

# Metadata-Driven Data Integration

*Ensuring Enterprise-wide Visibility and Transparency with Metadata Management*

WHITE PAPER



This document contains Confidential, Proprietary and Trade Secret Information (“Confidential Information”) of Informatica Corporation and may not be copied, distributed, duplicated, or otherwise reproduced in any manner without the prior written consent of Informatica.

While every attempt has been made to ensure that the information in this document is accurate and complete, some typographical errors or technical inaccuracies may exist. Informatica does not accept responsibility for any kind of loss resulting from the use of information contained in this document. The information contained in this document is subject to change without notice.

The incorporation of the product attributes discussed in these materials into any release or upgrade of any Informatica software product—as well as the timing of any such release or upgrade—is at the sole discretion of Informatica.

Protected by one or more of the following U.S. Patents: 6,032,158; 5,794,246; 6,014,670; 6,339,775; 6,044,374; 6,208,990; 6,208,990; 6,850,947; 6,895,471; or by the following pending U.S. Patents: 09/644,280; 10/966,046; 10/727,700.

This edition published January 2007

## Table of Contents

<b>Executive Summary</b> . . . . .	<b>2</b>
<b>Increasingly Regulated Environments, Yet Lacking Data Management?</b> . . . . .	<b>3</b>
<b>A Metadata-Driven Data Integration Architecture Enables Control</b> . . . . .	<b>4</b>
Metadata Services in an EAI Environment . . . . .	5
The Breadth and Scope of Metadata Services . . . . .	5
Metadata Services in Action . . . . .	6
<b>Informatica PowerCenter Provides Robust Metadata Management Capabilities.</b> . . . . .	<b>6</b>
Map and Verify: Data Lineage for Documenting Control. . . . .	7
Evaluate and Assess: Change Impact Assessment Across Multiple Systems. . . . .	8
Configure and Standardize: A Common Business Vocabulary and Data Dictionary for Data Governance . . . . .	9
<b>Towards Proactive Enterprise Data Management</b> . . . . .	<b>10</b>

## Executive Summary

To spot new opportunities, minimize costs, and manage risks, organizations need data that is accurate, validated, and reconciled. Many organizations are finding that a key mechanism for gaining this level of control is to automate the entire lifecycle for data integration. In particular, a metadata-driven integration approach enables organizations to use metadata as a unifying technical foundation for their data integration services—providing them with the ability to validate their data flows and relationships, while minimizing manual reviews.

Among the criteria to consider when evaluating an enterprise data integration architecture is whether the architecture enables a data-centric view. Pressures to increase earnings and comply with regulations have rendered the traditional application-focused methodology insufficient to deliver reports and alerts that produce actionable results. For example, an application-centric approach has often led to increased costs, inflexible deployment, and compliance failures, especially regarding data quality and governance. A data-centric approach, on the other hand, enables organizations to attain agility, auditability and consistency through the semantic reconciliation and source integrity validation provided by metadata on a service oriented architecture.

To make business decisions that drive revenue, detect material weakness, minimize risk, and deliver optimum business performance, enterprises and institutions need an in-depth and organization-wide understanding of their business that is rooted in complete and accurate data. To achieve consistency and repeatability with manageable cost structures, they are now taking a dedicated approach to the treatment of data as an asset. Solution architectures and aggregation methodologies must allow them to measure the effectiveness of their day-to-day activities based on data they can trust.

This white paper describes how Informatica® PowerCenter®, an enterprise data integration platform with a unique metadata-driven architecture, helps organizations proactively manage data. It examines how a metadata-driven data integration architecture uniquely meets data-intensive business demands. This paper also illustrates how PowerCenter delivers industry-leading metadata reporting and analysis features such as data lineage, impact assessment, and data standardization capabilities to support control activities. Finally, this paper summarizes why it makes sense to consider a metadata-driven data integration architecture for end-to-end business integration needs.

