Data Quality – is it happening to you?

Tom Breur
October 2009

Introduction

In some organizations, dealing with data quality issues always seems to happen in a "panic" mode. You get the impression of non-stop fire fighting. It's like a heath fire: whenever the fire is extinguished in one place, it reappears somewhere else. And you never know where.

In other organizations, data quality is pro-actively gauged and assessed. Ongoing data quality programs ensure conformance to pre-defined standards. Continuous improvement is carefully aligned with corporate strategy.

And yet others are blissfully unaware of *any* problems they might or might not have. Where would you rather work?

What are the symptoms?

Probably the clearest give away for a data quality "fire fighter" is that management *panics* over consequences or backlashes of data quality errors. These are typically publicity risks or imminent (financial) losses. This response in and of it self is only natural. What is worrisome is that BI gets involved. In the overwhelming majority of cases, the underlying data quality issues aren't new. Some report may have just *surfaced* them. What has BI got to do with that in the first place?

In most cases, it is the integration effort (probably ETL) that surfaced problems. Because the data warehouse (DWH) is the first place where data streams are confronted, that is where (previously existing) problems become apparent. You try to conform a dimension and the feeds don't match. Now BI gets to resolve a problem, as a reward for surfacing it. And probably without any management levers required to align (operational) stakeholders.

Also common are special taskforces that are put in place to deal with these problems. The (sudden) attention for data quality is fine. The solution, however, is not. Since responsibility resides with operational management, that is where you want to find a solution. And *not* within a project or program. The difference is incident management versus business as usual. And the latter is where attention for data quality belongs.

Another common symptom is that all "fires" appear equally important and urgent. Because there is no data quality monitoring (using data quality scorecards), no trends ever become apparent. A lack of business cases (see next section) precludes setting priorities. More fundamentally, because "fire fighters" typically have no BI nor data strategy, allocating scarce resources tends to happen on a first come, first serve basis.

The importance of business cases

What we've often observed is that "firefighters" are rarely able to come up with *any* kind of numbers as to how much poor data quality is costing them. "We've never done this", one of my clients said. He was suffering badly from data quality problems, but had no idea how to initiate improvement. Unless you *quantify* the impact, there is no telling where to begin. And neither what level of accuracy to strive for. This has everything to do with inability to set priorities, or plan (and invest in) sustainable improvement efforts.

It's easy enough to suggest that data quality should be prominent on the management agenda. But let's face it: data quality will *always* be competing with other topics for attention. So simply *stating* it is important, and that it deserves attention, doesn't work. What *does* work, is demonstrating exactly *how* data quality issues are related to business process breakdowns, and tying those to financial consequences.

For some reason, in each and every organization I have worked with so far, impact on bottom-line results invariably draws management attention. These are also the kind of facts that tend to "stick." Not only do they grab attention; they're remembered, too.

Of course making a business case for data quality is hard work. And it requires making (explicit!) assumptions. That's scary as well. My personal theory for the relative absence of business cases is fairly simple though: data quality, these days, remains predominantly in the

realm of DBA's rather than MBA's work. And that isn't right. The only way we can change that is by demonstrating unequivocally how seemingly "technical" data quality issues are tied in with general business process failures. And *that* is where MBA's could shine.

How to relate data quality to business processes

It is rare that data quality problems have a "purely" technical cause. In general, the cause lies somewhere else. As Michael Hammer noted as early as 1994 in his classic Reengineering the Corporation: "Seemingly small data quality errors are, in reality, important indications of broken business processes."

When two departments, say logistics and marketing, report different numbers on the inventory, (at least) one of them is wrong. Yet they are both convinced that the way they "see" it is right! In a perverse way that might just be the case. One company we worked with, was constantly haunted by these discussions. And even the accountant didn't agree with logistics about the counts!

One could argue that the accountant's view was "right", since their count would have been most likely to hold up in a court of law. But don't forget that the way an accountant arrives at his "definitive count" is as much the product of a model about reality as any other way of looking at the business.

