

JOINT DATA BASE ELEMENTS FOR MODELING AND SIMULATION

JAMES L. COLE
PETER VALENTINE

Electromagnetic Environmental Test Facility
U.S. Army Electronic Proving Ground
Fort Huachuca, Arizona 85613-7110, U.S.A.

ABSTRACT

This paper discusses the approach used in a proof of concept effort developing data elements and data models for specific applications in the modeling and simulation (M&S) functional areas of research and development, test and evaluation, analysis, and combat development. Specifically, this joint data base elements approach provides the M&S community with a reverse-engineering data modeling procedure using IDEF1X that is compatible with the Department of Defense Corporate Information Management initiative. This approach also includes the development of a data dictionary/directory, containing the data elements and their descriptions, information on M&S data bases, and a repository to support future developments.

1 INTRODUCTION

This paper presents the approach being used for the development of the Joint Data Base Elements for Modeling and Simulation (JDBE) project. The JDBE project promotes the effective interoperation of models and simulations by facilitating their interoperability and sharing of data through data modeling and standardization. JDBE attempts to increase integration and data sharing by providing a standard for the exchange of information in the M&S community. This effort includes developing standard definitions for data elements commonly used within the community and developing standard subject area information (SAI) models for information grouped by subject matter rather than by functional boundaries. JDBE also addresses mapping existing and future data bases to the standard elements and models, and maintaining a directory of the data bases identified during the course of the JDBE project. The JDBE project's foremost goal is to avoid duplication of effort with other ongoing standards efforts and the JDBE team will work with these efforts in a complementary fashion.

The JDBE project was proposed as a solution to the lack of common or standard data definitions in the M&S community. It is initially being sponsored by the Defense Modeling and Simulation Office as a way to address the current critical requirement for data interoperability among models and simulations.

For this proof of concept effort, data collection and data modeling are limited to one topic area, electromagnetic equipment. Since the JDBE information modeling methodology has been successfully demonstrated using the selected information area and a draft military handbook containing instructions for JDBE use has been produced, the information modeling effort can be expanded to include other information areas. Discussed in this paper are an overview of the JDBE process, analysis of the legacy systems, the JDBE integration scheme, legacy data mapping, development of a military handbook, the accomplishments completed to date, and conclusions.

2 JDBE METHODOLOGY

2.1 JDBE Approach

JDBE differs from many other standards efforts in that it uses a data-centered approach toward solving the problem of interoperability among existing systems. Many of the existing efforts, the Corporate Information Management (CIM) initiative for example, use both a process-oriented and data-oriented modeling approach. They study and model the processes in a system and use the data flows in their process models as a source of candidate entities for inclusion in a later data model. The approach used in the JDBE effort is significantly different. It focuses first on meeting the special data needs of the M&S community, while concurrently supporting the central CIM objectives.

A challenge with M&S is the richness of the data elements required. Because of the detailed, scientific nature of M&S, many more data elements are required than in traditional business-oriented data modeling. The

very nature of problem-solving in M&S involves discovering new data elements in an ad hoc fashion. A top-down approach to developing data requirements for M&S cannot identify all of the data elements required by a model because these tend to evolve during the course of development.

To meet these M&S peculiar requirements, JDBE is structured around a reverse-engineering approach to the problem. By approaching the M&S community requirements from the bottom up, we have focused our efforts on information that is already in use by the community. By standardizing this information, we can enhance the exchange of information within the community.

Because of this reverse-engineering focus on the data, the JDBE project requires very little process modeling. Unlike large standards efforts (CIM in particular), the JDBE role emphasis is to produce standard representations for the exchange of data. Compliance with JDBE standards will involve mapping existing legacy data elements to the standard data elements, and will not require the re-coding of legacy systems or the restructuring of existing data bases. Legacy data systems are existing data bases or model and simulation data sources. The candidate standard elements produced by the JDBE effort will be submitted to the CIM approval process, and the resulting approved elements mapped back to the legacy data elements.

JDBE is designed to be complementary to other data standardization efforts within the Department of Defense (DoD). The CIM effort is the DoD-wide standards program; however, its timeline for reaching the detailed data element level has been estimated to be 10 years out. Likewise, most of the other services' standards efforts are top-driven and are focused on data in support of managing the services and not on the detailed data used for M&S in support of research and development, test and evaluation (RDT&E) and combat development.