This paper is intended for enterprise and data architects, and practitioners interested in gaining insights into metadata-driven data integration solutions.

## Increasingly Regulated Environments, Yet Lacking Data Management?

Organizations have made substantial investments in modernizing data integration architectures, yet many have yet to realize the full value of those efforts due to:

- Lack of centralized data cleansing and profiling processes resulting in the inability to obtain accurate data and trust that data throughout the organization
- Limited automation evidenced by hand coding and repetitive manual labor by IT and business analyst staff, preventing them from focusing on tasks of higher value
- Inefficiencies in measuring historical change impact and controlling the impact of future changes to increase profitability, choose appropriate investment levels, and mitigate risks

For instance, if organizations deploy an integration platform that lacks a metadata management system, they cannot maximize the value of automation due to the lack of transparency and auditability inherent in a non-metadata driven approach. First, they are unable to trace how data is used across multiple systems and unable to validate whether the properties and descriptions accurately reflect business use. Second, they cannot quickly determine the impact of changes across multiple applications—for example, how will changes to customer data in a sales force automation system impact the organization's financial reporting system or the customer relationship management system used by employees whose duties rely on accurate customer data? Third, the staff from different functional areas—for example, finance, administration, IT, and line of businesses—need to use the same terminology to describe processes that have multiple dependencies. If the terminology is not consistent, it inevitably leads to time-consuming debates and manual reviews—rather than more productive activities that drive competitive advantage and business success. Last, the lack of updated documentation of the data flows and properties will hinder the organization's ability to reuse the data models and mapping specifications.

Automation plays a crucial role in data integration. IT organizations need a complete and accurate view of data relationships to automatically generate, document, and store them. For instance, to validate the internal controls for managing a revenue cycle, data from a general ledger application must be automatically synchronized with the rest of accounting operations. It takes significant manual efforts to access, process, and analyze data from multiple functional and business units. Even worse, manual processing is prone to errors and lacks consistency and repeatability, undermining control of financial reporting. The data rendered from control activities must be documented, tested, and presented to attest to the validity of internal control. Many IT organizations are realizing that they are still ill prepared to achieve operational efficiency and to reuse and automate data integration—common success measures for all IT projects.

IT organizations need to evaluate the capabilities of their enterprise data integration architecture, which serves as the foundation on which a data services layer is built. Common architectural weaknesses include:

**Limited ability to integrate data throughout its lifecycle:** Many IT organizations face challenges in developing an architectural framework in which they can monitor, analyze, and perform data management tasks. The lack of a consistent, uniform approach to data integration prevents easy extraction of meanings from the data to understand the full extent of business activities beyond day-to-day operations.

**Inability to break down data silos:** Because of the varying degrees of maturity and upgrade cycles of business applications and systems, IT organizations face a data silo problem. They also face roadblocks in design and implementation of a common data services layer. Existing data flows and relationships within an application as well as across applications often go unexamined. Consequently, duplication of data is common, while it can be expensive and time-consuming to root out redundancies and manually code mechanisms to access, integrate, and deliver data, as well as document data relationships.

**Difficulty in accelerating reporting cycles due to lack of data transparency and insight:** Companies have increased the number of control activities that need to be tested and demonstrated in their financial reporting cycles. Cross-functional collaboration is critical for keeping pace with the accelerated report cycles. If data integration methods cannot assess the impact of data relationships across multiple business units, a business can be burdened with lengthy cycle times in identifying, accessing, tracking, and explaining before completing the core task of reporting.

**Failure to standardize data collection, analysis, and investigation:** To attest to the validity of reporting, companies must ensure quality in data collection, analysis, and potential investigation tasks. Alignment across all levels of organizations can be attained only by providing the information in a format relevant to people according to their business functions and skill set. Insufficient standardization and untimely updates of business vocabulary and data dictionaries often lead to a lack of confidence in the data and in the business activities themselves.

