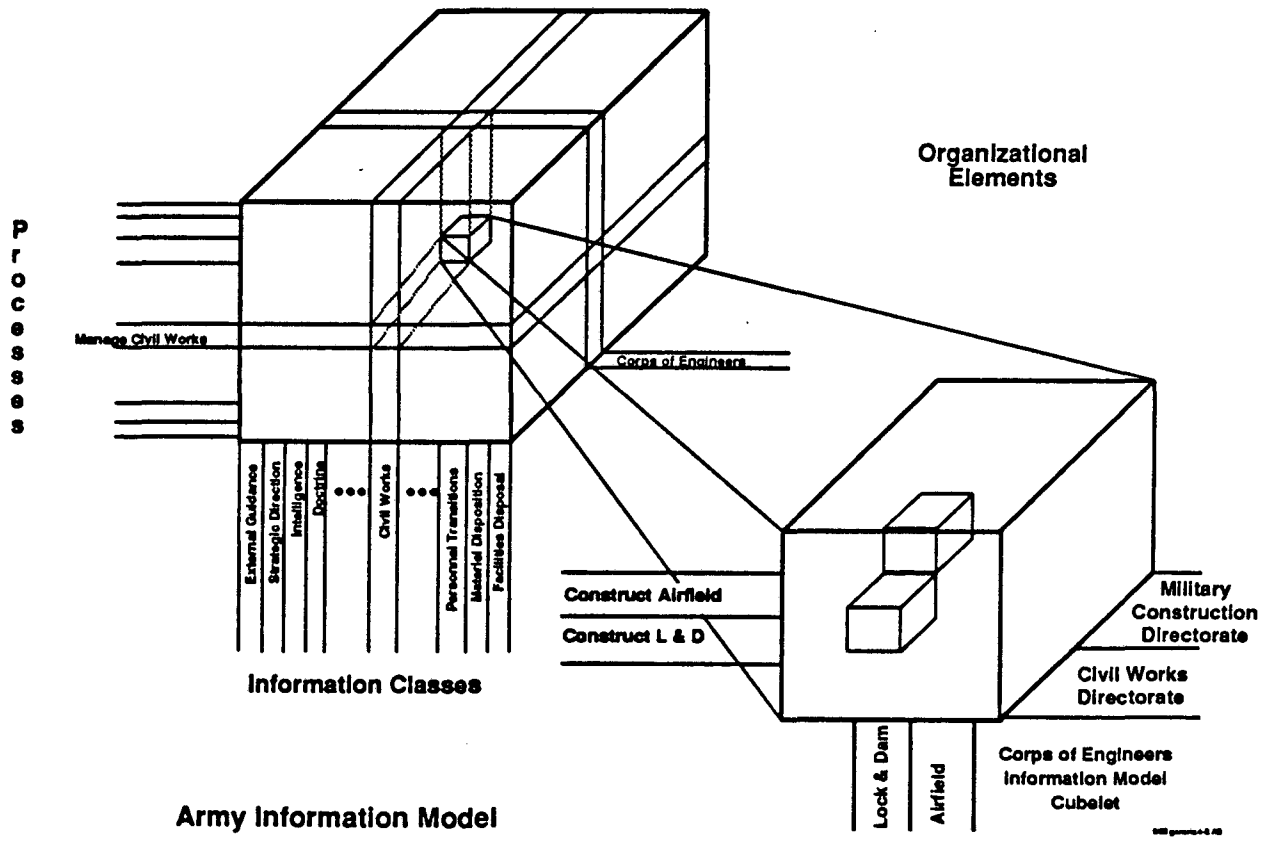


Figure 2: Information Model Cubic Substructure

Each cubelet defined by a particular process-information class- organizational element expands to define the information model of the subordinate organization. For example, the Manage Civil Works/Civil Works/Corps of Engineers cubelet can expand to an information model which describes the information requirements of the Corps of Engineers, in particular the subprocesses of Lock Construction, Airfield Construction, which are accomplished by the organizational sub-elements, the Civil Works Directorate and the Military Construction Directorate, respectively.

Using cubic substructure linkages, the ADE will support the information model through multiple levels of organizational abstraction.

Data Architecture -- To supplement the organization found in the Army Information Model, the Army has developed the Army Data Architecture to organize information stored in Army data bases and assure that it is aligned to the Army activities. The Army Data Encyclopedia will facilitate the direct links of subject areas and relationships in the data architecture to information systems development at all phases of the software development life cycle.

Applications Architecture -- The Applications Architecture defines in detail the manner in which Army information processes are to be implemented by informations systems support. The Applications Architecture quantifies requirements in terms of size, users, information descriptions, timeliness, and critically to operations in the Army. The encyclopedia will support the architecture by providing a mechanism for incorporation of metadata which support the information system design process: data flow diagrams, software processes, and data models as well as for collection of specifics about software systems, programs, modules, files, records, and database physical design information.

Geographic and Technical Architecture -- Information systems have both a software component and a hardware and telecommunications technological base upon which the software operates. Rarely does an information system operate on a dedicated technological base; usually several information systems share the same computer and mass storage configuration and operate over the same telecommunications network. It is the specification of this technological base, where it is located, and the communications between its components that form the Army's Geographic and Technical Architecture. The Army Data Encyclopedia will contain metadata to manage computer hardware configurations, network architectures and configurations, and the distribution of data bases across the technological components.

2.4 Resources Required

Implementing the ADE Version 1.0 will require about 7.5 staff for about 1 year, divided into the following areas (keyed to subsequent sections of this paper):

Section	Activities	Effort
3	Selection of DBMS and design of schema	0.5 staff year
4	Schema and data maintenance functions	1.0 staff year
5	Data element interface	1.0 staff year
6	Data content collection, verification, loading	2.0 staff year
7	Documentation and training	1.5 staff year
8	Procedures for remote user access	0.5 staff year
8	Mini/microcomputer read only version	1.0 staff year

2.5 ADE Related Activities in Progress

The Initial Data Dictionary (IDD) is a machine readable database of previously standardized data elements. While these data elements do not conform to recent AR 25-9 standards, they contain

many of the concepts and data elements in current use. In addition, many of them are derived from DoD and other higher level standards, and thus will continue to be included in a database of standard data elements. This database currently exists in DMD, stored in one or more commercial database systems. It is not being actively used, as of July 1988. The IDD data will eventually be reviewed and re-loaded into the ADE (or its precursor).

The Army Data Dictionary Pilot (ADD Pilot) is an interim prototype being implemented by Decision Support Management Agency (DSMA) in HQDA for use by DMD and for limited use by users. Its expected date of introduction is FY 88 Q4. It is being implemented in CSP on SQL/DS. It will include both a minimal user interface and a database of selected attributes for data elements. The data that is collected during the operation of the ADD Pilot will eventually be moved to the ADE (or its precursor).

