



helping build the smart and agile business

The Importance of a Single Platform for Data Integration and Quality Management

Colin White
BI Research

March 2008
Sponsored by Business Objects

TABLE OF CONTENTS

DATA INTEGRATION AND QUALITY: UNDERSTANDING THE PROBLEM	1
The Evolution of Data Integration and Quality Software	1
Building a Single Data Services Architecture	3
Applications	3
Service-Oriented Architecture Layer	4
Data Services Techniques	5
Data Services Management and Operations	5
Choosing Data Services Products	6
BUSINESS OBJECTS DATA SERVICES PLATFORM	7
BusinessObjects Data Services XI 3.0	7
Getting Started: Success Factors	9

Brand and product names mentioned in this paper may be the trademarks or registered trademarks of their respective owners.

DATA INTEGRATION AND QUALITY: UNDERSTANDING THE PROBLEM

Companies are fighting a constant battle to integrate business data and content while managing data quality in their organizations. Compounding this difficulty is the growing use of workgroup computing and Web technologies, the storing of more data and content online, and the need to retain information longer for compliance reasons. These trends are causing data volumes to increase dramatically.

The growing number of data sources is causing data integration problems

Rising volumes are not the only cause of data integration and quality issues, however. The growing numbers of disparate systems that produce and distribute data and content also add to the complexity of the data integration and quality management environment. Business mergers and acquisitions only exacerbate the situation.

Data quality management in many companies is immature

Most organizations use a variety of software products to handle the integration of disparate data and content, and to manage data quality. Often, custom solutions are required for complex and legacy data environments. Although data integration projects have grown rapidly, budget and time pressures often lead to data quality issues being ignored by project developers. The result is that data quality management has not kept pace with the growth of data integration projects, and its use in many companies is still immature.

Vendors are now providing consolidated data integration and data quality products

Business user complaints and compliance legislation are forcing IT groups to devote more energy and resources to solving data quality problems. Nevertheless, many data quality projects are still implemented separately from those for data integration. One reason for this is that in the past data quality tools have been developed and marketed by a different set of vendors than those that supply data integration products. This has led to fractured purchasing strategies and skills development in IT groups. Vendor acquisitions and mergers have led to consolidated solutions, but product integration issues still remain.

A data services architecture is required for enterprise-wide data integration and quality management

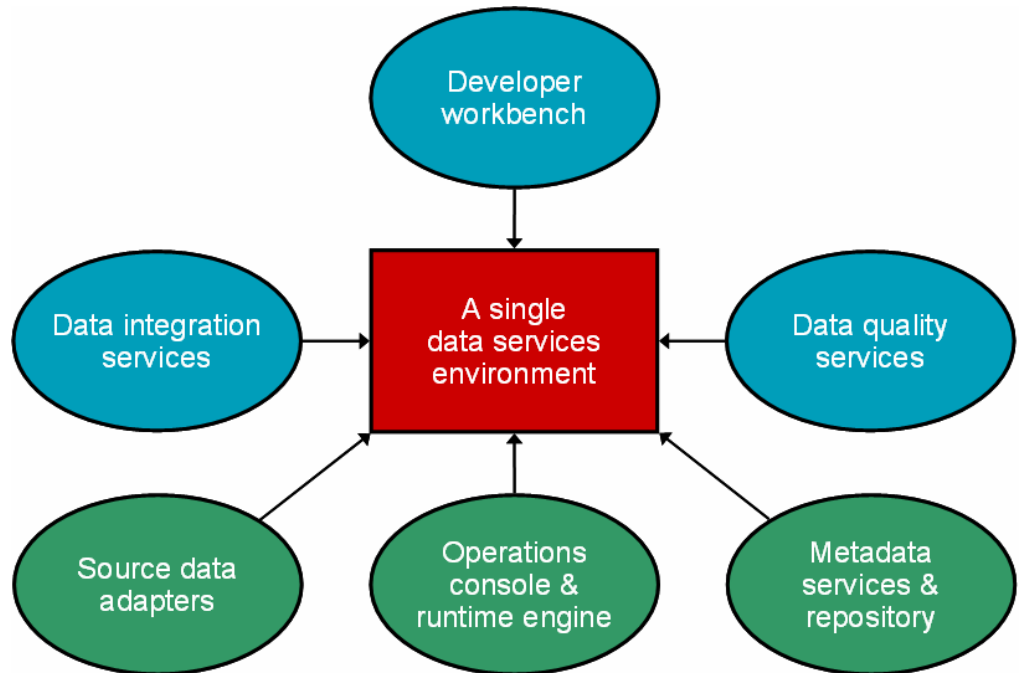
If companies are to manage the integration and quality of the ever-increasing information mountain in their organizations, they need to design and build a *data services* architecture that provides a single environment for enterprise-wide business data and content integration and quality management (see Figure 1). This paper examines the evolution of the data integration and quality industry, and explains the benefits of moving toward a single data services architecture. It outlines requirements for a software platform for supporting such an architecture, and, as an example, reviews the BusinessObjects™ Data Services XI Release 3 platform from Business Objects, an SAP company.