JDBE is working with the CIM effort to ensure that JDBE data models and data elements "fit" within the overall CIM effort. The primary reason for the formation of the JDBE project is to accelerate and integrate the development of standard data elements for M&S. This will result in immediate cost savings and increases in productivity and communication among the various DoD M&S efforts.

2.2 JDBE Process

JDBE begins by reverse-engineering legacy data systems and documents them using the IDEF1X methodology. As a number of project information models are stored in the JDBE repository, JDBE data analysts identify the commonalties in the data which determine subject areas. SAI modeling groups are formed to refine and validate

the SAI models. In the course of building the SAI models, new data elements will undoubtedly be discovered and proposed by the JDBE team as candidate elements to the Defense Information Systems Agency (DISA) effort. Once an SAI model has been validated, JDBE team members then map their local project information models to the standard, laying the foundation for the exchange of data with other compliant legacy systems. A final military handbook will be produced by the JDBE team that documents the process and contains an electronic appendix with the SAI models, standard elements, and data directory.

3 ANALYSIS OF LEGACY SYSTEMS

To document existing legacy data systems using a single common information modeling technique, we have chosen the IDEF1X information modeling methodology. IDEF1X was developed by the Air Force Integrated Computer Aided Manufacturing (ICAM) project. There were a number of reasons for selection of this particular methodology. Primarily, it is an evolving Federal Information Processing Standard (FIPS) that is currently used in support of the CIM effort. IDEF1X is a nonproprietary methodology in the public domain which is currently being adopted by National Institute of Standards and Technology (NIST) as a FIPS. It is a robust, mature, information modeling technique with a particular focus on relational data bases. Finally, there are some excellent, user-friendly tools available that aid the user in developing IDEF1X models.

IDEF1X offers a number of significant benefits to the JDBE project. First is its scalability from high-level to detailed modeling of systems. It supports multiple levels of abstraction of the information model. As a result, the methodology matches the reality of information modeling. IDEF1X permits the users to design/develop their information models in an incremental fashion, adding layers of detail as discovered.

In figure 1, taken from Bruce (1992), the various levels of IDEF1X models are shown in relationship to one another. The widely scoped models covering a large portion of the enterprise are called area models, and the narrowly scoped models are called project information models. IDEF1X differentiates between logical and physical models, where logical models exist independent of the physical dependencies of implementation and physical models are designed to address a particular physical implementation or system. IDEF1X is data centered, i.e., it focuses on data as a reusable asset, and the methodology supports viewing the information model from the corporate or business-wide perspective. IDEF1X supports the integration of multiple information areas or "views" under a single information model. This

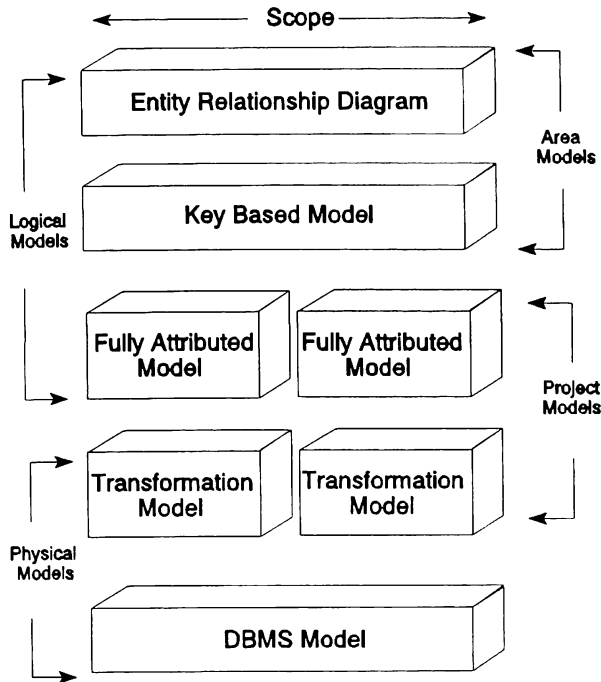


Figure 1: IDEF1X Models

integration approach is vital to the JD BE project and is reflected in JD BE's priority of integration with the DISA effort. It is also important to note the limitations of the IDEF1X methodology. It is a tool for modeling the structure of data and data relationships. Its focus is primarily on data and relational modeling and is not an object-oriented methodology. The need for an object-oriented modeling methodology, and the search for a nonproprietary, mature, object-oriented methodology continues. IDEF1X, however, allows modeling the 80 percent of the current legacy data systems which can be represented using relational information modeling. Practical rather than perfect solutions are the JD BE approach.