The Automated Dictionary Support System (ADSS) is a user interface module for creating standard data elements and for querying/retrieving existing elements. This module will supersede (in a compatible way) the user interface module in the ADD Pilot. Development of this module will begin after the ADD Pilot is operational. Phase I of this module is scheduled for implementation in FY 89 Q2. Phase I of this module will be based on the ADD (see below), while later phases will be based on the full functionality of the ADE. This module is intended to be the primary end-user interface to the ADE, and it is expected to undergo several revisions and extensions to take advantage of the capabilities of the ADE. (See "ADSS Attributes by Phases", by DMD, for further information.)

The Army Data Dictionary (ADD) is an interim database used to support the ADSS Phase I. It will be implemented using DB2. Development and implementation of Phase I of this module will proceed on the same schedule as the ADSS Phase I. When this module is operational the data from the IDD and the ADD Pilot will be transferred to it and those two systems will be phased out.

The Army Data Encyclopedia (ADE) will incorporate the ADD. The ADD is being developed to be a subset of the ADE in content and functionality. The ADE is a database and access system which provides the functionality needed by the Army to standardize and control its database applications. The overall architecture of the ADE is based on the pending FIPS IRDS standards and is described elsewhere in detail. The ADE is both a specification and an implementation; the intent is to allow multiple compatible implementations to co-exist. An initial implementation of an ADE will be sponsored by DMD and will be based on DB2. The ADE will be an extension of work on the ADD; it has not been decided whether actual code from the ADD will be reused.

An LBL Prototype is also in progress. The goals of this effort are to evaluate several prototypes available to LBL, to advise on the selection of data element attributes to be included in the ADE (and its precursors), to advise on database schemas for use in the ADD and the ADE, to develop the technology for the Information Resource Thesaurus, and to develop other advanced technologies needed for the ADE. The LBL Prototype is described in "Next Stages in ADE Prototyping", June, 1988.

Vendor efforts are also underway. In the long term, most of the ADE software will be vendor provided. These efforts must be monitored, and a suitable transition made from in-house efforts to vendor products. This will require extensive communication with vendors to provide them with information on Army needs, as well as to keep abreast of vendor progress and products.

Relationships among the ADE related efforts (except vendor efforts) is given in Figure 3.

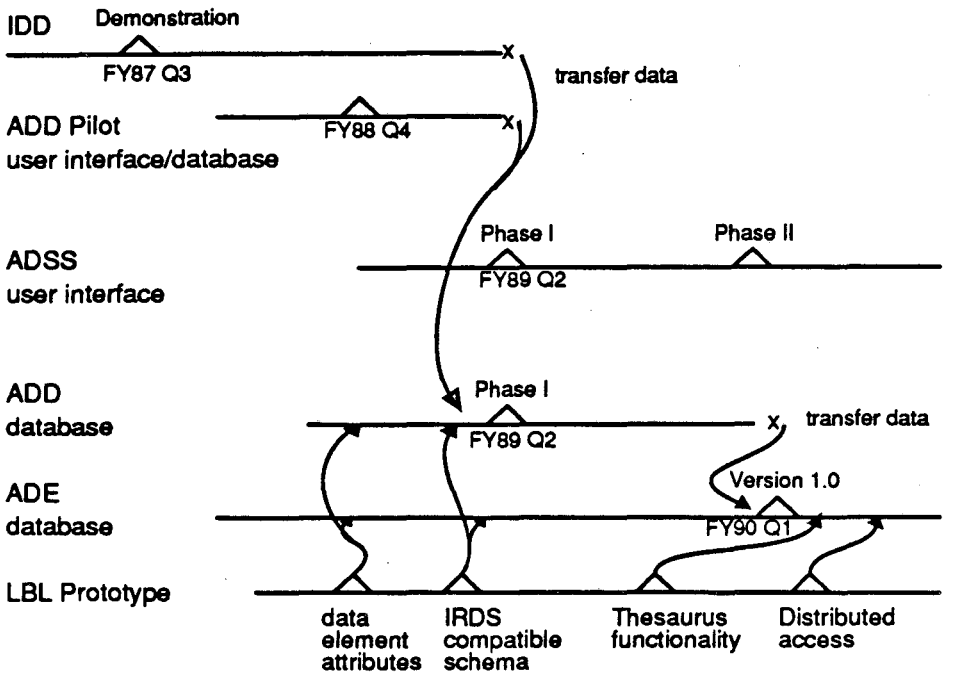


Figure 3: Relationships Among ADE Related Efforts

2.6 Items included/excluded in the ADE Version 1.0

Version 1.0 of the ADE will include the following capabilities:

- an IRDS framework,
- a Data Element search/retrieval/approval process, and
- relationship data for relationships between applications and data elements.

The following capabilities are planned as extensions in later versions of the ADE, but will not be present in version 1.0:

- a services (programming language) interface to IRDS,
- Import/Export modules for ADE metadata,
- Information Resource Thesaurus view of ADE entities and relationships,
- a sophisticated user interface for manipulating relationship data,
- integration with CASE tools, and
- integration with active distributed DBMSs.

The IRDS framework has been described in the Data Encyclopedia Architecture. The data element search/retrieval/approval process and the relationship data are described in sections below.

3. Software for the ADE Version 1.0

3.1 Software Acquisition Strategy

Given the time frame of the Army's acquisition strategy and the current state of technology and vendor offerings, a major decision facing the Army is whether to base the ADE on a commercial product or to fund internal development of an interim system. No commercial implementation of the IRDS standard has yet come to market. Thus the recommended course of action is for the Army to fund internal development of Version 1.0 of the ADE which will perhaps contain some limitations in its IRDS completeness, and then make the transition to a commercial IRDS as soon as the technology matures. An alternative course of action is to fund development by a vendor who is already involved in the development of IRDS capabilities. While this may establish the use of a commercial system somewhat earlier than a more internal strategy, the risk of getting "locked-in" to a particular vendor is substantial.

3.2 Use of Commercial (Relational) DBMS and IRDS Framework

The first step in the construction of the ADE is the acquisition of an IRDS framework. It is assumed here, for illustration and explanation, that a relational DBMS will be used; however, another type of DBMS, e.g. network or hierarchical, may be used in vendor implementations. If, as recommended, the Army funds development of the IRDS framework in the early stages of the ADE, the Army should require the use of a relational DBMS; only in this way can the Army assure the portability and interoperability that it requires. For a vendor provided IRDS framework, we assume that all accesses will occur through standardized access paths to the IRDS; thus, the underlying DBMS will not be directly visible. It is an open issue whether the Army should require that a vendor provided system be based on a relational DBMS.

For an internally developed IRDS framework, it should be specified which capabilities are most critical and which capabilities are not critical to Version 1.0 and thus could be deferred; the intent here is to achieve operational status with Version 1.0 as soon as possible.

3.3 Implementation Framework for ADEv1.0

The technical issues in moving toward version 1.0 are *efficiency, development time,* and *adherence to a standard interface*. Developing an IRDS Services Interface is not feasible within the given time-frame, and thus DMD must choose alternatives based upon advanced prototyping. Figure 4 shows an implementation framework upon which to develop ADEv1.0.

