



Whitemarsh
Information Systems Corporation

*Data Management Program:
Requirements for
Components, Costing and Schedule*

Whitemarsh Information Systems Corporation
2008 Althea Lane
Bowie, Maryland 20716
Tele: 301-249-1142
Email: Whitemarsh@wiscorp.com
Web: www.wiscorp.com

Table of Contents

Acknowledgments	iii
1.0 Architecture of the Data Management Program	1
2.0 Adherence to Frameworks	1
3.0 Adherence to Net Centric Data Goals	1
4.0 Components	1
4.1 Methodology	3
4.2 Metadata Repository System	4
4.3 Courses, Seminars, Workshops, Books and White papers	7
5.0 Operational Environment	8
5.1 Metadata Repository System	8
5.2 Evolution and Maintenance	8
5.3 Technical Support	9
5.4 Training and Awareness	9
6.0 Costing	9
7.0 Schedule	10



Acknowledgments

This material is an evolution of documents that were updated during the time frame: September 2003 through December 2004. The primary contributors were Bruce Haberkamp, James Blalock, and Michael Gorman of the Office of the CIO, United States Army. The foundational components of this work has been favorably reviewed by subject matter experts within the U.S. Department of Defense.



1.0 Architecture of the Data Management Program

The hallmark of a quality data management program is that data in either of its “real” or many metadata forms is able to be defined once and used many times. The overall goal of an enterprise data management program is to establish a data environment such that data and its supporting metadata is defined in the most efficient and cost effective manner possible throughout the enterprise. This program is squarely positioned within the net-centric environment.

A key characteristic of a quality data is that it is interoperable. Thus, organizations and systems are able to exchange data without any loss of meaning and precision. Because data interoperability is so critical to the data management program, focus to date has concentrated almost exclusively on establishing the data management program exclusively within the requirements of data interoperability.

In support of achieving the goals and objectives of the data management program there must be an overall data methodology and supporting computing environment as well as courses, seminars, workshops, and technical assistance that fully supports the data management effort.

2.0 Adherence to Frameworks

This data management program is set squarely within the Knowledge Worker Framework. This framework is specially designed for the knowledge worker and was specially designed to meet the many reasons for IT system failures enumerated from analyses of large scale IT system failures that were performed by the U.S. General Accountability Office (GAO).

3.0 Adherence to Net Centric Data Goals

The seven net-centric goals that relate to data are:

- Make data visible
- Make data accessible
- Institutionalize data management
- Enable data to be understood
- Enable data to be trusted
- Support data interoperability
- Be responsive to user needs



4.0 Components

The components of the data management consist of the following:

- Methodology
- Metadata repository system
- Courses, seminars, workshops, books and white papers

The data management program methodology is fundamentally data driven. Process models, as implemented through information systems exist within the context of data, that is, databases. Databases exist within the context of missions. Databases are not just inert, that is, they can contain data structures with embedded processes. Hence database objects.

In short, information systems, which are event triggered by the various functions performed by organizations cause changes in the database objects in support of the accomplishment of various missions.

This six part overarching set of models, that is, missions, database objects, information systems, events, functions and organizations form the basis for the methodology. These six models fully support the net-centric data goals and fully satisfy the requirements of the Knowledge Worker Framework..

The set of all artifacts that represent these six models across the enterprise are captured into, interrelated, updated, and reported through the metadata repository system. The metadata repository system is a database centered information system that both acts as a repository for the artifacts from these six models and also as a system through which these six models are actually created, employed, evolved, and reported. The metadata repository system is engineered to operate in a multi-user, shared database environment.

The methodology and the metadata repository system are conveyed through a series of courses, seminars, workshops, books and white papers. For example, the workshop that discovered and defined shared data databases empowered attendees to import database schemas of their legacy systems, and, armed with a definition of their shared data requirements that are set within their organizational mission functions to discover shared data and to create one or more shared data structures. The workshop then enabled the attendees to create a database about these shared data structures and to generate a software application that used this shared data database.