## **A Metadata-Driven Data Integration Architecture Enables Control**

A metadata-driven data integration approach is a dedicated, enterprise-wide approach to data integration using metadata as a common foundation. This approach:

- Provides data lineage for historical control demonstration
- Enables impact assessment for forward-looking assessment and validation
- Facilitates standardization and stewardship to embed data governance throughout enterprise operations

These capabilities are possible only if a metadata service layer is a part of the data integration architecture. The metadata services layer ensures that interdependent, underlying data supports required business activities. The metadata services layer also ensures that each service is working with consistent information in a verifiable manner. The data integration architecture uses the metadata service layer to guarantee that overlapping data, such as inventory or customer data, yields consistent and actionable intelligence. Data lineage capabilities enable analysts to track the iterations of data—for instance, what data was changed when and by whom—to support business and regulatory demands for data validation and auditing.

## Metadata Services in an EAI Environment

Metadata services can be seamlessly deployed within an enterprise application integration (EAI) environment. In fact, metadata services complement EAI, delivering a critical capability that enables the semantic foundation for the integration stack, which EAI lacks on its own. For instance, an enterprise service bus is concerned with application integration and work flows between the messaging layer levels without having the metadata foundation. In contrast, the data integration architecture is designed to enable holistic access to data, data-specific integration tasks, and interfacing capabilities across multiple services.

## The Breadth and Scope of Metadata Services

Metadata services extract, standardize, and maintain metadata from source systems, including relational databases, packaged and custom-built applications, legacy mainframe systems, messaging queues, flat files, and unstructured data in desktop applications and industry-specific formats. In addition, metadata services enhance data services capabilities for Web services, messaging, and data connectivity via open database connectivity (ODBC) with greater precision and efficiency, and orchestrate additional tasks, such as data profiling, data cleansing, and data federation. At the presentation layer, tools enable administrators, developers, and analysts to abstract the resulting data in the format most meaningful to their respective functions.

All of these services should be delivered on an enterprise-wide, data integration infrastructure that provides security, high availability, and scalability. Security related requirements including access control, segregation of duty, privilege management and reporting, and encryption can be implemented for the sensitive and regulated data management tasks. At the presentation-layer level, tools for administrators, developers, and analysts abstract the resulting data in a format most meaningful to their respective functions.

Figure 1 illustrates how the metadata services layer binds the three functional layers of the data integration architecture—data access, data integration, and data delivery services.

