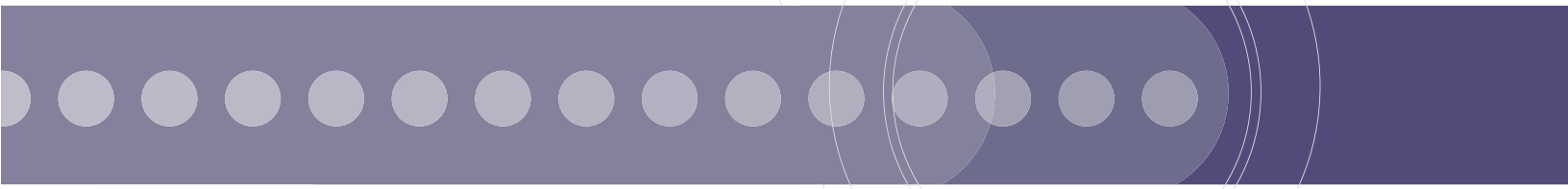
INFORMATICA®

WHITE PAPER

The Data Quality Business Case: Projecting Return on Investment

with David Loshin



This edition published June 2006

Table of Contents

ABOUT THE AUTHOR

David Loshin is the president of Knowledge Integrity, Inc., a consulting and development company focusing on customized information management solutions including information quality solutions consulting, information quality training and business rules solutions. Loshin is the author of Enterprise Knowledge Management - The Data Quality Approach (Morgan Kaufmann, 2001) and Business Intelligence - The Savvy Manager's Guide and is a frequent speaker on maximizing the value of information.

www.knowledge-integrity.com
loshin@knowledge-integrity.com

Executive Summary

The purpose of this White Paper is to outline the importance of data quality in today's business environment. It describes how an organization should tackle a data quality improvement process and where Informatica data quality software solutions fit into that process.

There is little doubt that the need for high quality data permeates most information-centric programs, whether they depend on transactional, operational, or analytical applications. In addition, new events take place and evolving techniques are introduced that affect the way our systems and business operate, all of which are dependent on the best use of quality data, such as:

- Regulatory compliance, in which organizations are required to account for the quality of the data they use and information they disseminate,
- Reengineering, Migration, and Modernization projects, in which older applications are updated and the legacy system data is migrated into new applications,
- Mergers and Acquisitions, in which multiple source data systems are merged into a single operational framework,
- Data Integration programs, such as Customer Data Integration, Product Data Integration, or any other Master Data Management program

Even though everyone fundamentally understands the need for high quality data, technologists are often left to their own devices when it comes to ensuring the high levels of data quality. However, at some point an investment must be made in the infrastructure necessary to provide measurably acceptable levels of data quality. In order to justify that investment, we must be able to articulate the business value of data quality in a way that will show a return on the investment made.

Often, developing a business case for the prevention of impacts may appear to be a challenge. Yet, even in the absence of critical events that necessitate senior management action, there is usually more than enough evidence available within an organization to develop a business case justifying the costs related to data quality improvement. This white paper will present a process for:

- Identifying key business dimensions impacted by poor data quality
- Reviewing approaches that can be used to improve data quality
- Assessing the actual historical costs related to data flaws
- Determining the costs to improve data quality
- Assembling this material into a business case justifying an investment in data quality tools and methodologies

We will look at case studies where these approaches are put into practice, and then conclude by summarizing the approach to developing a successful business case for investing in data quality improvement.

Establishing the Value of Information Quality

Business Expectations and Data Quality

There is a common notion that objective data quality improvement necessarily implies business value, and this notion often drives "golden copy," "single source of truth," or master data projects. This approach, though, does not take into account the fact that data quality is subjective, and relies on how data flaws are related to negative business impacts. Objective data quality metrics may not necessarily be tied to your business's performance, and raises some interesting questions:

- How do you distinguish high impact from low impact data integrity issues?
- How do you isolate the source of the introduction of data flaws to fix the process instead of correcting the data?
- How do you correlate business value with source data integrity?
- What is the best way to employ data integration best practices to address these questions?

This challenge can be characterized by a fundamental distinction between data quality expectations and business expectations. Data quality expectations are expressed as rules measuring aspects of the validity *of data values*:

- What data is missing or unusable?
- Which data values are in conflict?
- Which records are duplicated?
- What linkages are missing?