NIST is developing a standard definition of the IDEF0 and IDEF1X languages. The roots of the language come from the Air Force ICAM project, but the language has evolved considerably since then. The current draft NIST standard has been updated to reflect the current usage of the language and to open up IDEF1X to future enhancements.

Some of the potential additions to the IDEF1X language include the definition of a formal data modeling language to permit the modeling of complicated business rules and complex data domains. It is one small step from a data modeling language to modeling methods as

well as data, extending IDEF1X into the object-oriented arena.

The JD BE team, as members of the IDEF Users Group, has already been instrumental in updating the draft FIPS documents.

4 IDEF1X REVERSE-ENGINEERING PROCESS

The JD BE reverse-engineering approach starts with existing legacy data systems and strives to document those systems using IDEF1X. The first step in this process is to document the system, exactly as it exists, in the form of an IDEF1X data base management system (DBMS) model.

Level 1 - DBMS Model

An IDEF1X DBMS model is derived from an existing data base. This data base may be managed using a DBMS, but in many cases it is simply a set of structured files which contain data used in models and simulations. If the data are stored in a hierarchical DBMS, the data structure must be "flattened" to allow representation as a series of entities or tables.

After the physical model has been created, the detective work begins. The JD BE team member must interview experts and users of the legacy system, find and read whatever system documentation is available, and use this information to begin adding the logical components of the system to the model.

Level 2 - Project Information Model

The project information model seeks to extend the DBMS model into a logical picture of what the business requirements might have been when the system was developed. The DBMS model is normalized and, if possible, the business rules are added in order to transform the model to a logical statement of what is implemented.

Level 3 - SAI Model

SAI models comprise a logically grouped set of entities normally used together in the M&S community and represents a standard view of a particular subject area. Because of JD BE's reverse-engineering approach, subject area will be determined by analyzing the project information models and identifying commonly used information. Examples of subject areas are digital terrain data, weather data, and electromagnetic equipment characteristics, etc.

5 JD BE SCHEMA INTEGRATION

5.1 Overview

In accordance with the American National Standards Institute Standards Planning and Requirements Committee model, the three-schema view of data introduces the conceptual schema as a way of viewing data in an abstract way. This introduction of a conceptual step in the design methodology is a relatively recent development in computer science. It provides designers and developers with a high-level abstract view of their data, one which is independent of physical dependencies and the user view.

DBMSs are built around this conceptual view of the data, permitting an integrated view of the data across applications. During the design of a data base to manage an organization's data, the data base developers go through a process (consciously or unconsciously) of view integration. That is, the data base developer tries to integrate the various users' views of the data and the physical representations of the data into a single conceptual design, i.e., the data base schema. By doing this, the developer avoids duplication, problems of multiple updates, and inconsistencies in the data across applications.

The government and other large agencies suffer from incompatibility between data bases. It is the nature of large organizations to have large numbers of distributed data bases containing the data of use to the organization. The conceptual designs differ from one data base to the next, making the transfer of information between data bases difficult, costly, and time consuming. Accordingly, a methodology is required which will permit the integration of the various data base schemas.

Data base schema integration is a relatively recent problem which has surfaced in the context of distributed data bases. A distributed data base is a collection of data that logically belongs to the same system but which is physically located in a number of remote sites. Distributed data bases can be classified as homogeneous, where local data bases have the same data model and DBMS, and heterogeneous, which have different schemas and DBMSs.

This methodology will address the creation of a semantic data model which will facilitate the integration of multiple local data base schemas into a global schema. Local data models would then be mapped into the global schema to serve as a guideline for the transformation of data between distributed, heterogeneous nodes. Finally, the global schema could be used by a heterogeneous data management facility to map the requests of users from the semantic data model into the actual data bases.

