

Data Quality: the First Step on the Path to Master Data Management

SOLUTION BRIEF



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 June 2007

Table of Contents

Executive Summary	2
Initiate MDM by First Establishing Data Quality Baseline.	4
Profiling Reduces MDM Implementation Risks	4
Business-Relevant Data Quality Metrics	5
Completeness	6
Conformity	6
Consistency	6
Fostering a Data Quality Culture	7
Summary: For MDM Start with Data Quality	9
Informatica Underpins MDM.	9
Informatica Products	12

Executive Summary

More stringent regulation and market competition mean that organizations need to establish better control and consistency of their master data and the business processes responsible for the capture and maintenance of such shared data. Many organizations are implementing master data management solutions—under a number of different labels, including MDM, customer data integration (CDI), product information management (PIM), or global supplier management (GSM)—to control the definition and management of shared business entities, such as customer, product, and financial data. However, reconciling master data across the enterprise has proved to be a struggle for large organizations, especially those that have grown by acquiring new businesses, and companies whose operations and offices are spread across a wide geographic area.

Despite the effort required, the benefits can be manifold. Often touted as supporting data governance and compliance initiatives, MDM also holds out the prospect of generating significant business benefits, including cost savings and far-reaching operational efficiencies. By achieving a single view of product, customer, or vendor data, organizations can realize procurement efficiencies, improve customer service, and drive market penetration. And, at the same time, they can get better visibility of financial performance through more reliable enterprise reporting.

There are many approaches to master data management—as reflected in the numerous MDM solutions available in the market. The number of offerings can cause confusion because of the wide spectrum of capabilities provided. Where most of these solutions intersect, however, is in what they are missing: the generic data access, delivery, and quality technologies that the Informatica platform provides.

In this paper we will focus on the central role data quality plays in planning, developing, and maintaining a working MDM solution. The planning activity for MDM must encompass all types of master data and all reference data used in conjunction with it. And all business processes that create, maintain, or use master and reference data will need to be taken into account during the implementation planning stages. Even more crucially, however, the successful implementation and ongoing maintenance of any MDM solution depends on having a thorough understanding of data quality within the organization, and employing pragmatic and well-proven data discovery and data quality management strategies to ensure high levels of data quality are maintained.

According to a recent report from Forrester Research, “A core component to creating master data is the ability to first perform data quality profiling and then apply standardization, matching, merging, and enrichment logic.”¹

Most industry observers agree with this point of view. Having a single solution that can deal with data quality issues across all master data types, all applications, and all geographies (i.e., encompassing MDM, CDI, GSM, and PIM) is enabling organizations to maximize the value they can gain from their master data.

¹ “Demand For Master Data Management Software Is Timid But Growing Steadily Through 2010,” Forrester Research, Inc., March 6, 2007.

In a recent MDM study carried out by The Data Warehousing Institute (TDWI), 83% of respondents reported that their organizations have suffered problems due to poor quality master data, and 54% claimed to have derived benefits from good master data.²

Some of the problems associated with poor quality master data outlined in the TDWI report, and also cited by many large organizations that we have spoken to, include both technical and business issues:

- Poor customer service
- Inefficient marketing or purchasing
- New products delayed
- Inaccurate reporting
- Arguments over which data is appropriate
- Bad decisions based on incorrect definitions
- Data governance and stewardship limitations
- Limited visibility for data lineage
- No understanding of master data homonyms

Most of the benefits of high-quality data referred to by respondents to the TDWI survey were understandably the flip side of the problems listed above:

- Risk reduction
- Superior customer service
- Supply chain optimization
- Accurate reporting
- Better decision making
- Easier auditing of information's origins
- Consistent definitions

Data quality is sometimes built directly into MDM solutions rather than treated as a part of the infrastructure. In terms of ensuring the quality of the data within the MDM environment, there is no theoretical reason why this shouldn't work absolutely fine. However, the MDM environment is not an island.

The goal of a data quality exercise is to establish high-quality data in the first instance and then to maintain it. There is no point in cleansing all of your data and then allowing inefficient business processes to mess it up again. What is required is a set of processes that monitors all relevant data on an ongoing basis to help prevent future erosion of data quality levels, both inside and outside of the MDM infrastructure. Only standalone data quality platforms can achieve this; data quality only embedded within MDM is isolated and can leave much of the wider problem unsolved. It is better to treat data quality as one of the crucial enabling technologies for implementing and maintaining MDM rather than as a part of MDM per se.