Alternatively, business expectations are expressed as rules measuring performance, productivity, efficiency of processes, asking questions like:

- How has throughput decreased due to errors?
- What percentage of time is spent in scrap and rework?
- What is the loss in value of transactions that failed due to missing data?
- How quickly can we respond to business opportunities?

To determine the true value added by data quality programs, conformance to business expectations (and the corresponding business value) should be measured in relation to its component data quality rules. We do this by identifying how the business impacts of poor data quality can be measured as well as how they relate to their root causes, then assess the costs to eliminate the root causes. Characterizing both our business impacts as well as our data quality problems provides a framework for developing our business case.

Key Data Quality Dimensions

To be able to correlate data quality issues to business impacts, we must be able to both classify our data quality expectations as well as our business impact criteria. In order for the analyst to determine the scope of the underlying root causes and to plan the ways that tools can be used to address data quality issues, it is valuable to understand these common data quality dimensions:

- Completeness: Is all the requisite information available? Are data values missing, or in an unusable state? In some cases, missing data is irrelevant, but when the information that is missing is critical to a specific business process, completeness becomes an issue.
- Conformity: Are there expectations that data values conform to specified formats? If so, do all the values conform to those formats? Maintaining conformance to specific formats is important in data representation, presentation, aggregate reporting, search, and establishing key relationships.
- Consistency: Do distinct data instances provide conflicting information about the same underlying data object? Are values consistent across data sets? Do interdependent attributes always appropriately reflect their expected consistency? Inconsistency between data values plagues organizations attempting to reconcile between different systems and applications.
- Accuracy: Do data objects accurately represent the “real-world” values they are expected to model? Incorrect spellings of product or person names, addresses, and even untimely or not current data can impact operational and analytical applications.
- Duplication: Are there multiple, unnecessary representations of the same data objects within your data set? The inability to maintain a single representation for each entity across your systems poses numerous vulnerabilities and risks.
- Integrity: What data is missing important relationship linkages? The inability to link related records together may actually introduce duplication across your systems. Not only that, as more value is derived from analyzing connectivity and relationships, the inability to link related data instance together impedes this valuable analysis.

Using Data Quality Technology to Improve Your Data

Understanding the key data quality dimensions is the first step to data quality improvement. Being able to segregate data flaws by dimension or classification allows analysts and developers to apply improvement techniques using data quality tools to improve both your information, and the processes that create and manipulate that information. Let's briefly examine some data quality techniques, then review the kinds of tools employed for these techniques.

Data Quality Techniques

There are many policies and procedures that can be employed for ongoing, proactive data quality improvement. However, most successful programs make use of some simple techniques that enable the discovery, assessment, remediation, and reporting of baseline measurements and ongoing improvement.

Anomaly Analysis and Assessment

Before any improvements can be made to information, one must first be able to distinguish between "good" and "bad" data. The attempt to qualify data quality is a process of analysis and discovery. The analysis involves an objective review of the data values populating data sets through quantitative measures and analyst review. While a data analyst may not necessarily be able to pinpoint all instances of flawed data, the ability to document situations where data values look like they don't belong provides a means to communicate these instances with subject matter experts whose business knowledge can confirm the existence of data problems.

Data Standardization

Many data issues are attributable to situations where slight variance in representation of data values introduces confusion or ambiguity. For example, consider the different ways telephone numbers are formatted in Figure 1. While some have digits, some have alphabetic characters, and all use different special characters for separation, we all recognize each one as being a telephone number.

301-754-6350
(301) 754-6350
301.753.6350
1-866-BIZRULE
866 249-7853

Figure 1: Variant formats for representing telephone numbers

But in order to determine whether these numbers are accurate (perhaps by comparing them to a master customer directory), or to investigate whether duplicate numbers exist when there should be only one for each supplier, the values must be parsed into their component segments (area code, exchange, and line) and then transformed into a standard format.

Similarity Analysis (matching)

A common data quality problem involves two sides of the same coin: when there are multiple data instances that actually refer to the same real-world entity, or the perception by a knowledge worker or application that a record does not exist for a real-world entity when in fact it really does. These problems both are a result of approximate duplication. In the first situation, similar, yet slightly variant representations in data values may have been inadvertently introduced into the system, while in the second situation, a slight variation in representation prevents the identification of an exact match of the existing record in the data set.