5.2 Steps in the Integration Process

Preintegration

Analysis of the schemas to be integrated in order to determine the choice of schemas to be integrated, the order of integration, and assignment of preferences.

Comparison of the Schemas

Schema are analyzed and compared to determine correspondences and detect conflicts.

Conforming the Schemas

Resolution of conflicts. Automated conflict resolution is generally not possible; compromises must be worked out with experts.

Merging and Restructuring

Schema are merged and intermediate schemes are created and tested.

6 JD BE DATA BASE SCHEMA INTEGRATION AND CIM

The JD BE methodology for data base schema integration fits well with the CIM and other agency data standardization efforts like the Army AR 25-9. The approach is a bottom-up effort to achieve a working set of global schemas called SAI models. These SAI models formalize the development of global schemas, a process which is already occurring informally in the development of "super" data bases like those created by DIA, ECAC, TADS, and others. Super data base developers lump together a large number of smaller, existing data bases to form large super data bases, in effect creating a global schema. This is, however, an informal process which does not guarantee the usability of their newly created super data base by others in the government outside their original project's intent.

The JD BE approach is to provide a formal methodology for the creation of global schemas to share data between data bases, and to provide the tools and training needed to accomplish this. In addition to the creation of formal global schemas (SAI models), the JD BE project complements the ongoing data element standardization efforts like CIM and AR 25-9 by consolidating data elements and resolving conflicts at the local level. JD BE places emphasis on community involvement and resolution of conflict at the individual work group level so that the data elements submitted to AR 25-9 and CIM have already had many of their conflicts resolved by the experts in the field. See figure 2.

JDBE Proposed Methodology

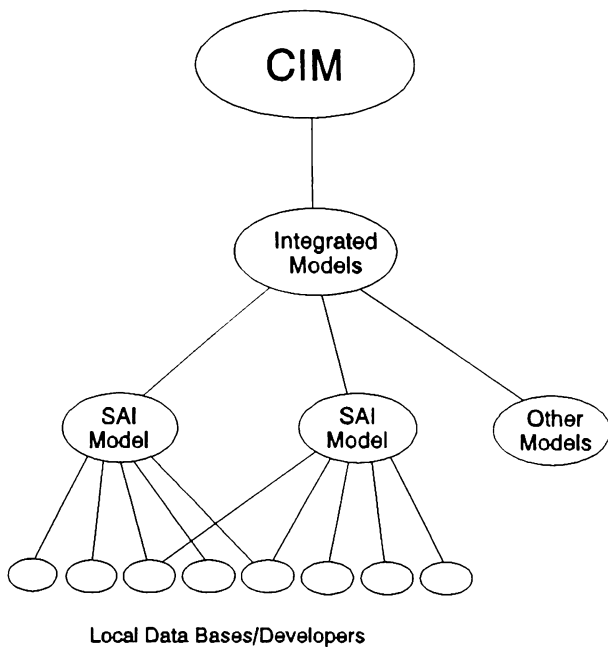


Figure 2: Inputs and Outputs of Schema Integration

The advantage of this approach is to make the process of integrating existing data bases into a developing standard smoother and reduce the load on the standards implementors. The primary thrust of JDBE is to facilitate the involvement of the community in developing standards which ease the transition from heterogeneous nonstandard data bases to a richly interconnected distributed data base environment of the future.

7 LEGACY MAPPING

7.1 What is Legacy Mapping?

After SAI models have been identified and standard data elements have been defined for all attributes in the information model, JDBE team members can then map their existing project information model to the SAI model. By identifying which of their model's attributes correspond to attributes in the SAI model (i.e., standard data elements), they can build a map of how to transform their data into the standard format. Conversely, they are also building a map of how to transform data stored in standard format into their legacy systems format. Individual attributes of entities in the project information model are mapped to corresponding attributes (standard

data elements) in the SAI model. If another project's attribute is mapped to the same SAI attribute, then the two are considered to be equivalent. For example, as shown in figure 3, Rdr_Freq equates to Radar_Tuned_Frequency. CurrFreq also equates to Radar_Tuned_Frequency; therefore, Rdr_Freq and CurrFreq are considered equivalent.