THE EVOLUTION OF DATA INTEGRATION AND QUALITY SOFTWARE

Although data integration and quality problems have been widespread in companies throughout the history of computing, they deteriorated noticeably when organizations moved away from centralized systems to using distributed processing involving

client/server computing, and more recently, Web-based systems. While there is no question that the move toward distributed processing systems improved access to data, which in turn enhanced business user decision-making and action-taking, it nevertheless increased the complexity of data integration and quality management tasks in organizations.

Figure 1.
Data Services: a
Single Environment
for Data Integration
and Quality
Management



Data warehousing and BI projects have helped improve data quality

Improvements in data integration and quality came with the introduction of data warehousing and business intelligence (BI). The business intelligence market has seen tremendous growth, and for many organizations business intelligence has become a key asset that enables them to optimize BI operations to reduce costs and maintain a competitive advantage. Business intelligence applications in these companies have become mission-critical because of the important role they play in the decision-making process. This reliance will grow as companies move toward using business intelligence, not only for strategic and tactical decision-making, but also for driving daily and intraday business operations.

Data integration and quality management is an enterprise-wide problem

The use of data warehousing and business intelligence has led to a much better understanding of how business data flows through the business and how it is used to make decisions. This is especially true for legacy system data, which is often poorly documented. This understanding is helping organizations deploy other data integration and quality projects that may not be directly related to business intelligence. Master data management is an example here. The result is that more organizations are now viewing data integration and quality as an enterprise-wide problem, not just an issue to be solved when building a data warehouse and business intelligence applications.

An enterprise-wide data services environment has remained elusive

Although companies have increased their spending on data integration and quality products, a single enterprise-wide data services solution has often remained elusive due to the complexity of the tasks involved, and also because of the lack of a consistent approach to information management across the enterprise.

The solution is to develop an enterprise data services architecture, deploy a single and open data services platform to support this architecture, fill any gaps in the platform with third-party or custom-built software, and gradually evolve existing data integration and data quality projects to support the new data services environment.

Six key aspects of an enterprise-wide data services architecture

The main characteristics of an enterprise-wide data services architecture are as follows:

- A single environment for data integration and data quality management
- A common developer user interface and workbench
- A single set of source data and content acquisition adapters
- Shared and reusable data integration and data quality cleansing transforms
- A single operations console and runtime environment
- Shared metadata and metadata management services

Many benefits to having a single data services environment

Although it will take time for organizations to move toward a single data services environment that supports both data integration and data quality management there are significant benefits to doing so:

- Organizations are more effective and competitive because they have access to consistent and trusted data
- IT architecture is simpler, which reduces IT maintenance and development costs
- Development cycle time is reduced due to a common data integration and data quality management environment
- Data standards are easier to enforce and maintain because data integration and data quality processes can be shared and reused across projects