Whenever two departments disagree, you can bet their misalignment is doomed to cause losses for the company. If marketing doesn't "count" the promotional materials, these (sellable) products are lost for the business. Either you are not selling inventory that you should, or products are offered that aren't available (causing expediting or other costly business disruptions). The point is not about a marginal cost of this inventory; the point is that a value stream is broken. And that *always* means a loss to the company.

The challenge for data quality professionals is to help the business "see" how upstream data quality errors cause process breakdowns further downstream. This can be "classic" sloppy data entry leading to lost shipments, duplicate mailings, ignored prospects, or what you have.

But this can also be a breakdown in the value creating primary process of the business (like a mismatch of inventory levels). Until you investigate these errors, it is hard to tell from looking at the bare

numbers. That, for one thing, is why accountants don't just look at the numbers; they also take sample counts from the warehouse, to triangulate their observations.

In both cases, you multiply number of errors times the cost of each loss. The challenge lies invariably in assessing "hidden" costs. Often it turns out that the downstream consequences of an "innocent" error can be dear. Let's take an example. One of our clients had a problem with too may returned shipments. This problem was noted in the call center, because that was where the majority of complaints came in.

When we investigated this case, it turned out that the *big* costs were somewhere else. Much to our surprise (as well), the costs of handling complaints, investigating where shipments had gone, whether they had already been sent, etc., was considerably higher than "merely" the lost packages. Once you factored in all the (over)time the call center had to put in, that was four times more expensive than shipments lost. Nobody saw that conclusion coming! Sometimes hidden costs are in plain sight. Note that their problem was not "excessive overtime in the call center" but rather "too many returned shipments"!

The way you investigate these cases is by gathering verifiable evidence. Simply report observable *facts.* Leave it up to management to interpret those, and set priorities. "cooking" this game, by making problems appear worse or more severe sets you up for an intriguing dynamic which can have only one outcome in the long run: you will loose your credibility and reputation. And once those are lost, there is really no way back, only a way out: out the back door.

Why do you need a "(data) strategy"?

The most important reason for BI departments to engage in a department and data strategy has to do with scarcity. In every company we have worked with, there were more things to do, more questions to answer than the available people in BI could handle. So every project you take on means saying "no" to something else. Unfortunately, that choice often remains implicit.

In order to maximize the impact and value of your BI efforts for the business, it is crucial to carefully align your BI efforts with the corporate strategy. One challenge you might face there is that corporate strategy need not be well defined. That makes seeking alignment awfully difficult. In other cases, there might well be "a" corporate strategy, but it is often vague and serves more to satisfy

public relations than to guide employee action. BI often finds itself more or less on its own, here. But a departmental BI strategy is needed nonetheless.

A pragmatic way to set strategy is to look at an existing model like Porter's (product versus price leadership), or Treacy & Wiersema (operational excellence versus product leadership versus customer intimacy, where the last two are basically an unfolding of Porter's product leadership). Then you plot where your company sits. You determine the "true" strategy of a company not by reading the corporate prospectus, but by determining what kind of results qualify for a manager's "excellent performance." BI should go out of its way to support these goals, and so should the BI strategy. From this strategy, in combination with capabilities and resources, you chart your feasible course of action.

Business alignment ensures your business gets the best possible value from BI. By "strategically" choosing your projects, you'll avoid many of the ever present "budget discussions", and will find much better support for data quality and business process improvement initiatives. Lead from the front, rather than trail behind.

Conclusion

Every company has data quality issues. In some places they may be small and well-managed. In other places they seem everywhere, coming after you like a bat in the dark. Instead of "waiting" for problems to "come to you", it makes far more sense to take a more pro-active stance. By assessing the extent and cost of data quality issues consistently, you stay "in control."

By bundling your BI activities in a departmental plan with accompanying BI strategy, you'll be able to hold your course. You will also get more mileage from your resources.

The soundest data strategies are invariably based on a direct relation between asset values form data and costs of data quality errors. Not only does this help in planning and sizing your projects, it also makes for clean and professional relations with senior management. Only BI professionals want to discuss "intangible value" of their work. The rest just wonders: "How much will I get for my investment?"

References

Michael Hammer (1994) Reengineering the Corporation. ISBN# 1857880560

