Evaluating the Business Impacts of Poor Data Quality

Submitted by:

David Loshin
President,
Knowledge Integrity, Inc.
(301) 754-6350
loshin@knowledge-integrity.com

1 Introduction

Establishing a business case for introducing and developing a data quality management program is often predicated on the extent to which data quality issues impact the organization and the return on the investment in data quality improvement. Today, most organizations use data in two ways: transactional/operational use ("running the business"), and analytic use ("improving the business"). When the results of analysis permeate the operational use, the organization can exploit discovered knowledge to optimize along a number of value drive dimensions. Both usage scenarios rely on high quality information, suggesting the need for processes to ensure that data is of sufficient quality to meet all the business needs. Therefore, it is of great value to any enterprise risk management program to incorporate a program that includes processes for assessing, measuring, reporting, reacting to, and controlling different aspects of risks associated with poor data quality.

Flaws in any process are bound to introduce risks to successfully achieving the objectives that drive your organization's daily activities. If the flaws are introduced in a typical manufacturing process that takes raw input and generates a single output, the risks of significant impact might be mitigated by closely controlling the quality of the process, overseeing the activities from end to end, and making sure that any imperfections can be identified as early as possible. Information, however, is an asset that is generated through numerous processes, with multiple feeds of raw data that are combined, processed, and fed out to multiple customers both inside and outside your organization. Because data is of a much more dynamic nature, created and used across the different operational and analytic applications, there are additional challenges in establishing ways to assess the risks related to data failures as well as ways to monitor conformance to business user expectations.

While we often resort to specific *examples* where flawed data has led to business problems, there is frequently real evidence of hard impacts directly associated with poor quality data. Anecdotes help to motivate and raise awareness of data quality as an issue. However, developing a performance management framework that helps to identify, isolate, measure, and improve the value of data within the business contexts requires correlating business impacts with data failures and then characterizing the loss of value that is attributable to poor data quality.

This requires some exploration into assembling the business case, namely:

- Reviewing the types of risks and costs relating to the use of information,
- Considering ways to specify data quality expectations,
- Developing processes and tools for clarifying what data quality means,
- Defining data validity constraints,
- Measuring data quality, and
- Reporting and tracking data issues.

Given these aspects of measurement, one can materialize a data quality scorecard that measures data quality performance.

Many business issues can be tied, usually directly, to a situation where data quality is below user expectations. Given some basic understanding of data use, information value, and the ways that information value degrades when data does not meet quality expectations, we can explore different categories of business impacts attributable with poor information quality, and discuss ways to facilitate identification and classification of cost impacts related to poor data quality. This paper considers types of risks attributable to poor data quality as well as an approach to correlating business impacts to data flaws.

2 Information Value and Data Quality Improvement

It is still premature to think that any organization lists data as a line item as either an asset or a liability on its balance sheet. Yet the significant dependence on data to both run and improve the business suggests that senior managers at most organizations rely on their data as much as any other asset. Data provides benefits to the organization, it is controlled by the organization, it is the result of a sequence of transactions (either as the result of internal data creation internally or external data acquisition), it incurs costs for acquisition and management, and is used to create value. Data is not treated as an asset, though; for example, there is no depreciation schedule for purchased data.

There are different ways of looking at information value. The simplest approaches consider the cost of acquisition (i.e., the data is worth what we paid for it) or its market value (i.e., what someone is willing to pay for it). But in an environment where data is created, stored, processed, exchanged, shared, aggregated, and reused, perhaps the best approach for understanding information value is its *utility* – the expected value to be derived from the information.

Utility value grows as a function of different aspects of the business, ranging from strictly operational to the strategic. Daily performance reports are used to identify and eliminate high-cost, low productivity activities, and therefore the money saved is related to the data that composed the report. Streamlined processing systems that expect high quality data can process many transactions without human intervention, and as the volume of processed transactions increases, the cost per transaction decreases, which represents yet another utility value for data.

