Meta Data Management for Business Intelligence Solutions

IBM’s Strategy

Data Management Solutions White Paper
Abstract

This paper presents IBM’s strategy for managing meta data for Business Intelligence solutions. The paper covers meta data concepts and reviews IBM’s strategy. Also covered is an overview of how IBM’s Visual Warehouse solution manages meta data.
Notices

References in this publication to IBM products, programs, or services do not imply that IBM intends to make these available in all countries in which IBM operates. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any of the intellectual property rights of IBM may be used instead of the IBM product, program, or service. The evaluation and verification of operation in conjunction with other products, except those expressly designated by IBM, are the responsibility of the user.

IBM may have patents or pending patent applications covering subject matter in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:

IBM Corporation
IBM Director of Licensing
208 Harbor Drive
Stamford, Connecticut 06904
U.S.A.

Version 1: November 8, 1996
Trademarks

DataAtlas, TeamConnection, VisualAge, and Visual Warehouse are trademarks or registered trademarks of International Business Machines Corporation in the United States and/or other countries.

Other company, product, and service names used in this publication may be trademarks or service marks of others.
Contents

Abstract ................................................................. ii
Notices ........................................................................ iii
Trademarks ................................................................. iv
Contents ....................................................................... v

Introduction ............................................................... 7
What is Meta Data? ..................................................... 7
Why is Meta Data Important? ....................................... 8
Integrating Meta Data - Centralized vs.
Decentralized Approaches ........................................... 9

IBM’s Business Intelligence Meta

Data Interchange Strategy ............................................ 11
Decentralized Approach to Meta Data Integration .......... 11
Support for Both Technical and End-user Meta
Data Using the Visual Warehouse Information
Catalog ........................................................................ 11
Visual Warehouse Information Models ............................ 12
Visual Warehouse Support of Open Interfaces and
APIs for Meta Data Interchange .................................... 12
IBM’s Meta Data Interchange Language, the Meta
Data Interchange Specification, and IBM’s VisualAge
Exchange product ....................................................... 12
UML and XML ............................................................ 13
OMG Committee for Common Warehouse Meta Data ........ 14
OMG Committee for XML/XMI .................................... 14
Visual Warehouse Automation of Meta Data
Interchange .................................................................. 15
Visual Warehouse as a Meta Data Integration Hub ............ 15

Meta Data Integration Between IBM’s

Visual Warehouse and Partner

Products ................................................................. 16
Integrating ETML Tools for Building and
Managing Business Intelligence environments ............ 16
Integrating Tools for Analyzing Information ................. 18
Brio Technology ........................................................ 19
Business Objects ......................................................... 19
Cognos ................................................................. 19
Hyperion Solutions .................................................... 19
Meta Data Integration Between IBM’s Visual Warehouse, IBM’s VisualAge DataAtlas Multiplatform, and PLATINUM’s ERwin

Summary .............................................. 23
Appendix A ............................................. 24
Additional Information .............................. 25
Related White Papers ................................. 25
Introduction

What is Meta Data?

Being able to consolidate and analyze your data for better business decisions can often lead to competitive advantage, and learning to uncover and leverage these advantages is what business intelligence is all about. All of this is possible if you have the right applications and tools to analyze data, and more importantly, if the data is prepared in a format suitable for analysis. Setting up this infrastructure for business intelligence solutions requires integrating a variety of tools -- tools to build and manage the infrastructure, databases to store the data, and tools for analysis. Integration can be facilitated by sharing certain types of information (meta data) across the components of the business intelligence environment.

Meta data, or information about data, provides administrators and business users with descriptions of the data or informational objects that they can access. There are two types of meta data: technical meta data and business meta data. Both types of meta data are important in building, maintaining, and using a datamart or data warehouse.

