AD-A261 331

USE OF THE IRDS
STANDARD IN CALS

DTIC
S ELECTE
FEB 2 6 1993
C D

David K. Jefferson
Cita M. Furlani

U.S. DEPARTMENT OF COMMERCE
National Institute of Standards
and Technology
National Computer Systems Laboratory
Gaithersburg, MD 20899

93-04048

NIST

93 2 25 054

USE OF THE IRDS STANDARD IN CALS

David K. Jefferson
Cita M. Furlani

U.S. DEPARTMENT OF COMMERCE
National Institute of Standards
and Technology
National Computer Systems Laboratory
Gaithersburg, MD 20899

September 1989

USE OF THE IRDS STANDARD IN CALS

David K. Jefferson
Cita M. Furlani

National Institute of Standards and Technology

Abstract

The objective of this point paper is to show
how the Information Resource Dictionary System
(IRDS) can fulfill critical design and
operational requirements for CALS Phase II.
First, a series of assumptions are made about
the data management services which are needed
by CALS Phase II. Next, these assumptions are
used to develop a series of requirements for a
dictionary system. The structure of the IRDS
family of standards is then described.
Examples are provided to illustrate how the
IRDS could meet the requirements. A schedule
is presented to show that the IRDS and other
data management standards will be available
when needed to meet the immediate requirements
of CALS. An architecture is presented to
illustrate additional standards required to
achieve longer-range goals of distributed
database. Finally, development tasks are
recommended.

TABLE OF CONTENTS

1. Assumptions

The following assumptions are made about the mechanisms needed to achieve CALS Phase II:

o   A _logically integrated database_ is absolutely necessary.

o   A single physical database is impractical--_multiple physical databases_ controlled by multiple DBMSs are required.

o   A _three-schema architecture_ is the only practical way of achieving control over such a database.

o   All activity of the DBMSs must be logically controlled by a _logically single representation of the three-schema architecture_.

o   A single physical representation of the three-schema architecture . is impractical--_multiple, coordinated, geographically distributed physical representations_ are required.

o   The development and maintenance of the three-schema architecture will be _extremely complex and require a very large volume of data_.

o   To add further complication, a three-schema architecture, since it represents a union of different disciplines (as represented by different external schemas), may be developed using _more than one modeling methodology_.

1

2. Requirements for a Dictionary

Given the preceding assumptions, the following may be concluded:

- o A mechanism is required to actively supply three-schema information to the various DBMSs.

- o A mechanism is required to coordinate separate sources of such three-schema information.

- o A mechanism is required to support data modeling and schema maintenance.

- o A mechanism may be required to support multiple methodologies for performing the data modeling process--i.e., there may be a requirement for a supporting mechanism that is neutral with respect to methodology.

Coordinated active dictionaries which are neutral with respect to data modeling methodology are one mechanism for satisfying the preceding requirements. Such dictionaries must support four distinct processes:

- o They must have a standard interface to provide three-schema information to a variety of database management systems.

- o They must support the process of coordinating communication to control the DBMSs.

- o They must support the process of describing the vocabulary needed by the chosen data modeling methodologies.

- o They must support the process by which each methodology populates a dictionary using its vocabulary.

2

3. Structure of the IRDS Family of Standards

3.1.   ANS X3.138

Most of the functionality of the IRDS is defined by the <u>American National Standard for Information Systems-Information Resource Dictionary System</u>, which is an American National Standard (ANS), X3.138-1988, and a Federal Information Processing Standard (FIPS), Publication 156. The parts of this standard most relevant to CALS Phase II deal with the creation, maintenance, and retrieval of two distinct levels of data. One level of data is used to <u>describe a vocabulary</u>, and the other level is used to <u>apply that vocabulary</u> in the data modeling methodologies. The former level is referred to as the <u>IRD Schema Level</u>, and the latter as the <u>IRD Level</u>. The IRDS provides for extensibility at the IRD Schema Level to satisfy the particular requirements of any set of methodologies and applications. This degree of generality is not free: the first task in the use of the IRDS to support concurrent engineering must be to use the IRD Schema Level to define a <u>standard vocabulary</u> which will serve as a union among the individual methodologies and applications. Fortunately, the different data modeling methodologies have small vocabularies (e.g., a few different kinds of boxes, lines, and arrows) so this should not be a particularly difficult task. For example, the following vocabulary covers the major concepts of the IDEF1X modeling methodology:

o   <u>entity</u> is an IDEF1X entity,

o   <u>entity-is-assembled-of-entity</u> is a type of IDEF1X relationship,

o   <u>(cardinality-1, cardinality-2)</u> is a property of an IDEF1X relationship,

o   <u>entity-contains-relationship</u> associates an attribute (or data element) with an entity; the association may have various properties (e.g., an attribute may be <u>used-as "pk"</u> to indicate that it is part of the primary key,

o   <u>category</u> is an IDEF1X category,

o   <u>entity-is-a-category</u> indicates that an entity is associated with a particular category, and

o   <u>category-set-is-total</u> indicates that the set of entities is exhaustive.