² The Data Warehousing Institute, "Master Data Management: Consensus-Driven Data Definitions for Cross-Application Consistency," Philip Russom, October 2006. To download the complete TDWI report, visit tdwi.org/research/reportseries.

Initiate MDM by First Establishing Data Quality Baseline

The path to MDM begins with data quality, and the starting point for any data quality process is discovery, profiling, and analysis. Ultimately, the data quality process should follow through to ongoing data quality monitoring and improvement if the MDM initiative is to generate measurable value for the organization.

Data Discovery/Profiling

- Identify the candidate operational systems that maintain the required master data records.
- Determine the operational systems that use these master data records and the business processes they support.
- Profile the data in candidate systems to identify attributes available within these systems and expose potential data quality issues such as obsolete or redundant data.

Whether the MDM solution is being implemented for a single master data object (e.g. customer or product) or for all operational master and reference data, we would argue that the MDM implementation approach can only be determined when the candidate source systems, the business processes they support, and data contained in those systems have been fully profiled, analyzed, and recorded. This is not something the IT department can be expected to do in isolation. It requires the full cooperation and involvement of both IT and business personnel—as do all phases of the data quality and MDM process development.

Profiling Reduces MDM Implementation Risks

Industry experience has shown that MDM projects are prone to the same challenges and problems that are common to all data migration or application implementation projects. If not planned correctly, there is a high risk of operational disruption during implementation and deployment. They suffer from extended time lines, uncertainty about phasing, transition and change management challenges—not to mention potential cost overruns, or outright project failures.

Understanding key issues about data gaps, misalignment, duplicates, and obsolescence from the start can dramatically reduce the time it takes to plan and execute these complex projects. With the level of understanding that can be rapidly achieved through upfront data profiling, the MDM project team will be better informed to make key decisions about data cleansing, implementation, and business transition. Exposing data quality issues early on also initiates the first stages of collaboration between business and IT, allowing business subject matter experts or ‘knowledge workers’ to play a key role in MDM planning and implementation phases.

Upfront profiling can help lower MDM project cost, reduce project risk, accelerate business value, and deliver better results by:

- Reducing direct project costs such as labor and resources
- Increasing project value through early project completion
- Decreasing project team size with consequential reductions in complexity and risk
- Avoiding unexpected costs and delayed benefits associated with overruns or cancellations

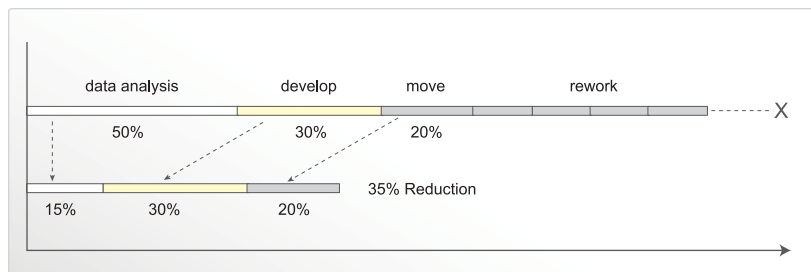


Figure 1: Faster master data management implementation. This graph shows how using Informatica Data Explorer immediately reduces project timelines by cutting the time needed for data analysis. Even more time savings result from less rework and delays in later stages.

Business-Relevant Data Quality Metrics

To ensure your organization has a proper understanding of data quality levels for key attributes in its source systems, business and IT must collaborate to define a common set of metrics to classify data quality and to describe defects and issues. This will enable better communications between business stakeholders and the IT project team when describing issues or cleansing processes and standards. The same metrics should form the basis for auditing and monitoring data quality before, during, and after the MDM project has been implemented.

Six metrics commonly used to measure good data quality are completeness, conformity, consistency, accuracy, duplication, and integrity. The attributes generally cover a multitude of sins that we most commonly associate with poor quality data: data entry errors, misapplied business rules, duplicate records, and missing or incorrect data values.

In mature organizations, data governance initiatives are now driving many of the requirements for data quality, consistency of data, and data definitions across the enterprise. New dimensions of data quality using different metrics may need to be measured. These metrics are not necessarily associated with master data management itself, but are associated with other governance priorities or process requirements such as security, roles and responsibility definitions, and timeliness and availability of data. However, it is worth considering them as part of the MDM planning process.