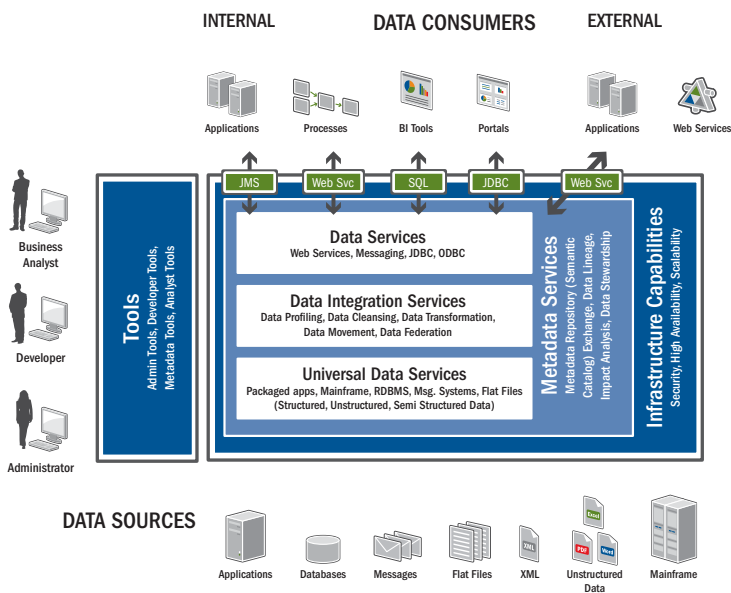


Figure 1: Service-Centric View on Data Integration

## CRITICAL REQUIREMENTS FOR METADATA-DRIVEN INTEGRATION INCLUDE:

- Metamodel design and mapping support for major applications and systems
- Transparent metadata analysis and modeling on an open, vendor-neutral platform
- Custom metadata configurator to address data from nonstandard sources
- Increased reuse through standardization of business vocabulary and data dictionary
- Metadata synchronization within and across heterogeneous environments
- Persistent metadata with grid capabilities used for capacity, throughput, or availability enhancement
- Graphical metadata modeling tools for design, mapping, and development for accelerated development
- Metadata repository to capture management of data integration life cycle

Metadata services support a broad set of data services that benefit all types of data integration initiatives:

- Broadest data access, enabled by the Access Services
- Reusable data services, enabled by the Integration and Delivery Services
- Foundation for enterprise-wide consistency, enabled by Metadata Services
- Data security and business resiliency through Infrastructure Services layers
- Highest productivity, enabled by a robust set of Tools that help deliver automation and reusability

## Metadata Services in Action

Let's look at an example of how these metadata-driven capabilities benefit a company engaged in a data integration initiative. A diversified insurance company is deploying a single general ledger instead of multiple general ledgers and using a unified metadata repository to consolidate systems used by multiple business units.

A single metadata-driven solution uniquely enriches the data integration processes for procure-to-pay, fixed-asset analysis, travel and expenses, and accounting operations. Without metadata management, the company would never have been able to automate data validation processes in such diverse systems. With it, more than 200 users have insight into data flows and can track data lineage to help ensure consistency and accuracy.

The capabilities provide greater flexibility and efficiency by driving processes from a common infrastructure, primed to support mission-critical applications. The company is also able to resolve visible data governance issues, such as misalignment in data definitions and standards, taxonomies, development guidelines, and data models.

## Informatica PowerCenter Provides Robust Metadata Management Capabilities

Informatica PowerCenter is an enterprise data integration platform designed at an architectural level to deliver robust metadata management capabilities. PowerCenter's easy-to-use graphical user interfaces allow IT architects and developers to:

- Rapidly identify information assets, redundancies, process flows, and data lineage across data sources, transformation mappings, and reports
- Predict, assess, and manage the impact of change to enterprise data with a visual map of data dependencies
- Automatically deliver an audit trail that documents changes to data as it flows throughout various enterprise systems
- Exchange metadata with a host of different applications to facilitate interoperability

The following section describes those key PowerCenter metadata management capabilities in more detail.



## Map and Verify: Data Lineage for Documenting Control

PowerCenter allows users to rapidly identify information assets, redundancies, process flows, and data lineage across data sources, transformation mappings, and reports. As Figures 2 and 3 show, data flows for booking and invoicing are tracked backward and forward in the data lineage diagram. A user can easily invoke the data lineage from the PowerCenter Designer while mapping the sales invoice definitions. Organizations can use this tracking to validate the data flows associated with monthly sales and revenue field data. Information can be exported or sent by email for team reviews. The where-used reports validate object properties in a verifiable format.

