Thomas C. Redman

The Impact of Poor Data Quality on the Typical Enterprise

Poor data quality has far-reaching effects and consequences.

As practitioners know, creating awareness of a problem and its impact is a critical first step toward resolution of the problem [5]. The needed awareness of poor data quality, while growing, has not yet been achieved in many enterprises. After all, the typical executive is already besieged by too many problems, low customer satisfaction, high costs, a data warehouse project that is late, and so forth. This article aims to increase awareness by providing a summary of the impacts of poor data quality on a typical enterprise. These impacts include customer dissatisfaction, increased operational cost, less effective decision-making, and a reduced ability to make and execute strategy. More subtly perhaps, poor data quality hurts employee morale, breeds organizational mistrust, and makes it more difficult to align the enterprise. Poor data quality and its underlying causes are potent contributors to an "information ecology" [2] inappropriate for the Information Age. Further, leading enterprises have demonstrated that data quality can be dramatically improved and the impacts mitigated. Readers are referred to other articles in this section and to [7] for case studies and techniques for doing so.

Naturally enough, the particulars vary from enterprise to enterprise. Perhaps the most important point is that many of the problems facing today's executive have poor data quality at their roots. A practitioner can often create awareness by showing how poor data quality contributes to the better-known problems. This is often more effective than estimating error rates and following errors to determine the consequences.

An estimate of data accuracy levels in a typical enterprise is given in the following section; subsequent sections describe impacts at the operational, tactical, and strategic levels.[1] A number of examples from Fortune 100 companies are cited (though the identities of these companies are not given).

Data Quality Issues and Data Accuracy

Over the last several years, more and more references to poor data quality and its impact have appeared in the news media, general-readership publications, and technical literature [7, 11]. An enterprise may have a wide array of data quality problems. One way to categorize these issues is as follows [7]:

- Issues associated with data "views" (the models of the real world captured in the data), such as relevancy, granularity, and level of detail.
- Issues associated with data values, such as accuracy, consistency, currency, and completeness.
- Issues associated with the presentation of data, such as the appropriateness of the format, ease of interpretation, and so forth.
- Other issues such as privacy, security, and ownership.

The science of data quality has not yet advanced to the point where there are standard measurement methods for any of these issues, and few enterprises routinely measure data quality. But many case studies feature accuracy measures. Measured at the field level, error rates range wildly, with reported error rates of 0.5–30%. Naturally there are difficulties in comparing these error rates. For the purposes of this article, the following statements may be useful:

- Unless an enterprise has made extraordinary efforts, it should expect data (field) error rates of approximately 1–5%. (Error rate = number of erred fields/number of total fields.)
- That which doesn't get measured, doesn't get managed, so the enterprise should expect that it has other serious data quality problems as well. Specifically, enterprises are bedeviled by redundant and inconsistent databases and they do not have the data they really need.

Impact on Operations

Poor data quality impacts the typical enterprise in many ways. At the operational level, poor data leads directly to customer dissatisfaction, increased cost, and lowered employee job satisfaction.

Customers have every right to expect that their names and addresses will be correct, that the products and services they order will be properly delivered, that they will be billed properly, and that their accounts will be properly serviced. But simple errors get in the way—customers are not correctly addressed, they are sent the "medium" rather than the "small," and they are forced to spend time straightening out billing errors. Many customers simply expect the details associated with their order to be correct and they are especially unforgiving of data errors.

Poor data quality increases operational cost because time and other resources are spent detecting and correcting errors. The cost incurred by the customer service organization to correct customer addresses, orders, and bills is a typical example. I once worked with an organization whose sole purpose was to detect errors in the invoices of the enterprise's largest suppliers. Their annual budget amounted to tens of millions of dollars per year (see Chapter 5 of [7]). Similar costs are incurred throughout the typical enterprise. Thus, the shipping department spends time correcting errors it receives from the order entry department, the human resources department spends time correcting data about employees, and the supplier management department spends time correcting errors about the supplier base.

This article would be enhanced with an estimate of the total cost of poor data quality, but studies to produce such estimates have proven difficult to perform. I am aware of three proprietary studies that yielded estimates in the 8–12% of revenue range. More informally,

[1] Operations, tactics, and strategy represent an informal hierarchy of work performed. Loosely, operations are day-to-day tasks such as order entry, customer support, and billing; tactics are decisions made by (usually) mid-level managers that have consequences in the short-term to mid-term; and strategy is long-term business directions.