**Technical meta data** is used by administrators and software tools and provides the technical descriptions of data and operations. Technical meta data includes information about source data, target data, and the rules that are used to extract, filter, enhance, cleanse, and transform source data to target data. Technical meta data could be created by a relational database management system (e.g. database statistics), by warehouse and transformation tools (e.g. descriptions of transformations), by the warehouse manager (e.g. schedules), etc.

**Business meta data** is used by business analysts and end users and provides a business description of informational objects. It assists end users in locating, understanding, and accessing information in the datamart, data warehouse, or other informational sources. Business meta data might include the calculation used to create a particular value, the date and time that a report was created, or a description of the approval state of the company forecast.
Business meta data not only applies to the data in the warehouse but equally to information about a broader class of informational objects such as a graph or chart viewed through a presentation tool, a query or report returned from a decision-support or OLAP tool, or a Web page retrieved from the Internet. For example, a business information directory could contain meta data about all the reports having to do with sales in the southeast region.

Why is Meta Data Important?

Meta data helps to achieve two major objectives:

1. **It provides a means to improve the productivity of administrators and the reliability of solutions**

   Many components are used when building a system supporting Business Intelligence. They include database management systems, modeling tools, transformation tools, process managers, data mining and decision support tools, etc. These tools leverage many different platforms, data formats, and vendor suppliers. And each tool typically has its own meta data store and administrative interface. Making a mix of tools work together, as part of a data warehouse process, requires passing data and meta data between them. Administrators must create and maintain such data bridges. Often, identical information (i.e., meta data) must be fed to multiple tools, creating meta data consistency problems. Thus, meta data interoperability between tools simplifies integration tasks for administrators.

2. **It provides a means to assist business analysts and users in locating and understanding data.**

   Users need to understand where a given piece of information comes from -- its lineage or genealogy -- and how current the information is. For example, how current is sales data? Or, what calculations were used to derive profit? Making users self-sufficient is critical given the growth in the number of users accessing datamarts or data warehouses. Otherwise, deployment will become gated by the bandwidth of the help desk staff to answer user’s questions about finding or understanding information available to them.
Integrating Meta Data - Centralized vs. Decentralized Approaches

If administrative efficiency is the objective, then a means must be found for tools from different vendors to share common information. However, this can be a very difficult task. Each tool defines the same object in a different manner (the tool’s information models are different) and therefore the objects are not easily shareable. Furthermore, the same meta data is redundantly defined and stored in many meta data stores - data dictionaries, repositories, encyclopedias, database catalogs, copy libraries, etc., each with a different Application Programming Interface (API). Typically, the tools cannot exchange meta data among themselves because the meta data is stored in a proprietary format understood only by a specific tool. As a result, changes to meta data in one meta data store are not easily changed in the others.

There are two primary approaches to meta data integration: centralized and decentralized. In a **centralized approach**, each tool stores and accesses meta data using a central repository. This is an appealing approach for most customers as it can simplify tool integration and operations. However, it is generally problematic because:

- Each tool has different private meta data that must be stored to support unique product features. Replacing a tool’s proprietary meta data store is a costly undertaking with little added competitive differentiation to warrant the investment.
- Tool meta data is continually evolving so a complete centralized information model of all relevant meta data is difficult to achieve.

This approach has only been successful where a set of products or components from a single vendor have been designed to work together **from the beginning**.

With a **decentralized approach**, individual tools use portions of a common model definition and then exchange this meta data using an agreed-to interchange language. In this approach, a central repository can still fulfill the role of an information source but the repository is not a component upon which other tools rely during execution mode.

- This approach provides the widest range of tool interoperability, allowing each tool to be autonomous with its own proprietary meta data store while still sharing common common meta data.
• This approach can be used between two or more tools, allowing for simple to complex scenarios as to the number of tools involved.

• Each tool can be used “as is” without any need to replace a tool's existing meta data store.

• Because the tool is not changed, execution and performance should not be affected.

