Understanding the Financial Value of Data Quality Improvement

Prepared by:

David Loshin
Knowledge Integrity, Inc.
January, 2011

Sponsored by:

INFORMATICA®

www.knowledge-integrity.com
(301) 754-6350

Introduction

Despite the many years of lip service paid to high quality data, the difficulty in establishing quantifiable value often relegates data quality improvement to the eleventh spot on the CEO's list of top ten business imperatives. And as opposed to the technical aspects of data validation and cleansing, it appears that the biggest challenge is effectively communicating the business value of data quality improvement. Yet with a well-defined process for considering the costs and risks of low-quality data in relation to an iteratively-refined set of business impact categories not only provides a framework for putting data quality expectations into a business context, it also enables the definition of clear metrics linking data quality to business performance.

As an example, it is often suggested that data errors prevent the sales team members from up-selling and cross-selling products, and this claim is used to justify the need for a data quality improvement effort. However, a more comprehensive quantification of the number of sales impacted or of the total dollar amount for the missed opportunity is much more effective at showing the value gap, especially when it can be directly associated with specific data flaws.

The communication gap between technical analysts and business managers often impedes the articulation of the value of data quality improvement. This article looks at a hierarchical categorization of the financial dimension of business value drivers, corresponding performance measures, linking those measures to specific data issues, and a process for evaluating the relationship between acceptable performance and quality information. This article is targeted to those technical analysts who seek to understand the connection between information and optimal business performance so that the case for data quality improvement can be rooted in quantifiable measures.

The Value of High Quality Data

Both operational and analytical business applications rely on high quality data. And those organizations that lack processes for identifying and managing data quality issues introduce risks to the usability and trustworthiness of the data upon which many applications depend, leading to the types of negative financial impacts we describe in this article. This suggests that there is value in instituting processes for assessing, measuring, reporting, reacting to, and controlling different aspects of poor data quality. Data is an asset that is created or acquired and then repurposed multiple times, and process flaws introduce risks to successfully achieving business objectives. The dynamic nature of data adds to the challenges in establishing ways to assess risks as well as ways to monitor conformance to business user expectations.

Are Anecdotes Enough?

Resorting to specific anecdotes and examples of business problems linked to bad data may raise awareness temporarily, but this is no substitute for demonstrating real evidence of hard impacts. This is common in organizations that "institutionalize reactivity" by regularly engaging staff members to correct data errors when their catastrophic impacts have already occurred, instead of proactively assessing data quality assessment, performing root cause analysis, and eliminating the sources of the errors. The first step begins with developing a performance management framework that helps to identify, isolate, measure, and improve the value of data within the business contexts, which requires

- Correlating business impacts with data failures and then
- Characterizing the loss of value that is attributable to poor data quality.

This means reviewing the types of risks and costs relating to the use of information and categorizing the business impacts as a prelude to asserting data quality expectations and metrics. Business issues are directly tied to missed data quality expectations, and the framework described in this article explores categories of business impacts that may be rooted in poor data quality. By identifying and classifying business impacts and establishing the connection to reliance on high quality data, technical analysts gain the tools for effectively communicating the value of improved data quality to key stakeholders in the organization.

Financial Impacts

There are essentially two sides of the coin when evaluating financial impacts, both ultimately connected to increased profitability: increasing revenues while decreasing costs. The process of soliciting, categorizing, and measuring financial impacts involves these basic steps:

1. Clarifying and prioritizing the financial business expectations, and working with the business clients to understand where the financial expectations are not being met;
2. Finding specific examples where failed expectations has lead to known impacts;
3. Categorizing the business impact at a fine-enough level of granularity in a way that ensures that the impacts can be measured;
4. Formalizing the dependence on specific data sets and data errors;
5. Researching the history of occurrence, probability of recurrence, and cumulative impacts; and
6. Understanding and then eliminating the root causes.

Organizational Objectives

Even in organizations whose leaders have articulated well-defined high-level strategic goals, the role each staff member plays is often lost in translation. Therefore to ensure that the information technology and business function teams are aligned, it is important for the data quality analyst to establish a good rapport with business data consumers to

- Identify those business processes that are dependent on high quality data;
- Understand how those business processes use data; and
- Solicit the data consumers' expectations in a way that can be translated into defined business rules for data validity.

