



"turning data into dollars"

Tom's Ten Data Tips – June 2006

Data Quality

Data quality gives a competitive edge. Everybody agrees how important good data quality is. And everybody has been agonized by erroneous data. We've all lost a lot of time working with crappy data, and "Garbage In, Garbage Out" is probably the most commonly cited proverb in IT. Then how come it is always so hard to find volunteers to do something about it?

Because the consequences of non-quality data are propagated throughout the organization, one seemingly innocent problem upstream can easily cause a dozen problems downstream, and sometimes even more! The accumulated costs of dealing with the resulting errors can become staggering. Tackling and resolving the issues that cause data quality problems is one of the most high-leverage investments a company can make, in a world that is increasingly relying on digital information.

Why do these problems exist, and why do they live on? It often appears to be business *misalignment* of the worst kind when many 'bystanders' realize there are indeed data problems, but nobody "owns" these problems. This commonly recurring phenomenon lies at the heart of the omnipresent challenge to find resources (both money and time) to overcome such data quality problems.

1. What Is Data Quality?

Data quality is determined not only by the *accuracy* of data, but also by relevance, timeliness, completeness, trust and accessibility (Olson, 2003). All these "qualities" need to be attended to if a business wants to improve its competitive advantage, and make the best possible use of its data. Data quality implies its fitness for use, *including* unanticipated future use. Accuracy takes up a special place because none of the others matter at all if the data is inaccurate to begin with! All other qualities can be compromised, albeit at your peril.

2. Data *Non-Quality* Is Expensive

"Reports from the Data Warehousing Institute on data quality estimate that poor-quality customer data costs US business a staggering \$611

billion a year in postage, printing and staff overhead" (Olson, 2003). There are many ways in which non-quality data can cost money: typically these costs remain largely hidden. Senior management either doesn't notice these costs, or even more likely: is grappling with problems of which it never becomes clear that they are *caused* by poor-quality data.

3. Quantifying The Cost Of Non-Quality Is Very Important

Since data quality has such a strong tendency to go unnoticed, it is even more important to translate the consequences of poor-quality data to the one dimension each and every manager understands so well: dollars. This also gives a perspective on the kinds of investments that are appropriate to make in order to resolve such issues. Also, a mechanism for prioritizing improvement programs is desirable. You want to begin picking the low-hanging fruit first, but you certainly *also* want to know where the whoppers are! According to Gartner, Fortune 1000 enterprises may lose more money in operational inefficiency due to data quality issues than they spend on Data Warehouse and CRM initiatives.

4. Data Quality Issues Typically Arise When Existing Data Are Used In New Ways

In my experience as a data miner, where I am very often looking for new ways of using existing data, this is where many problems originate. The data itself hasn't changed, but it are new uses for existing data that make problems *apparent* that were already there. So what constitutes "data quality" needs be considered *in relation to its intended use*. And change of usage then brings up new ways to evaluate the quality and hence may bring up concerns. The reason these problems didn't surface before is usually because the business adapted to the data, the way they are. People and processes avoided the consequences of inaccurate entries. Which incidentally, is also why legacy system migrations can be so painful.

5. Many CRM Projects Collapse Under Data Quality Issues

Gartner and Forrester have estimated that 60-70% of CRM implementations fail to deliver on expectations. That is not to say that these projects are all abandoned halfway; it's foremost that expectation aren't met. One of the biggest reasons for the 'technical' challenges in bring CRM projects to completion is that disparate data sources are getting merged to create a 360° customer view. Often,

this is the first time that customer records of disparate systems are merged. There is typically tremendous “fallout”, and records that do get merged contain many inconsistencies. This then often leads to disappointed end-users, and unmet expectations.

6. Data Quality Is A Management Issue, Not A Technology Issue

The typical situation in the overwhelming majority of organizations I have visited is like this:

- there is *low awareness* of the embedded cost of their data quality issues
- management has no idea of the potential value in fixing data quality issues “upstream”
- those who have insight in data quality issues have little or no incentive in bringing these issues out

Hence, the problems have a nasty habit of perpetuating themselves. For sure, subordinates need to carry their weight and take responsibility. But notice how far all three of these issues, essentially the final responsibility for bringing these “unwelcome surprises” out in the open lies with *management*. What is the culture like in *your* company? My experience has been that managers may or may not be motivated to bring such issues out in the open, sometimes depending on the time horizon they consider for their own tenure.

7. Manage Data For What It Is: A Strategic Resource

Data is not merely a byproduct of business processes, but something that has value beyond its immediate processes. Finding new uses for existing data makes it more valuable, at no capital investment! Future changes to the way the data are to be used cannot be predicted, yet are guaranteed to happen! This proliferation of data usage needs to be anticipated, and calls for flexible data models. Good database design is resilient in the face of *unanticipated* changes. This means flexibility in hardware/infrastructure on the tangible side (avoid vendor or platform lock-in). On the intangible side, you want to avoid aggregating or any other data commitments that can not be reversed within the data scheme. It is fundamentally impossible to find a generic “right” way to aggregate inconsistencies in data. That is why flexibility calls for *late* commitments in the data model.

8. Higher quality data lead to far more flexibility for your corporate strategy

Fast access to accurate data *not only* gives a competitive advantage. What is even more important is the flexibility such companies enjoy in adjusting to changes in market conditions. So over time, as market changes *will* occur, the gap with the competition can grow even further. Also, changes in legislation or market regulation can be much more easily exploited and turned into an opportunity rather than 'suffered'.

9. Data Quality Improvement Is A Process, Not An Event

In many ways, one can draw parallels between Total Quality Management efforts, and the issues surrounding data quality. The Japanese use a word "Kaizen" that denotes both an incremental improvement method as well as a philosophy. What is crucial is that it's an on-going, never-ending effort to keep raising the bar. Data quality is never "perfect" as every new application of existing data is likely to bring up new issues. And the proliferation of data usage is not ending any time soon. So data quality issues are guaranteed to stay with us for a while.

10. Collecting Data Is Only A Few Decades Old

No wonder we're dealing with "growing pains". Few corporations actually planned their data strategy, and their IT infrastructure grew in a time when data were being handled in silos. As data are being shared and warehoused increasingly, we need to think through the goals and objectives of the enterprise with regards to the data. This is all fairly new, and few if any 'established' standards exist. A sort of 'global plan' or 'road map' as to where and how to expand on existing capabilities is a sound investment to manage project risks. Also, this 'road map' needs to conform to the existing IT strategy. Time and money will only be invested if project goals are in line with the overall corporate strategies. The road is littered with unsuccessful BI projects, many of which started without a clear business case. A well-conceived data strategy greatly leverages the considerable investments that are needed to get the best mileage from your data.