

GETTING TO GRIPS WITH DATA QUALITY

A strategic data quality programme can serve as a catalyst to bring business and IT closer in the organisation's quest to boost overall business efficiencies, cut costs and improve profits.

By Bryn Davies

A BUSINESS IMPERATIVE

We have treated data as a by-product of the business processes that create it for far too long. In so doing, we have built not databases, but data dumps, and this has been compounded by the ease with which we have spawned new systems with copies of the same data across our empires. The fact is that these multiple systems all have to interact as they play their role in business operations, and an entire industry has grown up to manage the integration of these silos. We might even have eventually got the integration part right, only to realise that we had just moved data non-quality up another notch, and introduced non-aligned data across multiple systems.

Coupled with this, IT has evolved largely as a means of automating business processes, and as a result, the industry has become fixated on the 'application', while the 'data' that drives it has received secondary status. This oversight, compounded by our almost uncontrollable rate of data collection, has resulted in a growing desire to manage it as it justifiably should be. Its quality is one aspect, but it is an aspect that is critical to overall business success.

ACCEPTING THE UNACCEPTABLE

Just how serious is the problem? In something like the building industry, you need to get things right first time – a bridge or a shopping complex simply can't be built with structural defects. The potential consequences are obvious, and the inherent understanding of the time,



Reporting, analytics and predictive analysis are all inherently flawed from the outset without sound data quality.

effort and costs of having to work around problems and redo the initial effort usually ensure that an acceptable level of quality is adhered to from the start.

Unfortunately, this is not the case in the data creation business – i.e. your business. We all, to a lesser or greater degree, quite happily allow incorrect, invalid and inconsistent data to be entered into our databases. The consequences go largely unnoticed, or are simply accepted as part of normal

➔ A MEASURED APPROACH

There are a number of core activities associated with data quality, but none as important as measurement. Before embarking on an exercise to improve things, it is critical to first find out just how good or bad your data actually is. This usually takes the form of a data profiling exercise where the deliverables are a view of data quality levels expressed in relative terms, as well as a scope of the effort required. Inevitably, in order to get ongoing buy-in from senior levels, data quality needs to be expressed in business impact Rand terms as well, and an exercise to prove the potential cost savings and increased revenue potential needs to be performed.

practice. Because high quality data is in general not expected in the first place, poor quality data is tolerated, and we simply accept the fact that we need to spend time every day in hunting down correct, trustworthy information to complete or fix a business transaction.

There are clearly many direct and indirect costs associated with poor data quality, but if you're not looking for them, they're not always obvious. In general, the effects of poor data quality manifest further down the line in other departments or at higher levels in your organisation, or even within your customer base. For example, an irate customer calling to complain about incorrect billing, or incorrectly recorded package dimensions which only get noticed when the products don't fit onto the shelves.

CALCULATING THE COSTS

So what are the types of costs that result from a lack of quality in our corporate databases? At a high level, the two major categories are (a) the costs of redoing things due to initial errors, and (b) the cost of lost opportunities. The first category takes people's time and other costly resources



to continuously fix what manifest as business transaction errors. To make matters worse, this often occurs only once the company has been alerted by the customer on the receiving end of the mistake, incurring additional costs in ensuring ongoing customer loyalty.

But there are other types of more pernicious consequences, with attendant costs, to low quality data: frustrated employees who have to deal with the resultant inefficiencies, a need to perform frequent verification of figures in reports, re-coding of computer applications and database fixes, and the costs of multiple redundant copies of the same data. All of these add to the cost of lost opportunities, and of course, diminished business agility.

NEW STICKS AND CARROTS

There are also new factors driving data quality. Perhaps one of the biggest sticks has been substantially increased regulatory compliance requirements: there is nothing more effective than the threat of hefty penalties and/or a stint behind bars to get resources mobilised. Banks, for example, have had to invest heavily in Basel II initiatives, as well as focus on Sarbanes-Oxley (SOX), Anti Money Laundering (AML), FICA and RICA, all of which need to be underpinned by high quality, trusted data.

On the carrot side, there is a growing realisation that data is an asset that influences every area of the business, and therefore needs much better management. CRM and gaining a 'single view of the customer' is impossible if unreliable data is replicated across the organisation. Similarly, BI and BPM have become top priorities for most companies, but reporting, analytics and predictive analysis are inherently flawed from the outset without sound data quality. After all, dirty data leads to dirty decisions. And because BI relies not only on the data itself but also on its meaning and presentation, it has also become necessary for organisations to re-assess how they model, store and present their information.

HOW DIRTY IS DIRTY?

Data is multi-faceted, making any attempts to measure its quality a potentially difficult and complex task. However, bearing in mind that the quality of any product is determined by its users, shift the focus from merely measuring the actual object (in our case, data) to measuring the impact of its lack of quality on users (in our case, our employees, partners and customers). This provides objectivity to the measurement of data quality, in that it serves as a guide to which aspects of the data should be assessed. This means understanding the information value chains that span our business processes, and the value of quality data to those who depend on it. It also means that what is measured will be different for different data sets, divisions, companies and end users, and might well also be different from one month or year to the next. *Bottom line: Don't just measure because it can be measured – measure what matters*

Another more recent trend has been to profile and cleanse data within legacy systems, before moving it to, for example, a new ERP package. For years, data has simply been 'mapped then moved', only to cause significant problems in the new system because quality levels are unable to support enhanced functionality. This is a bit like moving house but taking all your junk with you. Better to thoroughly clean, consolidate and enhance data before trying to shoe-horn it into its new home, and so data profiling, standardisation and matching has become common practice as sub-projects of ERP implementations.

TACKLING THE CHALLENGE

Whether you see it as a carrot or a stick, all paths eventually lead to an imperative to do something about improving the quality of the only re-useable resource any organisation possesses – its data. But the fact that your data is not up to scratch is merely the symptom – the

cause will be found in defects in the processes and activities that create the data in the first place.

It's natural to want to tackle dirty data by simply 'cleansing' it, in itself a daunting task for some organisations. However, in order to sustain high levels of data quality, it is also necessary to fix the problems with its creation and modification on an ongoing basis. This generally not only involves additional validation checks in computing applications, but must also include a concerted effort to examine and address inefficiencies in the human, transactional and workflow elements involved in collecting your data – in other words the root causes of the bad data that winds up in your databases. Remember too that poor data quality also stems from poorly designed databases, and so the structure, not just the content, of your databases must also be closely examined and improved, by properly aligning the data model to the business requirements.

Granted, this all significantly broadens the scope of any data quality initiative, and takes it to perhaps unanticipated levels in the company. But it is this very subtlety of a data quality programme that can serve as a catalyst to bring business and IT closer in the organisation's quest to boost overall business efficiencies, cut costs and improve profits. Indeed, a data quality initiative can help to promote issues such as information ownership and accountability, drive efforts to identify and document agreed business rules and metadata, and ultimately form the backbone of corporate data governance or master data management programmes.

Business rules are data quality rules, making data quality neither an IT issue nor a business issue – it really is *everyone's* issue. 🇨🇦



Bryn Davies is Practice Manager of BI Practice Cape Town