Communications Gap

Unfortunately, it is much easier to identify gaps in meeting financial expectations than to quantify them in measureable terms, especially in relation to their dependence on data. This is complicated even more by the communications gap that often exists between the business and technical teams. The second step can be simplified by providing a framework for categorizing business impacts at a level of granularity that can be discretely evaluated and measured. Developing a taxonomy of financial impact

classifications enables the technical analysts to map identified business issues into a specific category, which is more likely to be linked to specific data sets and their expectations for data quality.

Once the most critical impacts have been identified and linked to associated data flaws, the analysts can define data quality rules that reflect the business user expectations. Continuous inspection of conformance to formally-defined data quality rules not only provides discrete quantification of the scale of existing data problems, its results can contribute to a performance scorecard linking data quality to specific financial value.

To continue our earlier example, revenue generation may be negatively impacted through missed up-selling and cross-selling opportunities. If this is caused by inability to have complete customer visibility as a result of duplicate entries in the customer database, then measuring the number of duplicates provides a good indicator that sales opportunities are being missed. More comprehensive analysis can link the specific number of duplicates to one or more missed sales, completely closing the loop for defining business-oriented data quality rules.

The following sections provide some examples of this classification and categorization for two high-level categories of financial impacts (revenue growth and decreasing costs), subcategories, and descriptions of potential business measures. For each set of impact areas, examining the corresponding measures and considering their dependence on data will help the analyst to link the financial impact to high quality data.

Revenue Growth

One might say that the primary goal of any business is profitability, but it would be difficult to achieve profitability without generating revenues. There are many areas associated with growing revenue, and this section looks at four:

- Customer acquisition
- Customer retention
- Leveraging income-generating opportunities
- Cross selling and up-selling

Customer Acquisition

Companies rely on a community of customers to purchase their products and services. Table 1 provides examples of areas of impact and corresponding measures for new customer acquisition.

Area of Impact	Measures
Customer segmentation	Effectiveness of targeting individuals for customer acquisition, retention, and breadth of geographic and demographic coverage
Quality of sales leads	The degree to which sales lead data records contain complete and accurate contact information
Meeting sales targets	Determining whether sales team members are meeting or missing their sales targets
Sales channel partner effectiveness	The degree to which channel partners are successful in promoting and selling to new customers

Area of Impact	Measures
Sales channel effectiveness	Comparisons to gain of market share by competitors through the same and/or alternate channels
Sales cycle time	Monitoring how quickly leads and prospective customers progress through the stages of the sales cycle
Qualification rate	Determining the percentage of leads that can be qualified as prospective customers
Sales closure rate	Determining the percentage of prospective customer that commit to purchasing a product or service
Time spent selling	Measuring the percentage of sales team member time is spent selling

Table 1: Areas of business impact for new customer acquisition.

Customer Retention

Once the company has acquired the customer, there is a continuous drive to ensure that the customer's business is retained rather than allow the customer to purchase products or engage services from the company's competitors. Table 2 provides examples of areas of impact and corresponding measures for customer retention.

Area of Impact	Measures
Focus on high value customers	Ability to calculate customer value
Lifecycle transitions	Employing knowledge of customer life cycle events and transitions for proactive retention
Loyalty management	Enrollment in loyalty programs, allocation of loyalty rewards, and redemption of loyalty rewards
Response to customer needs	Degree of responsiveness to customer requests, inquiries, service needs
Customer portfolio management	How effective is the organization at culling out those customers that are not profitable?
Attrition	Churn rates, effectiveness in making barriers to defection
Contractual compliance	How effective is the company at monitoring adherence to contractual details
Feedback and satisfaction	Response to requests for feedback, measured customer satisfaction

Table 2: Areas of business impact for customer retention

Cross Selling and Up-Selling

Once a prospect has transitioned into being a customer, there is a desire to encourage the customer to buy additional products and services. Cross selling is a strategy for suggesting products that are complementary to the ones already purchased, such as offering french fries along with the purchased hamburger or a set of speakers with the purchase of an audio receiver. Up-selling is a strategy to entice the customer to increase the scale of size of the purchase, such as offering a larger-sized soda or an upgraded version of the audio receiver. Table 3 provides examples of areas of impact and corresponding measures for cross selling and up-selling.