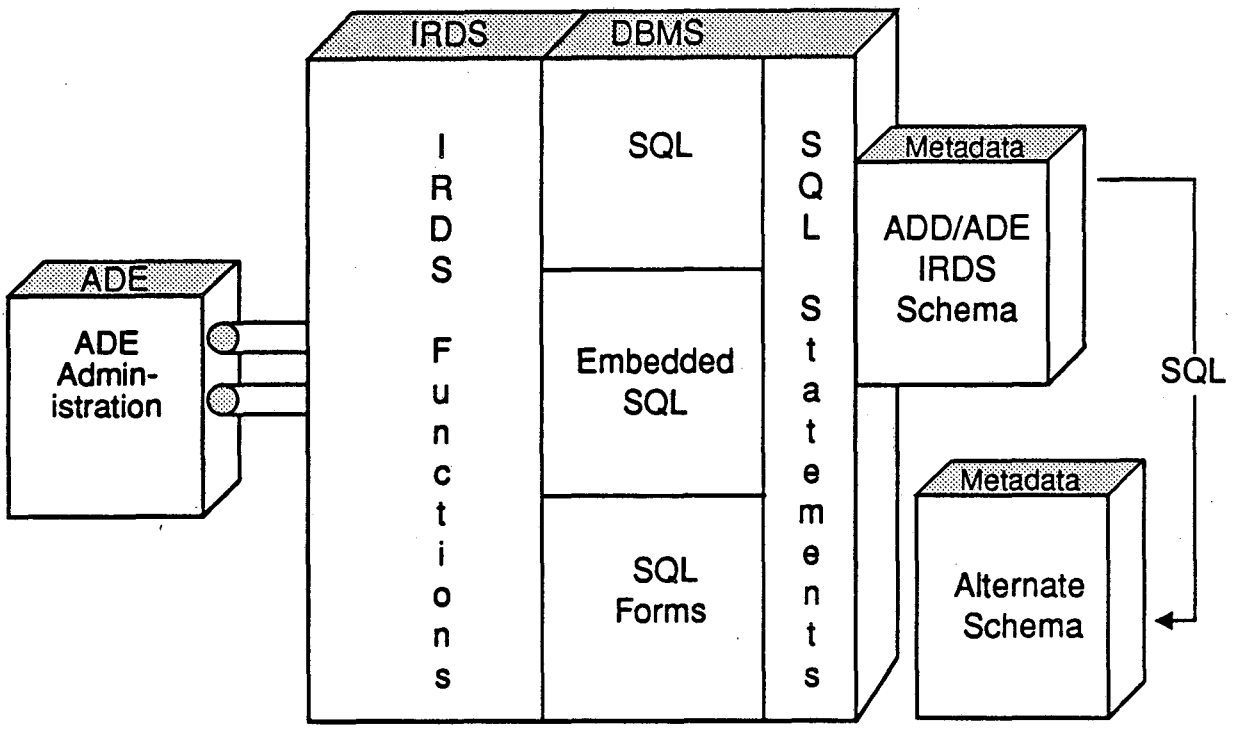


Figure 4: ADE Implementation Framework

ADE administration and other interfaces are built upon functions which interface to the IRDS schema using one or more of SQL command language, Embedded SQL through module calls, or an SQL forms package supplied by a dbms vendor. The tripartite interfaces all utilize the same SQL statements to actually access the IRDS schema and ADE data. This framework buys portability for internal development, and flexibility for schema modification; that is, it only takes SQL statements to unload the schema and data and reload it into an alternate schema, should one be deemed beneficial. The use of a forms interface is vendor-dependent but may substantially reduce development time over developing screens in the a third generation language language which calls embedded SQL. The use of embedded SQL, on the other hand, gives greater portability and independence from a particular vendor.

3.4 IRDS Framework Required for ADE Version 1.0

The IRDS framework required for the initial ADE includes the basic elements to maintain entities and relationships, and the elements needed to support extensibility. Thus, the IRD Schema must contain at least the following entities:

- Entity Type,
- Relationship Type,
- Attribute Type,
- Relationship Class Type,

and the following relationships:

- Entity_Type contains Attribute_Type,
- Relationship_Type connects Entity_Type, and
- Relationship_Type contains Attribute_Type.

The following elements of the IRDS standard are not critical to a first implementation of the ADE:

- Attribute Group Types,
- plural (repeating) attributes,
- life cycle management,
- views, and
- security.

By relaxing the requirements for attribute groups and repeating attributes, the necessary framework can be easily mapped into a set of tables in the relational model. It is not required that each component of the schema be stored in a separate table, they might be combined and retrieved via views, as illustrated in both the Dolk [DOLK 1987] and the NBS [GOLD 88a] prototypes. The requirement that attributes can be dynamically added requires a DBMS that allows dynamic extension of its tables. A minimal set of tables, with examples, to support this IRDS framework is given in Figure 5 below.

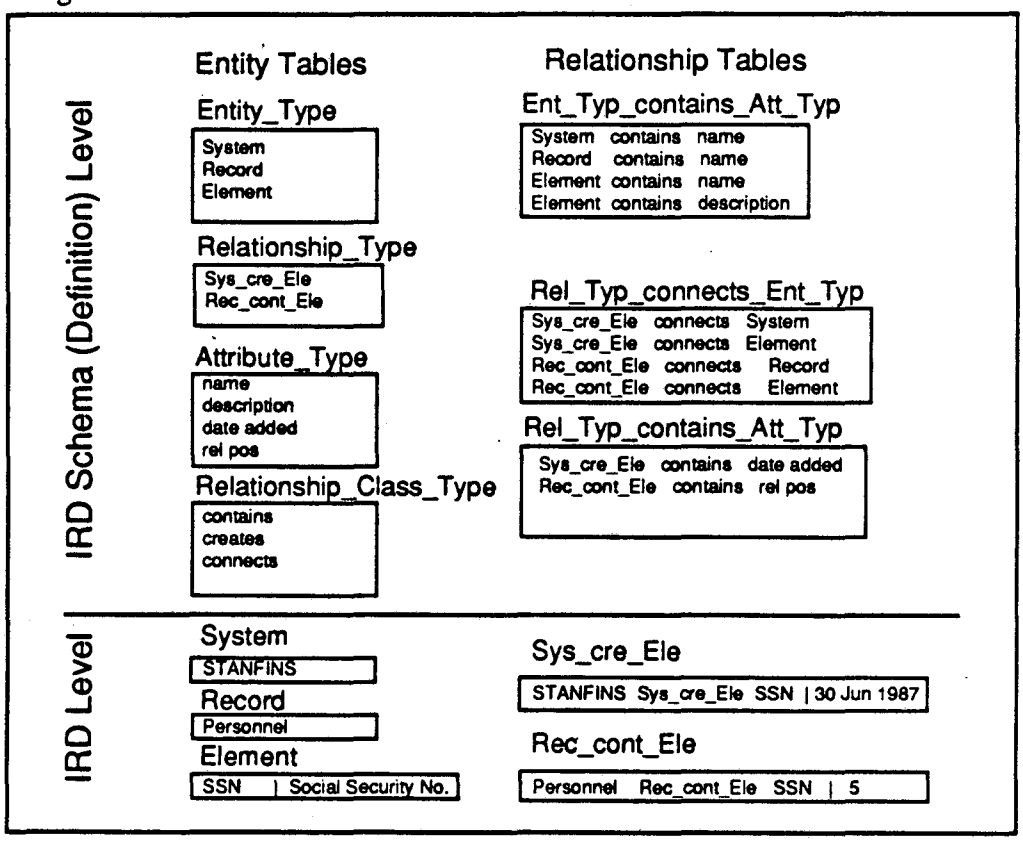


Figure 5: Relational Tables Required for an Initial ADE (with examples).