It may be difficult to directly assign a monetary value to a piece of data, but it is possible to explore how the utility value changes when the data does not meet business client expectations. One can analyze how data is being used for achieving business objectives, and how the achievement of those goals is impeded when flawed data is introduced into the environment. To do this, we must consider:

- What the business expectations are for data quality,
- How business can be impacted by poor data quality, and
- How to correlate business impacts to specific data quality issues.

3 Business Expectations and Data Quality

There is an expectation that objective data quality improvement implies business value, but limited awareness and understanding of what data quality improvement can truly imply often drives technical approaches that don't always translate into improving the business. In reality, data quality is *subjective*, and relies on how data flaws are related to negative business impacts within your own organization.

If objective data quality metrics (such as number of invalid values, or percentage of missing data elements) are not necessarily tied to organizational performance, then we must ask these questions:

- How do you distinguish high impact from low impact data quality 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 quality?
- What is the best way to employ data quality 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 reworking failed processes?
- What is the loss in value of transactions that failed due to missing data?
- How quickly can we respond to emerging opportunities?

The value added by data quality improvement must be tied to meeting business expectations, and measured in relation to its component data quality rules. This involves identifying business impacts, their related data issues, their root causes, and then a quantification of the costs to eliminate the root causes. Characterizing both the business impacts as well as the data quality problems provides a framework for developing a business case.

4 Identifying and Categorizing Impacts

A straightforward approach to analyzing the degree to which poor data quality impedes business success involves categorizing business impacts associated with data errors within a classification scheme. This classification scheme begins with defining a simple taxonomy that lists primary categories for either the negative impacts related to data errors, or the potential business improvement resulting from improved data quality, including the following areas:

- Financial impacts, such as increased operating costs, decreased revenues, missed opportunities, reduction or delays in cash flow, or increased penalties, fines, or other charges.
- Confidence and Satisfaction-based impacts, such as customer, employee, or supplier satisfaction, as well as decreased organizational trust, low confidence in forecasting, inconsistent operational and management reporting, and delayed or improper decisions.
- Productivity impacts such as increased workloads, decreased throughput, increased processing time, or decreased end-product quality.
- Risk and Compliance impacts associated with credit assessment, investment risks, competitive risk, capital investment and/or development, fraud, and leakage, and compliance with government regulations, industry expectations, or self-imposed policies (such as privacy policies).

This categorization is intended to support the data quality analysis process and help in identifying risks related to data quality issues and then differentiating between data issues that have serious business ramifications and those that are benign. And while these areas of risk (and their originating sources) differ, they are similar in the need for mandating high quality information and the means to demonstrate adequacy of internal controls governing data quality. Organizations must be able to assess, measure, and control the quality of data as well as have the means for external auditors to verify those observations. Ultimately, the objective is to maximize the value of the information based on reducing the negative impacts associated with each set of potential problems. In turn, determining when and where poor information quality affects one or more of the variables that contribute to these categories becomes the core task of developing a business justification for data quality improvement.

5 More on Impact Classification

Classifying the business impacts helps to identify discrete issues and relate business value to high quality data, so it is valuable to examine these impact categories more closely. Impacts can be assessed within enumerated subcategories, which help to refine a means for quantification. A data quality scorecard reporting the value of high quality data can be rolled up as a combination of separate measures associated with how specific data flaws prevent the achievement of business goals.

5.1 Financial

Financial impacts are associated with missing expectations associated with costs, financial management, and revenues. In Table 1, we show some of the subcategories of financial impacts, along with some specific examples.