### Map - Source to Target Definition for Sales Invoice

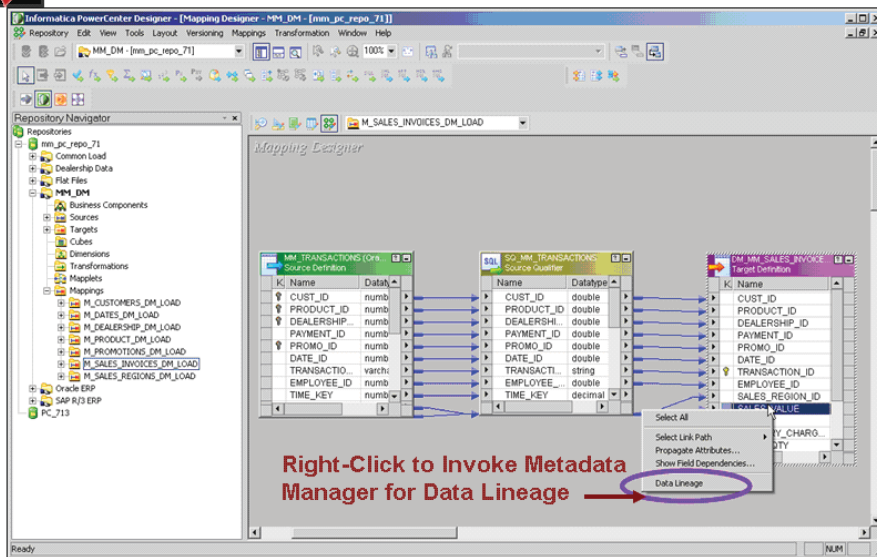


Figure 2: Example of Invoking Data Lineage from PowerCenter Designer

### Verify - Data flows for booking and invoicing

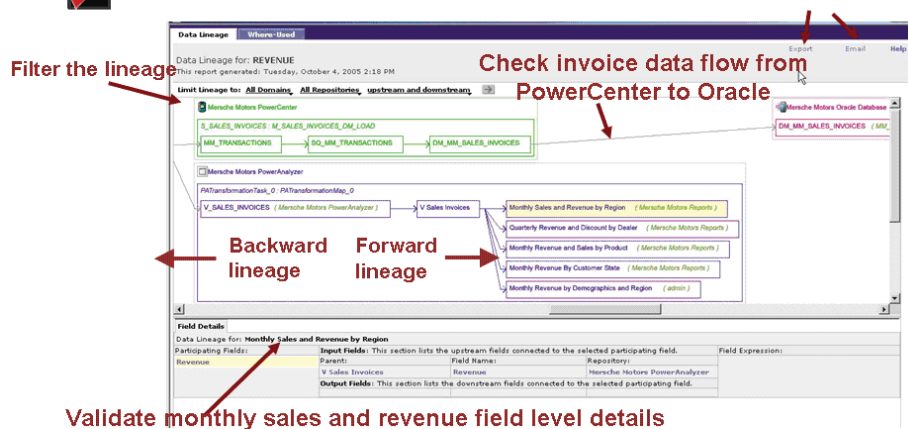


Figure 3: Example of Data Lineage for Documenting Control

## Evaluate and Assess: Change Impact Assessment Across Multiple Systems

PowerCenter allows users to predict, assess, and manage the impact of change to enterprise data with a visual map of data dependencies. Figure 4 illustrates PowerCenter's capability to toggle between data and metadata views. Users can quickly review revenue data by region and validate the definitions via drill-down features to data lineage for critical transactional flows. They can also drill down to see the data lineage view by a single click, which invokes a graphical representation of data flows and relationships.

Once users are in the impact assessment view, the impact of changes in data relationships can be readily presented across the data connectivity options, called xConnect in Figure 5. This predictive facility can be added to the dashboard for continuous oversight of target activities. A hotlink data lineage enables quick review of data lineage.