These tables support the basic IRDS concepts of entities and relationships, with both entities and relationships having attributes. In IRDS, relationships are restricted to relating entities to entities, not entities to relationships or relationships to relationships. The current IRDS standard supports only binary relationships, not the more general n-ary relationships provided by some entity-relationship models. The tables at the schema level support these concepts by providing several tables in which new objects and their attributes are recorded. (At the schema level, entries in these tables are called "meta-entities" or "meta-relationships"). The Entity_Type table allows new kinds of objects, e.g. a "document," to be created. An authorized user enters name, description,

authorizing agency, etc. for the new object; then the attributes of the new object, e.g. name of document, description, length of document, author of document, etc. are assembled using the "Ent_Typ contains Att_Typ" relationship. Each attribute of the new object is an instance in this relationship. After a new object is defined, the IRDS can record instances of this object, e.g., "Gettysberg Address," "battlefield speech," "270 words," "Abraham Lincoln," and so forth. Before an attribute can be associated with an object, the attribute must be defined in the Attribute_Type table. This allows the characteristics of the attribute itself to be specified, e.g., the "rel_pos" attribute will be numeric with a field width of 5 digits. Each attribute can be associated with as many entities or relationships as desired.

To create a new relationship, an additional step is required to declare the entities which are connected by the relationship; this information is entered into the "Rel_Typ connects Ent_Typ" table. Since only binary relationships are supported by the IRDS, there are exactly 2 instances in this table for each relationship. One of the examples of a relationship in Figure 5 is "Rec_cont_Ele." This relationship was created by

1) adding it to the "Relationship_Type" table,
2) recording the entities related in the "Rel_Typ connects Ent_Typ" table, and
3) specifying the attributes of this relationship in the "Rel_Typ contains Att_Typ" table.

In the Dolk implementation of the IRDS model, relationships are recorded in a separate table which contains the name of the relationship and both of the entity types which it connects. The "Relationship_Type" table and the "Rel_Typ connects Ent_Typ" table are then implemented as views of this underlying table.

The basic functionality required for schema maintenance is the ability to retrieve and edit the contents of the schema tables. The system should apply consistency checks to these edits, such as requiring that entities exist before they are referenced in relationships (this consistency check is also appropriate at the IRD level in most cases). The fields of the schema tables themselves can be described in the IRD schema, allowing software for user level edits to be reused at the schema level.

IRDS functions beyond those required for schema maintenance include export/import commands, consistency checks, and others described in the standard.

4. Functionality of ADE Version 1.0

The following functionality is to be incorporated into the ADE Version 1.0:

- ADE/IRDS Maintenance Functions
- Online query (by attribute and subject index)
- Impact of Change and other reports
- Word (subject) indexing from descriptive text
- Integrity mechanisms (to assure that types & instances conform)
- Audit mechanisms

4.1 ADE/IRDS Maintenance Functions

Functions available to the ADE administrator in version one should be:

CREATE/REMOVE ADE To initiate or delete the ADE

OUTPUT IRD-SCHEMA To report the contents of the IRDS schema for the ADE

Functions for ADE Schema Administration:

ADD META-ENTITY/META-RELATIONSHIP
MODIFY META-ENTITY/META-RELATIONSHIP
DELETE META-ENTITY/META-RELATIONSHIP
MODIFY META-ENTITY ACCESS-NAME

Functions for ADE Administration from an IRDS viewpoint:

ADD ENTITY/RELATIONSHIP
MODIFY ENTITY/RELATIONSHIP
DELETE ENTITY/RELATIONSHIP
MODIFY ENTITY ACCESS-NAME/ENTITY DESCRIPTIVE-NAME
COPY ENTITY
OUTPUT ADE To report the contents of the ADE

These functions can be implemented either using a command language interface as provided by the IRDS standard, or by building a panel interface as described in the standard (see Figure 6).