Both of these issues are addressed through a process called similarity analysis or matching, in which the degree of similarity between any two records is scored, most often based on weighted approximate matching between a set of attribute values between the two records. If the score is above a specific threshold, the two records are deemed to be a match, and are presented to the end client as most likely to represent the same entity. It is through similarity analysis that slight variations are recognized and data values are connected, and subsequently cleansed.

Data Quality Auditing and Reporting

It is difficult to improve a process without having a means to measure that process. In addition, it is difficult to gauge continuous improvement without being able to track performance on a regular basis. To this end, it is necessary to define relevant data quality metrics that can be measured through an auditing process, with the results captured and reported to the relevant stakeholders.

Data Quality Tools

To address these remediation needs, we can employ the following data quality tools/technologies to achieve our quality objectives:

Data Profiling

Data profiling is a set of algorithms for statistical analysis and assessment of the quality of data values within a data set, as well as exploring relationships that exists between value collections within and across data sets. For each column in a table, a data profiling tool will provide a frequency distribution of the different values, providing insight into the type and use of each column. Cross-column analysis can expose embedded value dependencies, while inter-table analysis explores overlapping values sets that may represent foreign key relationships between entities, and it is in this way that profiling can be used for anomaly analysis and assessment.

Data profiling can also be used to proactively test against a set of defined (or discovered) business rules. In this way, we can distinguish those records that conform to our defined data quality expectations and those that don't, which in turn can contribute to baseline measurements and ongoing auditing for data quality reporting.

Parsing and Standardization

Our innate ability to recognize familiar patterns contributes to our ability to characterize variant data values belonging to the same abstract class of values. Continuing our example in Figure 1, people recognize these all as telephone numbers because these are all frequently used patterns. Luckily, if we can describe the format patterns that can be used to represent all other data

objects (e.g. Person Name, Product Description, etc.), we can use a data quality tool to parse data values that conform to any of those patterns and even transform them into a single, standardized form that will simplify the assessment, similarity analysis, and cleansing processes. Pattern-based parsing can automate the recognition and subsequent standardization of meaningful value components.

Similarity and Linkage

Attempting to compare each record against all the others to provide a similarity score is not only ambitious, but also time-consuming and computationally intensive. Most data quality tool suites use advanced algorithms for blocking records that are most likely to contain matches into smaller sets, whereupon different approaches are taken to measure similarity. Identifying similar records within the same data set probably means that the records are duplicated, and may be subjected to cleansing and/or elimination. Identifying similar records in different sets may indicate a link across the data sets, which helps facilitate cleansing, knowledge discovery, reverse engineering, and master data aggregation.

Auditing and Monitoring

The value of using these tools to quantify the existence of data flaws is increased when there is a well-defined framework for collecting and reporting those data quality statistics. Some tools interface well with standard query/reporting tools to populate a data quality dashboard that can be drilled-through by data quality analysts for root cause analysis and remediation.

Identifying Impacts of Poor Information Quality

Fundamentally, the return on your DQ investment is based on the real pains incurred by data flaws in running your business. If the goal of your business is to optimize productivity and profitability while minimizing costs and risks, then we can characterize business impacts across these four dimensions:

- Increased Costs
- Decreased Revenues
- Decreased confidence
- Increased Risk

Our goal is to maximize the value of the information based on impacts associated with each dimension, and our task in developing the business case is to determine when and where poor information quality affects one or more of these variables. These impacts are summarized in Figure 2. This characterization allows one to classify impacts and subsequently determine a formula for assessing actual costs.