The decentralized approach facilitates the integration of tools from multiple vendors with lower cost and more flexibility ... and therefore, with greater likelihood of success. This approach creates the need for meta data synchronization.
IBM’s Business Intelligence Meta Data Interchange Strategy

Following are the key elements of IBM’s Business Intelligence meta data interchange strategy:

Decentralized Approach to Meta Data Integration

Since it is unlikely that a totally new set of tools can be developed which would satisfy all the flavors of Business Intelligence one must assume that specific customer solutions will require different tools, some from different vendors. Therefore, for Business Intelligence, IBM has chosen the decentralized approach to meta data integration. Here, each tool can be autonomous with its own proprietary meta data store, however, it shares common meta data, that is, it is part of a federation of meta data stores. In essence, the approach consists of a meta data interchange facility as well as a meta data integration hub used to validate the objects being interchanged and provide specialized services in a Business Intelligence environment.

Support for Both Technical and End-user Meta Data Using the Visual Warehouse Information Catalog

The Visual Warehouse Information Catalog contains meta data for both technical and business users.

• Administrators are able to use the meta data to help manage the data warehouse or datamart.

• Business analysts and end users are able to use the meta data to identify available information in a business context.

Typically, data warehouse products collect meta data for the use of the warehouse administrator. This meta data is neither available nor oriented to the business user. The Visual Warehouse Information Catalog supports meta data both for the administrator and for the end user. End-user oriented meta data is made available through an interface tailored specially to end users, including the ability to navigate and search using business terms, to present data lineage and currency information, and to automatically launch applications associated with the data. Additionally, the Visual Warehouse Information Catalog spans a breadth of informational objects allowing web pages, spreadsheets,
presentations, and other objects to be represented along with information about data in the datamart or warehouse.

Visual Warehouse Information Models

Interchange of Business Intelligence meta data is significantly enhanced if a common interchange information model is used to avoid private interchanges between pairs of tools.

IBM provides externally published information models (used by the Visual Warehouse Information Catalog) for relational tables, file records (including DL/I segments), multidimensional cubes, reports, spreadsheets, pictures as well as a process and lineage model. The Visual Warehouse Information Catalog metamodel allows easy creation of additional information models. Using this extensibility feature, IBM has developed support of the Meta Data Interchange Specification (MDIS) models defined by the Meta Data Coalition. See http://www.MDCinfo.com for more information about the Meta Data Coalition.

A metamodel consists of the constructs used to develop information models. The Visual Warehouse Information Catalog uses its own generalized metamodel to allow administrators to describe new information models. Once defined, the Visual Warehouse Information Catalog will create the necessary tables used to store the new objects.

Visual Warehouse Support of Open Interfaces and APIs for Meta Data Interchange

Interchange is significantly enhanced if a common interchange syntax is used to avoid private interchanges between pairs of tools.

IBM’s Meta Data Interchange Language, the Meta Data Interchange Specification, and IBM’s VisualAge Exchange product

The Visual Warehouse Information Catalog is designed to integrate with a wide variety of products. It comes with extraction technology for a variety of database management systems and end-user tools. Source and target meta data can be imported directly from RDBMS catalogs. Additionally, Visual Warehouse can exchange meta data with any system that conforms to IBM’s Meta Data Interchange Language, the Meta Data Interchange Specification (MDIS) adopted by the Meta Data Coalition, or IBM’s VisualAge Exchange product. See Appendix A for a list of products with which the Visual Warehouse Information Catalog can exchange meta data.
Also, the Visual Warehouse Information Catalog has an externalized C type functional API which is mapped to a series of relational tables. Upon request, IBM will provide the model of these relational tables so they can also be accessed via SQL.

The following diagram illustrates the meta data flows just discussed.

![Diagram illustrating meta data flows](image_url)

**UML and XML**

**UML** or Unified Modeling Language is designed to give application developers a common language for specifying, visualizing, constructing, and documenting distributed objects and business models. Data Warehouse meta data models can be designed and visually represented in UML. For example models could be defined for warehouse constructs such as a data transformation, an OLAP schema, or an extract process.

