



## Why nobody has data quality issues, yet everybody seems to suffer

Tom Breur  
April 2009

### Introduction

Data quality is like an iceberg. It's an egregious, albeit largely hidden issue. The Data Warehousing Institute estimated in 2006 that poor-quality customer data costs US businesses a staggering \$611 Billion a year in postage, printing, and staff overhead. Yet we observe another fascinating phenomenon: *no* executives ever seem to have these problems! They all must assume their competitors have such problems, because they don't. Why is that?

Joe from corporate headquarters is well respected throughout the business, but feared is an adjective that also comes to mind. He compiles the reporting for board meetings each month. Everybody knows Joe needs his input on time so he can assemble and sanitize the numbers. When the board comes back with questions he chases people up.

Many companies have their "Joe." How else could they get consistent and timely reporting to the board? And because nothing ever seems "wrong" about reports that executives get, they never knew there was an issue with data quality. Much less will they know how much time and effort was spent "straightening out" numbers to produce these reports.

### Who wants to *hear* about problems?

In many, if not most organizations, there is infinitely more support for people bringing good news, than bad. So employees who (repeatedly) bring up problems like data quality issues don't do their career any good. At least that's the way it usually works. Young professionals quickly learn that if you want to advance, you don't bring up any problems. Problems don't exist, only opportunities.

What makes addressing data quality concerns even trickier is that the relation between cause (data non-quality) and effect (business process breakdown) is often not entirely clear, certainly not to non-technical business partners.

One organization we worked with ran a promotional campaign where customers could get an mp3 player for signing up (much in vogue at that time). Due to the larger than expected success of the campaign, the call center was flooded with traffic. In their haste (and because of a poor front-end application) they did a less than perfect job of recording customer details, which then led to complaints about late or undelivered mp3 players.

When we investigated this “straightforward” case of poor data entry, we were in for a surprise. The symptom here was a call center that got inundated with delivery complaints. The data entry application had been slapped together in a hurry for this campaign. Hey, isn’t marketing always leading the charge? Because the tool lacked basics like zip-code/address verification, and was woefully unergonomic, this was quickly identified as the cause for “lost” shipments of mp3 players.

Further analysis revealed that, yes, some premiums had been sent to wrong addresses, which were then not returned. Small wonder, these missent shipments contained a desirable bit of technology. However, that technically constituted “merely” a \$50 loss per error. Upon further analysis it turned out that the preponderance of costs for this organization were hitting somewhere else.

Since the campaign was a much bigger success than anticipated, it was difficult to keep up with the demand for premiums. The ad had promised 7-day delivery, which wasn’t met. The overwhelming majority of calls for this campaign dealt with sorting out why delivery hadn’t happened. And *that* was not so often because the premium had been missent. Agents spent an awful lot of time track-and-tracing the shipment, which often simply hadn’t been sent at all. At least, not yet.

There were some peripheral issues like possible fraud in the warehouse when empty boxes had been sent. Negative publicity costs were hard to quantify but deemed serious. But the conclusion was that call-center staff working overtime (extra cost), spending disproportionate amounts of time to investigate shipment status was costing exorbitant amounts. The \$50 per premium that was lost as a result of data entry errors paled in comparison.

Because the relation between problem (overload in the call center), and cause (supposedly data entry, but more accurately poor expectation management with customers) isn't quite clear, this precludes removing root causes. They expected "complaints due to non-delivery of premiums (mp3 players)", attributable to data entry staff that isn't doing their job very well. But instead they found the root cause to lie somewhere else. This makes the initial problem statement sound an awful lot like ordinary whining!

#### The importance of business cases

Senior executives will only turn their attention to data quality if someone comes up with a solid business case. This is hard, and (too) few IT professionals know how. IT is crowded with DBA's, not MBA's. Creating a business case requires making assumptions, which many professionals loathe. All these factors combined result in few, if any business cases for data non-quality being made.

The surprising thing about this phenomenon is that in those cases where we did go through the effort of calculating costs associated with non-quality, the numbers were always surprising. Most often they were (much) larger than imagined, and in other cases the assumed costs of data non-quality was small, but it turned out there were significant other hidden costs (like with this call center example).

If you want involvement from senior management, big cases draw a lot of attention. In general, management wants to see hard evidence, numbers, preferably tied to financials.

#### Who wants to *talk* about problems?

It is only natural that people *downstream* draw attention to "problems" that are caused by poor data quality. They are the ones who are suffering the consequences of poor quality, by the time this results in business process breakdown. Even with the best of intentions, and describing organizational processes in terms of cause and effect (rather than seeking blame), all too often the person bringing up problems that originate somewhere else is not seen as a "team player."

Another reason that prevents people from raising data quality issues is that subordinates who (repeatedly) bring up data quality issues will often be the ones who get assigned the nasty job of trying to resolve

them. This management practice never fails to work: when you raise an issue, you get assigned the duty to resolve it. To make the job even more interesting, usually without the means nor leverage to become successful...

Although many people *acknowledge* the issue of data quality and it's importance, it is mighty hard to find volunteers who are willing to do something about it. Why? Everybody has a role and place in the current "status quo." Also, the people suffering the "pain", the consequences of poor data quality, are often far removed from those who control the resources to resolve it. This lack of organizational alignment is probably the most fundamental cause for the perseverance of data quality problems in so many organizations.

## Conclusion

Data quality issues are pernicious, and given a number of dysfunctional management practices we have outlined in this paper, it is exceedingly difficult to drive out these problems. So probably data quality problems are here to stay with us for quite some time. Data volumes worldwide are growing at a staggering rate: IDC estimated in 2007 an annual compounded growth rate of 60%. Given the astronomical magnitude of costs that have been associated with data quality this topic merits more attention than it is currently getting. We have outlined some of the reasons for "sweeping it under the rug."

There are many more colleagues to be found who love to listen to good news and wonderful opportunities, than hear what is going wrong and how much that is costing the business. In particular if the business is (still) making money, data quality problems can be seen as a natural cost of doing business. Of course it isn't. Doing things right the first time around is always cheaper (in the long haul), which is why "Quality is Free" (Crosby, 1980).

Many professionals don't like to come forward with their insights about existing problems. Either because they are afraid they'll get assigned the unrewarding duty of resolving it, or because they haven't gone through the trouble of quantifying the magnitude and extent of the problem. That is plain hard work, and requires making "risky" assumptions. Setting up a business case is a skill that can be learned. Maneuvering it onto the corporate agenda might be more of an art.

Taken together these forces explain why data quality problems are so often underestimated at the boardroom level. The reports they are

getting look just fine, although it may appear to take remarkably long to produce them. Very few staff members are willing to present a business case that demonstrates the costs resulting from data non-quality they are incurring. After all, if the business is doing fine, why change anything?