

DATE	March 2009
CIRCULATION	6,299

**DATA
STRATEGY**

DATA QUALITY REVIEW

Upstream data blocks that flood the downstream

Far from being the afterthought that is once was, data quality is rapidly moving up the agenda. **David Reed** reports

When data quality gets discussed, it is usually in the context of its impact on local operations, such as lists for direct mail, email or telemarketing. Errors and duplicates reduce performance of campaigns and create waste.

Upstream from these areas of activity, less recognition exists about the potential costs and harm that can stem from poor data quality. Yet as companies increasingly move towards data-driven decision-making, confidence in the underlying data sets is starting to emerge as a key concern.

Unless the single view of the customer is clean and accurate, the company may draw up plans based on false assumptions about the size of opportunities or level of risk. Equally, predictive analytics can only be accurate if the analytical data set is not skewed or full of missing or inaccurate variables.

That is why Richard Lees, chairman of Database Group, says: "We are seeing a shift. Marketing has tended to pigeonhole data quality as purely a concern in communications. When companies are making decisions about their product ranging, estate management or risk, however, the quality of their data takes on an extra dimension."

In the context of multi-channel companies, this shift is even more marked. Companies that set up their dot.com operation as a standalone division have been rapidly integrating their e-commerce with their physical world operations. Ensuring that individual customers do not get identified as duplicates across each channel is essential to this process.

"Most operations haven't had a view of those multi-channel customers," says Lees. "Yet customers are often researching online, then going offline to buy. You need to know about the journey they are taking."

A good quality single customer view is a fundamental requirement for companies in this scenario. While the focus of SCV projects is usually on the data asset which results from customer data integration, data quality has to be identified as a core and ongoing element of the programme.

"There are only two ways to increase your number of customers – get more new ones or increase your rate of retention," says Lees. He notes that since the size of the existing customer base is a function of attrition rate, companies are starting to realise how vital it is to be able to communicate accurately with those customers. "That is when data quality comes into play," he says.

This is often the first time that an organisation has thought about data quality as a standalone issue that needs to be resourced in its own right. That throws up a range of challenges, from introducing new processes to support the quality of data on the way in to investing in cleaning and validation technology.

"Ownership of the data budget has to be decided. If the business has set out to use customer data across a range of strategic initiatives, then it will begin to value data and will therefore need to apply a budget to it," he says.

Return on investment into customer data quality is usually relatively easy to calculate, from reducing churn to increasing multi-product holding. But data quality projects do not have to be limited to customer records – other data types can demonstrate even bigger returns and equally larger downside risks from errors.

Lees points to the example of pricing data for a petrol retailer. Each 1p increase in the price of fuel can lead to a 40 per cent increase in margin. So having accurate data on your own sales and competitor



DATE	March 2009
CIRCULATION	6,299

DATA STRATEGY

DATA QUALITY REVIEW

IN ASSOCIATION WITH



pricing is essential for the price elasticity modelling that is used to decide whether to increase or decrease pump prices.

Alan Thorpe, commercial director at G2 Data Dynamics, agrees that confidence in data is becoming a critical issue. "Whenever we produce analytics, we have to have it sense-checked by the person who is going to take those findings to board level. They want to be certain there are no red herrings," he says.

As a result, he is seeing a growing emphasis on data governance and master data management projects. In many cases, this is the first time that organisations have made a concerted effort to clean and maintain their core data assets.

"One client is excluding 30 per cent of its customer base because there are fundamental elements missing that would unbalance the analysis," notes Thorpe. Not having date of birth is one example of a variable that can be critical, but which is routinely under-populated.

The process of mergers and acquisition is often a major reason for the gaps in data quality. Each time a new operation is added to a group, data feeds come on stream that may be in entirely different formats and at a different standard to the host data. Little effort has been expended in standardising this data in the past.

Legacy systems are another cause of poor data quality. "The company might have sold an investment product and not had any further contact for 20 years," he says. When the purpose of an analytical project is to identify cross and upsell opportunities, those assumptions need to be based on valid data. "That may justify retrospective activity to improve data quality," says Thorpe.