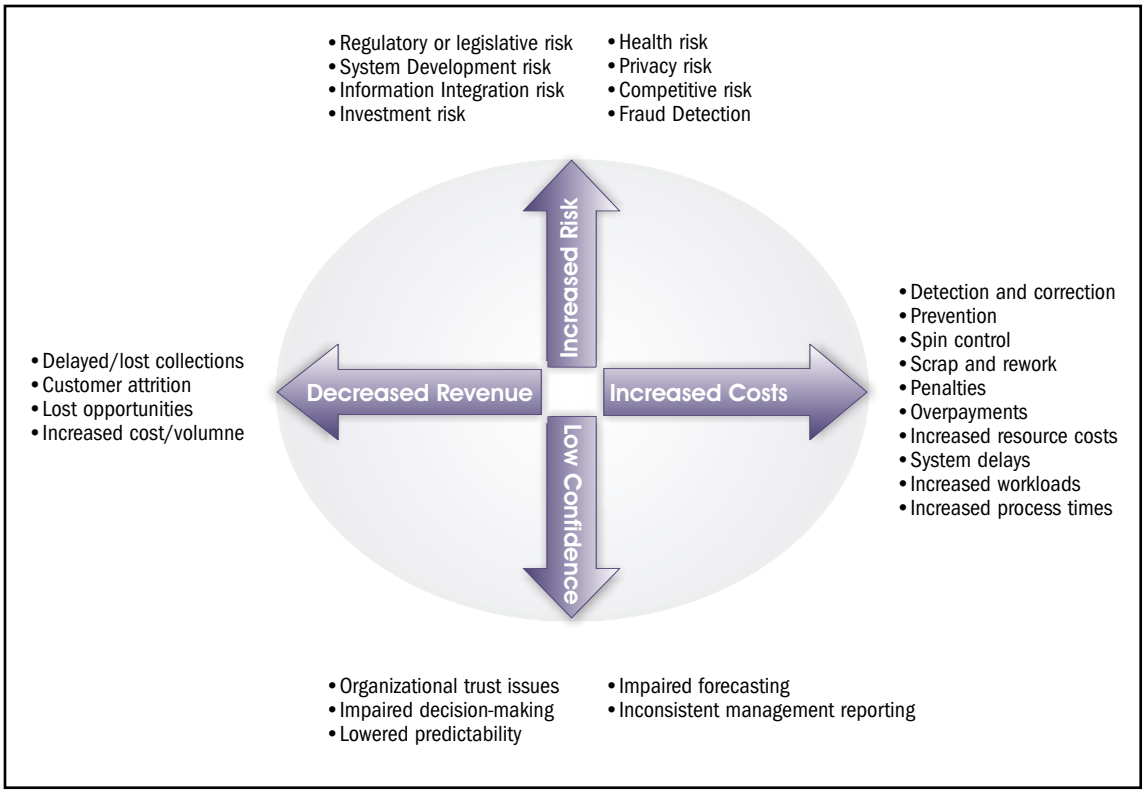


Figure 2: Impacts of poor data quality

Increased Costs

Costs may be incurred when addressing information quality issues or by ignoring them. For example, **detection** and **correction** costs are incurred when a problem has been identified, and these may be relatively large but infrequent. Alternatively, **prevention** costs may be incremental costs that are ongoing, and may diminish in expense as time goes on.

Spin control costs are associated with ensuring that data quality impacts exposed outside of the organization are mitigated, such as the discovery (by an external organization like a newspaper) that decisions about which medical procedures are approved by a health insurer are based on faulty data, indicating that the needs of the member are not always being met properly. The cost of spin control includes the costs of publicity to address the discovery, plus any acute costs incurred to immediately modify procedures in place to close the perceived gap exposed by the discovery.

Scrap and rework refers to costs associated with rolling back computations, undoing what had been done, and starting again. Information quality problems can impact levels of application service; if there are well-defined service-level agreements that are not being met, **penalties** for missing objective targets may be incurred. The inability to properly track all representational and contact information related to financial agreements with business partners can result in accounting failures, leading to potential **overpayments** or duplicated invoice payments. Ultimately, many of these issues may roll up into **increased workloads** on system, as well as human resources, leading to **system delays** and **increased process times**.

Decreased Revenues

The inability to resolve uniquely identifiable parties within a financial system can result in accounting entries that cannot be found without searching for the exact variation, potentially leading to **delayed** or **lost collections**. At the same time, the inability to resolve uniquely identifiable customer/client records ultimately reduces the effectiveness of any party relationship management system, which in turn may lead to **customer attrition**. Bad data may result in a delay in exploiting information at the proper time, leading to **lost opportunities**, such as not being able to execute transactions in a timely manner, inaccurately deactivating resources, or missing up-sell or cross-sell opportunities.

Lastly, data quality problems interrupt streamlined workflow, reducing throughput and volume, resulting in **increased cost per volume**. In environments that rely on high volume, predictable volumes relate to determining the average costs per transaction, which contribute to predictability across resource planning, volume pricing, and other ways to increase profit margin.

