
Data Quality First: It's Just Logical

- ▶ *The Importance of Trusted Information to Successful Data Integration and Business Intelligence*

Author: Mike Schiff

Contributors: Michelle Bergevin, Darren Cunningham, Jennifer Meegan, Philip On,
David Townley

Contents

Introduction	iv
Warning Signs: Does Your Company Suffer from Poor Data Quality?	1
Data Quality and Data Integration	3
Elements of a Successful Data Quality Solution	4
The Business Objects – Firstlogic Combination	6
Industry Reactions	7
Summary and Conclusions	8
About MAS Strategies	9
About Business Objects	9

Introduction

Trustworthy information is based on trustworthy data. And in order to better serve customers, increase profitability and revenue growth, improve operations, comply with government regulations and reporting requirements, and make better decisions—organizations need to access, integrate, and analyze data from a variety of sources. For this to be effective, the data must be accurate. In fact, the reliability of analysis is directly related to the quality of the underlying data. This is simply common sense and is reflected in one of the best-known adages of the computer industry: “garbage in, garbage out.”

The importance of data quality should come as no surprise. Almost all manufacturing and service organizations have strong quality control functions to minimize the possibility of selling defective goods or providing substandard services. Ensuring high-quality data for an organization’s operations and decision-making processes is a logical extension of the quality control function.

Innovative organizations recognize the importance of building an integrated, high-quality data foundation and have adopted enterprise information management (EIM) initiatives to establish a solid data foundation—one that is timely, integrated, and trustworthy. A comprehensive EIM strategy will ensure accurate and trusted information availability while providing a holistic view of the organization. Data quality and data integration are cornerstones of any EIM initiative.

The lack of a trustworthy data foundation can lead to serious and expensive problems. In a January 2006 InformationWeek research survey of business technology professionals and their plans to expand the deployment of business intelligence (BI) tools within their organizations, 38% responded that more than half of their company’s employees will use BI tools within the next two years. As to why more employees are not currently using BI tools, 51% cited integration issues with existing systems and 45% cited data quality issues.

In The Data Warehousing Institute’s (TDWI) 2002 “Data Quality and The Bottom Line” report, it was estimated “that poor quality customer data costs U.S. businesses a staggering \$611 billion a year in postage, printing, and staff overhead.” This is in addition to the even greater costs associated with alienating customers or perhaps even losing them altogether. On April 4, 2006 TDWI published a follow-up study, “Taking Data Quality to the Enterprise through Data Governance,” in which 53% of the respondents surveyed answered “yes” to the question: “Has your company suffered losses, problems, or costs due to poor quality data?”

Warning Signs: Does Your Organization Suffer from Poor Data Quality?

Organizations often struggle to understand many business problems may be directly related to poor data quality. The following list demonstrates common situations that could benefit from using data quality technology.

Inability to compare data from different sources.

As a result of an out-of-stock condition for a critical part, one division of an organization must reorder the part at a premium price. After the parts arrive, the organization discovers another division, which uses a different part numbering scheme, has excess inventory of the same part.

Data entered into the wrong fields.

When it came time to generate year-end 1099-INT tax statements, a bank discovered many of its records did not have entries in the social security number field. After reviewing the records, the bank discovered while each record contained a social security number, the number was sometimes entered into fields intended for phone numbers, in-care-of, or even as part of a free-form name field.

Lack of consistent data definitions.

In order to improve supervisory skills, a company's training department developed compulsory courses for all of its managers. However, one department used the "manager" title as a way of promoting individual contributors and had to do a lot of explaining when it tried to get these employees exempted from the compulsory supervisory training.

Inability to consolidate data from multiple sources.

A multi-divisional company keeps an internal list of its top 25 accounts; these customers receive preferred treatment anytime they contact the company. The COO of the company's tenth largest account complains to the vice president of sales that her people are being treated very rudely whenever they contact the company. The vice president of sales checks the internal list and finds that the customer's organization is not on it. It was later determined that the company did business with four separate divisions of the organization, each of which identified the organization by a different name and different account number.

Inability to track data across time.

A major women's clothing retailer assigned an SKU number to green sweaters in a particular line and size, but then used this same number to represent a different shade of green each season; for example hunter green in the winter season, mint green in the spring. Worse yet, the SKU color shade coding was sometimes not even consistent for the same season in two different years. This greatly limited its ability to analyze sales by color, especially between seasons when the company's retail stores could concurrently carry both shades.