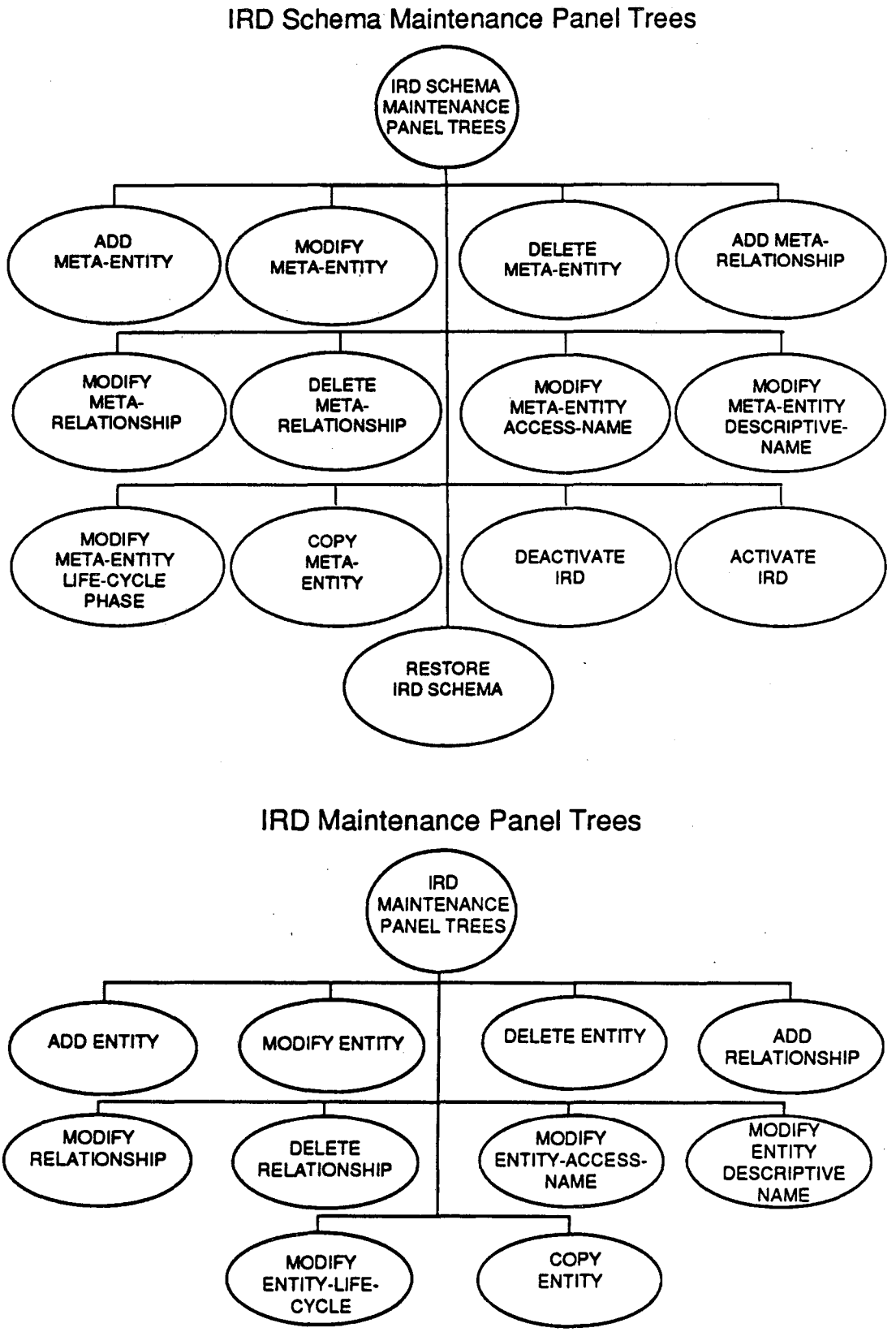


Figure 6: ADE Administration/IRDS Maintenance Panels

4.2 Online query

The ADE must support on-line selection of both IRD schema information (to be utilized the the ADE data administrators) and IRD entities relationships and attributes. Querying by subject index (to search for existing standard elements which might be used instead of proposing a new element) and attribute (find all systems described by information class "CIVIL WORKS") are the minimum query functions which should be supported.

An important function is that of navigational *browsing* of lists of entities and relationships. This is best accomplished using a multi-level windowing environment with horizontal scrolling and highlighting. More will be said about this in the user interface section below which describes initial screens for the ADE Version 1.0. The important functional aspect is the ability to highlight and choose an entity, then open a window that displays all relationships which that entity participates, or to open another window which displays all the attribute values for that entity.

4.3 Impact of Change reporting

One of the central features of the Encyclopedia will be able to assess the physical (and economic) impact of changes made to encyclopedia objects. For example if the post office has moved from 5 digit to 9 digit zipcodes, how many files, programs and modules will be affected in the operation of Army information systems? The ADE Version 1.0 must implement functionality equivalent to the impact-of-change command of the IRDS standard [ANSI 88]:

```
OUTPUT {INDIVIDUAL | CUMULATIVE} IMPACT
  SELECT <ENTITY>
  SHOW {RELATIONSHIPS [OF TYPE <RELATIONSHIP_TYPE> |
       ENTITIES | ATTRIBUTES}
```

which will display all entities, relationships, and attributes affected by change to a selected entity or (in the case of cumulative impact) will count the number of such impacts.

4.4 Word (subject) indexing from descriptive text

A useful feature for grouping ADE objects with similar characteristics is to do word indexing of descriptions and comments. For example the system STARCIPS (STandard Army CIvilian Payroll System) has the description

> Is designed to provide sophisticated and comprehensive *payroll* and accounting support to Army installations worldwide.
>
> A word index and search on the work **payroll** would yield other systems which deal with payroll:

HPPS	Health Professions Payroll System
JCPS	Job Corps Payroll System
JUMPS	Joint Uniform Military Payroll Systems
PLACE	Production, Labor Accounting and Cost Estimating
RCPS	ROTC Cadet Payroll System
WPPS	West Point Payroll System

Independent indexing and search improves over substring search of textual descriptions in the following ways:

- Only content words are stored in the index; words without content (which comprise most of the textual description) are filtered out by the indexing process.
- The SQL query statements are shorter and can include more complex information search strategies.
- The index can be enhanced to become a basis for more advanced information access techniques.

Word indexing can be used for identifying existing standard elements which might be used instead of proposing new ones, for software reuse of programs and modules performing the same function, and for identifying duplicated functions performed by separate systems.

4.5 Integrity Constraints

A major function of the ADE is the preserve the integrity of Army metadata and provide for consistency of the information objects and relationships stored in the Encyclopedia.

4.5.1 Type-Instance Integrity

The IRDS framework within the ADE must be implemented to support the type-instance mechanisms for IRD Schema/IRD Data layers. Thus for example, the ADE will ensure that:

> A relationship type connects two entity types (and not two relationship types or a relationship type to an entity type).
>
> The entities participating in a relationship are of the correct type described by the relationship type defined in the schema, e.g. if the schema defines the relationship type "system creates element" then the construct:
>
> > STANFINS CREATES SOCIAL-SECURITY-NUMBER
> > is legal, while
> >
> > HOURS-WORKED CREATES SOCIAL-SECURITY-NUMBER
> > is not, assuming HOURS_WORKED is an entity-type="element"

Note that this enforces type-instance integrity, but still allows potentially incorrect entries such as

> TEMP-MEASUREMENT-SYSTEM CREATES SOCIAL-SECURITY-NUMBER

Additional integrity mechanisms, dependent upon thesaurus development, can be incorporated in later versions of the Encyclopedia.

4.5.2 Completeness and Restriction Integrity

Completeness is a check applied to relationships and relationship types such that if two entities(entity-types) are to be related, they must already exist and be entered into the IRD (IRD-Schema). In this way incomplete relationships with no referent such as

STANFINS CREATES garbage

will not be legal, if "garbage" is not an entity.

Among the *restrictions* inherent in the IRDS standard is that an attribute-type attached to a relationship-type cannot be an entity.

Semantic overide is a feature by which an integrity constraint can be by-passed by explicit command of the Encyclopedia administrator. For example in the

SYSTEM COMES-FROM SYSTEM
(denoting information interchange between system entities)

relationship type, there may be existing data such as found in the STAMIS MOD database as

STARCIPS COMES-FROM TREASURY
(STARCIPS is the standard Army civilian payroll system)

where "TREASURY" is not a system, but an external regulating agency. It may be important to capture this data even though it has some flaws, and correct the inconsistencies later.

4.5.3 Deletion Integrity

Deletion integrity refers to those mechanisms which insure that consistency is maintained when entities are deleted from the encyclopedia. For example if the entity STANFINS were removed from the encyclopedia, then all relationships (for example of SYSTEM COMES-FROM SYSTEM type) will also be removed, in order to perserve consistency of relationships.