Decreased Confidence

Organizational management is a world that is highly influenced by information – its use, its guardianship, and its ownership model. Flawed data introduces **organizational trust** issues leading to suspicion and anxiety. As more data is propagated into business intelligence platforms for decision support services, invalid or incorrect data leads to **impaired decision-making** and **forecasting**.

Impaired forecasting can reverberate across the organization – improper staffing, reduced investments in capital expenditures for hardware and software, incorrectly setting service rates and product prices, etc. However, the impact of poor decision-making is related to the actions taken based on that bad decision. When assessing the impact, it is valuable to :

- Identify specific business decisions that are "impaired" (suggestion: come up with a word better-suited to your organization)
- Determine if those decisions are directly attributable to bad data (and which data!)
- Only once you can directly attribute bad decisions to bad data, then have your business partner explain the actual cost impact of the decision

Increased Risk

Regulatory risks (e.g. HIPAA, Sarbanes-Oxley) introduce constraints associated with information quality with well-defined penalties for noncompliance. **System development** risk refers to the investment in building systems that cannot be rolled out until the end-clients are satisfied with the levels of information quality. When attempting to assemble a master data management program, it is important to be able to integrate data from across many different systems; flawed data can cause an **integration** impact when they are left unaddressed.

In the health care world, there can be serious **health risks** associated with incorrect data. An obvious severe example is the case of heart transplant patient Jesica Santillan, where inaccurate information regarding blood-typing resulted in a botched heart-lung transplant, which not only led to the girl's death, but also prevented other critical patients from receiving needed donated organs.

As an example of **privacy risk**, HIPAA's regulatory aspect underscores health insurers' ethical responsibility to ensure patient privacy. Properly capturing and managing data related to with whom an individual's health information may be shared impacts many aspects of that person's life, ranging from protection against physical abuse to employment discrimination, among others.

Fraud risks may be masqueraded throughout your applications when fraudulent behavior is performed to exploit information failures within the system.

Developing the Business Case

A data quality improvement program is a serious commitment on behalf of an organization. Its importance deserves to be effectively communicated to the business managers who will sponsor both the technology and the organizational infrastructure in order to ensure a successful program. And as we have already identified key impact dimensions and corresponding impact categories associated with poor data quality, some additional research can document:

- The quantification of the identified financial impacts,
- Actual root causes in the information processing that are correlated to those impacts,
- The costs to remediate those process failures, and
- A way to prioritize and plan the solutions of those problems.

We will accumulate this information into an Impact Template, which documents the problems, the issues, the business impacts, the quantifiers, all of which will enable the determination of a yearly incurred impact.

Identifying Impacts

Most likely, there will already be some awareness of some of the existence of impacts of poor data quality. Using our impact taxonomy, we can begin to determine how the results of different data quality events can be grouped together, which simplifies the research necessary to determine financial impact.

Researching Financial Impacts

The next step in the process is to get a high-level view of the actual financial impacts associated with the problem. This step combines subject matter expertise with some old-fashioned detective work. Because we are trying to get a high-level impact assessment, we have some flexibility in exactness, and in fact much of the information that is relevant can be collected in a relatively short time.

Anecdotes are good starting places, since they are indicative of high-impact, acute issues with high management visibility. Historical data associated with work/process flows during critical data events can provide cost/impact information. To understand the actual cost impact, delve deeper into the core of the story; ask these kinds of questions:

- What was it about the data that cause the problem?
- How big is the problem?
- Has this happened before?
- How many times?
- When this happened in the past, what was the remediation process?
- What was done to prevent it from happening again?

At the same time, consult issues tracking system event logs, management reports on staff allocation for problem resolution, and review external impacts (e.g., stock price, customer satisfaction, management spin) to identify key quantifiers for business impact. The answers to the questions combined with the research will provide insight into quantifiable costs, which will flow into the Impact Template (see Figure 3).

Problem	Issue	Business Impact	Quantifier	Yearly Incurred Impact
Missing product id, inaccurate product description at data entry point	Inability to clearly identify known products leads to inaccurate forecasts	Slower turnover of stock	Increased cost	$30,000.00
		Stock write downs	Increased cost	$20,000.00
		Out of stocks at customers	Lost revenue	
		Inability to deliver orders	Lost revenue	$250,000.00
		Inefficiencies in sales promotions	Speed to market (and lost revenue)	$20,000.00
		Distribution errors and rework	Staff time	$24,000.00
		Shipping costs	Increased shipping costs	$78,000.00
		Unnecessary deliveries	Staff time	$23,000.00

Figure 3: An example impact template

The Data Quality Impact Matrix

The template in Figure 3 reflects an example of how invalid data entry at one point in the supply chain management process results in three impacts incurred at each of three different client applications, Inventory Management, Fulfillment, and Logistics. For each of these business areas, the corresponding impact quantifiers are identified, and then their associated costs are projected and expressed as yearly incurred impacts.