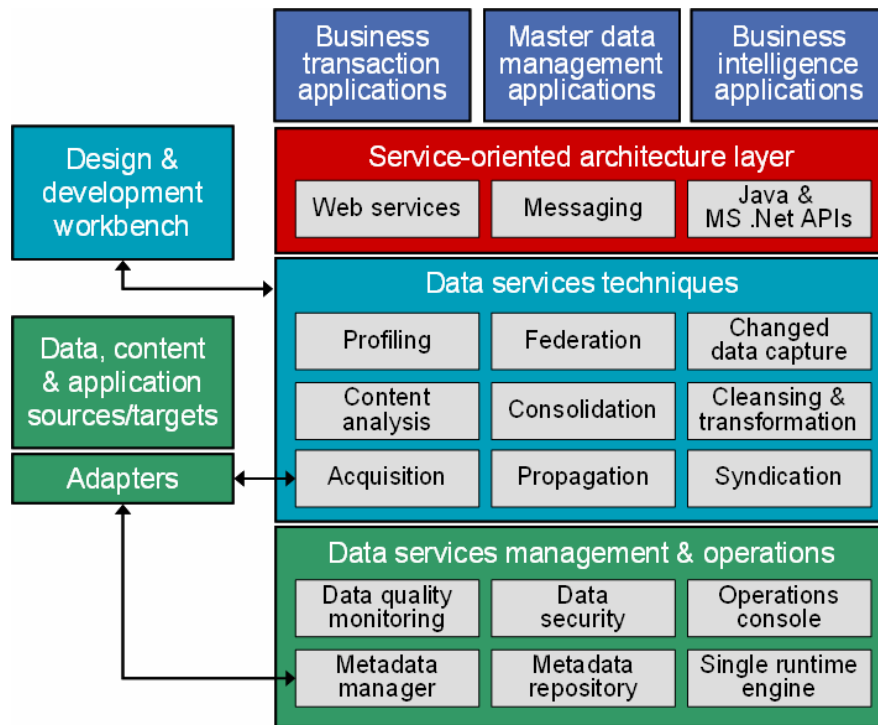
BUILDING A SINGLE DATA SERVICES ARCHITECTURE

Figure 2 illustrates the key requirements for building a single enterprise-wide data services environment. These requirements fall into four main areas: applications, application interfaces, techniques, and management.

Applications

The applications component represents those business applications that require data services for improving data quality and integrating data and content. Business transaction processing, master data management and business intelligence are key examples here. The move toward a service-oriented architecture (SOA) based on Web services is adding applications such as business content management and business collaboration to the applications mix.

Figure 2.
Data Services
Requirements



Service-Oriented Architecture Layer

Most data quality and integration projects involve batch applications that gather data from multiple sources, clean and integrate it, and then load the results into a target data file or database. With demand growing for lower-latency data and a services-based architecture, this model of data integration processing must be enhanced and made more dynamic.

Developers need a set of dynamic and shared data services

Developers now want to build applications that can use data services interactively, rather than in batch mode. Web-based business transactions applications can use such a dynamic interface to validate and correct address information as it is entered by customers. The correction of data as it enters the system, rather than after the fact, reduces data integration development effort and improves data accuracy. For companies that offer products through multiple channels, these data services can be shared by each of the applications that support those channels. The sharing of services improves data consistency, and thus data quality.

Web services is becoming an important interface

Not all IT organizations use the same development approaches or programming environments. Dynamic interfaces therefore need to support the main service-oriented architecture development approaches in use today, including Web services, messaging protocols such as JMS, and the Java and Microsoft .NET environments.

Data Services Techniques

Data profiling can be used to monitor data quality

The range of potential techniques that can be supported by a data services architecture has grown since the early days of the batch extract, transformation, and load (ETL) tools used for data warehousing. Data profiling, for example, now enables developers to examine proactively source data to determine what data quality issues need to be fixed during the data integration process. Profiling can also be used to monitor data quality on an ongoing basis.

Advanced content exploration and analysis techniques are making it easier to incorporate unstructured business content into a data integration project to enhance the value of the information delivered to business users.

Data may need to be acquired from unstructured content and packaged applications sources

Data services adapters for acquiring source data and content now support access to an increasing number of sources, regardless of whether the source is structured data, unstructured business content, an application package, or a Web service. Supplied software developer kits (SDKs) support a common architecture that enables supplied and custom-developed adapters to be shared by multiple data services components, which reduces adapter redundancy.

In addition to the standard data consolidation techniques supported by ETL products and technologies, other approaches such as data federation, propagation, and syndication are becoming equally important for supporting data integration. As data volumes increase, the ability to capture data changes to source system and propagate them to target systems is a major improvement compared to the earlier approach of taking snapshots of source data and reloading the target file or database after each snapshot.

Shared and reusable transforms are required for data integration and data cleansing

Techniques for doing data cleansing and transformation have also improved. Many products now come with prebuilt transforms for handling specific types of data quality improvement and supporting industry standard vocabularies. Address information and product description cleanup are prime examples here. When these pre-built functions are inadequate, developers can create their own reusable functions in a scripting language, or standard programming language, and then incorporate these custom functions into the system using documented and supported product interfaces.

A developer workbench is a important requirement

Given this wide range of capabilities available in products today, it is important that a data services architecture include a design and development workbench with a consistent interface for implementing each of these various techniques. The workbench then enables developers to select the technique that best suites the technical tasks and business requirements involved in any given project.

Data Services Management and Operations

Management and operations is a key distinguishing area between products

Data services management and operations covers a wide range of capabilities from metadata management and reporting to the scheduling and monitoring of data integration jobs. Key requirements here are a documented metadata architecture, a shared metadata repository for the many types of metadata involved in a data quality and integration project, and a single console and common runtime environment for all aspects of data services operations. These aspects of data services are often key distinguishing features between competing products.