Few organisations are in the position to start with a clean sheet and create their customer data assets to the right standard. Instead, there is usually a host of tactical activity to keep data clean at point of usage and, occasionally, at the point of entry. That does not add up to a coherent data strategy, however.

What is different in the current climate is that companies are looking for a more solid solution that keeps data fit for purpose, rather than fixing it each time. SCV builds are one opportunity to make that happen.

"When we're managing data for somebody, we start out from the point of view that the client will want to achieve x, y and z and design the solution to take that into account," says Richard Higginbotham, marketing manager at CDMS. Forecasting what the managing information and analytical needs might be ahead of time is never easy for a client. That does not mean the data architecture should not be optimised for unforeseen requirements.

One of the features which Higginbotham says is becoming more important is version management. "We build a file and track changes in those files. That provides both a current view and an historical view. That can be very important in reporting," he says. "It is about the data being fit for purpose – even if we have not been briefed to do that."

Confidence is one of the elements missing from financial and housing markets at the moment, with the result that few transactions are taking place. Unless confidence exists with an organisation about

the quality of its data, business users will not want to act on it. Just ask any sales team that has been passed hot leads, started to make contact and found that many of the phone numbers do not work.

This confidence has to be built from the ground up, by ensuring that all data entering systems is validated for both content and format, as well as from the top down, by demonstrating that the board sees customer data as a critical business asset. Fail to get both aspects right and belief in the project evaporates.

"It is like going to war in Iraq based on evidence of WMDs," says Martin Doyle, president and CEO of DQ Global. He points out that there is often a knee-jerk reaction when an error is found. "There is a view that if one piece of data is wrong, all of it is wrong."

The dictum that one bad apple spoils the whole barrel does not apply to a customer database. Every database, even those subject to the highest levels of data governance, will have missing, invalid or incorrect fields and records. This does not stop the business from operating or making decisions.

Equally, the most accurate data and forecasts may be presented to decision makers without them taking any notice. Doyle points to warnings given to RBS that the ABN Amro deal was too big, too expensive and too risky which were evidently ignored.

What concerns him about current approaches to data quality is that they are often post-hoc. "Data is often the last thought in a marketing campaign, but the first thought for blame," says Doyle.

One example of this he cites was a project to match email addresses. The file was provided in CSV format with each data cell separated by a comma. When imported into Access database software, however, every comma was identified as the start of a new record.

"In the body of the text of some variables you might find a comma, such as the business type 'agriculture, farming, fisheries. The default operation of that software is to separate those into cells, unless you pre-qualify it. Most marketers don't understand how that can happen," he says.

If targeting data is being processed and merged in a rush, which is the usual state of affairs, it is easy for this sort of predictable error to creep in and not be recognised. The outcome is a higher than expected level of undeliverable messages, leading to complaints about poor quality data.

Doyle likens this to catching a bus. "If it is nearby, you run for it. But there comes a point when it is too far away, so you don't bother. Data quality is often considered to be beyond marketers' effort because they are not used to working at that level of detail," he says.

His company has developed a new data cleaning solution that is based on the "right first time principle". To be effective, the technology needs to be aligned to processes that ensure data is correct, complete and consistent.

At Harte-Hanks Trillium, senior vice-president of marketing and services, Len Dubois, is seeing similar shifts in the way clients deploy and support his company's application. "We use the term 'veracity' to help our clients understand the business aspects of the challenge of data accuracy. Veracity directly impacts their business."

“
When we're
managing data
we start out from
the point of view
of the client



DATE	March 2009
CIRCULATION	6,299

DATA STRATEGY

DATA QUALITY REVIEW

IN ASSOCIATION WITH



In the financial services industry, regulatory pressures are making this concept ever more germane, since decision makers have to be able to demonstrate the grounds on which they made a particular deal. Accurate data is a fundamental when dealing with a SEC or FSA enquiry.