4.6 Extending the IRDS Schema

On top of a commercial DBMS, the IRDS will provide the components and services described in the ANSI/FIPS standard (or, for a developmental system, the required subset). Extensibility of the IRD will be used to provide additional entities and relationships (and their attributes) to support the specific entity and relationship types used by the Army. These will include entities necessary to store the Army's IMA Architectures, to store design information and/or the location of design information, and to store implementation, operation and maintenance information. An initial list of attributes for Data Elements is contained in "ADSS Attributes by Phases," a DMD internal document.

The IRDS will also be extended to include domain information, keyword indexes and other information necessary to provide a convenient and effective user interface.

5. Develop a Search/Retrieval/Approval Application Module

A key component of Version 1.0 is an application module which supports search, retrieval, and approval operations for standardization. This module will provide a convenient user interface for system designers and data base administrators, and it will enforce the standardization policies of the Army to the extent possible.

This module should be interfaced to the Services interface of the IRDS framework, if that is possible. If the Army is developing an IRDS framework internally, then the Services interface may not be available. In this case, it will be necessary to interface the module to the DBMS directly. This should be undertaken with great caution and concern that the module be designed so that it can be moved to the IRDS Services interface as soon as it is available. Figure 7 shows an ADE software configuration.

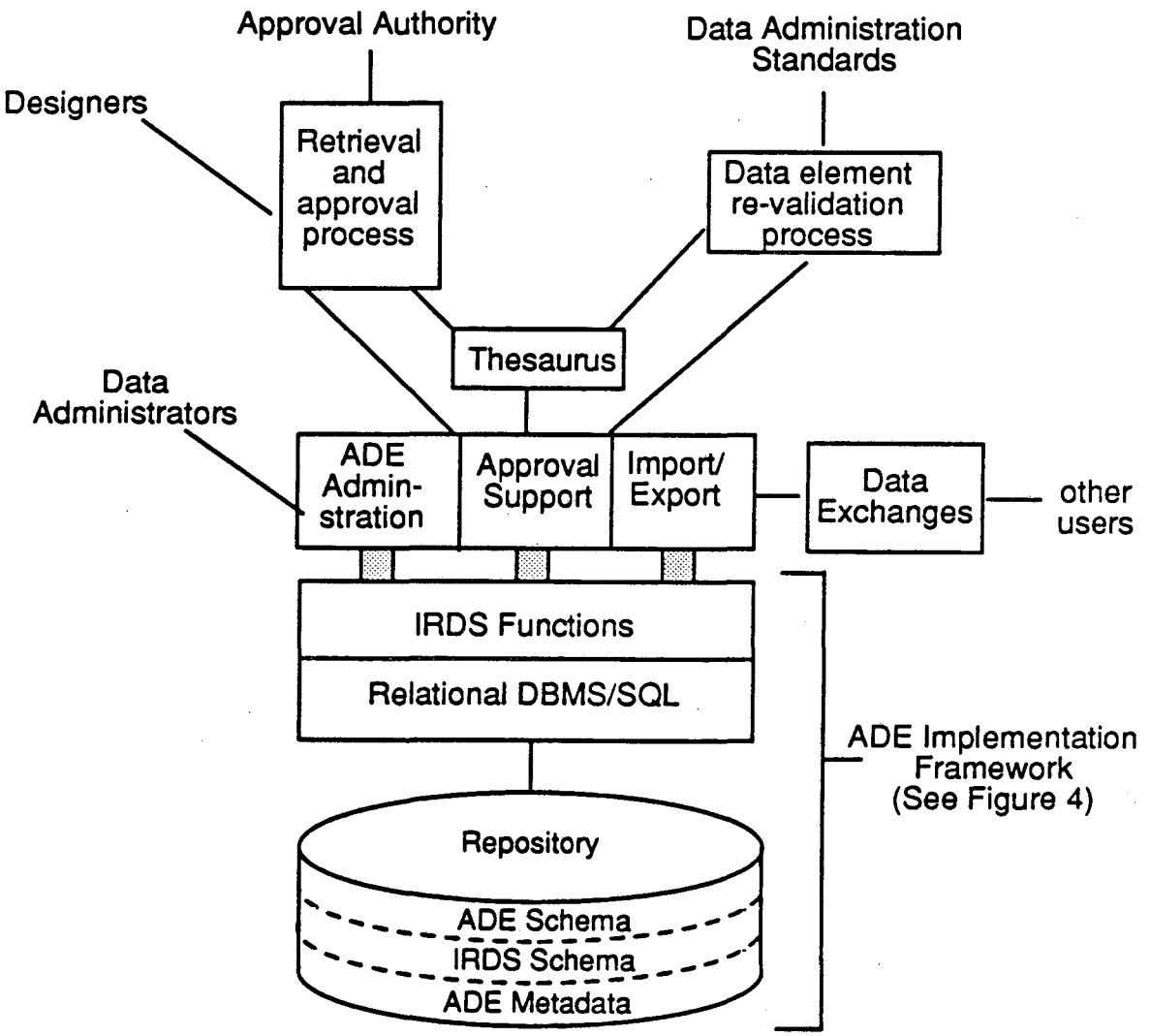


Figure 7: ADE Version 1.0 Software Configuration

This module should use a commercial screen management utility to promote portability and minimize the coding effort required. The following screens should be provided:

- An opening screen with choices for searching or maintenance by:
  Data Element name (exact or approximate),
  index term,
  attribute with specified value, or
  participation in specified relationship with specified entity.

After the initial choice, a window will open to accept the name, term, attribute and value, or relationship and entity. The system should accept specifications for inexact matches. If no exact match was found, a window should open with candidate items and descriptions so that one can be selected.

- Selection screen showing candidate items and one line descriptions.
- Display screen (item oriented) showing the most relevant information about an item, with pull down menus to access the less commonly used information, including less important attributes, comments and long descriptions (in a scrolling window), and relationships that this item participates in.
- Display screen (relationship oriented) showing an item and a user selected class of relationships and related entities, again with pull down menus to access less commonly required information.

These screens are illustrated in Figure 8.

