THE CASE FOR A DECISION SUPPORT METADATA STRATEGY
THE CASE FOR A DECISION SUPPORT METADATA STRATEGY

Businesses today have access to greater quantities—and a greater quality—of decision support information than ever before. But to make the most of this opportunity, each organization must formulate an effective strategy for putting to use the various business intelligence tools that are essential to decision support.

The way to do that is to formulate a strategy for leveraging metadata. Metadata is the core internal documentation that forms the lifeblood of all decision support resources. By planning and expediting the flow of metadata among software tools, IT management can build a foundation for tomorrow’s decision support infrastructure that will maximize the efficiencies, and minimize the risks, of the most ambitious business intelligence tool deployments.

This white paper details the state of metadata today, and presents an overview of the key functions and components behind an effective business intelligence metadata strategy. These include:

- Metadata organization and administration, which promotes sharing and central management of metadata in a distributed repository architecture;
- Content creation and integrity, to maintain consistency of metadata that may be passed among various tools throughout the phases of the project;
- Component-based metadata sharing, which includes facilities for exchanging metadata among upstream design/modeling tools and downstream analytic applications;
- Planning for the future, necessary for ensuring compatibility with emerging metadata and interoperability standards.

Finally, the paper reviews Informatica’s role in supporting and advancing the cause of metadata in distributed enterprise decision support architectures.

METADATA: GROWING FUNCTIONALITY, NEW USERS

Metadata has always been critical to the workings of relational database technologies, but only recently has it become visible outside the world of database administrators and systems programmers.

A decade ago, metadata was used chiefly to describe and catalog the organization of relational database tables for the enlightenment of those database administrators and systems analysts charged with updating, modifying and extrapolating information from relational tables.

As organizations turned to data warehousing and other techniques for extracting and organizing operational data for user analysis, the importance of metadata grew. Data warehouses and line-of-business data marts required new tables—and business logic geared to decision support analysis. Also, they demanded that functions such as data cleansing and data transformation be applied to operational data before it was ready for analysis.

Metadata took on an expanded role among day-to-day knowledge workers. Today it serves as the main catalog, or map, to the data warehouse. It portrays everything from logical and physical definitions of tables and columns for source and target systems to extract histories, aging and purging parameters, transformation formulas, business rules and other types of detailed information. Metadata is no longer the sole property of the database administrator; now it is used by knowledge workers for purposes of designing and modifying analytical queries.
In fact, metadata is critical in helping knowledge workers understand the meanings behind the data, and reconcile conflicts in analytical results. Without access to metadata, users have to rely on IT management to explain the data definitions and describe the logic paths behind their analyses. This can place a burden on IT management and staff; more important, it can reduce the analytical effectiveness of the user.

**An Expanding Role**

In the years ahead, the role of metadata will continue to expand. This is largely because of the major trend toward greater distribution of function, toward the use of best-in-class component technologies for everything from enterprise modeling and extract/transform/load routines to front-end query and analytic applications. Gone are the days when a single, galactic warehouse could solve the diverse needs of a large enterprise, and gone are the days when a single repository could maintain the corporate metadata.

Increasingly, organizations will rely on distributed metadata repositories to feed and inform their distributed component-based infrastructures. Metadata will have to be able to pass easily among heterogeneous components and component-based repositories; and it will have to be controllable centrally, through a distributed architecture and via an overriding metadata strategy.

**Strategy Components**

An effective metadata strategy should cover architecture organization and administration, content creation and integrity, and component-level sharing. Finally, it should have a lookahead aspect, rendering it compatible with future initiatives by major industry players.

**Metadata Organization and Administration**

The core aspect of a decision-support metadata strategy involves the organization of the metadata itself. Organization reveals the relationships among the various metadata abstractions—logical schemas, tables, columns, business logic—that populate the enterprise’s repositories. Organization also makes possible the creation of a virtual repository as a user-friendly overlay to the many repositories that may hold metadata throughout the enterprise.

