Informatica Data Quality and Data Integration

How to ensure business-focused data quality, enterprise wide

Symbiotic Cycles of
Data Profiling
Integration, and Quality

(poster)

TDWI, 2006

"Data profiling, data quality, and data integration are three business practices that go together like bread, butter, and jam. . . data management professionals and their business counterparts need to coordinate efforts and design projects that integrate all three practices effectively."

Despite nearly 40 years of information technology advances, it is only now that some organizations are realizing the true value of their corporate data. Increasing reliance on automated business processes, more stringent regulation, and the ever-increasing pitch of competition are driving organizations to focus on the data they use to run their business as they strive to improve customer service, comply with government regulations and customer mandates, and streamline global operations.

Data Must Be Easily Accessible Enterprise Wide and Certifiably Accurate

To support today's business processes and goals, all corporate data needs to be universally accessible, flexible, certifiably accurate, and reusable. Organizations need to know more about what is in their source systems, they need to be able to integrate data from multiple systems into new, more productive data-intensive applications, and they need to be able to cleanse and enhance data as well as monitor and manage the quality of data as it is used in different applications.

Data quality and data integration technologies have evolved to support organizations grappling with fragmented and defective data. There is no doubt that the two technologies support intrinsically symbiotic activities, but where does one end and the other begin? The answer is that the two functions should work together seamlessly.

BENEFITS:

- Universal connectivity to profile and cleanse data from any system

- Performance and scale for enterprise data quality deployment

- High availability and fault tolerance for ongoing high-volume data quality

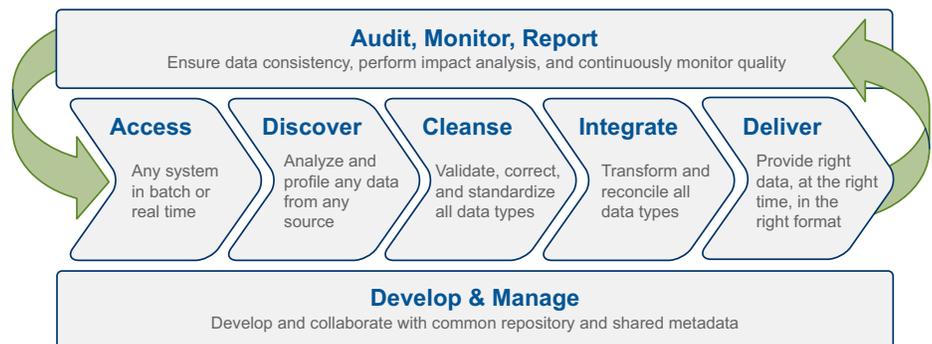- Shared metadata analysis and reuse of business rules for data quality

Figure 1: Data quality and data integration support intrinsically symbiotic activities.

DATA INTEGRATION AND DATA QUALITY: ENABLING BUSINESS AND IT COLLABORATION

Informatica's data integration technology provides connectivity to virtually any system, including structured data (e.g., relational databases, enterprise applications, mainframe systems), unstructured data (e.g., Word and Excel documents, email, binary files), and semi-structured data (e.g., XML and industry-specific standards such as EDI, HIPAA, ACORD, FIX, SWIFT). The Informatica platform enables an organization to automate the lifecycle of accessing, discovering, and integrating all its data assets in any format, historical and operational, across the enterprise. In addition,

Informatica provides a platform that is reusable to reduce cost and implementation risk.

Informatica, the company that has set the standard for data integration, is also a leading provider of data quality products and services. Data quality analysis and cleansing have tended to be more business-focused activities where profiling technology is used to reveal hidden data quality issues, gaps, and inconsistencies and incompatibilities within data sources. Once the data sources have been profiled, the next phase of the data

quality process uses the results to design and automate data cleansing strategies aimed at improving, enriching, and monitoring the content based on business rules and reference data. Informatica's data integration and data quality solutions enable business and IT to work effectively together to achieve real value for your organization's data assets. An organization can gain the advantages of minute-by-minute decision making, cross-business unit visibility, synchronization of mission-critical operations, and transparency for regulatory compliance.