The specific attributes you choose to measure data quality will depend on what your organization is trying to achieve and may depend on the stage of your project. For example, understanding the completeness, consistency, and conformity of the data within and across each system is critical at an early stage in the project. This will give visibility to the key issues and challenges to be faced and the business users within the organization who must be engaged and consulted in this process.

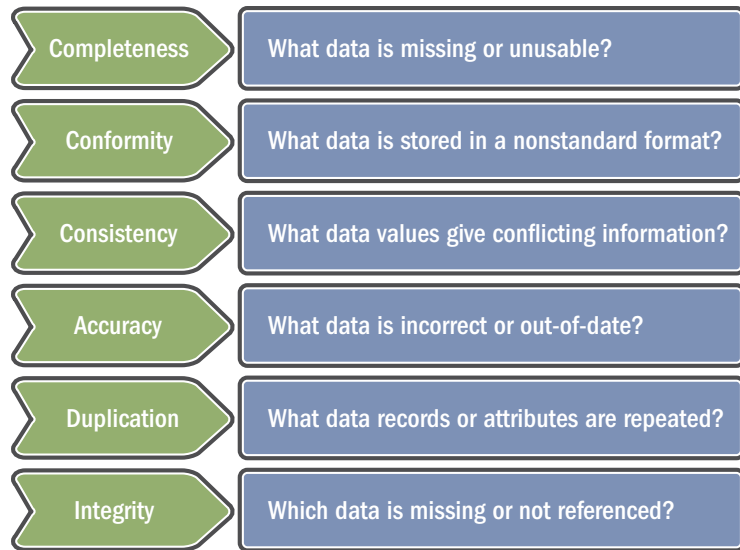


Figure 2: Example of data quality metrics.

Completeness

(e.g. date of birth not null, is a date, and is in the past)

The process to create one master record may mean compiling or consolidating data from many records into single or multiple systems. Opportunities to perform consolidation or de-duplication must be considered.

Conformity

(e.g. phone number correct length and no alphabetic text)

Differences in data models across systems will lead to challenges with alignment of the structure of data and the actual data model to be used.

Consistency

(e.g. gross weight equal or greater than net weight)

There will always be challenges around alignment of data that need to be taken into account. Different codes and values used for the same purpose in different systems can cause issues, and all too often different systems use different hierarchies and classifications leading to conflicts.

Only when these and other similar issues have been uncovered can the business and IT project team make the key decisions about how to manage an MDM project.

Some of the issues the team should be informed about include:

- The key cleansing and alignment tasks and the resources that will be needed
- The prerequisites needed to change the master data maintenance processes
- Harmonization and change capture management
- Realistic implementation phases and the interim processes or interfaces required to support the operational business processes during the project

The major IT and business process implementation challenges can now be considered. Some of these include:

- Decoupling master data creation and maintenance routines from operational systems
- Harmonizing operational systems with the new master data management repository
- Interim processes that will need to be implemented during the transition period
- Interim interfaces must be implemented to support the smooth running of operations during the transition period

Fostering a Data Quality Culture

Major MDM projects require data profiling and data quality analysis not only during the discovery and preparation phases of a project, but also during and after the implementation project.

Monitoring and coordinating data cleansing and data alignment activities are major challenges in any operational environment going through a significant change initiative. Data quality is best managed by the data owners: the people who really understand the data in the context of how processes interact with and use that data. These subject matter experts need to be empowered to track progress and address challenges in cleansing initiatives. Without this visibility to issues, issue resolution, and ownership of data readiness, the organization cannot ask these people to sign off on the state of readiness at the time to go live.

After implementation, an MDM solution will significantly help data quality in an organization by aligning processes and practices, providing a single application for the capture of master and reference data and enforcing a single data model for key master data objects. However, it is not practical for most large organizations to completely decouple master data capture and maintenance from all systems. For example, most organizations will have master data created in legacy systems, or will be dependent on external parties for some of their master data objects. In some organizations, the level of local or regional requirements will mean it is not practical to implement a single data model across the whole organization. Thus, a repository or a registry is maintained for some of the key data attributes, and other attributes are maintained locally.