Inability to comply with government regulations.

A financial institution was concerned about possible penalties and sanctions associated with being unable to comply with the "know thy customer" rules. They had no program to identify—

through interdiction list matching—possible terrorist-related financial activities and money laundering schemes, and therefore were non-compliant with regulations imposed by the U.S. Patriot Act and the Office of Foreign Assets Control (OFAC) Bank Secrecy Act. The organization considered this to be a substantial risk requiring an immediate solution.

Delayed or rejected reimbursement from third party providers.

A hospital that prided itself on its strong emergency patient care was having cash flow issues when reimbursements from insurance companies were being delayed or even denied because it was submitting claims with incorrect patient ID numbers or incorrect medical procedure and medical supply codes. It was discovered that emergency room personnel, in an effort to expedite patient care, were using codes such as “99999” and not adequately checking insurance cards. Once this was corrected, reimbursements picked up and cash flow greatly improved. A similar situation existed in the service department of an automobile dealership because warranty work was frequently miscoded and the automobile manufacturer was therefore rejecting valid claims.

The inability to determine important relationships.

While from a customer relationship management (CRM) perspective much attention has been placed on “householding,” or the ability to link members of the same household together and market to them as a unit, the Department of Homeland Security has Border Patrol initiatives requiring it to determine if two individuals were members of the same “social network.” In order to accomplish this, it was necessary to go beyond traditional name-and-address matching and examine additional criteria such as passport IDs, phone numbers, email addresses, retail transactions, bank routing information, and other data. While some companies speak of “bet your business” applications, data quality technology was successfully deployed in what was most certainly a “bet your borders” operation.

Data Quality and Data Integration

Data quality relates to the correctness, consistency, and completeness of an organization's data. While data correctness is of paramount importance, data must also be complete and consistent. For example, to determine total sales revenue for a multi-divisional organization, the data from every division must be included and the numbers consolidated. However, if one division reported in dollars and another in pounds or euros, it would first be necessary to convert each division's revenues to a common unit-of-measure.

Data integration is of obvious importance in a data warehouse environment as most data warehouses integrate data from many sources. Data integration is also important in operational environments. While the number of integration points may be less in an operational environment, most application systems are fed from, and feed to, other application systems.. Just as a chain is only as strong as its weakest link, integrated data is only as valid as its least accurate component.

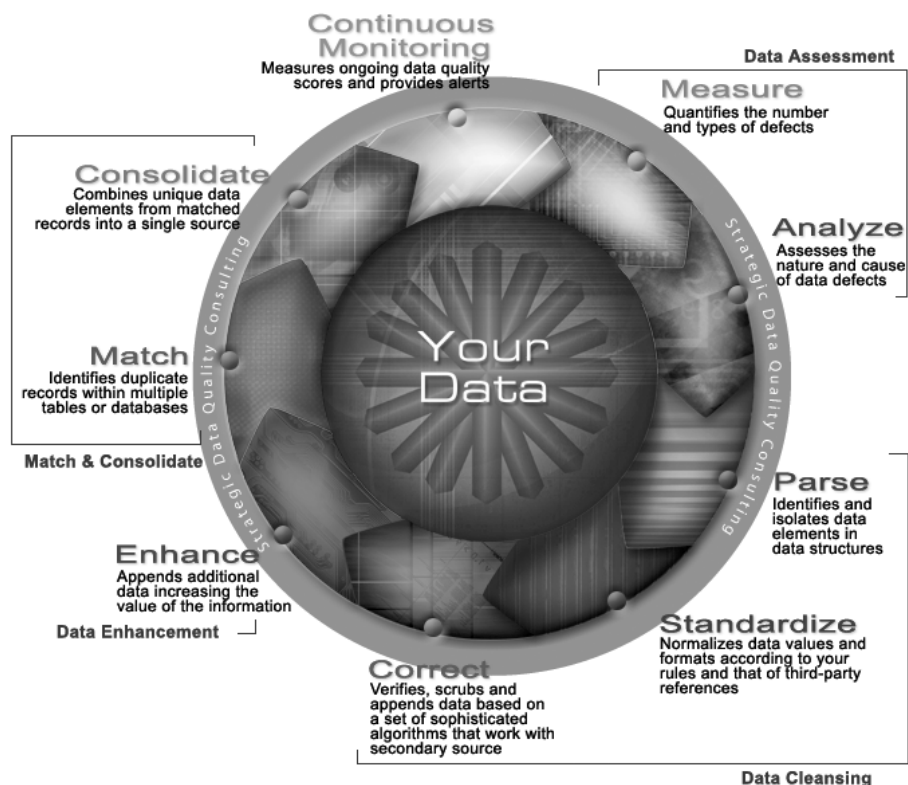
While successful data integration involves consolidating data from multiple sources and is highly dependent on the quality of the source data, data quality is a necessity even for a single data source. High-quality data is a necessity in both analytical and operational environments. No one would trust a data warehouse if it contained erroneous data. And organizations that ignore data quality when populating their warehouses run the risk of building a data dump. On the operations side, an organization's operations are likely to run amuck if the underlying application databases contained incorrect data values such as invalid part numbers, erroneous shipping addresses, or incorrect prices.