Equivalency, however, is not enough to establish the mapping between attributes. CurrFreq could be represented as a Long Integer with units of hertz, where Rdr_Freq might be a floating point number with units of megahertz.

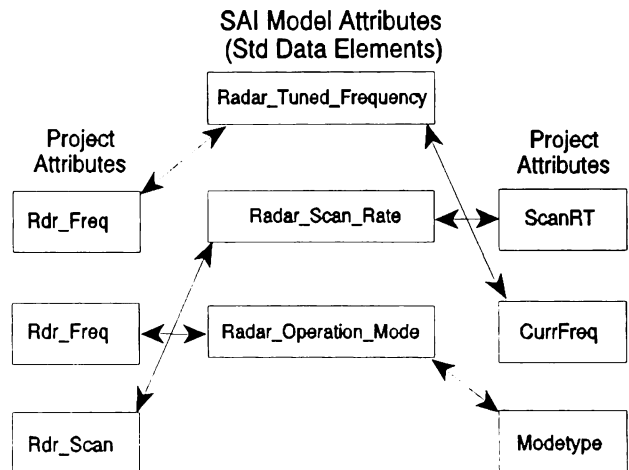


Figure 3: Example of Data Mapping

7.2 Data Conversions

When the data map between project information models and the standard SAI model has been fashioned, it is theoretically possible to build/provide an automated tool that can transform data from one format to another. While this capability is not envisioned as being provided during the initial phases of this project, it will be addressed by the JDBE project, and the feasibility of using commercial off-the-shelf software or developing custom software to perform this function will be evaluated.

7.3 Data Mapping Methodology

A formal methodology for defining the mappings between data elements is currently being developed by the JDBE team and will be refined and tested during this first phase of the JDBE project.

8 DEVELOPING A MILITARY HANDBOOK

8.1 Scope

The scope of the JDBE project encompasses all data used in support of the RDT&E and combat development process. The draft military handbook developed by this project consists of two parts, the body of the handbook and appendixes containing the meta-data.

The body of the handbook describes the JDBE methodology and where it applies and lays out the process by which legacy systems are identified and modeled. It also discusses how the JDBE project is a highly focused effort to model existing legacy systems and map these systems into the standards being developed by the CIM program. Finally, the handbook discusses the process for developing SAI models and how these models can be used to generate candidate CIM standard data elements.

8.2 Appendixes

The handbook contains appendixes which are available in both hard-copy and electronic form. The electronic version lends itself to being updated more often than the hard-copy version and should be the preferred source for the latest SAI models and legacy system information.

The data dictionary contains the data elements, prime elements, and generic elements used and referenced by the SAI models. The data directory also contains information about existing data bases, models and simulations, and associated points of contact.

NOTE: Due to the large storage requirements, the direct inclusion of project information models and mappings may be restricted to special requests, or possibly distributed using CD-ROM media only. In any event, all information will be available by accessing the JDBE repository directly.

9 ACCOMPLISHMENTS AND CONCLUSIONS

We have completed the methodology for implementing the reverse-engineering process, the SAI model for electromagnetic equipment, and a draft version of the military handbook. Also included are the data dictionary, with elements from electromagnetic equipment data model, and the data directory, populated with the initial set of data base information. We have also demonstrated data mapping between the project models and the SAI models. These methods may also be applied to other areas of mutual interest within the M&S community.

ACKNOWLEDGMENTS

The author thanks Mr. Stephen Matsuura, Ms. Janet McDonald, and Mr. Rafael Anton of the Electromagnetic Environmental Effects Division, U.S. Army Electronic Proving Ground, for their contributions.

REFERENCE

Bruce, Thomas A. 1992. *Designing quality data bases with IDEF1X information models*. New York: Dorset House Publishing.

AUTHOR BIOGRAPHIES

JAMES L. COLE is Chief of the Electromagnetic Environmental Effects Division of the U.S. Army Electronic Proving Ground, Fort Huachuca, Arizona. He received his B.S. degree from Texas A&M University in 1975. His research interests include electromagnetic propagation modeling and the application of simulation modeling to test and evaluation.

PETER VALENTINE is a systems analyst for the Electromagnetic Environmental Effects Division of the U.S. Army Electronic Proving Ground, Fort Huachuca, Arizona. His research interests include data base design, data modeling, and software development.