System of Origin	Centralized	Centralized	Centralized
System of Record	Centralized	Centralized	Distributed
System of Reference	Centralized	Distributed	Distributed
System of View	Distributed	Distributed	Distributed
	Update Process	Master Copy	Authoring

Figure 3: Different approaches to master data management means that master data capture and maintenance may not be completely decoupled from all systems.

With this in mind, ongoing data quality degradation is a constant threat. Data quality must be measured, audited, and monitored at multiple points throughout the organization to ensure conformity and consistency to the standards set by the business are maintained.

Knowledge is power. Always knowing where your organization stands in terms of data quality is perhaps the most important ingredient in any data quality management process. Firms that take steps to measure data quality are more likely to successfully address the problem. Being able to measure and monitor data quality throughout its lifecycle and compare the results over time are essential ingredients in the proactive management of ongoing data quality improvement, data governance and master data management.

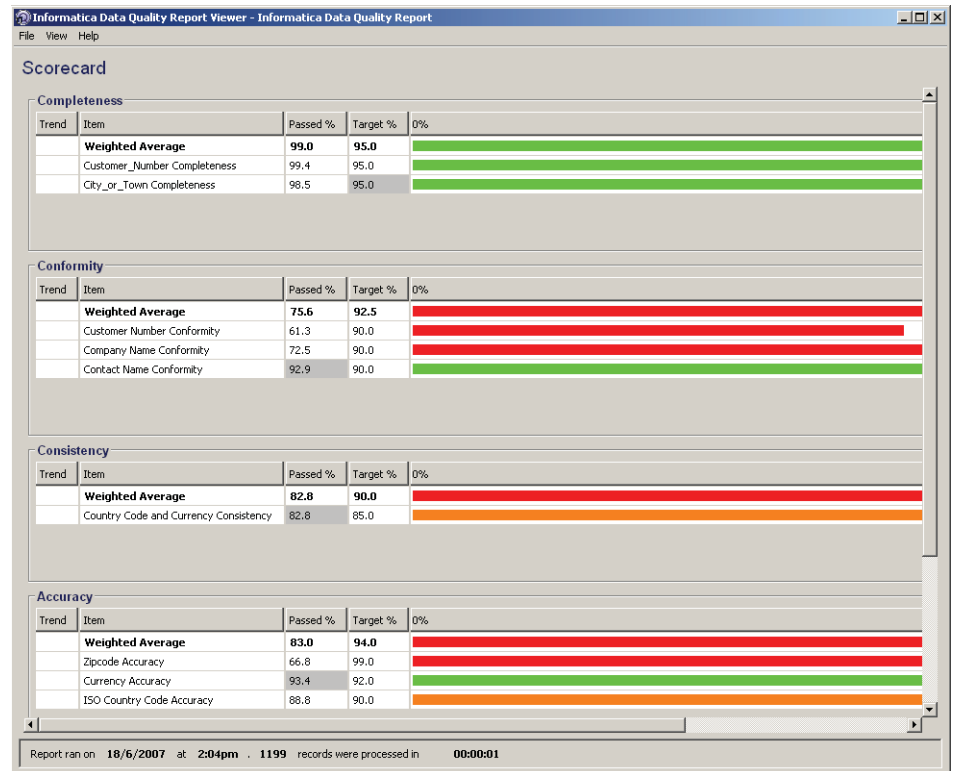


Figure 4: Scorecard tracks data quality metrics.

Data quality metrics defined and used during the planning and implementation phases of the MDM project are the key to data quality monitoring. Inherent in the metrics-driven approach is the ability to aggregate company-wide results into a data quality scorecard or scorecards. A scorecard is the key visual aid that helps to drive the data quality process in the right direction, empowering data analysts to set accurate and focused quality targets and to define improvement processes accordingly, including setting priorities for data quality improvement in upstream information systems.

Metrics and scorecards that report on data quality—audited and monitored at multiple points across the enterprise—help to ensure data quality is managed in accordance with real business requirements. They provide both the carrot and the stick to support ownership, responsibility, and accountability.

Summary: For MDM Start with Data Quality

Data quality is the critical first-step of MDM—without high-quality data, no MDM solution can be expected to deliver value to the business. Without the data quality processes firmly in place, an MDM solution will not only deliver poorer results than expected but could also end up costing an organization dearly in terms of repairing the processes that lead to data problems.