To function effectively in a multi-user, multi-tool environment, metadata thus requires an organizational architecture that can facilitate metadata sharing and re-use among users of local enterprise repositories. Such an architecture should possess these functions:

*Global repository*—This serves as a central point for managing the metadata that will be shared across the enterprise’s local repositories, and that will be made available to other business intelligence component tools.

*Metadata synchronization facilities*—The global repository should support strong synchronization features. One example: the system of shared folders and hotlinked references, or pointers, employed in Informatica’s PowerCenter product. In PowerCenter, metadata can be “promoted up” for sharing from local repositories registered with the global repository, or it can be designated for sharing—with registered local repositories—from the global repository. In either case, that metadata is placed into PowerCenter’s store of shared folders, and associated hotlink references are created in other repositories.

Shared-folder metadata thus appears to each user as if it were local to that user’s repository; and because there is only a single copy of the metadata, subsequent modifications are automatically reflected through the hotlink references. (PowerCenter users also get the option of copying shared metadata; in that case, the copies do not reflect subsequent changes.)

*Central administration*—The global repository should also present a central platform from which to manage the distributed metadata architecture. Management responsibilities range from physical aspects—load-balancing transformation sessions among servers, for instance—to multiple-user access issues and permission controls.
Organized this way, metadata sharing brings substantial benefits to enterprise developers as well as analytical users. Sharing and re-use works to minimize complexity and reduce metadata quantity. Also, metadata synchronization contributes to repository-to-repository consistency. This is important in simple metadata definitions, and critical in complex transformations.

**Content Creation and Integrity**

Another strategy aspect involves the production and quality of metadata content.

In earlier-generation decision support architectures, systems developers were routinely forced to write their own metadata in order to document key software processes, or to populate their companies’ central information repositories. Today and in the future, systems developers, database architects and users should expect far greater utility and sophistication from their metadata content. A metadata strategy should contain these next-generation content features:

*Self-recording capabilities*—More and more, business intelligence tools are progressing from simple code-generation devices to higher-level engines capable of creating their own metadata. Object-oriented, rule-based tools such as Informatica’s PowerMart deployment suite record their own processes as they are used, and so automatically populate repositories with self-descriptive content.

*Native support for diverse and complex abstractions*—Business intelligence tools should have access to comprehensive, well-defined metadata semantics so they can “speak” the diversity of languages required in different business intelligence situations.

For instance, metadata should support business-level as well as system level abstractions. System-level abstractions are used by system analysts to evaluate basic performance and design quality. Business-level abstractions include less-technically oriented information to help CIOs, business analysts and other end users organize and frame queries. Examples: the business logic used to define “Customer,” or the specific business calculations used to define “Revenue.”

Also, metadata should support the higher-level abstractions of the analytic applications themselves. Such abstractions bring a new degree of automation to the applications, and make it easier for users to become productive with them.

Further, the native metadata grammar should support complex structures such as the database schemas and various transformations used in data warehouses and data marts. This way, it’s not up to the user to work with an external language—to write supplementary Visual Basic or C++ scripts, for instance—in order to support aggregations, target updates, or other complex transformations.

Native support for complex transformations benefits users as well as system developers. For instance, a user might want to drill into a “scrub” expression to see precisely how a target field might have changed while it was in the source-to-target pipeline of a deployment system. With native support for the associated transformations, the user could see clearly what logic had been applied to the data. Without native support, the user might have to request the aid of a system developer to decipher an arcane programming language.

*Lineage awareness*—Metadata content should also be capable of possessing critical information about itself: when and from what source it was created; how and when it was updated; what business rules were used in creating it; what dependencies exist. This self-awareness pays off in metadata integrity, since it gives the user a tool for checking likelihood of accuracy. Lineage awareness also permits the database architect to perform impact analysis, evaluating the enterprise-wide impact that might result from, say, modification of a specific source column.
Newer generation business intelligence tools are now appearing that promise to make high-level decision support largely a plug and play exercise.