"It is not just data quality at the point of use, it is data quality as a service at an enterprise level," says Dubois. Delivering that service should be the task of a unit that combines the technical skills of an enterprise data architect with the knowledge of a business analyst.

"That also helps with risk mitigation. Organisations are trying to understand what data is putting them most at risk. They need to make that transparent to the business using data profiling solutions, dashboards and reporting tools to give a view of the data," he says.

A current example of how data quality connects with risk is credit rating agency data. A financial services provider may base its risk assessment on seeing a AAA rating for a customer from Moody's. The terms and price of the deal will be set based on that rating.

“
Ownership
of the data
budget has to
be decided

However, if the ratings agency downgrades the customer to AAB, the risk factor has increased. "The KPI that the business user wants is to understand is how much of their loan book is AAA and how many have changed to AAB and is at risk of default," says Dubois. Tracking changes of that sort is one example of how Trillium is being put to work by clients.

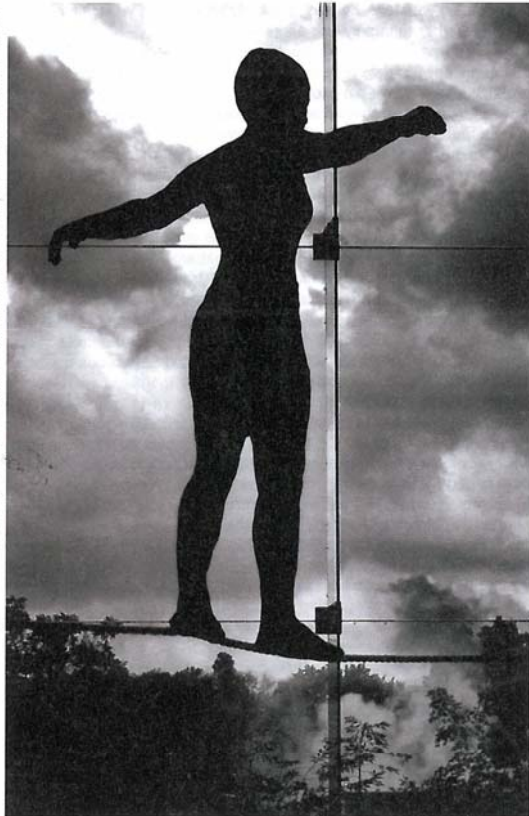
It is also why the application often migrates across a business once it has been adopted by one function. "We tend to see our solution grow within client organisations. Most have four or five applications they use based on one Trillium engine," he says.

Having data as a visible asset in this way is a very new concept to many businesses. So is putting in place personal and department-level data quality KPIs. Yet these are what connect the owners of data to the impacts on the business of what they are doing.

That is why Database Group offers its Cognition process to kick off data quality projects. A team of senior people from customer services, account management, marketing and other functions work through a structured process to identify the critical key performance indicators in the business. These may be simple indicators, like revenue and margin, or more complex ones, like customer profile and lifetime value by identifying the KPIs that really matter to the business and how these are affected by data issues. This leads to a set of ranked actions that can be taken which are known to be achievable, without the risk of running into problems because of missing or flawed data.

"The workshop process searches for the drivers of that revenue, for example getting customers to spend more, then it looks at what might create that change in behaviour, such as increased frequency of product use, pricing, or channel – things the business can influence," says Lees.

Some of these drivers may be easier to influence than others.



Balancing act: Allocating the correct amount of resource to data quality may mean that other areas of the business lose out

For example, there may be very little price elasticity in a highly-competitive marketplace, but the same company could have a very low level of cross-selling that might be increased. Often drivers emerge that the business hadn't recognised as central to its operating model.

From this, a ranked list of priorities is developed for the business to take action on. At this stage, goals are identified which are achievable based on the resources and data available to avoid the risk of running into problems because vital information is flawed or missing.

"Data quality almost always falls out as a key issue. That is the reason why we tie KPIs to the availability of data, in order to identify whether the data allows you to do something or not," says Lees. For decision makers, knowing that an asset is available for exploitation is the corner stone of confidence. ●