**XML** or Extensible Markup Language is an open standard of the World Wide Web Consortium (W3C) designed for defining, validating, and sharing document formats on the Web. In a data warehousing context, one could exchange instances of UML models using XML as the interchange language.
As these standards gain widespread usage, IBM can be expected to use UML for representing warehouse meta data models and to use XML for exchanging instances of meta data represented in those models between tools. For example, IBM could enhance IBM’s Team Connection repository to store a warehouse meta data model and support exchange of instances of warehouse meta data using XML. Visual Warehouse could be enhanced to exchange meta data instances with any source that supports XML interchange via a common data type definition.

**OMG Committee for Common Warehouse Meta Data**

IBM, in conjunction with Oracle and Unisys, is sponsoring an OMG (Open Management Group) subcommittee for the standardization of Common Warehouse Meta Data. The objectives of this committee are to establish an industry standard for common warehouse meta data interchange and to provide a generic mechanism that can be used to transfer a wide variety of warehouse meta data. The intent is to define a rich set of warehouse models to facilitate the sharing of meta data, to adopt open API’s (Java and Corba) for direct tool access to meta data repositories, and to adopt XML as the standard mechanism for exchanging meta data between tools. The subcommittee, chaired by IBM, has requested vendor proposals for the above objectives.

**OMG Committee for XML/XMI**

A related OMG subcommittee has been formed to standardize XML Meta Data Interchange (XMI). IBM, Unisys and other industry leaders are also involved in this work. IBM and Unisys have submitted a proposal co-submitted by Oracle, DSTC, and Platinum Technology and supported by numerous other vendors. The proposal for an XML Meta Data Interchange Format specifies an open information interchange model that is intended to give developers working with object technology the ability to easily interchange meta data between modeling tools and between tools and meta data repositories. In a data warehousing context, the proposal defines a stream-based interchange format for exchanging instances of UML models.
Visual Warehouse Automation of Meta Data Interchange

Maintaining independent meta data stores introduces the risks that the meta data stores will become inconsistent. IBM addresses this problem by automating meta data interchange within IBM’s Visual Warehouse solution. For example, Visual Warehouse can automate the scheduling of updates for:

- Definitinal meta data such as the addition of a new column to a table
- Operational meta data such as time stamp information about the refresh of warehouse data
- OLAP meta data such as information about a multidimensional database


Visual Warehouse as a Meta Data Integration Hub

Although meta data interchange can occur directly between pairs of tools, the interchange is significantly enhanced if the interchange goes through a meta data integration hub. The primary role of the of the hub is to parse and validate the meta data being interchanged, reconcile it with other meta data, and detect name collisions. The hub also provides export/import functions from/to itself in a similar fashion as the participating tools do. In general, the hub assumes the role of “master” for the shareable objects being interchanged. The ideal choice for a meta data integration hub is Visual Warehouse as it fulfills the requirements just discussed.
Meta Data Integration Between IBM’s Visual Warehouse and Partner Products

IBM has partnered with several companies with the objective of providing an integrated and scaleable environment for Business Intelligence solutions. At the core of these partnerships are development projects to provide integration between Visual Warehouse and partner products. These partnerships cover two principal areas -- ETML (Extract, Transformation, Move and Load) tools and Analysis tools.

Integrating ETML Tools for Building and Managing Business Intelligence environments

To prepare data for analysis, it is usually extracted from a variety of sources, transformed (e.g., summarized, derived, cleansed, etc.), then moved and loaded into a data store suitable for analysis. IBM’s Visual Warehouse has specific integration points with products from Evolutionary Technologies International Inc. and Vality Technology Inc.

- ETI•Extract, a tool suite from Evolutionary Technologies International, Inc. extends Visual Warehouse capabilities in the area of data extract and transformation generation.