Subcategory	Examples
Direct Operating Expenses	Direct labor, materials used for fulfilling contractual obligations,

	subcontractor costs associated with fulfilling contractual obligations.
General Overhead	Rent, maintenance, asset purchase, asset utility, licensing, utilities, administrative staff and general procurement.
Staff Overhead	Staff necessary to run the business such as clerical, sales management, field supervision, bids and proposals (B&P), recruiting, and training.
Fees and Charges	Bank fees, service charges, commissions, legal fees, accounting, penalties and fines, bad debt, merger a acquisition costs.
Cost of Goods Sold	Design of products, raw materials, production, cost of inventory, inventory planning, marketing, sales, and customer management, advertising, lead generation, promotional events, samples, order replacement, order fulfillment, and shipping.
Revenue	Customer acquisition, customer retention, churn, missed opportunities.
Cash Flow	Delayed customer invoicing, missed customer invoicing, ignored overdue customer payments, quick supplier payments, increased interest rates, EBITDA.
Depreciation	Property market value, inventory markdown,
Capitalization	Value of equity
Leakage	Collections, fraud, commissions, inter-organizational settlement

Table 1: Subcategories of financial impacts.

5.2 Confidence and Satisfaction

Confidence and satisfaction impacts are associated with missing expectations associated with meeting expectations in the consumer marketplace or satisfying the internal ability to execute against strategy. In Table 2, we show some of the subcategories of confidence and satisfaction impacts, along with some specific examples.

Subcategory	Examples
Forecasting	Predictability of staffing, financial, material requirements, spend vs. budget
Reporting	Timeliness of reports, currency of reports, availability of reports, accuracy, need for reconciliation
Decision-making	Time to decision, predictability of results
Customer satisfaction	Sales costs, retention, purchases per customer, products per customer, sales costs, service costs, time to respond, referrals, new product suggestions
Supplier management	Optimized procurement, reduced commodity pricing, simplified acquisitions
Employee satisfaction	Costs to recruit, hiring, retention, turnover, compensation

Table 2: Subcategories for confidence and satisfaction.

5.3 Productivity

Productivity impacts are associated with hard measurements of operational efficiency. In Table 3, we show some of the subcategories of productivity impacts, along with some specific examples.

Subcategory	Examples
Workloads	Increased need for reconciliation of reports

Throughput	Increased time for data gathering and preparation, reduced time for direct data analysis, delays in delivering information products, lengthened production and manufacturing cycles
Output quality	Mistrusted reports
Supply chain	Out-of-stock, delivery delays, missed deliveries, duplicate costs for product delivery

Table 3: Subcategories of Productivity.

5.4 Risk and Compliance

These impacts, with some examples in Table 4, are associated with the ways that data issues can increase exposure to various risks, whether they are compliance risks, financial risks, or increased ability to execute in the marketplace.

Subcategory	Examples
Regulatory	Reporting, protection of private information.
Industry	Processing standards, exchange standards, operational standards
Safety	Health hazards, occupational hazards.
Market	Competitiveness, goodwill, commodity risk, currency risk, equity risk, demand.
Financial	Loan default risk, investment depreciation, noncompliance penalties.
System	Delays in development, delays in and deployment,
Credit/Underwriting	Credit risk, default, capacity, sufficiency of capitalization.
Legal	Legal research, preparation of material.

Table 4: Subcategories of risk and compliance impacts.

6 Business Impact Analysis

The business impact analysis process documents the impacts that are attributable to data quality. First, there may already be some existing documentation within an incident reporting systems, so determine if there is a logging and tracking system for data quality issues. If there is, review each logged issue to note:

- The business function/role reporting the issue
- How and at what point in the business process the issue was discovered
- Who discovered the issue
- Why this was reported and which business impacts are related to the issue
- How the issue was remediated
- Any business rules or data quality checks recommended

In addition, or if there is no data quality issue logging and tracking system, subject matter experts and representatives of the business functions should be interviewed to solicit descriptions of business impacts. The discussion with the subject matter expert is centered around "points of pain" and further drill-down to determine whether noted impacts are related to the use of data. An interview may take this form:

1) Open-ended request to describe most critical business application problems

2) For each problem, ask:
    a. What makes this a critical business problem? (This is to assess level of criticality)
    b. What are the business impacts? (This is to help is quantification of the issue)
    c. How is the business problem related to an application data issue? (establish a relationship with data)
    d. How often does the data issue occur? (both time frame and as a percentage of interactions or transactions, to gauge frequency and probability of occurrence)
    e. When the data issue occurs, how is it identified? (determine if there is an inspection process in place)
    f. How often is the data issue identified before the business impact is incurred? (determine if there is an existing data stewardship and data issue remediation process)
    g. What remediation tasks are performed? (note the effort for remediation)

Specific business impacts can be collected and then collated within the defined impact categories. An example of collected issues for review is shown in Table 5.

Impact Category	**Examples of issues for review**
Financial	• Lost opportunity cost • Identification of high net worth customers • Increased value from matching against master customer database • Time and costs of cleansing data or processing corrections • Inaccurate performance measurements for employees • Inability to identify suppliers for spend analysis
Confidence	• Improved ease-of-use for staff (sales, call center, etc.) • Improved ease of interaction for customers • Inability to provide unified billing to customers • Impaired decision-making for setting prices
Productivity	• Decreased ability for straight-through processing via automated services
Risk/Compliance	• Inability to access full credit history leads to incorrect risk assessment • Missing data leads to inaccurate credit risk • Regulatory compliance violations • Privacy violations

Table 5: Example areas of business impacts related to data quality.

7 Additional Impact Categories

These impact categories are not inclusive. Since every industry is different, every business is different, and every organization is different, there may be many ways that poor data quality affects operations or achieving goals. Understand not just what the impact categories are, but how they relate to different business activities, since this will help in both aggregating impacts and communicating those impacts to the right people in the organization.

Identifying the organization's impact categories depends on examining how data is being used for achieving business objectives, and how the achievement of those goals is impeded when flawed data is introduced into the environment. Business goals typically revolve around maximizing the way that the organization's customers are served, and that suggests a process such as this:

1. Enumerate the products and services that the organization provides;
2. For each product or service, evaluate what data is being used as input and how the output is employed;
3. For each product or service, determine the top 3-5 data objects that are critical to successful operation;
4. For the most important data items, consider what kinds of errors can be introduced into the system;
5. For each introduced error, review how the introduction of the error affects the ability to provide optimal value.

Applying this process should expose data quality impacts that related specifically to your organization. Here are some examples:

- Incomplete insurance claim forms not only introduce additional staff resources needed to collect the right amount of information, missing diagnostic codes ultimately lead to inaccurate actuarial computations, which may be used to incorrectly set provider rates.
- Inaccurate contact information used in locating debtors may delay the time to initiate income withholding, reducing overall collections.
- Poor identity management results in conflicts over how physical assets are distributed.
- Improperly documented mechanical components are viewed as non-working; proper documentation allows them to be put back into operation.
- Inaccurate inventory allows commodity components to depreciate when the application systems are not aware of the correct numbers of available items.

8 Impact Taxonomies, Iterative Refinement, and Measurement

The four high-level categories in this paper are not necessarily inclusive. If within your organization there are additional high level classes with their own impact subcategories, it is desirable to document the new class and the subcategories that belong to that class. Alternatively, the categories provided in this chapter may still be at too gross a level, requiring further segmentation. An iterative approach uses the categories listed in this paper as a model and allows the introduction of additional categories and classifications within a taxonomic hierarchy that shows how small impacts will roll up at different levels, and ultimately feed into a data quality scorecard.

The objective of designing a business impact hierarchy for data quality issues is two-fold. First, incrementally classifying impacts into small pieces for analysis makes determining how poor data quality impacts our business processes a much more manageable task. Second, the categorical hierarchy of impact areas will naturally map to future performance reporting structure for gauging improvement. As one identifies where poor data quality impacts the business, one also can identify we will also be identifying where data quality improvement will improve the business, and this provides a solid framework for quantifying measurable performance metrics that will eventually be used to craft key data quality performance indicators.