Elements of a Successful Data Quality Solution

Data quality is an ongoing discipline, not a one-time task. In order to address data correctness, completeness, and consistency, a data quality solution must include tools and technology, services, and methodologies. A data quality solution consists of data profiling to determine existing status, data cleansing and validation to identify and correct any errors, data auditing, and continuous monitoring for quality assurance.

A data quality solution should encompass the following data quality processes:

- ▶ **Measuring:** Quantifying the number and types of defects
- ▶ **Analyzing:** Assessing the nature and cause of data defects
- ▶ **Parsing:** Identifying and isolating data elements in data structures
- ▶ **Standardizing:** Normalizing data values, units-of-measures, and formats according to business rules and third-party reference data
- ▶ **Correcting:** Verifying and correcting data values, and appending data based on a set of algorithms that work with secondary data sources
- ▶ **Enhancing:** Appending additional data (e.g., credit ratings, demographics, geocoding information, email addresses, etc.) to existing data in order to increase its overall utility
- ▶ **Matching:** Identifying duplicate records within the same or even differing databases
- ▶ **Consolidation:** For a given entity occurrence, consolidating unique data elements from multiple records into a single master data source
- ▶ **Reporting:** Presenting data quality information at every step of the process



A data quality solution should make use of best practices and common sense. For example:

- ▶ Correcting potential errors at their source to ensure valid entries are being made within transaction systems.
- ▶ Designing data quality into operational systems so, for example, values are entered into the appropriate record fields (e.g., separate first-name, middle initial, last-name fields rather than one freeform field).
- ▶ Establishing common data definitions across the organization so everyone speaks the same language. For example, if two employees each work 20 hours a week, do they represent one headcount or two? If two divisions define this differently, can you really compare their revenue/headcount metrics?
- ▶ Being aware that data quality is not limited to name-and-address data but extends to all of an organization's data.
- ▶ Identifying master data or reference files (e.g., products, customers, employees, vendors, financial accounts, etc.) are common to many operational and analytical systems, and proactively establishing master data management initiatives to provide a single consolidated and trustworthy view of this data that can be shared across these systems.
- ▶ Continuous monitoring and auditing applications, alerts, and analytical systems to ensure data quality is being maintained and not deteriorating. Finding the cause of new errors and taking corrective measures. Continuous monitoring is also necessary as many data values are subject to change. For example, according to the United States Postal Service, more than 44 million Americans change their addresses each year. This makes address data that was once valid, now incorrect.

The Business Objects – Firstlogic Combination

As leading business intelligence and data integration vendor, Business Objects has long recognized the importance of data quality and has offered data cleansing as an integrated option within BusinessObjects™ Data Integrator, an industrial-strength extraction, transformation, and loading (ETL) tool. Data quality capabilities such as cleansing and match merge are delivered through tight integration with Firstlogic's market-leading data quality platform. Because the success of any data integration effort is directly dependent on the quality of the data being integrated, this combined solution allows organizations to audit, validate, cleanse, standardize, enhance, and consolidate data during ETL processing.