Area of Impact	Measures

Area of Impact	Measures
Channel effectiveness	Measure effectiveness of advertising media channels, partner sales, and other sales channels
Customer touch points	Measure customer experience across all touch points including purchase, delivery, support, and customer service
Identifying cross sell and up-sell opportunities	Analysis that exposes opportunities for cross selling and up-selling
Instituting cross selling and up-selling processes	Developing approaches, campaigns, processes, and training for cross selling and up-selling
Process effectiveness	Measuring and improving effectiveness of cross selling and up-selling processes
Staff effectiveness	Evaluate staff effectiveness and properly incentivizing success

Table 3: Areas of business impact associated with cross selling and up-selling.

Leverage Income-Generation Opportunities

Alternatively, organizations often control existing assets that can be leveraged into generating additional income. Table 4 provides examples of areas of impact and corresponding measures for making use of existing assets to generate additional income.

Area of Impact	Measures
Exploiting existing assets	Sell existing assets whose values have appreciated, lease or sell extra capacity to other organizations
Intellectual property	Sell or license intellectual property
Cash reserves	Improve investment income from existing funds
Manage taxes	Manage and/or defer federal, state, and local tax payments

Table 4: Areas of business impact associated with leveraging income-generating opportunities.

Decreasing Costs

The other side of profitability is reducing the expenses paid for operations. Again, there are many areas in the business in which costs are incurred, but this section provides some details regarding these areas:

- Overhead and administrative costs
- Cost of goods sold
- Fees and charges

Overhead and Administrative Costs

Almost every organization utilizes a physical facility and incurs costs associated with running the business such as rent and facility maintenance. Even virtual organizations have overhead and administrative costs such as telephones, internet, furniture, hardware, and software purchase/leasing and maintenance. Table 5 provides examples of areas of impact and corresponding measures for overhead and administrative costs.

Area of Impact	Measures
Rent	Costs of rented office space

Area of Impact	Measures
Maintenance	Costs associated with building, furniture, machinery, software, and grounds maintenance
Asset purchase and licensing	Costs associated with purchasing or leasing assets and equipment, and the effort for choosing one option over the other
Utilities costs	Costs for telecommunications, energy, water, gas, etc.
Administrative staff	Number of staff members dedicated to administrative activities, percentage of time spent in overhead and administrative activities
Office supplies	Cost and use of paper, pens, notebooks, etc.
General procurement	Costs associated with spend and procurement processes

Table 5: Areas of business impact associated with overhead and administrative costs.

Cost of Goods Sold

The cost of goods sold (often referred to as COGS) comprises the costs and expenses associated with the production and sales of products, including the costs of materials, the direct labor expenses incurred in manufacturing, and the direct labor expenses incurred in selling the products. Table 6 provides examples of areas of impact and corresponding measures for the cost of goods sold.

Area of Impact	Measures
Product design	Staff and materials costs for designing new products
Raw materials	Costs of acquiring, storing, and using raw materials
Cost of production	Costs of manufacturing and finishing products
Sales staff base costs	Base salaries paid to sales account executives
Product quality	Percentage of manufactured items within acceptable quality guidelines

Table 6: Areas of business impact associated with the cost of goods sold.

Fees and Charges

For many of the services critical in running a business (such as banking , legal, and accounting) there are fees and charges incurred. Table 7 provides examples of areas of impact and corresponding measures for fees and charges.

Area of Impact	Measures
Bank fees and service charges	Bank service fees, transaction fees, interest charges, missed payment fees
Legal fees	Costs paid for legal services
Accounting fees	Costs associated with maintaining accurate financial accounts
Document fees	Fees and charges associated with application and document filing
Commissions	Payments exceeding purchase costs paid to agents or brokers
Bad debt	Costs associated with payments and debt that is difficult to collect or is uncollectable
Penalties and fines	Penalties and fines for noncompliance of regulations, penalties associated with failure to observe agreements
Merger and acquisition costs	Costs associated with process of merging operations and application systems

Table 7: Areas of business impact associated with fees and charges.

Formalizing the Dependence on Data: An Example

In order to demonstrate the process, let's look at one example from our categories, sales channel partner effectiveness. The example measure is the degree to which channel partners are successful in promoting and selling to new customers. There are at least two data sets associated with the example measure: the set of channel partners and the set of prospective customers. The analyst must consider the types of errors that may exist in those data sets and in turn, how those types of errors might lead to decreased success in channel partner success. Examples might include missing or incorrect prospect data, incomplete or in accurate partner data, or incorrect tracking of channel sales transactions.