- Vality Integrity, a data re-engineering tool from Vality Technology Inc., extends Visual Warehouse in the area of data quality.

The following diagram shows an administrator’s view of flows of information between Visual Warehouse, ETI•Extract, and Vality Integrity.
Note: there are several possible scenarios using Visual Warehouse, ETI-Extract, and Vality Integrity. In the scenario depicted above, several meta data related activities occur:

- A1 - a data conversion specialist using ETI-Extract registers the “stage 1” data conversions with Visual Warehouse. As each data conversion is registered, meta data is imported into Visual Warehouse that defines business views which will invoke the data conversion programs. Also imported into Visual Warehouse is meta data that describes the data conversion source(s) and target(s), and transformations. (i.e., the ETI data conversion targets).

- A2 - the Integrity programmer imports, from Visual Warehouse, the meta data that defines the input files.

- A3 - the Integrity programmer registers the Integrity process with Visual Warehouse. The Integrity registration process automatically transfers meta data that defines a single business view that will invoke the Integrity process as well as meta data that describes the input and output files and the relationship of the input and output files to the Integrity business view.
- A4 - a data conversion specialists using ETI Extract registers the “stage 2” data conversions with Visual Warehouse. As in step A1, meta data is imported into Visual Warehouse.

- A5 - the Visual Warehouse administrator defines cascading relationships between the last business view in the “stage 1” extract process to the Integrity business view and from the Integrity business view to the first business view of the “stage 2” data conversions. To schedule the entire sequence, the Visual Warehouse administrator defines a schedule for the first business view in the “stage 1” extract process.

---

**Integrating Tools for Analyzing Information**

Once data is prepared for analysis, there are a variety of analysis tools that can be used. IBM has partnered with Brio Technology, Business Objects, Cognos and Hyperion Solutions to provide meta data integration between respective partner products and the IBM visual Warehouse solution. The diagram and descriptions below characterize the meta data flows:

![Diagram of meta data flows](image-url)
**Brio Technology**

Brio Technology provides powerful query, analysis and reporting functionality for Web environments with the Brio Enterprise Server and End-User Tools. IBM and Brio Technology are in the process of defining and implementing integration points between Brio Enterprise Server and End-User Tools and Visual Warehouse.

**Business Objects**

Business Objects delivers integrated query, reporting, and OLAP capability with BusinessObjects.

1. Technical and business meta data can be extracted from the Visual Warehouse Information Catalog and used to build a BusinessObjects Universe.
2. The Visual Warehouse Information Catalog can extract meta data about BusinessObjects Universe, Classes, Objects, and Report types.
3. BusinessObjects reports or queries can be launched from within the Visual Warehouse Information Catalog.

**Cognos**

Cognos provides a data query and reporting tool -- Cognos Impromptu, and an OLAP tool for multidimensional analysis -- Cognos PowerPlay.

1. Mappings can occur between business descriptions in the Visual Warehouse Information Catalog and Cognos Impromptu.
2. Mappings can occur between business descriptions in Cognos Impromptu and the Visual Warehouse Information Catalog.
3. Visual Warehouse subject areas can be mapped to Cognos Impromptu folders.
4. Cognos Impromptu and PowerPlay reports and cubes can be registered in the Visual Warehouse Information Catalog with program association.
5. Meta data synchronization can be scheduled and automated by Visual Warehouse or initiated from Cognos Impromptu.

**Hyperion Solutions**

Hyperion Solutions and IBM integrate the Hyperion Essbase OLAP engine and API with the IBM DB2 family of relational databases.
1. The Visual Warehouse Information Catalog can extract metadata about the multidimensional model from Hyperion Essbase or the DB2 OLAP Server.