### Evaluate – Revenue exposure and risk

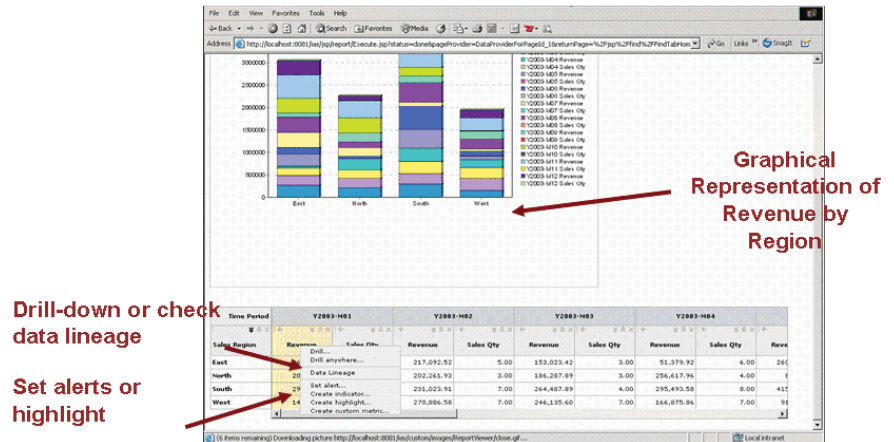


Figure 4: Example of Revenue Exposure and Risk Analysis

### Assess – Impact of changes in data relationships

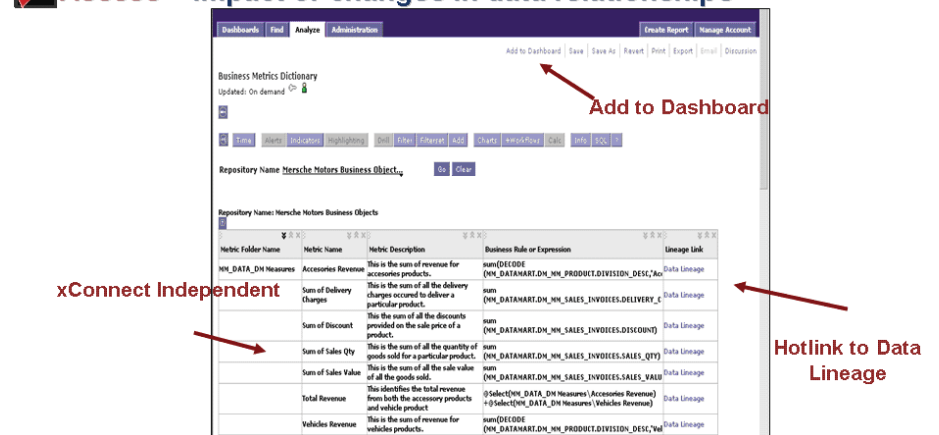


Figure 5: Example of Impact Assessment

## Configure and Standardize: A Common Business Vocabulary and Data Dictionary for Data Governance

Data governance—developing and implementing standards, policies, and processes for data management—fundamentally depends on a common business vocabulary and data dictionary that is shared consistently across the enterprise. A metadata foundation is essential to establishing the semantic and syntactical underpinning that ensures data is uniformly defined and understood across multiple units.

Informatica provides packaged metadata connectivity options as well as custom configuration capabilities to enable universal access to metadata models. Users can select and associate custom metadata to configure repository types, classes, associations, and hierarchies, as Figure 6 illustrates.

PowerCenter's metadata management capabilities accelerate the data standardization processes required to develop and manage a common business vocabulary and data dictionary. As Figure 7 shows, browsing, filtering, and search capabilities help IT users and business analysts to perform assessments efficiently.