The success or failure of any MDM implementation project hangs in the balance until both business and IT stakeholders fully understand the issues and challenges for that project. On too many projects, the business stakeholders are left to worry about business processes and change management issues, while data quality, integration, consolidation, and migration are left to IT to resolve and implement.

Exposing the data quality and data alignment issues and challenges early in a project, empowers the business stakeholders to play a central role in the decision processes for the resolution of these data challenges and enables the project team to thoroughly address the key risks that can lead to project failure.

Just as data quality is the first step for MDM, data profiling (or more specifically a data quality audit and scorecard) is the first step for data quality. During a data quality audit you should measure the level and nature of data quality problems across multiple data sources—to identify, quantify, and categorize current and potential data quality issues and adherence to business rules.

The audit should provide answers to questions such as “How bad is my data?” and “How much impact is it having on business performance?” Knowing the size of the problem provides the organization with a starting point from which to kick-off data quality enhancement and MDM. But measurement doesn’t stop there; the audit process needs to be ongoing to quantify data quality improvement over time and indicate areas where more work can be done.

Informatica Underpins MDM

Implementing a data management initiative involves a combination of people, processes, and technology. The Informatica platform empowers the right people in your organization to implement effective and lasting data management processes that support MDM, CDI, GSM, and PIM. The platform features the key enabling technology required to underpin MDM solutions—including data profiling, data quality, data integration and metadata management.

Informatica’s data profiling solution, Informatica Data Explorer™, enables the upfront data profiling of multiple data sources that is necessary for organizations to understand the truth about data in all internal and external data sources when planning an MDM project. Informatica Data Quality™ provides the functionality needed to further analyze content, standardize and cleanse data for use in data hubs, leverage fuzzy matching to identify relations to consolidate or eliminate duplicates, and deliver data quality monitoring to ensure that the quality of all master data is tracked over time across the enterprise. Informatica Metadata Manager provides metadata visualization and reporting for full data lineage and impact analysis.

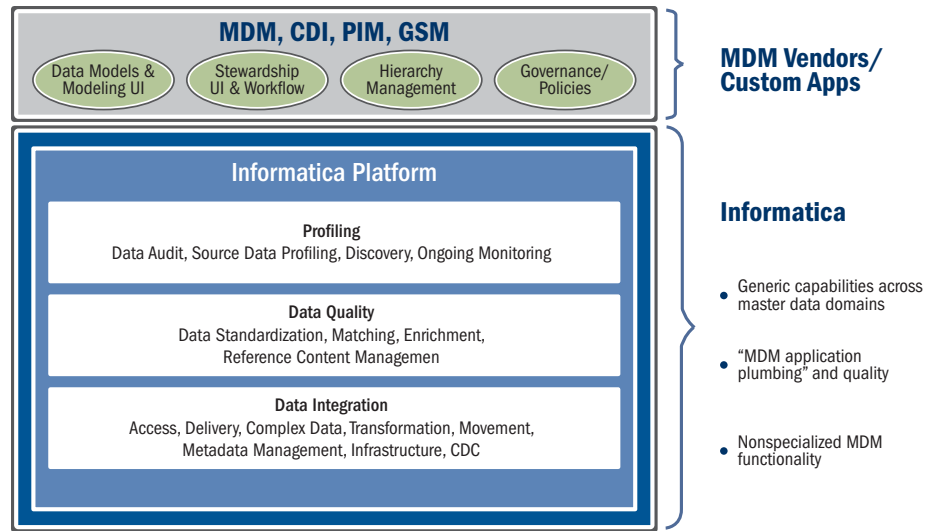


Figure 5: The Informatica platform underpins master data management.

Together, these products offer a comprehensive solution for supporting MDM implementations from data analysis and profiling to data cleansing and data integration of all data types. Informatica provides a single enterprise data integration platform to help organizations access, transform, and integrate master data from a large variety of systems and deliver that information to their MDM applications or custom solutions. The Informatica enterprise data integration platform offers the key capabilities that businesses need to access, integrate, migrate, and consolidate master data, which reduces complexity, ensures consistency, and empowers the business.

Informatica's data quality, data integration, and metadata management solutions work together in concert as the foundation of MDM solutions, enabling organizations to harmonize, cleanse, and certify as "fit for purpose" all of the common information assets that must be shared across the enterprise.