Choosing Data Services Products

The data services architecture shown in Figure 2 provides a starting point for each data integration and quality management project. The actual techniques used are determined by the business and technical requirements of each project, but the data services architecture provides the basis from which these selection decisions can be made.

It is difficult to build a single data services environment by integrating best of breed products

The features of the architecture illustrated in the figure cover a broad range of capabilities and no single vendor product at present supports all of those capabilities. It is possible to build the architecture using best-of-breed products from multiple vendors, but it is becoming increasingly more difficult to do this because of overlapping functionality between products and the difficulties involved in integrating them. Metadata duplication and data services management are also major issues with this approach.

Some vendors are moving toward providing a single data services platform that encompasses many aspects of the architecture presented here, but progress varies by vendor. Many of these vendors are expanding their data services platform by acquiring other companies, and the level of integration between the products in the platform is less than perfect in many cases.

It is better to use a data services platform from a single vendor

In choosing products to support a data services architecture it is better usually to select a data services platform from a single vendor and support missing functionality by integrating best-of-breed products as required. In choosing such a platform, however, it is important to not only evaluate the current capabilities of the platform, but also to review product directions and determine how open the platform is for integrating the vendor's other products, tools from third parties, and custom capabilities. Key areas to consider here are the developer user interface and workbench, metadata management, operations management, and data source adapter architecture.

There are pros and cons to relying on a single vendor to provide IT infrastructure and services components. Vendor viability, vendor lock-in, and product cost are important considerations here. The alternative to a single vendor solution, however, is significant in-house integration development and maintenance effort, which diverts resources from building applications and supporting business needs.

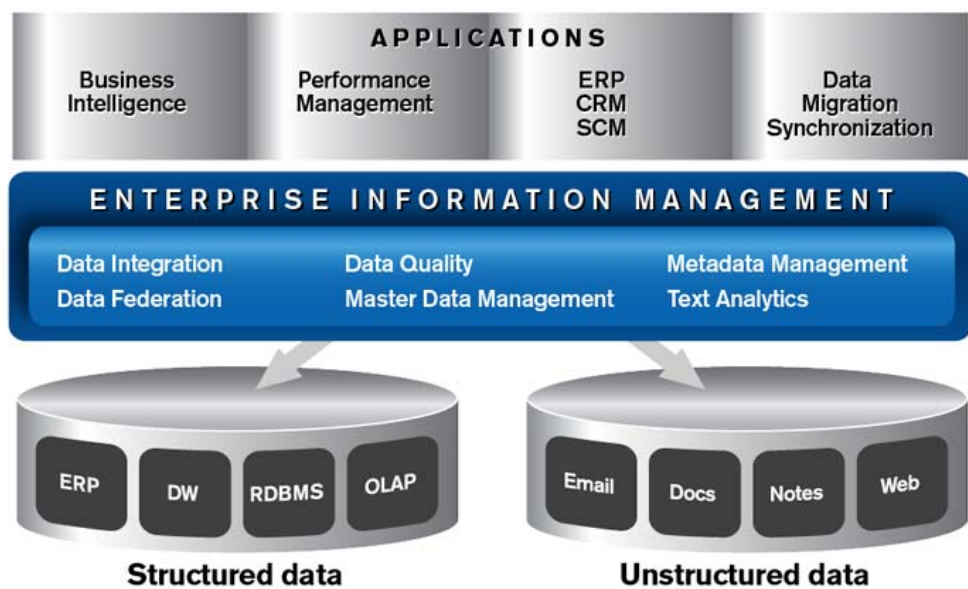
The data services platform should be open and support the integration of additional tools and components

With the current move toward product and vendor consolidation, it is becoming easier to consider one of the main vendors for software infrastructure. The important thing is to choose a vendor that provides a sound data services platform at a reasonable cost, while at the same time supporting an integrated but flexible architecture that makes it easier to add in additional tools and components.

BUSINESS OBJECTS DATA SERVICES PLATFORM

Business Objects has evolved from being an early provider of business intelligence products to offering an extensive set of tools and applications covering business intelligence, business performance management, and enterprise information management (EIM) – see Figure 3. The cornerstone of the Business Objects strategy for EIM is represented by *BusinessObjects Data Services XI 3.0*. The objective of this data services platform is to provide a single environment for supporting enterprise-wide data quality and data integration projects.