Access, Discover, Cleanse, Integrate, Deliver

At Informatica we believe data quality is central to all data integration initiatives. Before the data integration platform can deliver data to the data warehouse or to the customer relationship management, the enterprise resource planning, or the business analytics applications, the data needs to be analyzed and cleansed. Analyzing and profiling the data prior to data integration are essential steps in the planning process and dramatically speed up the development of data integration workflows and mappings. This initial profiling helps organizations to identify and understand their source data and ultimately to reconcile the source data and the target systems.

But data quality is not a one-off activity. The problem with data is that its quality degenerates over time. So just as data integration is an ongoing process, so too is data quality. Data quality encompasses more than finding and fixing missing or inaccurate data. It means delivering comprehensive, consistent, relevant, fit-for-purpose, and timely data to the business regardless of its application, use, or origin.

Enterprise Data Quality Needs Data Integration

As data quality reaches across the enterprise, the distinctions between data quality and data integration blur even further. For just as the data integration platform takes advantage of data quality, so too enterprise data quality leverages data integration technology. As data quality initiatives are getting bigger and becoming more deeply embedded in operational systems and processes, they require the underlying data integration platform to deliver performance and scale to help organizations realize the true value of their information assets.

Informatica for Enterprise Data Quality Management

Informatica offers an enterprise data integration and data quality platform that allows your organization to implement a comprehensive, full-lifecycle approach to addressing data quality, both upstream and downstream. This enterprise platform includes the leading profiling, quality, and integration products Informatica PowerExchange®, Informatica Data Explorer™, Informatica Data Quality™, and Informatica PowerCenter®.

The Informatica data quality and data integration platform offers capabilities far beyond most standalone data quality offerings on the market. It includes

- the ability to access data from and deliver data to almost any system, application, or database

- the performance, scale, high availability, and fault tolerance needed for the ongoing processing of high-volume tasks

- a shared metadata environment for metadata analysis and reuse

The Informatica Data Integration and Data Quality Platform

Companies of all sizes rely on Informatica to access data they can trust throughout their enterprise based on our data integration and data quality technologies. The data integration suite integrates and delivers data from virtually any business system, in any format, and delivers that data throughout the enterprise at any speed. The data quality products provide profiling and cleansing of data to ensure that the data delivered is not only accurate, but certifiably accurate. The products in the Informatica platform work together at different stages of the data integration and data quality process providing robust, enterprise-class, quality data from a single integrated environment.

Access: PowerExchange
Packaged Apps., Mainframe, RDBMS, Msg. Systems, Flat Files, Complex Data (Structured, Unstructured & Semi-Structured Data)

Discover: Informatica Data Explorer
Data Audit, Source Data Profiling, Discovery

Cleanse: Informatica Data Quality
Data Standardization, Matching, Enrichment, Reference Content Management, Ongoing Monitoring

Integrate & Deliver: PowerCenter
Data Transformation, Data Movement, Data Federation, Web Services, Messaging, JDBC, ODBC

Tools
Admin Tools, Developer Tools, Metadata Tools, Analyst Tools

Metadata
Metadata Repository (Semantic Catalog), Exchange, Data Lineage, Impact Analysis, Data Stewardship

Infrastructure
Security, High Availability, Scalability

Business Analyst

Developer

Administrator

DATA TARGETS / SOURCES

Applications Database Messages Flat Files XML Unstructured Data Mainframe
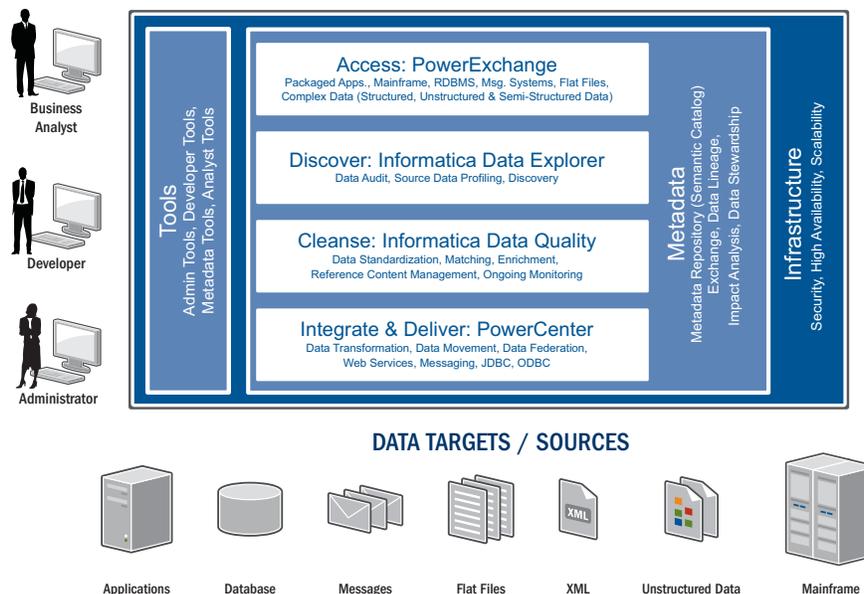
Figure 2: The Informatica Platform is an enterprise data integration and data quality solution that enables organizations to implement a comprehensive, full-lifecycle approach to addressing data quality.