Like its prior acquisition of Crystal Decisions, Business Objects' acquisition of Firstlogic is certainly a "marriage of winners." The acquisition was extremely logical for Business Objects as Firstlogic's robust data quality solution clearly supports the company's focus on EIM. While the existing Data Integrator built-in data quality features were successful, the acquisition of Firstlogic will allow Business Objects to directly control the future direction and evolution of Firstlogic's data quality technology and even more deeply integrate it into Business Objects' overall product portfolio. Consequently, Business Objects' customers will be able to use a single vendor, "best-of-breed" data integration, data quality, and BI solution. Customers will benefit not only from superior functionality, but also from the lower cost of ownership, faster and easier implementations, broader deployments, and reduced training requirements that standardizing on a single-vendor can provide.

In addition to its product set, Firstlogic is a data quality thought leader as evidenced by its charter sponsorship of the Massachusetts Institute of Technology's Total Data Quality Management program, and has considerable expertise and experience in many industries including education, financial services, government, healthcare, and retail. It offers a full complement of educational, consulting, and implementation services.

Firstlogic's functionality extends well beyond the limited name-and-address capabilities of some of its competitors and includes pattern matching and recognition features for use with, for example, product codes, SKUs, vehicle identification numbers, phone numbers, email addresses, social security numbers, or business-rules defined user data fields. Firstlogic provides data enhancement options such as email and phone append, geocoding, and move-related change of address updates. It also provides interdiction list matching for use in tracking client records against the Office of Foreign Assets Control (OFAC) list.

Business Objects clearly recognizes data quality is an important enabler of EIM. While EIM includes data quality as part of a data integration process, it goes well beyond that. A trustworthy data foundation requires data quality supporting both enterprise applications and transaction systems. Accordingly, Business Objects is committed to continuing to offer Firstlogic's data cleansing technology as a standalone offering. Customers who have previously acquired data integration technology other than BusinessObjects Data Integrator for ETL—or perhaps even developed their own in-house solutions—will still be able to derive substantial benefit from the Business Objects data quality solutions (from Firstlogic).

Industry Reactions

The industry reaction to Business Objects acquisition of Firstlogic has been overwhelmingly positive as evidenced by comments from the analyst community. Industry analyst comments include:

Keith Gile, principal analyst at Forrester Research in an interview with SearchDataManagement.com (Update: Business Objects sweeps up Firstlogic, By Hannah Smalltree, News Writer, Feb 8 2006 SearchDataManagement.com):

Keith Gile, principal analyst with Cambridge, Massachusetts-based Forrester Research Inc., agreed the acquisition will help current Business Objects customers and said it could potentially help the company win deals against large competitors like Microsoft, Oracle Corp., and IBM. This acquisition puts pressure on everyone, from other BI vendors to ETL vendors, and will have "interesting and huge ripple effects."

"BI isn't just about reporting," Gile said. "The value of the reporting and analysis that is done is dependent on the quality of the data that's being analyzed. This [acquisition] makes Business Objects more attractive to more buyers. Now their data integration mechanism is stronger."

Philip Russom, senior manager of research at The Data Warehousing Institute:

"Unlike other business intelligence vendors, Business Objects has a deep commitment to data integration, as seen in its world-class ETL tool and recent acquisition of enterprise information integration (EII) technology. But data integration inexorably reveals data quality issues, whether problems to be fixed or opportunities to be leveraged. Business Objects' acquisition of Firstlogic recognizes that best practices demand tight interoperability between data integration and data quality."

Robert Lerner, senior analyst at Heavy Reading:

"Business Objects' acquisition of data-quality vendor Firstlogic offers a lot of benefits for Business Objects. The company gains leading data-quality technology to augment its EIM efforts, which will enhance the value of these efforts while, at the same time, put pressure on some of its major competitors to respond. Furthermore, given Firstlogic's position in the data-quality market, Business Objects can now make a play in that important market from a position of strength."

Dan Vesset, research Director, Analytics and Data Warehousing at IDC:

"My reaction to Business Objects' acquisition of Firstlogic is two-fold. First, customers are increasingly looking for a more complete stack of functionality in the overall business analytics space and adding data quality to their data integration software is a step in the right direction. Secondly, as Business Objects is looking to expand its presence in the packaged analytic applications space, the data quality component will serve as a strong basis for an entry into the CRM analytics market. In this case, I think the Firstlogic acquisition will serve as a building block for other analytic applications in the future."

Summary and Conclusions

Most organizations recognize the importance of data integration and data quality for their operational and analytical systems. While “quality is job number one” may have once been a motto of an automobile manufacturer, it is definitely a major requirement of any operational application or data warehouse.