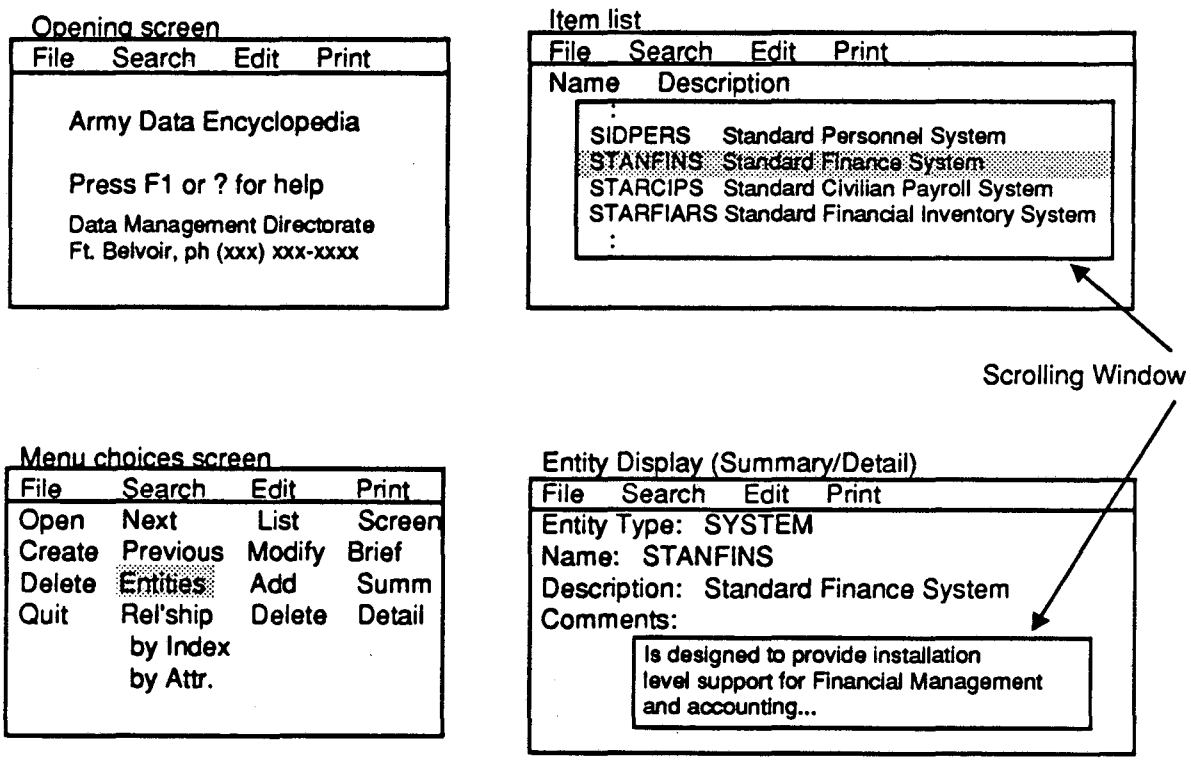


Figure 8: Screens for the ADE Version 1.0

The screens shown in the figure build upon one another to provide a navigational facility throughout the Encyclopedia information base. The *Menu choices* screen (lower left) lists possible options for accessing searching, updating, and deleting information. If the choice *Entities* (highlighted) is made, another screen (not shown) would open which lists the possible entity types stored in the ADE. Among these would be the *SYSTEM* entity type, which, if chosen, would open the screen *System list* (upper right) which displays a scrolling window of system names and descriptions. A cursor (or mouse) can highlight a particular system *STANFINS* which opens a detail screen about that system (lower right). Within that screen, the *Comments* form yet another scrolling window through which the ADE user can view running commentary which is too large to fit in a fixed size screen. Similar navigation mechanisms can be used to traverse the relationships which are attached to each entity.

At each screen it should be possible to navigate through entities and relationships by pointing to an entity or relationship of interest and jumping to the display screen for that item.

6. Initial Content of the ADE Version 1.0

This section describes the data to be initially loaded into the ADE. We recommend that the information be loaded into the ADE in the order of the following sections; each portion of information will thus provide a foundation for subsequent data as described in each section.

Each portion of information will require extensions at both the IRD Definition level -- additional entity types and relationship types -- as well as at the IRD level -- new entities, relationships and attributes -- before the IRD can be populated. The complete list of attributes is derived from the basic list established by the IRDS standard, from AR 25-9, from DoD and NATO regulations, and from national and international standards (X3L8 and others).

6.1 Army Data Architecture

The Army Data Architecture will be included in the ADE in the form of Subject Areas (28 areas), Processes (63), and Information Classes (66). In addition, approximately 15 relationship types will contain the 100 or so relationships defined in the Army Data Architecture. (See Figure 9.)

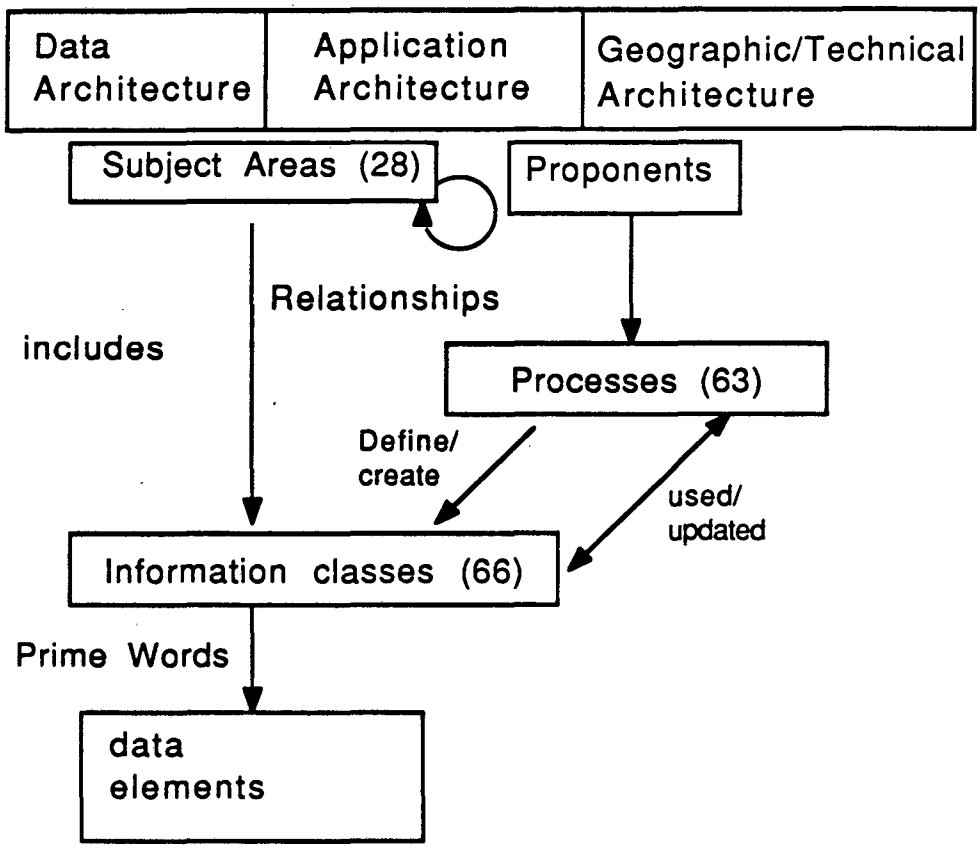


Figure 9: ADE Support for Army Information Architecture

These entities and relationships will be the machine readable repository for the Army Data Architecture. They will be the basis for consistency checks to ensure that data added to the ADE conforms to the Army Data Architecture, and they will be the starting point for information systems requirements definition, planning and implementation. For new projects, the Army Data Architecture will be loaded from this source into CASE tools or other environments for use in design and implementation. This will ensure that such designs are consistent with the ADA.