Researching Impacts and the Value Gap

The fourth step of the process is researching the history of occurrence, probability of recurrence, and cumulative impacts. There is a story behind each perceived business problem, and the analyst's goal is to ask the right kinds of questions to fully assess the level of criticality. For example, for each business problem, the analyst might ask the business stakeholders these types of questions:

- What makes this a critical business problem?
- What are the measurable impacts?
- How is each impact classified?
- How is the impact measured?
- Is there a data management behavior that we are looking to influence to achieve better results?

The answers reveal the aspects of the problem that are used to determine the scope of its impact, its prevalence, its frequency, and its probability of occurrence. Together, these variables can be used to prioritize the business problems and direct the second phase of the analysis to assess the relationship to flawed data. For each business problem, the analyst can ask questions such as these:

- How is the business problem related to an application data issue?
- How often does the data issue occur?
- When the data issue occurs, how is it manifested within the business process?
- Who are the individuals who recognize the existence of the problem?
- How often is the data issue identified before the business impact is incurred?
- What rules and metrics can be defined to validate the quality of the data?
- Have we seen this happen before, either internally or in the market? If so, can we directly link the data issue to specific business impact?

Documenting the quantifiers associated with the business impact may involve some detective work – perhaps reviewing change control logs, requests for business process modifications, or requests for changes to underlying data models and application code. The requests often correlate with known business issues and will point to quantifiable measures for the impacts. Linking those measures to specific data validity metrics provides the hard link between business impact and flawed data.

Establishing the Value of Data Quality Improvement

The last step in developing the business case involves understanding the root causes of the issues and determining how those issues can be addressed. This typically means reviewing the information flow from the data creation point through the business process to determine where in the process the error was introduced. Many data quality issues are due to process failures, so correcting the process where the error is introduced will lead to greater improvement than correcting bad data downstream. With enough research into the granularity of the problem and its relationship to flawed data, one might even be able to amortize the cumulative costs related to the business impacts over the number of times an error occurs factored within the probability that the error will occur, thereby providing an estimated cost of each error.

Once the source of the introduction of the error is identified, the data analyst can consider alternatives for eliminating the root cause, instituting preventative techniques, and/or taking some corrective action. Each, or perhaps all, of these alternatives require an investment of both money and resources for the acquisition of any appropriate technologies, staffing for designing, developing, and implementing solutions, training, and ongoing maintenance of the solution. Providing a conservative estimate at this point establishes a baseline cost for remediation. At this point we have an estimate of the cost impacts associated with each specific issue, and we can call these cost impacts the *value gap.* This is the conservatively estimated costs that are attributable to data quality issues, and provides a quantification of the corresponding loss of value to the organization.

The result of this process is a list of financial impacts directly related to measurable data failures prioritized by the breadth of the value gap. And because the data metrics are associated with known performance measures, improving the quality of data as measured by the data metrics increases the probability that changes to the data environment will result in measurable business improvement.

About the Author

David Loshin, president of Knowledge Integrity, Inc, (www.knowledge-integrity.com), is a recognized thought leader and expert consultant in the areas of data quality, master data management, and business intelligence. David is a prolific author regarding BI best practices, via the expert channel at www.b-eye-network.com and numerous books and papers on BI and data quality. His book, "Business Intelligence: The Savvy Manager's Guide" (June 2003) has been hailed as a resource allowing readers to "gain an understanding of business intelligence, business management disciplines, data warehousing, and how all of the pieces work together." His most recent book is "The Practitioner's Guide to Data Quality Improvement," and his insights on data quality can be found at www.dataqualitybook.com. His book, "Master Data Management," has been endorsed by data management industry leaders, and his valuable MDM insights can be reviewed at www.mdmbook.com.

David can be reached at loshin@knowledge-integrity.com

About the Sponsor

Informatica Corporation (NASDAQ: INFA) is the world's number one independent provider of data integration software. Organizations around the world gain a competitive advantage in today's global information economy with trustworthy, actionable, and authoritative data for their top business imperatives. More than 4,100 enterprises worldwide rely on Informatica to access, integrate, and trust their information assets held in the traditional enterprise, off premise, and in the cloud.