The next task is then to use that vocabulary within the data modeling methodologies to define the three types of schemas.

3

For example, the following objects serve to group entities into the indicated schemas:

o conceptual-schema-view collects together the entities in the conceptual schema, and

o LSA-external-schema-view collects together the entities in the external schema used for LSA.

The following are examples of entities, relationships, and attributes in a schema:

o product-item is-assembled-of product-item-usage cardinality = (parent="1", child="n") indicates that one item may be assembled from many simpler items,

o product-item contains part-number used-as = "pk" indicates that the primary key of item is the part number, and

o geometric-model is-a geometric-model-category geometric-model-category can-be-a wire-frame-model indicates that a geometric model can be a wire frame model or some other type of model which may or may not be enumerated explicitly.

A particularly important feature of the IRDS is the ability to provide user-oriented names. For example, associated with product-item may be the following:

o identification-names = (alternate-name = "LSA product item", alternate-name-context = "LSA"), which could provide a special name for product-item in the LSA external schema,

o descriptive-name = some-very-long-descriptive-name-which-is-readable-but-not-easily-typed-or-accepted-by-programming-languages-or-DBMSs, provides an alternate to the normal short access name, and

o some-general-name (LSA-SQL-variation: 2) indicates that this is revision 2 of a general object which has been specialized to the use of SQL in an LSA application.

3.2. IRDS Services Interface

ANS X3.138 defines two interfaces which can be used by people to define and populate a repository for schema definitions, but there are no standard capabilities for constructing analyses, diagrams, and so on for the schemas. This is appropriate, given that there are no standard methodologies. Other tools, dependent on particular methodologies, must request data from

4

the IRDS to provide these capabilities. The draft standard for the IRDS Services Interface defines functionality for communicating between the IRDS and other software, such as a diagramming program or a DBMS. The Services Interface is a program call interface which complements the human interfaces defined in ANS X3.138. This interface will provide the active dictionary support that is required to control the DBMS, as well as allowing continued use and integration of existing support tools for IDEF0, IDEF1X and other methodologies.

3.3. IRDS Export/Import File Format

ANS X3.138 also specifies functionality for exporting all or part of a dictionary into a file, checking a second dictionary for compatibility with that file, and optionally importing that file into the second dictionary. However, ANS X3.138 does not specify the format of the file; this is the subject of the final critical element in the IRDS family of standards, the draft standard for the IRDS Export/Import File Format. This defines an efficient and reliable format for communicating mass data, such as an entire database schema or collection of database schemas, from one dictionary to another. ANS X3.138 in conjunction with the IRDS Export/Import File Format standard can be used to compare one dictionary with another, as well as to populate an empty dictionary. The standards will therefore provide the required capability of dictionary coordination.

3.4. Categorization of the IRDS and Other Standards

The standards needed by CALS Phase II can be divided into two general categories:

o   Data standards represent the "business rules" of CALS, and are developed by the CALS Office, the military Services and Agencies, and defense contractors. They are represented by the contents of two layers of the IRDS:

    oo   IRD Schema Layer represents modeling rules, and

    oo   IRD Layer represents the schemas and related information.

o   Technical standards represent the standard "tools" for building CALS database systems. These are currently the following:

    oo   IRDS ANS X3.138,

    oo   proposed IRDS Services Interface,

    oo   proposed IRDS Export/Import File Format,

oo   <u>SQL</u> for relational database management at a
     particular site, and

oo   proposed <u>Remote Database Access (RDA)</u> for executing
     a query at a remote site.

4. Using the IRDS to Meet Dictionary Requirements

A series of technical reports is being prepared at the National Institute of Standards and Technology to assist people in the use of the IRDS. General references include the following:

o A Technical Overview of the Information Resource Dictionary System (Second Edition), NBSIR 88-3700, by Alan Goldfine and Patricia Konig,

o Using the Information Resource Dictionary System Command Language (Second Edition), NBSIR 88-3701, by Alan Goldfine,

o Guide to Information Resource Dictionary System Applications: General Concepts and Strategic systems Planning, NBS SP 500-152, by Margaret Henderson Law, and

o The ICST-NBS Information Resource Dictionary System Command Language Prototype, NBSIR 88-3830, by Alan Goldfine and Thomasin Kirkendall.