### ✓ Configure Repository types, classes, associations, and hierarchies

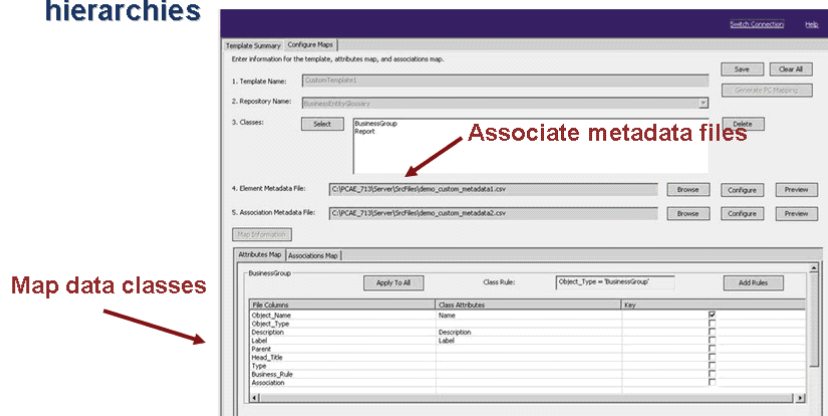


Figure 6: Example of Custom Metadata Configuration

### ✓ Standardize – Business vocabulary and data dictionary

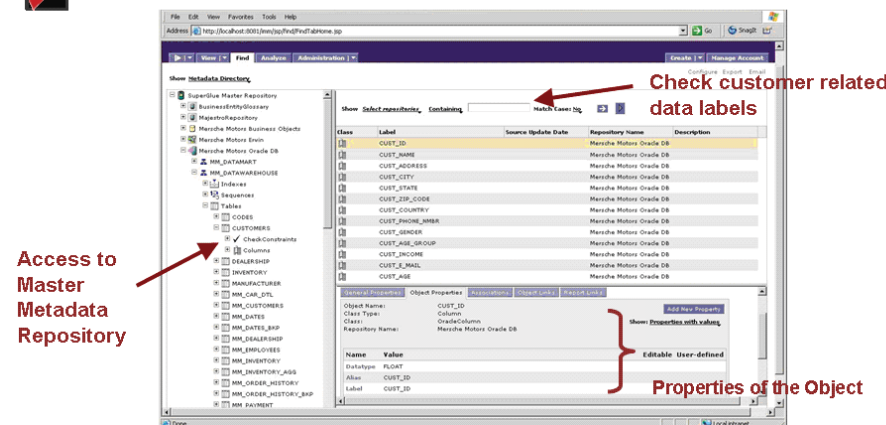


Figure 7: Example of Business Vocabulary and Data Dictionary for Data Governance

## Towards Proactive Enterprise Data Management

Today's highly regulated climate requires a re-examination of data integration architecture to demonstrate internal control of all levels of enterprise operations. Informatica delivers a new approach using a metadata service layer as a mechanism to bind core data integration tasks to address multiple data silos, inefficiencies of reporting, and limited standardization. To effectively manage data, organizations need an enterprise data integration platform that offers robust metadata management capabilities.

Informatica aids organizations to perform enterprise-wide data management tasks via its metadata-driven architecture and key capabilities. The Informatica data integration platform offers metadata management functions, such as data lineage verification, forward-looking impact assessment, and data stewardship, demonstrating the platform's readiness to perform the essential data integration tasks. The platform's powerful graphical user interface for design and review readily exposes the interdependencies and relationships of systems and applications and increases user productivity. Informatica uniquely provides these metadata management capabilities as part of its leading data integration platform with unparalleled strengths in universal data access, performance, and security.

The Informatica platform extends the enterprise data integration infrastructure beyond communication at the messaging level and complements existing application-oriented frameworks. The platform transforms organizations' ability to proactively integrate data, substantially curtailing the need for manual reviews and reactive management of data flows and ensuring data consistency across shared services. These metadata management capabilities serve as a unified foundation to govern data integration-related tasks. Organizations worldwide are realizing substantial value from the deployment of Informatica's data integration platform – a platform that helps drive the transformation in enterprise data management toward a more proactive and predictive model.

For more information about the metadata management capabilities of the Informatica data integration platform, please visit [www.informatica.com](http://www.informatica.com) or call 1.800.653.3871.









Worldwide Headquarters, 100 Cardinal Way, Redwood City, CA 94063, USA  
phone: 650.385.5000 fax: 650.385.5500 toll-free in the US: 1.800.653.3871 [www.informatica.com](http://www.informatica.com)

Informatica Offices Around The Globe: Australia · Belgium · Canada · China · France · Germany · Japan · Korea · the Netherlands · Singapore · Switzerland · United Kingdom · USA

© 2008 Informatica Corporation. All rights reserved. Printed in the U.S.A. Informatica, the Informatica logo, and PowerCenter are trademarks or registered trademarks of Informatica Corporation in the United States and in jurisdictions throughout the world. All other company and product names may be trade names or trademarks of their respective owners.

6687 (09/17/2008)