soltius

WHITEPAPER

Data Quality: How data governance addresses minor errors that cause major problems

Small problem, big impact: A real-world example

Minor data entry problems can quickly snowball. Here's how:

In a customer address record, the city of Christchurch is misspelled or missing altogether. This happens because a business system has no ability to verify addresses or enforce correct entry.

What are the consequences?

- At the executive level:

The organisation doesn't get accurate insight into performance per sales area. The sales associated with the customer may be overlooked (empty city) or attributed to a non-existing location. In the best case scenario, an executive notices the issue in a sales dashboard and requests resolution from Business Intelligence and Analytics.

- At the operational level:

Goods shipped to this customer may not arrive on-time or at all and invoices that cannot be delivered will not get paid. These problems impact the efficiency of the supply chain as goods are returned to the organisation and must be reshipped. Even worse, these problems can lead to immediate customer dissatisfaction and possible penalties if the organisation is unable to meet contractually agreed delivery terms.

- At the customer level:

The organisation's goodwill is affected. The customer may not receive the ordered goods in time, or at all. Advertising sent to a wrong address is wasted if it doesn't arrive or could offend if details are incorrect.

Executive summary

It is an issue which is almost as old as the technology industry itself, but one which persists to this day: poor data quality. No matter how good the technology, the system or the process, if it is based on inaccurate information, the outcome cannot be trusted. Yet, despite acknowledgement that data quality is essential for effective operational processes, and becomes incrementally more critical for business intelligence and analytics, it either isn't perceived as a problem at all or languishes in the 'too hard basket' for a great many organisations.

As a result, when business intelligence projects are initiated, they often falter. When it is discovered that the data upon which analysis is to be performed cannot be trusted, the costs balloon and delivery schedules are affected. The very success of such projects is jeopardised – and all because of a simple, well understood fact: garbage in equals garbage out.

A 2013 SAP-sponsored whitepaper, 'The State of Data Revisited' reveals that despite considerable investments in data quality initiatives, little had improved. 'There has been an almost doubling of data quality initiatives since 2009, but no marked improvement in the perception of data quality in the eyes of organisations. Hence data quality must be perceived as a big challenge and one which has largely defeated companies', it notes.

Closer to home, a micro-study performed by Soltius at the CIO Summit reveals that most organisations have a data quality strategy; many have projects in the pipeline where data migration is necessary and most recognise that data migration presents a significant risk to the project.

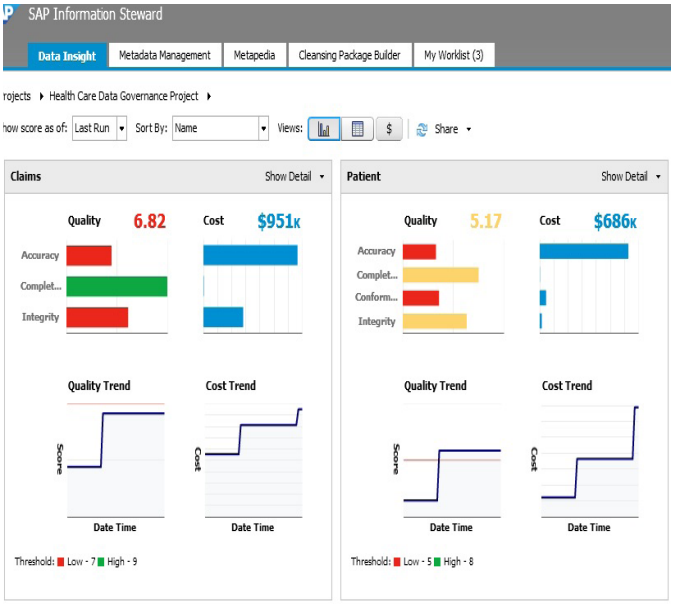
In this white paper, the problem of data quality is discussed and approaches suggested for how organisations can ensure better data gathering, improve the quality of existing data, and why data governance is necessary for ongoing quality control of enterprise information.

1. Data quality at source – and why most companies don't believe it is a problem

Data quality, or the absence of quality, often starts at the very front line of the organisation. Typically, data is gathered by people typing it into forms, giving rise to the obvious potential for 'finger trouble' (simple keying errors), or, where system checks require field

completion, the propensity to make a meaningless entry to keep the process moving.

Such errors are typically tolerated as they are seen in isolation. Correcting a single customer record, after all, takes perhaps five minutes. Most business managers view this as so inconsequential that the very idea of investing in technology or training to find and fix the problem isn't worth consideration.



However, a single record is one thing; thousands of records multiplied by five minutes is another. Furthermore, a minor data problem can have far larger implications on time and effort further up the value chain: an error in a bank account number, for example, doesn't just mean re-entering data. It may mean a missed payment and considerable administration to identify and resolve the issue. All that takes time – and the time of Full Time Employees can be measured in dollar terms.

Data quality at the point of entry is an issue so minor it generally isn't even perceived to be a problem. However, it is akin to sand filtering into a big machine. Individually, the small grains of sand don't cause great damage, but when enough grains accumulate, the machine eventually breaks down. The big machine is the broader organisation and as customer data is elevated through the ERP system and ultimately into data warehouses and Business Intelligence systems, minor inaccuracies can mean major problems. It is death by a thousand cuts, as these advanced data initiatives effectively have poor foundations.

And even at the operational level, it means the organisation is bleeding money and efficiency, as employees must revisit records to repair improperly entered data.

1.1 Existing data: 'Pretty Good'?

When performing data projects, such as a business intelligence or analytics initiative, a first step taken by a BI consultant is to ascertain data quality. Typically, customers consider their data to be 'pretty good'. Indeed, the SAP white paper confirms this view, with 63% of those surveyed claiming their data to be 'good'.