In our impact matrix, the intention is to document the critical data quality problems, review the specific issues that occur within the enterprise, and then enumerate all the business impacts incurred by each of those issues. Once the impacts are specified, we simplify the process of assessing the actual costs, which we also incorporate in the matrix. The resulting matrix reveals the summed costs that can be attributed to poor data quality.

Correlating Impacts to Root Causes

The next step in developing the business case involves tracking the data flaws backward through the information processing flow to determine at which point in the process the data flaw was introduced. Since many data quality issues are very likely to be process failure, eliminating the source of the introduction of bad data upstream will provide a much greater return on investment than just correcting bad data downstream.

In our supply chain example, the interesting thing to note is that each of the client application users would assume that their issues were separate ones, yet they all stem from the same root cause. The value in assessing the introduction of the flaw into the process is that when we can show that one core problem has multiple impacts, the return on our investment is remediating the problem will be much greater.

Costs to Remediate

An ROI calculation doesn't just take into account the benefits – it also must factor in the costs associated with the improvements. Therefore, we need to look at the specific problems that are the root causes and what it would cost to fix those problems. In this step, we evaluate the specific issues and develop a set of high-level improvement plans, including analyst and developer staff time along with the costs of acquiring data quality tools.

Problem	Issue	Solution	Software Costs	Staffing
Missing product id. inaccurate product description at data entry entry	Inability to clearly identify known product leads to inaccurate forecasts	Parsing and standardization. Record, monitoring linkage tools for cleansing	$150,000.00 for license 15% annual maintenance	.75 FTE for 1 year .15 FTE for annual maintenance

Figure 4: Example solution investment assessment

Figure 4 shows an example solution investment assessment, documenting the cost of each solution, which also allows us to allocate the improvement to the documented problem (and its associated impacts). Because multiple problems across the enterprise may require the same solution, this opens up the possibility for economies of scale. It also allows us to amortize both the staff and technology investment across multiple problem areas, thereby further diluting the actual investment attributable to each area of business impact.

Projecting Return on Investment

We now have two artifacts that can be used to project the return on investment: the impact matrix and the solution assessment. Deploying a proposed solution will eliminate some number of yearly incurred impacts, and therefore, the return on investment can be calculated as the difference between the sum of those yearly incurred impacts and the yearly resource and staffing requirements.

In turn, we can use these results to prioritize our investment and program growth. By reviewing the criteria for providing value (e.g., biggest bang for the buck, fastest results, lowest up front costs), management can select and plan the project investments that grow the data quality program in the most strategic way.

Case Studies

In each of these case studies, individuals were aware of the existence of data quality problems, but were challenged by understanding how those problems impacted the achievement of their business objectives. The process of identifying data flaws, correlating them to business impacts, and establishing a priority for remediation contributed to the organization's allocating budget for data quality tools, and more importantly, data quality management.

Pharmaceutical Company

A subsidiary of a pharmaceutical company was developing a new application to be rolled out to the sales staff to assist in the sales and marketing process. However, the sales representatives hesitated in accepting a new sales application, and while they contended that the quality of the data was insufficient to meet their needs, no one was able to effectively communicate the basis of their hesitance to senior management.

A high level assessment of the data revealed a number of potential anomalies in the data, including:

- Duplicated entries for customers, suppliers, and research grantees
- Missing telephone and address contact information
- Significant use of represented null values (e.g., "N/A," "none," "????")

The assessment revealed a number of impacts related to these deficiencies, including:

- **Investment Risk** – As the suspicions of the sales staff regarding the quality of the data was confirmed, in that duplicate data and missing contact information would not improve the sales and marketing process. This made it clear that their reluctance to use the new application indicated that the significant investment in development of the new application was put at risk.
- **Regulatory Compliance** – The assessment revealed that a number of customers were also research grantees, which exposed a potential inadvertent risk of violating the federal Anti-Kickback statute.
- **Decreased Service Capability** – Duplicated and/or missing data contributed to a decrease in the ability to satisfy negotiated client-side service level agreements.