4.1 Methodology

The data management methodology consists of a number of integrated methodologies. The set of work products created, evolved, or maintained in each methodology conforms to the artifacts of the metadata repository system. The six classes of database architectures in the data management environment that are the subject of the methodologies are:



- Reference databases (Authoritative Data Sources)
- Original data capture databases
- Transaction data staging area databases (Information Exchange Standards Specifications (IESS))
- Subject data store databases
- Warehouse databases

The various methodologies that cause the creation and/or maintenance of these database architecture classes are:

ADMINISTRATIVE AND MANAGEMENT PROJECTS

- Administrator Documentation
- DBMS Selection and Evaluation
- Repository Selection and Evaluation
- Enterprise Management Audit
- Information Systems Plan
- Repository Development
- Standard Estimation

SPECIFICATION PROJECTS

- Conceptual Specification
- Database Process Model
- Data Element Development
- Detailed Data Model
- Functional COTS Selection
- Functional Prototype
- High-level Data Model
- Impact on Enterprise Model
- Implementation Strategy
- Mission Model
- System Control Requirements

IMPLEMENTATION PROJECTS

- DBMS Database Specification and Implementation
- DBMS Schema and View Development
- Functional Area Implementation
- Functional COTS Implementation
- Interrogation Development
- System Control Implementation
- System Documentation
- Testing

OPERATION AND MAINTENANCE PROJECTS

- Emergency Maintenance



- Repository Evolution
- Standard Maintenance
- Application Optimization
- System Evolution

MULTIPLE CATEGORY PROJECTS

- Data Collection and Validation Systems
- Data Conversion
- Training
- Warehouse

4.2 Metadata Repository System

The metadata repository system is a database centered software system. Given the multi-user nature of this system, the key functional requirements are that:

- The data environment infrastructure supports minimum redundancy and maximum reuse of all the components that comprise the managed interoperable data environment infrastructures.
- The data environment infrastructure is engineered to support the identification of common business facts, concepts, processes, business rules, value domains, and data structures across the providing or consuming business information systems.
- The data environment infrastructure supports the identification, management and integration of common business facts including their meanings, value domains, and other supporting semantics across all uses of these facts.
- The data environment infrastructure is engineered to manage value domains employed exclusively within single business facts or across multiple business facts, and must manage the evolution value domains including value domain transformation mappings.
- The data environment infrastructure is engineered to identify, manage, and deploy standard data structures of standard business facts.
- The data environment infrastructure is engineered to support the discovery of redundant and/or conflicting business information systems that generate or process interoperable data.
- The data environment infrastructure is engineered to support the identification, exposition, and interrelationship of data generation and/or use within the operational schedules of business information systems included within the interoperable data environment infrastructures.



- The data environment infrastructure is engineered to identify, manage, and integrate authoritative sources of data for standard data structures.
- The data environment infrastructure is engineered to map standard data structures, business facts, and value domains across variant definitions and deployments of those data structures, business facts and value domains.
- The data environment infrastructure is engineered such that the generation of XML schemas and XSLTs is automatic.
- The data environment infrastructure is engineered to support the importing and exporting of SQL schemas for relational databases.
- The data environment infrastructure is employable by proponent developers within the existing work efforts of quality-engineered database projects.
- The data environment infrastructure must be employable as an integrating mechanism within communities of interest, and to identify conflicts across communities of interest, and across the enterprise..
- The data environment infrastructure supports the identification and management of projects that are creating and/or evolving aspects of the data environment infrastructure.
- The data environment infrastructure supports uploading of work products to appropriate repositories.
- The metadata environment is contained within a SQL database of explicit tables so that the metadata can be accessed, explored, and reported through data mining, report writers, and Internet publishing tools.
- The metadata environment supports access through a variety programming languages which, in turn, can be employed to create embedded metadata processing rules that control and/or govern the fundamental processes of the metadata repository system.
- The metadata environment is not within proprietary data formats, and when the metadata is exported it must be in a SQL “values” format on a table by table basis.

Beyond these core mandatory functions, the following additional functionality are also supported:

- **Administrative information.** Enables the association of organization, necessary creation and modification dates, and if appropriate, person with metabase artifact trees.