Figure 3.
The Business
Objects Product Set
for Information
Management



BUSINESSOBJECTS DATA SERVICES XI 3.0

Prior to the release of BusinessObjects Data Services, data quality and data integration processing was supported by two separate products: *BusinessObjects Data Integrator* and *BusinessObjects Data Quality*. The most important aspect of the Data Services offering is the unification of the Data Integrator and Data Quality products into a single platform. This platform offers a common developer user interface and management console, a single metadata repository, a common set of data transforms supported by a scalable runtime engine, and shared adapters for capturing data from source files, databases, and industry-leading application packages.

Figure 4.
Business Objects
EIM Support for Data
Services

	BusinessObjects Data Services XI 3.0	Related EIM Products
Applications		<i>BusinessObjects Rapid Marts™</i> - integration solutions for common ERP/CRM applications
SOA Layer	Web services (WS-I compliant) JMS-based messaging	
Techniques	Data integration capabilities: - profiling - consolidation - propagation - changed data capture - integration transforms Data quality capabilities: - cleansing transforms - Universal data cleansing option Acquisition adapters - relational DBMSs - flat files including Excel - application packages - mainframe systems (via Attunity technology integration) SDK scripting languages - Python - proprietary	<i>BusinessObjects Data Insight</i> – additional profiling capabilities <i>BusinessObjects Data Federator</i> - federation <i>BusinessObjects Text Analysis</i> - content analysis
Management and Operations	Desktop development UI and workbench Metadata repository (MySQL or user selected RDBMS) Web-based management console Parallel runtime engine for online and batch jobs with grid computing support	<i>BusinessObjects Data Insight</i> - data quality assessment and monitoring <i>BusinessObjects Metadata Management</i> – open metadata management facility for Business Objects solutions and other products

The data integration and data quality features can be purchased separately for a phased implementation

Although Data Services is a single platform, Business Objects customers may separately purchase the data quality (with or without the universal data cleanse) and data integration functionality of the platform. This provides a very flexible development environment because customers can start with the platform's data integration capabilities and then easily add its data quality features, and vice versa.

Figure 4 shows the main capabilities of Business Objects EIM software and Data Services XI 3.0. As the figure shows, the Data Services platform provides a significant number of the requirements outlined in Figure 2. Optional products from Business Objects fill in most of the remaining gaps. Data Insight extends the data profiling capabilities of Data Services, while Data Federator and Text Analysis extend its data integration support. In addition, Metadata Management provides a open metadata environment not only for handling metadata from Data Services, but also from Business Objects BI solutions and third-party products.

Other EIM products will be added to Data Services XI in future releases

The architecture of Data Services has been designed so that it provides a solid foundation for future product development. This architecture enables products such as Data Insight, Data Federator and Text Analysis to be integrated and take advantage of the capabilities of the Data Services platform in subsequent releases.

In summary, BusinessObjects Data Services XI 3.0 offers a single integrated platform that brings together two key Business Objects products Data Integrator and Data Quality. This platform offers customers a scalable and open environment for supporting enterprise-wide data integration and data quality management.

GETTING STARTED: SUCCESS FACTORS

Data integration and data quality are not separate entities

The key to success in data integration projects is to consider data integration and data quality management as a single entity. In the past this has not been the case, which is one reason why data quality management in many organizations has not kept pace with changes in data integration requirements and technologies.

An integration competency center reduces project risks and developments costs

Another important success factor is the creation of an integration competency center to bring together the key data integration, data quality management, and applications integration IT staff of the organization. The center reduces political, budget, and technology issues, and enables the sharing and reuse of technical skills and best practices, which reduces project risks and reduces IT costs. The competency center is responsible for designing the data services architecture, selecting the data services platform, and helping application and line-of-business groups implement and gradually move toward using a single data services environment.

Data integration projects require senior management support and funding

As with any IT project, there is more to solving data integration and data quality management problems than simply installing software and implementing new technologies. Senior management needs to realize data is a valuable corporate asset and a sound data integration and data quality management strategy is the key to leveraging this asset for business success. Data integration projects must be funded adequately, and senior management must endorse and fully support the move toward a single data services environment managed by a data integration competency center.

About BI Research

BI Research is a research and consulting company whose goal is to help companies understand and exploit new developments in business intelligence and business collaboration. When combined, business intelligence and business collaboration enable an organization to become a smart and agile business.

BI Research

Post Office Box 398

Ashland, OR 97520

Telephone: (541)-552-9126

Internet URL: www.bi-research.com

E-mail: info@bi-research.com

The Importance of a Single Platform for Data Integration and Quality Management

Version 1, March 2008

Copyright © 2008 by BI Research

All rights reserved