Soltius' micro-study reinforces this view among local CIOs. Of the 28 individuals surveyed, 85% indicated that their organisation has strategies and technology in place to manage integrity and quality of data on an ongoing basis. Half those surveyed answered the next question 'do you think you need a strategy for data integrity?' with just 47% of that half saying 'yes'.

However, while it appears that CIOs believe their data is sound, it belies the issue as in reality most organisations have no way of measuring data quality. Until it can be measured, any comment on data quality is itself merely qualitative. In effect, most organisations are operating in the dark inasmuch as data quality is concerned.

However, upon project execution, a clearer picture of data quality inevitably emerges and it typically isn't good. Since BI depends on input data for its outputs, such a discovery puts the project at risk.

These projects typically start with high expectations, which become tempered by the reality of poor quality data. If masses of data are to be corrected, overruns are inevitable. If the quality isn't addressed, the project cannot deliver value.

While poor data quality has a quantifiable impact on operational performance, the impact is amplified with enterprise data projects, becoming much more expensive to resolve and even putting such projects into jeopardy. For example, ambitious projects like predictive analytics cannot succeed if the data on which they are built isn't reliable. And more often than not, data quality problems are often not discovered until such projects are well underway.

With a better measure of data quality, such projects can (and should) be deferred until reliable data is available,

avoiding capital and other expenditure until such time as those investments are likely to achieve the desired results.

2. Data governance: Solving the data quality conundrum

Data quality generally occurs as a result of process issues, as data is the output of a process. It is often the case that processes are not engineered with sufficient checks and balances to ensure quality output; instead, people are measured on volume rather than accuracy. This approach is, however, short sighted, as is made clear in this white paper.

Therefore, the first initiatives to improve data quality should be targeted at the source of the problem: by improving processes for data collection, which may include training for operatives.

In addition, tools for data quality management help detect individual data issues across all business systems, while also correctly attributing the financial impact of each issue to provide insight into the total impact of poor data quality across the entire organisation.

2.1 Better processes should be matched with appropriate tools

Tools are also available to drive improved data governance. They include SAP Information Steward, SAP Data Services and SAP Master Data Governance solutions.

SAP Information Steward is targeted at business users, not technical people, who understand the processes and systems from which information is drawn.

SAP Information Steward combines data profiling and metadata management to provide continuous insight into enterprise information quality, supporting operations, analysis and governance. It provides transparency into data origin, with the ability to consolidate, integrate, and audit metadata from all sources, while improving data governance with consistent validation rules and guidelines.

Using this tool, business analysts can build individual rules for data. Even basic rules can have a big impact: from validating addresses against the national postal system dictionaries, to more complex rules, such as ensuring stock positions across systems and facilities, to managing discount grading thresholds.

Perhaps most importantly, SAP Information Steward also allows the analyst to investigate the cost of each incident of inaccurate data and define scorecards over time to aggregate and score data quality.

The ability to put a price to data quality is a breakthrough, as it helps organisations assess the true cost. Instead of being seen in isolation, and therefore not worthy of investing tens of thousands of dollars to correct, data quality is provided with a 'big picture'. When that big picture represents costs of millions, it is possible to justify investments to improve data quality.

The primary tool to build and populate the data warehouse is SAP Data Services, an integration tool for large batch processing in particular for solutions that require complex data integrations or resolution of data quality issues.

Rules defined in SAP Information Steward can easily be integrated into SAP Data Services, allowing for the identification of data which might be faulty (which forewarns that the output of the BI system may be faulty) and influencing decision-making habits.

Importantly, corrections effected by SAP Data Services should be loaded into front-end systems, eliminating the operational issues which result from inaccurate data.

Together with SAP Information Steward, which can graphically present data quality trends, SAP Data Services provides the ability to detect violations of data entry rules so these issues can be nipped in the bud.

SAP Data Quality Management provides the capability to automatically correct postal address data upon entry in SAP business systems.

For example, it corrects addresses to the New Zealand Post SendRight™ standard upon entry in the core SAP business systems, thus improving the delivery rate of physical items or effectiveness of marketing campaigns. The SAP Data Quality Management can also detect duplicate entries of customer, vendor and supplier data; and guide business users to select the appropriate data, preventing erroneous creation of new variations of existing customers, vendors or suppliers.

Finally SAP Master Data Governance ensures ongoing data quality management once data is in a clean state. It ensures central management of master data manipulation through an approval workflow which integrates business rules and subject matter domain expertise so any changes made are understood in the broader context.

By centrally maintaining financial, supplier, customer, material and other master data for consistency across the enterprise and providing an audit trail of when, why, and by whom master data is changed, data is continually validated, ensuring its quality.

3. Summary

Solving data quality issues is best achieved at the point at which data is captured. For many organisations, that point is the call centre or other customer-facing situations, such as in-store or via online self service. The nature of the data itself is often quite simple: physical and email addresses and other customer information. Most organisations don't view inaccuracies in data at this level as a particularly costly problem, or even a problem at all: corrections to individual records, after call, can be solved with a phone call or an email. However, when thousands or hundreds of thousands of records are inaccurate, the costs very quickly balloon.

Naturally, data quality initiatives have to be applied to existing data stores, too, either at source or within the data warehouse. This process is invariably challenging, time consuming and costly; however, once executed, 'clean' data should not only be forwarded to the business intelligence system, but also repopulate production systems to ensure accuracy and consistency on an ongoing basis.

Ensuring data quality on an ongoing basis is a process, not a destination and falls under the broader concept of data governance. With effective data governance, organisations are assured not only that information gathering is accurate within operational procedures, but that data which is used for strategic insights can be trusted, resulting in business intelligence initiatives which deliver true value.