- **Authoritative data sources.** Enables the identification of the source for a particular class of information such as reference tables.
- **Automatic definition creator.** These are for data elements, attributes, columns, DBMS columns and view columns so that no intellectual effort need be expended beyond the definitions associated with the constituent parts of these metabase artifacts.
- **Automatic name abbreviator.** This is necessary for DBMS columns and view columns based on name length parameter for those DBMSs that have a restricted name length.
- **Business Rules** (a.k.a. data integrity rules). Enables the central definition and allocation of data integrity rules to each other and to appropriate metabase artifacts.
- **Business Terms.** Enables the definition and allocation of business terms to each other and to appropriate metabase artifacts.
- **Configuration management.** Enables the isolation of development from test, and test from production. There will be multiple production model instances and one development and test instance.
- **metadata repository system Project management.** Complete integration of project management and project deliverables, tasks, and earned value reporting.
- **Internet Presentation Layer.** Provides access to the metadata repository system database through the Internet for reporting purposes.
- **Reference data** Forms the basis for establishing, tracking changes to, and cross referencing all restricted value domains including the organizations that establish and maintain them.
- **Transactions.** Enables the formalization of standard transactions that are to form the basis of data interoperability.

High level descriptions of the functional model of the metadata repository system is provided within its documentation. Broadly, the functions of the metadata repository system enable its users to:

Forward engineer by building, employing, interrelating and/or maintaining:

- Mission Models
- Function Models
- Organization Models
- Database Domain Models
- Conceptual Data Models



- Database Object Classes
- ISO 11179 Data Elements and Supporting Metadata
- Logical Data Models
- Physical Data Models

Or Reverse Engineer legacy environments into a modern interoperable environment by:

- Discovering legacy metadata
- Importing metadata into physical data model
- Reverse engineering to build the logical data model layer
- Reverse engineering to build the conceptual data model layers
- Building the data element layer
- Forward engineering to build standard-data based applications

4.3 Courses, Seminars, Workshops, Books and White papers

The data management program is conveyed through a series of courses, books, white papers, seminars and workshops. The major categories for these materials are:

- Data Quality
- Database Design
- Database Objects
- Database Projects
- DBMS
- metadata repository system
- Enterprise Database
- Process Modeling

A listing and brief descriptions of these materials is contained in Attachment 7.



5.0 Operational Environment

The operational environment for the data management program is provided as follows:

- Metadata repository system
- Evolution and Maintenance of the metadata repository system and supporting materials
- Technical Support
- Training and Awareness

5.1 Metadata Repository System

It is projected that the metadata repository system operated within a multi-user environment.

5.2 Evolution and Maintenance

The metadata repository system is conveyed to the enterprise under a limited use license. That is, the metadata repository system and all supporting materials are able to be used by any personnel and/or contractors and other organizations directly involved with enterprise programs. Further, the metadata repository system is able to be delivered in source code form such that it can be modified by qualified personnel and/or contractors. The SQL DBMS upon which the metadata repository system database is constructed is delivered to the server environment.

Metrics must be provided that can be used to predict the time and cost for metadata repository system evolution.

5.3 Technical Support

The metadata repository system is to be supported through a website that allows Email questions, interactive dialogs, remote access of client machines, and through direct telephone contacts. There shall be one technical support staff per 15 installed servers.

5.4 Training and Awareness

All training and awareness the materials exists as book, paper, course, or workshop. The length in class days or hours is provided along with an enumeration of any required audio visual supports. Student materials are available in a form so that they can be reproduced and maintained. If specialized software is required to maintain any training materials then one copy of that software is provided. Prices are provided for additional copies of the software. Training is available either through collaborative facilities or through on-site classes.



6.0 Costing

The costing must provided as follows:

- Cost per server for the metadata repository system
- Cost to provide the metadata repository system
- Cost to develop the courses, books, seminars, and workshops.
- Costs for copies of any integrated development software environments for the metadata repository system, the SQL DBMS for the metadata repository system database, and for any of the training and awareness materials.
- Cost to provide technical support for the metadata repository system.
- Cost to evolve and maintain the metadata repository system.
- Cost to provide training.



7.0 Schedule

The scheduling information that must be provided is as follows:

- 90 days to provide the core functionality of the metadata repository system.
- 120 days to provide the extensions to the metadata repository system.
- 60 days to be able to provide the training and awareness materials and to begin training.