Access

Informatica PowerExchange

Informatica PowerExchange automates the access and transformation of virtually all forms of data to make it available, usable, and meaningful across the enterprise. This includes virtually all forms of electronic data from legacy and ERP data to industry-specific standards to unstructured data such as email and Word documents.

Discover

Informatica Data Explorer

Informatica Data Explorer is breakthrough software that puts powerful, easy-to-use data discovery and profiling capabilities in the hands of the business specialists and IT. Data analysts and data stewards use the software to create a complete and accurate picture of the content, quality, and structure of enterprise data, which is used as the foundation for addressing data quality throughout the enterprise.

Cleanse

Informatica Data Quality

Informatica Data Quality is unique, easy-to-use software that is specifically designed to put the control of data quality processes in the hands of the business. The software delivers powerful data cleansing, matching, and reporting and monitoring capabilities in a single solution, empowering business information owners to implement and manage effective and lasting data quality processes enterprise wide. Informatica Data Quality has nearly 50 easily customizable components for parsing, standardization, and matching. The flexibility of the Informatica Data Quality Workbench enables data analysts to build data quality plans that include practically any combination of the components for optimum efficiency and effectiveness.

Integrate and Deliver

Informatica PowerCenter

Informatica PowerCenter is a single, unified enterprise data integration platform for accessing, discovering, integrating, and delivering data throughout the enterprise in any format, at any speed. 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.

Maximize Return from Enterprise Data

Achieving and maintaining high-quality data is critical to efficient IT and business operations, as well as to the success of strategic business initiatives and to your company's long-term competitive advantage. Traditionally, data integration and data quality were addressed as part of data warehousing projects, because moving data out of operational systems into the warehouse offers an opportunity to refine the data as it is moved. As the data integration technology that powers this process has increasingly been adopted to support other initiatives —such as data migration, data consolidation, data synchronization, master data management, and outsourcing — these initiatives now offer opportunities to improve data access and quality on a continuous basis.

Informatica's data quality capabilities allow organizations to analyze, standardize, validate, correct, enrich, and monitor any data source. Doing so maximizes the integrity and value of your organization's most important information assets and provides users with accurate, business-relevant information. With Informatica data integration capabilities, this data quality functionality can be deployed across the enterprise based on the robust access, connectivity, metadata management,

parallel performance, and linear scalability provided by the Informatica suite of products. Together, they deliver the infrastructure and processing power necessary to meet enterprise data quality requirements.

The Informatica platform also safeguards process quality by ensuring the use of standard definitions, providing development standards, and enforcing access security. It supports operational quality through capabilities for audit and control, system management, and error handling.

Through comprehensive and powerful data quality capabilities, Informatica is able to support a large number of data quality projects and speed the overall project lifecycle to give organizations confidence that the data they use for strategic business decisions is accurate, current, and consistent. Informatica PowerExchange, Informatica Data Explorer, Informatica Data Quality, and Informatica PowerCenter combine to form the ideal platform for addressing data quality at the enterprise level.

For more information, see the Informatica white paper "Making Data Work: Addressing Data Quality at the Enterprise Level."

Informatica Corporation delivers data integration software and services to solve a problem facing most large organizations: the fragmentation of data across disparate systems. Informatica helps organizations gain greater business value from their information assets by integrating their enterprise data. Informatica's open, platform-neutral software reduces costs, speeds time to results, and scales to handle data integration projects of any size or complexity. With a proven track record of success that extends back to 1993, Informatica helps companies and government organizations of all sizes realize the full business potential of their enterprise data. That's why Informatica is known as the data integration company.

6787 (08/26/2008)