2. Visual Warehouse can schedule automatic updates for changes to the metadata extracted from Hyperion Essbase or the DB2 OLAP Server.
Meta Data Integration Between IBM’s Visual Warehouse, IBM’s VisualAge DataAtlas Multiplatform, and PLATINUM’s ERwin

VisualAge DataAtlas Multiplatform is IBM’s data modeling, database design, and data dictionary product. DataAtlas generates IBM Meta Data Interchange Language for relational objects (Databases, Tables, and Columns). The Visual Warehouse Information Catalog can then import these files.

IBM and PLATINUM Technology have a partnership agreement that will integrate technologies from VisualAge DataAtlas Multiplatform and the PLATINUM ERwin data modeling tool to provide a joint solution that capitalizes on the technical strengths of each company’s tool. Additional development work will deliver interoperability between ERwin and the Visual Warehouse Information Catalog.

1. The Visual Warehouse Information Catalog can import meta data about relational objects from VisualAge DataAtlas.
2. PLATINUM ERwin will be enhanced to export meta data to the Visual Warehouse Information Catalog.
3. DataAtlas can import DDL (Data Definition Language) from ERwin and DataAtlas can generate IBM Meta Data Interchange Language to export that meta data to the Visual Warehouse Information Catalog.
Summary

IBM advocates a decentralized approach to the exchange of technical and end-user meta data for Business Intelligence solutions. Today, IBM’s Visual Warehouse supports integrated and automated meta data management in conjunction with products from key partners (Brio Technology, Business Objects, Cognos, Evolutionary Technologies International, Hyperion Solutions, and Vality Technology). In addition, the Visual Warehouse Information Catalog can interchange meta data with dozens of sources using IBM’s Meta Data Interchange Language and/or the MDIS Interchange Language. Even more sources can be accessed through bridges in IBM’s VisualAge Exchange product.

IBM firmly believes that widely-adopted industry standards are necessary to address warehouse meta data issues. IBM is leading work with other industry vendors on OMG standards for Common Warehouse Meta Data and standards for interchanging meta data.

Currently available IBM solutions, combined with IBM’s sponsorship of evolving meta data standards, clearly position IBM as the data warehouse meta data leader.
Appendix A

**The IBM Meta Data Interchange Language** provides the ability to exchange meta data between the Visual Warehouse Information Catalog and:

- Arbor Essbase
- Bachman DBA
- IBM DB2 Family Catalogs
- IBM DataJoiner
- ODBC Sources
- Oracle RDBMS Catalogs
- Sybase RDBMS Catalogs

- Business Objects
- CorelDraw!
- Harvard Graphics
- IBM QMF
- Lotus 1-2-3
- Lotus Approach
- Microsoft Excel
- Microsoft Word
- Quatro Pro
- Word Perfect

**The Meta Data Coalition Interchange Language** is supported by the Visual Warehouse Information Catalog as depicted earlier in the paper. For more information about the Meta Data Coalition go to: [http://www.MDCinfo.com](http://www.MDCinfo.com)

**IBM’s VisualAge Exchange** product provides the ability to exchange meta data between a variety of CASE tools and Visual Warehouse. VisualAge Exchange includes bridges for:

- Bachman Analyst
- Bachman Groundworks
- EIA CDIF (1994)
- IBM OS/VS DB/DC Data Dictionary
- Intersolv Excelerator IS
- Intersolv Excelerator SSADM
- Logic Works ERwin/ERX
- Oracle CASE Dictionary
- Oracle Designer/2000

- Popkin System Architect
- Powersoft S-Designer
- SELECT Software Systems Engineer
- SELECT Software SSADM Engineer
- Sterling IEF
- Sterling ADW
- Sterling Key for Enterprise
- Texas Instruments IEF
- Texas Instruments Composer
Additional Information

Related White Papers

This white paper is one of a suite of Data Management papers available from IBM.

To view IBM data management white papers or to download files for viewing and printing, access this web site: http://software.ibm.com/data

There you will find information about the IBM data management products and solutions. Selecting the “Library” link on this page will guide you to the “Data Management White Papers”.