In an operational environment no one wants to ship the wrong part to the wrong address, deliver 100 tons of a product when only 100 pounds were ordered, provide a patient with the wrong medication, or deposit a wire transfer into the wrong bank account. In a data warehouse environment, no one wants to make decisions based on incorrect, incomplete, or inconsistent data. The deployment of data quality products and services can ensure improved accuracy in operations and a high level of trust in information that comes from them.

Long a leader in both the business intelligence and data integration markets, Business Objects has offered data cleansing as an integrated option with BusinessObjects Data Integrator. Capitalizing on the success of this offering, and wishing to increase its overall investment in data quality, Business Objects decided to take its relationship with its OEM partner, Firstlogic, to the next step: it acquired the company. While the acquisition is beneficial to Business Objects as it will allow it to own this important technology, it is also beneficial to Firstlogic and its customer base who now find their technology in the hands of a highly successful vendor with extensive resources to invest in its continuing evolution. Given Business Objects’ successful history of integrating acquired products and companies such as Acta Technologies and Crystal Decisions into its fold, the successful integration of Firstlogic is a certainty.

While many organizations speak about their desire to implement systems that consist of multi-vendor “best-of-breed” components, they are also aware of the problems this can create when something goes wrong; each vendor can point the finger elsewhere. Business Objects recognizes the need for, and benefits associated with, standardization. Now users can obtain an integrated best-of-breed data integration, data quality, and business intelligence solution from a single source and focus on making better decisions, rather than trying to fit all the pieces together or worrying about whom to call if there is a problem.

Business Objects is focusing its efforts, and its acquisitions, on building out its EIM framework. The Firstlogic acquisition is a quality move that provides a critical component of that framework. After all, organizations need to focus their efforts on their on-going operations and on making better decisions; they should be creating a trustworthy data foundation, not wasting time trying to reconcile “multiple versions of the truth.”

About MAS Strategies

Michael A. Schiff is the founder and principal analyst of MAS Strategies. MAS Strategies specializes in helping vendors market and position their business intelligence and data warehousing products in today's highly competitive market. He has over 30 years of experience in the information technology industry.

Michael was the vice president of the Data Warehousing and Business Intelligence service at Current Analysis, Inc., an industry analyst firm where he provided tactical market intelligence and analysis while managing the company's E-Business analyst team. He was the executive director - Data Warehousing and Advanced Decision Support for Oracle Corporation's Public Sector Group and director of software AG's Data Management program. In 1984, while at Digital Equipment Corporation, he formulated the architecture for one of the first successful data warehouse implementations. In previous positions as IT director and systems and programming manager he acquired practical, first-hand, knowledge of the technical, business, and political realities that must be addressed for any successful systems implementation or product launch.

Michael earned his Bachelor and Master of Science degrees from MIT's Sloan School of Management where he specialized in operations research as an undergraduate, and in information systems as a graduate.

For further information about MAS Strategies, visit its web site at: www.mas-strategies.com.

About Business Objects

Business Objects is the world's leading business intelligence software company. With more than 35,000 customers worldwide, including over 80% of the Fortune 500, Business Objects helps organizations gain better insight into their business, improve decision making, and optimize enterprise performance. The company's business intelligence platform, BusinessObjects XI, offers the BI industry's most advanced and complete platform for performance management, planning, reporting, query and analysis, and enterprise information management. BusinessObjects XI includes Crystal Reports®, the industry standard for enterprise reporting. Business Objects has built the industry's strongest and most diverse partner community, and also offers consulting and education services to help customers effectively deploy their business intelligence projects.

► www.businessobjects.com

For a complete listing of our sales offices, please visit our web site.

Business Objects owns the following U.S. patents, which may cover products that are offered and licensed by Business Objects: 5,555,403; 6,247,008 B1; 6,578,027 B2; 6,490,593; and 6,289,352. Business Objects and the Business Objects logo, BusinessObjects, Crystal Reports, Crystal Enterprise, Crystal Analysis, WebIntelligence, RapidMarts, and BusinessQuery are trademarks or registered trademarks of Business Objects SA or its affiliated companies in the United States and other countries. All other names mentioned herein may be trademarks of their respective owners. Copyright © 2006 Business Objects. All rights reserved.