Warehouse design and enterprise modeling tools will be used to create basic mappings for preparing and analyzing enterprise operational data. Extract/Transform/Load tools will perform the high-speed functions needed to populate data warehouses and analytical tools. And analytical tools, from basic OLAP query engines to newer generations of task-oriented analytic applications, will give users powerful capabilities for drilling into decision support data.

Each of these resources will play a role in the new world of business intelligence. Together, they will share a common need for metadata.

In earlier generations, system developers or database architects were forced to re-write metadata for each decision support tool; typically, metadata couldn’t be exchanged between different vendors’ products. The result was severe redundancy, and an ever-present risk of inconsistent metadata. Next generation architectures—and a forward-looking metadata strategy—will go a long way toward solving those problems. These functions will be required:

- **Standards-based repository programming interfaces**—Through COM (Microsoft’s Component Object Model) or COM-compatible programming interfaces, newer-generation metadata repositories and tools will be able to exchange metadata that’s vital both to upstream and downstream users.

  - Upstream tools include enterprise-modeling and data warehouse-design systems; with standards-based programming interfaces, metadata produced by these tools will be available for importing into the distributed repository architecture. Downstream tools include the various analytical products, from OLAP to analytic applications. Metadata stored in the distributed repository architecture—including metadata imported from upstream applications—will be available for import here, too, as a means of giving users easy, point-and-click capabilities for verifying the facts, dimensions and aggregations used in their analytical calculations.

- **Compatibility with the Unified Modeling Language**—UML has become an important standard for modeling metadata abstractions. With UML, even the most abstract enterprise models and database schemas can be represented, then imported easily into metadata repositories.

- **Compatibility with the Extensible Markup Language**—XML is becoming a standard as a high-level presentation-layer language for Web-based information access—in this case, for accessing the metadata repository. With XML, users get exponentially greater navigational and manipulative control over Internet/intranet tools and data than is possible with HTML.

- **Support for variable metadata granularity**—Through their interfaces, the new distributed metadata architectures should be capable of communicating business-layer abstractions, as well as the technically oriented metadata. Business-layer abstractions, such as replacing “Emp ID” with “employee number,” make it easier for business users to interpret metadata without requiring intervention by database administrators or IT technicians. Also, business-language definitions are typically entered through the upstream data modeling software; support for business-layer abstractions ensures that they will be passed through the entire architecture.

- **Support for multi-dimensional metadata**—Multi-dimensional structures are employed increasingly by downstream analytical tools to perform highly sophisticated data analysis. Multi-dimensional structures may range from complex cubes with multiple dimensions that appear in multiple hierarchies to various schemas—star, snowflake and redundant. A distributed metadata architecture and its associated programming interface should support the various metadata hierarchies that go into forming these structures. That way, a multi-dimensional environment can be defined early in the upstream design phase, then passed through the
Acting on a Metadata Strategy

To turn strategy to implementation, organizations can put in place this four step methodology:

1. Identify and prioritize the decision support business need(s)—IT management should create a process to 1) educate users in the basics of business intelligence, and in the specific tools that may become available to them, and 2) gather input from users on how they might put those tools to use. Importantly, assessing the business needs of decision support must begin with the users—including executive management as well as analytical knowledge workers. Once the users have expressed their preferences, it is up to IT management to prioritize those needs and turn them into an enabling infrastructure.

2. Evaluate infrastructure requirements—This begins with an accounting of the operational databases that will act as decision support resources, as well as other physical needs, such as networking equipment. From here, IT management should sketch out the likely paths of decision support data to the various user groups. By analyzing these paths—and by noting data relationships that may be common to multiple groups of users—IT management can evaluate infrastructure needs from an enterprise-wide viewpoint.

3. Integrate business intelligence tools within the decision support infrastructure—Next step is to select and integrate component business intelligence technologies. These include software for designing and managing the information stores to be used as staging areas for extracting and cleansing data, as well as the query tools themselves. Key to these decisions will be the capabilities of these components to create, reuse and exchange metadata.