As a result of the data quality assessment and subsequent impact analysis, senior management recognized that the financial impacts and compliance risks warranted an investment in data quality improvement. The organization has defined a staff position whose role is to oversee a data quality program, and has purchased data quality tools to be used in data quality analysis, improvement, and monitoring.

Health Insurance Company

Ongoing data flaws that had propagated into this health insurance company's data warehouse had resulted in flawed decision making related to premiums, provider rates, financial forecasting, as well as member and group management issues such as underwriting, sales and marketing, and member attrition. In addition, inaccurate and untimely data impacted regulatory reporting, exposing the organization to compliance risk.

As part of the data quality program, a straightforward system was developed to proactively test a number of data quality rules and assertions before the data was loaded into the data warehouse. The approach described in this paper was used to assess the cost impacts of the data quality issues that had occurred to provide a baseline for determining return on investment.

Over a one-year time period, this proactive validation captured a relatively large number of data flaws, whose impacts were prevented due to early intervention. Subsequent analysis revealed that the return on the proactive validation was 6-10 times the investment in the development and maintenance of the validation application.

Government – Department of Defense

As described in the Department of Defense Guidelines on Data Quality Management , an assessment of the cost impacts of data quality was described, with expectations set by specifying that resultant costs should be quantified wherever possible. As described in the report,

"... the inability to match payroll records to the official employment record can cost millions in payroll overpayments to deserters, prisoners, and "ghost" soldiers. In addition, the inability to correlate purchase orders to invoices is a major problem in unmatched disbursements. In the DoD, resultant costs, such as payroll overpayments and unmatched disbursements, may be significant enough to warrant extensive changes in processes, systems, policy and procedure, and AIS data designs."

In one specific instance, the assessment approach identified specific issues with Bill of Material (BoM) data, identified approaches for solutions, with the recommendation of purchasing particular data quality tools. The benefits of the solution included "the ability to quickly produce meaningful results that were easily understood by functional users and management," as well as projected cost savings as high as 40% in both time dedicated to reacting to a diagnosing data quality problems and re-entering incorrect data. In one specific case, the projected net savings exceeded $3.7 million. Importantly as well, "the automated data quality analysis tool and the structured levels of analysis enabled the project team to quantify the benefits of the project into a format for high-level discussions."

Telecommunications Company

Inaccurate and invalid data can plague service-oriented companies, such as in the telecommunications industry. For one telecommunications company, an assessment of the quality of their component inventory, billing records, and customer records revealed a number of financial impacts:

- Numerous high-bandwidth components had been misconfigured, resulting in decreased service bandwidth
- Discrepancies existed between services provide and services actually billed
- Inconsistent and invalid information about the system created excess capacity in some areas while simultaneously capacity was strained in others

As a result of this analysis, near-term funding for data quality efforts was inspired by the more traditional approach of revenue assurance through the detection of underbilling. The telecommunications company was able to derive the following benefits as a result of data quality improvement:

- Revenue assurance/underbilling analysis indicated revenue leakage of just over 3 percent of revenue attributable to poor data quality
- 49 misconfigured (but assumed to be unusable) high-bandwidth circuits were returned to productive use, thereby increasing bandwidth
- Cleansing of customer data revealed more than 200,000 "unknown" potential customers

Summary: How to Get Started

As we can see from our case studies, successful business cases can be developed for investing in a data quality. By investing a small amount of analyst time, and with senior management sponsorship, here is an outline to develop your Data Quality Business Case:

1) Identify 5 business objectives impacted by the quality of data

2) For each of those business objectives:

   a. Determine cost/impacts areas for each flaw

   b. Identify key quantifiers for those impacts

   c. At a high level, assess the actual costs associated with that problem

3) For each data quality problem:

   a. Review solution options for that problem

   b. Determine costs to implement

4) Seek economies of scale to exploit the same solution multiple times

At the conclusion of this exercise, you should have the right information to assemble a business case that not only justifies the investment in the staff and data quality technology used in developing an information quality program, but provides baseline measurements and business-directed metrics that can be used to plan and measure ongoing program performance.

Notes

INFORMATICA®
The Data Integration Company™

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

J50954 6731 (06/20/2006)