40–60% of a service organization's expense may be consumed as a result of poor data. These ranges are "good working estimates" of the cost of poor data quality.

Finally, at the operational level, poor data quality lowers employees' job satisfaction. One simply cannot expect the hotel clerk dealing with tired travelers whose reservations have been lost to exhibit a high level of positive morale.

Impacts at the Tactical Level

There is no evidence that the data needed and used by managers is any better than the data used by customer-service employees.[2] And the impact is far-reaching. First, poor data quality compromises decision-making. It is a widely accepted maxim that decisions are no better than the data on which they are based. And since any decision of consequence depends on thousands of pieces of data, the chance that a decision is based only on good data is extremely small. The slightest suspicion of poor data quality often hinders managers from reaching any decision. One executive explained it to me this way: "We spend about half our (decision-making) time just arguing about whose data is better!" And of course the most relevant data may be simply unavailable. While all decisions involve some amount of uncertainty, it is clear that decisions based on the most relevant, complete, accurate, and timely data have a better chance of advancing the enterprise's goals.

A more subtle way that poor data impacts decision making is that it makes implementation of data warehouses, whose purpose is to help an organization make better decisions, more difficult [1].

Second, at the tactical level, poor data quality makes it more difficult to reengineer. One way of looking at many reengineering projects is that they aim to put the right data in the right place at the right time to better serve a customer. But, as noted previously, one simply cannot serve customers when the data is not correct. It is interesting that many of the reengineering case studies cited by Hammer [4] involve data quality in one way or another.

Finally, just as poor data decreases employee job satisfaction, poor data quality also increases the mistrust that internal organizations may have for one another. A manufacturing facility I visited provides a good example. Two departments, call them A and B, each needed data about parts. Their needs overlapped considerably, but each needed a few fields that the other did not. Initially the data was maintained by department A, but the quality wasn't high enough for department B, so it developed its own database. Soon the databases became horribly inconsistent. The issue became a "lightning rod" and it became impossible for the two departments to work together.

Strategic Impacts

There is less direct evidence of the impact of poor data quality at the strategic level, as practitioners are only beginning to work there. But consequences that stem directly from operational and tactical issues are becoming clear. First, since selecting (or developing, evolving) a strategy is itself a decision-making process, we should expect strategy making to be adversely affected. Indeed, since strategy has much longer-term consequences and requires data from outside the enterprise that may be harder to acquire and that may be of uncertain quality, we should expect the impact to be at least as great at this level. Indeed, the lack of relevant, complete, accurate, and timely data—about customers, competitors, technologies, and other relevant features of the strategic landscape—may be the single biggest hindrance to developing sound strategy. Of course, the impact will also be more subtle and the effects will be much more difficult to observe directly.

Second, as a strategy is rolled out, specific plans are deployed, then modified as results are obtained. If the reported results are inaccurate, late, or in some other way of poor quality, execution of the strategy is much more difficult. One executive explained it this way:

"In our corporation, each unit makes P/L commitments to the Chair. In the fall, we get agreements on those commitments for the following year. Then we develop and implement plans to meet those commitments. And early each month, we get summaries of the previous month's and year-to-date progress toward those commitments. So if we're not on target, say, in February, I can adjust my strategy to meet my commitments. But I don't really get January results in early February, for example, I don't know what my true expenses are. Instead I get an estimate or forecast. Let's suppose the January forecasts show I'm on track. January actuals trickle in for many months. Not until July do I close the books on January. Sometimes in July I'll receive a real January expense record that is nowhere near the forecast. Now I'm way off track in meeting commitments and, since half the year is gone, I have virtually no chance of adjusting my strategy in any significant way. Of course, I may get lucky. The February actuals, which I finally close in August, may make up the difference. But even when I'm lucky, this is no

[2]A *Harvard Business Review* study [8] cites poor data in spreadsheets as the first solid evidence of which I am aware that data used in decision making is no better than data used in operations.

way to run a business."[3] [6]

Third, implementing a strategy requires operations. Thus the enterprise that selects a strategy of customer intimacy [10] must operationalize that strategy. But a company simply cannot become intimate with customers who do not trust that company because of billing errors. And an enterprise pursuing operational efficiency will be plagued by the costs associated with poor data quality described earlier. Similar factors negate an enterprise's ability to pursue product leadership.