Other reports, more specialized and still under development, include:

o "IRDS Support of the 3-schema Architecture," by Alan Goldfine,

o "IRDS Support of Data Modeling Methodologies," by Elizabeth Fong,

o "Storing Diagrams in the IRD," by Alan Goldfine, which addresses one aspect of the support of data modeling methodologies,

o "IRDS Support for DBMSs," by Alan Goldfine, and

o "Active Interfaces to the IRDS," by Alan Goldfine, which addresses the issues of how to control a DBMS and coordinate dictionaries.

5. Availability of the IRDS and Data Management Standards

The diagram on Timelines of Data Standards Needed by CALS and PDES is intended to demonstrate that data management standards and commercial implementations are expected to be available when they are needed to satisfy CALS requirements. SQL is evolving through SQL2 to SQL3, which is expected to be object-oriented and capable of effectively managing text, graphics, and other CALS data types. Remote Database Access (RDA) will provide a primitive capability for sending a query to a remote database and receiving data and status information in response.

	1989	1990	1991	1992	1993	1994	1995
LSAR	SQL Schema			LSAR/PDES Integration			
PDES	PDES Integration		PDES 1.0 Std.				
SQL	SQL Software		SQL2 Std. and Software				SQL3 Std.
RDA			RDA Std. and Software				
IRDS	Std.	Software	Services Interface and Export/Import Standards and Software				
	Information Modeling			Use of Information Models			

Timelines of Data Standards Needed by CALS and PDES

8

6. An Architecture for Distributed Database

RDA is insufficient for true distributed database management, since it does not deal with such issues as query decomposition (where a single query requests data from multiple databases) or transaction management (needed to resolve simultaneous accesses and updates to the same data). The following is a possible layered architecture for standards that would provide the required services:

6. DBMS syntax and presentation transparency

5. location and performance transparency

4. transaction transparency

3. communication transparency

2. mapping from global conceptual schema into local internal schema

1. database access

The users interact with Layer 6, which isolates them from details of names, formats, and DBMS syntax at the various databases. Each user seems to be dealing with a single, dedicated database specific to his or her requirements. IRDS support of a three-schema architecture is essential to map from the user's local external schema into the global conceptual schema. Layer 5 performs services such as query decomposition and optimization, in order to effectively deal with data distribution. Directory and schema information is essential. Layer 4 provides for detection and resolution of conflicts in accessing and updating data. Layer 3 handles the details of communicating with remote sites. RDA is positioned in Layer 3. Layers 2 and 1 provide services at a particular remote site to satisfy a particular piece of the original query or update. IRDS services are again required to provide support for the three-schema architecture. New standards will be required for Layers 6, 5, 4, and 3, while IRDS and SQL may require extensions to deal with Layer 2. Layer 1 provides operating system services that are identical for distributed or centralized databases.

7. Development Tasks

The following are some of the near-term tasks required for CALS Phase II:

o   Development of the vocabulary at the IRD Schema Layer to support methodologies for modeling the various applications.

o   Development of the vocabulary at the IRD Schema Layer to support modeling distributed databases.

o   Population of the IRD Layer with application models.

o   Population of the IRD Layer with distributed database models.

o   Interfacing Computer-Aided Software Engineering (CASE) tools to the IRDS, in order to assist in development and analysis of models.

o   Development of standard ways of using the IRDS to support integrated process and data modeling (e.g., through a standard based on IDEF0 and IDEF1X).

The following are some of the longer-term tasks required for CALS Phase II:

o   Completion of the architecture for distributed database.

o   Development of standards based on that architecture.

The conclusion is that the IRDS is a good basis for CALS Phase II, but a large amount of work is needed to develop other standards that will be required.

4. TITLE AND SUBTITLE

Use of the IRDS Standard in CALS

5. AUTHOR(S)

David K. Jefferson and Cita M. Furlani

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.)

The objective of this point paper is to show how the Information Resource Dictionary
System (IRDS) can fulfill critical design and operational requirements for CALS Phase II.
First, a series of assumptions are made about the data management services which are
needed by CALS Phase I'. Next, these assumptions are used to develop a series of
requirements for a dictionary system. The structure of the IRDS family of standards
is then described. Examples are provided to illustrate how the IRDS could meet the
requirements. A schedule is presented to show that the IRDS and other data management
standards will be available when needed to meet the immediate requirements of CALS.
An architecture is presented to illustrate additional standards required to achieve
longer-range goals of distributed database. Finally, development tasks are recommended.

12. KEY WORDS (6 TO 12 ENTRIES; ALPHABETICAL ORDER; CAPITALIZE ONLY PROPER NAMES; AND SEPARATE KEY WORDS BY SEMICOLONS)

Computer-aided Acquisition and Logistic Support (CALS); data modeling; distributed data
management; Information Resource Dictionary System (IRDS); Remote Database Access (RDA);
SQL