6.2 Standard Data Elements from Existing Sources -- Uncontrolled

The existing collection of DoD and other (prior) standard data elements should be loaded. These will be loaded into the IRDS as uncontrolled elements, where the IRDS facilities can be used to expedite their standardization. Data elements, such as the DoD elements, may be moved immediately to "Controlled" status if the Army does not wish to add any information to them. Each Data Element will be related to the proper Subject Class through the use of a Prime Word (which is the same as its Subject Class), and to the proper Proponent through the use of a specific Proponent attribute (or indirectly via Information Class => Process => Proponent). Other data elements will be loaded into the ADE as they are standardized. They will come from the standardization of elements in other existing systems, and from the standardization of new elements using the ADSS (or its successors).

6.3 Software Systems, Programs, and Modules

Metadata about software systems, applications and module information will be stored in the ADE Version 1.0 including STAMIS MOD data from SPOD (Software Policy and Operations Directorate) and other systems information which should be obtained for standardization from whatever sources are available in the Army.

6.3.1 Organization of Software Metadata

The organization of metadata about software objects within the ADE version 1.0 will follow the *basic functional schema* of the Information Resource Dictionary System. In this organization, the major entities are

SYSTEM	collections of processes & data
PROGRAM	automated processes
MODULE	logical subdivisions of PROGRAMs

Major attributes (independent of universal attributes such as DATE-ADDED which apply to all entities) for software entities include

DURATION
NUMBER-OF-LINES-OF-CODE
SYSTEM-CATEGORY

to which additional attributes should be added to more fully specify these software objects. Some of these have come from the STAMIS inventory and have been incorporated into the initial ADE prototype are

NUMBER-OF-PROGRAMS
FUNCTION (e.g. STAMIS)
STATUS (e.g. "in redesign" or "operational")

6.4 Relationships Between Software Systems and Data Elements

As information becomes available, the relationships between data elements and the software systems, files, etc. which use them should be added to the ADE. This will begin to provide an integrated view of the Army's Data Management Program. Major relationship-types which software entities participate are

CONTAINS relationship-class-type:

SYSTEM-CONTAINS-{SYSTEM | PROGRAM | MODULE}
PROGRAM-CONTAINS-{PROGRAM | MODULE}

PROCESSES relationship-class-type:

{SYSTEM | PROGRAM | MODULE}-PROCESSES-{FILE | DOCUMENT | RECORD |ELEMENT}

GOES-TO relationship-class-type:

SYSTEM-GOES-TO-SYSTEM
PROGRAM-GOES-TO-PROGRAM
MODULE-GOES-TO-MODULE

Other relationships (following the ADE prototype) to be added include a new relationship-class-type USES

USES relationship-class-type:

SYSTEM-USES-{HARDWARE | LANGUAGE | DBMS | NETWORK}

which more fully capture the geographic and technical architecture under which a particular information system operates.

6.5 Information Resource Thesaurus Relationships

As the ADE accumulates a substantial body of information, the relationships which should be captured in the Information Resource Thesaurus will become more clear. This information should come from the Army's various Architectures, from keywords provided or added when the ADE data is loaded, and from observations of usage patterns and other significant correlations among subject areas, as observed by the data administrators.

7. Documentation and Training for the ADE Version 1.0

Documentation and training are the key ingredients to making the ADE accessible and useful to its intended audience. Instructions for the data base administrators will be based on the IRDS command language interface or a panel equivalent and will include Army policy and regulations related to the ADE, as well as instruction in the use of specific ADE functions. Instructions for designers, developers and other end users will be focused on the application module for standardization, with less emphasis on direct use of the IRDS command language.

7.1 Data Administrator Capabilities

The IRDS command language of the ADE, or its (limited) functional equivalent in a panel interface, will provide capabilities for retrieval, and for IRDS administration. The general capabilities of these interfaces have been described in the "Data Encyclopedia Architecture," in the IRDS standard, and in other documents.

7.2 The Approval Process Module: Capabilities

The facility for searching, retrieving and approval provides for keyword searches, for subject area searches via the use of keywords, and for other attribute controlled searches, e.g., find terms related to volume measure. This facility operates by searching lists of keywords, subject terms, domain classes, and linkages between these terms and information categories. The database of terms and linkages is maintained by the IRDS. Until the Thesaurus is fully operational, this module will provide a limited subset of the Thesaurus capabilities.

8. Providing User Access to the ADE from Remote Sites

In the initial implementation of the ADE, access to the ADE should be provided via network and dialup access for standardization functions and via mass replication and distribution of the contents of the ADE for retrieval functions. This dual distribution mechanism recognizes that most accesses will be retrievals, with a far fewer number of accesses for update. By replicating

the ADE and distributing copies of it, users are provided with faster access for retrievals and most of the load on the network and the host is avoided. The critical issue then becomes: how are these copies of the ADE data to be kept up to date? Distribution of updates should proceed through three phased implementation plans:

(1) In the early phases (when the ADE database is less than 20 Mbytes), the database should simply be replicated and distributed in its entirety. This can take place via magnetic tape or other media, or the data could conceivably be sent over the network if a re-distribution plan were used to avoid undue network traffic.

(2) In the medium term, users should be provided with periodic updates to the ADE, with the periodic updates being distributed via a magnetic media or the network.

(3) In the long term, an automatic caching scheme should be used. Applications which access the ADE will maintain a timestamp for each data item, and retrieve a new copy of an item when they encounter an item whose timestamp has expired.

Plan 3 is most appropriate when the ADE database is too large to be stored in its entirety at a site which uses it. If a site is able to store all of the ADE data, then plan 2 is probably more efficient. Plan 2 could be automated such that updates could be sent out over the network daily (or less often) and automatically incorporated into the copies of the ADE data.

9. References

ANSI 88 American National Standards Institute, Committee X3H4, *Information Resource Dictionary System Standard, Modules 1-6,* Document X3.138-1988, 850 pages.

CSR 88 Computer Science Research Department, Lawrence Berkeley Laboratory, *Encyclopedia Architecture for Army Information Management,* Report LBL-25137, June, 1988.

DOLK 87 Dolk, Daniel, and Robert A. Kirsch II, "Relational Information Resource Dictionary System," *Comm ACM 30, 1 (January 1987), pp. 48-61.*

GOLD 88a Goldfine, Alan, and Patricia Konig,*Technical Overview of the Information Resource Dictionary System (Second Edition),* National Bureau of Standards, NBSIR 88-3700, U.S. Department of Commerce, Gaithersburg, MD, January 1988.

GOLD 88b Goldfine, Alan, and Thomasin Kirkendall,*The ICST-NBS Information Resource Dictionary System Command Language Prototype,* National Bureau of Standards, NBSIR 88-3830, U.S. Department of Commerce, Gaithersburg, MD, August 1988.

LAWRENCE BERKELEY LABORATORY
TECHNICAL INFORMATION DEPARTMENT
1 CYCLOTRON ROAD
BERKELEY, CALIFORNIA 94720