Finally, issues that lower employee job satisfaction and breed organizational mistrust have strategic consequences as well. Data and enterprise culture are inextricably entwined. Data captures the enterprise's values and ideals, "filling-in the white space" in the organization chart and defining the enterprise's internal language. And poor data contributes to difficult, in some cases irresolvable, political situations. Seemingly simple exercises like developing a common definition of "customer" or arranging for one organization to access another's data can become no-holds-barred battles.[4] Indeed, the politics of data ownership are among the most brutal in many enterprises [3, 9]. The net result is to divert management attention from customers and the competition, sap the enterprise's strength, and make it more difficult to align departments and organizations toward common goals.

Table 1. The impact of poor data quality on the typical enterprise

TYPICAL ISSUES:
Inaccurate data: 1–5% of data fields are erred
Inconsistencies across databases
Unavailable data necessary for certain operations or decisions
TYPICAL IMPACTS:
Operational Impacts:
Lowered customer satisfaction
Increased cost: 8–12% of revenue in the few, carefully studied cases
For service organizations, 40–60% of expense
Lowered employee satisfaction
Typical Impacts:
Poorer decision making: Poorer decisions that take longer to make
More difficult to implement data warehouses
More difficult to reengineer
Increased organizational mistrust
Strategic Impacts:
More difficult to set strategy
More difficult to execute strategy
Contribute to issues of data ownership
Compromise ability to align organizations
Divert management attention

Concluding Remarks

The accuracy level and impacts presented in this article are summarized in Table 1. Creating awareness of these issues within the enterprise is the first obstacle that practitioners must overcome when implementing data quality programs. The more tangible impacts, such as customer dissatisfaction, increased cost, ineffective decision making, and the reduced ability to make and execute strategy, are bad enough. The softer impacts, including lower morale, organizational mistrust, difficulties in aligning the enterprise, and issues of ownership, may be even worse. Taken together, they contribute to an "information ecology" [2] that is simply inadequate for the Information Age. As noted earlier, it doesn't have to be that way—the techniques needed to improve are becoming available and have been effectively applied by a few leading companies [7]. C

References

1. Celko, J. Don't warehouse dirty data. *Datamation* (Oct. 15, 1995), 42–52.
2. Davenport, T.H. *Information Ecology*. Oxford University Press, New York, 1997.
3. Davenport, T.H., Eccles, R.G., and Prusak, L. Information politics. *Sloan Management Review 33*, (Fall 1992).
4. Hammer, M. and Stanton, S. *The Reengineering Revolution*. HarperCollins, NY, 1995.
5. Kotter, J.P. *Leading Change*. Harvard Business School Press, Boston, MA, 1996.
6. Redman, T.C. Improve data quality for competitive advantage. *Sloan Management Review 36*, 2 (Winter 1995), 99–107.
7. Redman, T.C. *Data Quality for the Information Age*. Artech House, Boston, MA, 1996.
8. Ross, J.A. Spreadsheet risk. *Harvard Business Review 74*, (Sept.–Oct. 1996), 10–12.
9. Strassman, P. *The Politics of Information Management*. The Information Economics Press, New Canaan, CT, 1995.
10. Treacy, M. and Wiersema, F. *The Discipline of Market Leaders*. Addison-Wesley, Reading, MA, 1995.
11. Wang, R.Y. and Strong, D.M. Beyond accuracy: What data quality means to data consumers. *J. Manage. Info. Syst.* To appear.

Thomas Redman (tomredman@aol.com) is President of the Navesink Consulting Group, a Rumson, NJ-based company that advises and assists organizations in improving their data quality.

[3] This example can be used to illustrate at least two other points about data quality, beyond the importance of data quality to strategy. First, it underscores the importance of timeliness or currency as a dimension of data quality. Second, it illustrates that finding errors and correcting them can take an inordinately large amount of time. Data errors are difficult to detect and correct—it is much better to create accurate data in the first place [6, 7].

[4] This is not to say that common definitions or unlimited access are necessarily good.