The Informatica platform provides key capabilities that allow businesses to tackle the challenges faced when planning, designing, deploying, and maintaining MDM, including:

Data Analysis and Profiling. To understand how master data housed in existing disparate systems relates to each other and to the MDM application, there is a need to analyze and profile the master data in all systems—often starting with the MDM application and the desired results and working back to the existing systems. The data analysis and data profiling capabilities of the Informatica Platform help IT departments understand the current state of their company's master data in existing systems and plan MDM application implementations accordingly.

Data Correction, Cleansing, and Enrichment. Master data is often of questionable or poor quality. Therefore, master data needs to be corrected and cleansed before it can be used within an MDM application or solution. The Informatica platform's data correction, cleansing, and enrichment capabilities enable MDM to meet business needs by delivering high-quality, fit-for-purpose data to the MDM application. A key capability of Informatica's data quality solution is its ability to standardize and cleanse all types of data, including name and address data, product/materials data, and vendor data. This is a key requirement of data quality support for MDM implementations.

Data Quality Monitoring. Monitoring and tracking the quality of master data on an ongoing basis is an important consideration in the maintenance of any MDM initiative. Informatica Data Quality enables the ongoing monitoring of data quality across all key attributes to ensure master data is managed to the highest standards (internal and external).

Metadata Management. Metadata management provides an enterprise-wide view of data lineage that services the needs of both business and technical users—enabling fast, accurate impact analysis and supporting enterprise-wide data stewardship.

Data Access, Delivery, and Transformation. Master data is likely to be stored in various systems in different formats, all of which are apt to be different from the format inside the MDM application. The master data needs to be transformed from each system format to the MDM application's format and vice versa. The Informatica platform's robust data transformation capabilities accelerate MDM application implementations by providing the right master data in the right format. Only PowerCenter 8 is capable of handling missioncritical, enterprise-wide data integration from a single platform. No other product manages a broader range of complex data integration initiatives, such as migrating from legacy systems, synchronizing databases, or consolidating applications.

ABOUT THE AUTHORS

Tommy Drummond, VP Product Marketing, Data Quality Informatica

Tommy Drummond is a leading data quality expert and plays a central role in planning the future direction of Informatica's data quality solutions. He has nearly 20 of years experience helping large organizations to solve key business and IT problems. For more than seven years, he has focused on the area where IT and business converge—data quality. He has a deep understanding of the techniques and processes for tackling quality at an enterprise level.

Paul Woodlock, Senior Director Business Solutions

Paul Woodlock is a business process and management expert with nearly 20 years of experience in delivering business-focused information technology solutions to the manufacturing and financial services sectors. He has a profound knowledge of data quality and works closely with many of Informatica's top customers to help them implement lasting and effective data quality management programs.

Informatica Products

Informatica Data Explorer is breakthrough software that puts powerful data profiling and mapping capabilities in the hands of the business and IT. Data analysts and data stewards work with Informatica Data Explorer to search and profile data from any source. With unparalleled ease of use, the software reveals hidden data quality issues, gaps, and inconsistencies and incompatibilities within and across data sources. Accurately profiled data accelerates and streamlines data integration initiatives and other business or IT projects. Organizations use Informatica Data Explorer to create a complete and completely accurate picture of the content, quality, and structure of their enterprise data that can be used as the foundation for addressing data quality throughout the enterprise-wide.

Informatica Data Quality is specifically designed to put the control of data quality processes in the hands of business professionals. With unparalleled ease of use, the software delivers powerful data cleansing, data matching, reporting, and data quality monitoring capabilities in a single solution. Data analysts and data stewards use Informatica Data Quality to easily design, manage, deploy, and control individual and enterprise-wide data quality initiatives. By providing a complete platform for ongoing measurement, monitoring, tracking, and data quality improvement at multiple points across the organization, Informatica Data Quality empowers business information owners to implement and manage effective and lasting data quality processes.

Informatica PowerCenter is a single, unified enterprise data integration platform that enables companies and government organizations of all sizes to access, discover, and integrate data from virtually any business system, in any format, and deliver that data throughout the enterprise at any speed. Organizations rely on PowerCenter to serve as the foundation for all mission-critical enterprise data integration initiatives, including data warehousing, data migration/consolidation, data synchronization, master data management, and cross-enterprise data integration. PowerCenter helps organizations derive business value from all their data so that they can reduce IT costs and complexity, streamline operations, and drive revenue growth.



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

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 The Data Integration Company 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.

6785 (09/30/2008)