4. Enrich metadata over time—IT management should strive to improve the quality and effectiveness of decision support metadata by monitoring user experience regularly, perhaps through a program of email surveys. At the same time, IT management should track the metadata-oriented aspects included in industry development of new tools and technologies, and should pay close heed to progress in metadata standards activities.

User extensions will be necessary for metadata organization—Metadata APIs and other extensions will allow users to create meta-metadata tables; that is, to form new logical organizations of metadata, as well as new grammar, to suit their business rules and feed their analytical applications.

Informatica’s Role

As a builder of deployment software for business intelligence applications, Informatica is positioned at the heart of the decision support infrastructure. Companies in all major industries rely on Informatica as a strategic decision support partner, and trust Informatica to continue its leadership in building critical enabling technologies.
Just a few years ago, Informatica pioneered the concept of off-the-shelf, business intelligence data marts as a solution to the problems of galactic data warehousing. Most recently, Informatica has introduced the industry's first enterprise architecture for distributed metadata, through its PowerCenter product.

Today, Informatica deployment platforms are at the state of the art for addressing the key functions necessary for setting metadata strategy.

**Metadata organization**—Informatica’s PowerCenter architecture creates a means of organizing enterprise repositories into a coherent, dynamically linked architecture for metadata sharing and re-use. PowerCenter permits organizations to grow their distributed infrastructures incrementally; it lets them choose the status (dependent or independent) of each new local repository, and it lets users select which metadata they wish to share, and which to keep private. Moreover, PowerCenter’s hotlink reference facilities help users maintain metadata currency, and minimize overhead and redundancy.

**Content creation**—Informatica’s software product are both self-describing and self-recording, so they automate the process of metadata creation. And because Informatica’s deployment engines employ object-oriented, rule based design, the metadata they create is stored in clear, business-logic compatible terminology. Informatica’s metadata is also comprehensive: it supports various abstraction levels, complex structures such as multi dimensional schemas, and it maintains full knowledge of its own lineage.

**Component metadata sharing**—Through its next-generation metadata exchange (MX2) architecture, Informatica is pioneering the interchange of metadata with popular design, modeling and analytical tools. Currently, Informatica has metadata-exchange partnerships with more than a dozen leading component makers, including Microsoft, Hyperion Solutions, Seagate Software, Microstrategy, Brio Technology, Business Objects, Cognos and PowerSoft.

**Metadata Benefits in An Age of Componentization**

A sound metadata strategy will ensure freedom of movement and exchange among various vendors’ business intelligence tools and repositories. It will maximize sharing and re-use of important metadata, and will minimize redundancy. These are critical needs as business intelligence infrastructures grow larger, and support more users. Finally, it will ensure metadata integrity, from tool to tool, and user to user.

Without these metadata qualities, users could be caught in a morass of close ended, undependable decision support endeavors. With them, users and organizations will be able to grow their business intelligence resources to their full potential, and take advantage of coming generations of analytical tools.

**FOR MORE INFORMATION, CONTACT:**
Informatica Corporation  
3350 West Bayshore Road  
Palo Alto, California 94303  
Tel: 650.687.6200 or 800.653.3871  
Fax: 650.687.0040  
Web: www.informatica.com

Informatica Corporation  
Chancery Court, Lincolns Inn  
Lincoln Road, High Wycombe  
Buckinghamshire HP12 3RE  
United Kingdom  
Tel: 0870 9070200  
Fax: 0870 9070201

Informatica GmbH  
Ridler Strasse 11  
80339 Munich, Germany  
Tel: 089 502 00 912  
Fax: 089 502 45 40

Copyright © 1999 Informatica Corporation. All rights reserved. Printed in the USA.  
Informatica, PowerCenter, PowerConnect and MX2 are trademarks of Informatica Corporation. All other company and product names may be trade names or trademarks of their respective owners.