



VNUNet – The Importance of Data Quality

By Clive Longbottom, Service Director, Quocirca Ltd

Ever since I first started following the customer relationship management market in the dim and distant past, the lack of attention paid by companies to the quality of the data they hold has always galled me.

From a customer point of view, I'm sure that all of us have received letters sent to a misspelt name, or where the gender title has been assumed, and has been assumed wrongly.

Just as bad (or worse) is multiple copies of the same missive just because the company in question has multiple records of us - C Longbottom, Mr C Longbottom, Clive S Longbottom, and so on.

Not only does this affect customer satisfaction, it has a direct cost on the company concerned.

For those using paper-based communications, the cost of creating, enveloping and sending such communications soon adds up, and for those in areas such as catalogue marketing the cost is considerable for each catalogue sent.

Beyond such simplicity, there are other data issues that should also be weeded out before any harm is caused.

I was reading about a woman who found that her tax code had been changed to zero. On calling the Inland Revenue, she found that this was because as far as the IR was concerned, she had been declared bankrupt.

An IR employee had been going through the bankruptcy court records, had seen this person's name and her town, and that was enough information.

That there were many people with the same name in the same town was neither here nor there, and it seems that a stab in the dark was enough.

However, if intelligent data cleansing had been used, the IR employee would have been able to more closely match the available information from the court record with the record in the IR database.

In fact it could have been automated, saving time and money while being more accurate and less prone to any issues.

Historically, data cleansing has been a batch job, often involving the use of an external bureau.

Turnaround could be a matter of days, something which just isn't effective in the market where the master dataset may have changed substantially between the original extract sent to the bureau and the subsequent load of the cleansed data back to the application.

Even if such an approach were feasible, all that happens is that you get a saw-tooth effect: you have clean data for a while but, as soon as you start working against it, it is going to get dirty again unless you have measures in place to prevent it happening.

What can be done? Companies such as Experian and UK Changes provide services for cleaning data, de-duping it and ensuring that postcodes and addresses match, as well as cross-matching against multiple different external data services (such as Acorn, TPS, FPS, OCIS) to add additional value to the data prior to it being used.

Some of these services can also be provided as quasi real-time services, but you might not want to be dependent on 'quasi real-time' for your core business.

Many other software vendors provide systems that perform direct file-level comparisons, leaving the user to decide which is the master record and which is the wrong record.

Others provide a system of polling where there are more than two records available; if two records are the same, the third is deemed to be wrong.

Another company, Datanomic, provides an on-site approach that enables full data cleansing across multiple data sources. It can also be used to report against data and for data migrations, and provides a basis for compliance and governance.

Datanomic's system can run in real time against a set of data sources. A graphical front-end hides the complexities of a powerful rules engine that is the heart of the system.

Pre-packaged rules enable the identification of data, the extraction, transformation and load of the data, and determine how data comparisons are carried out.

Not only does the system work on a 'black/white' basis, where obvious records can be weeded out, but it can identify any shade of grey (records that may be wrong), and the user can decide where the trigger point should be for any exception to be raised for human intervention.

'Fuzzy' logic, involving 'soundexs' (e.g. 'Reading' and 'Redding'), thesauri (e.g. Ms/Miss/Fraulein/Mlle) and other techniques enable a far more in-depth level of cleansing.

The system is very powerful, but Datanomic faces a few problems. Firstly, it is a UK-based company with little visibility overseas or even on its home turf.

But this issue could be resolved through branding and advertising, and the US market (where data quality is a huge issue) could be broken into by taking an Autonomy/Staffware type approach of setting up in the US as if Datanomic was always a US company, with a US chief executive and so on.

Secondly, the main problem is that many organisations still do not understand issues surrounding data quality and what poor data quality actually costs them, not only in hard money but in the harm to brand and profile that can be caused by poor service.

Dealing with this issue is harder. Datanomic has some solid customers, many of which would not like to be named, as the product creates real added value for them. Therefore, case studies are not quite as thick on the ground as Datanomic would like.

As many organisations seem to prefer to wait until it is too late before taking any action, i.e. after the problem has shown itself through customer relationships being harmed, Datanomic is often a retro-fit solution taken as a White Knight saviour.

Getting on to more shopping lists has to be the focus for Datanomic, something that it seems to be very aware of.

Quocirca's view is that data cleansing should be built in as a callable service for enterprise IT systems, ensuring that not only is data created clean but that it remains clean as time progresses.

For those who are dependent on the use of large volumes of customer data, Datanomic is well worth a look.

About Quocirca

Quocirca is a primary research and analysis company specialising in the business impact of information technology and communications (ITC). With world-wide, native language reach, Quocirca provides in-depth insights into the views of buyers and influencers in large, mid-sized and small organisations. Its analyst team is made up of real-world practitioners with first hand experience of ITC delivery who continuously research and track the industry and its real usage in the markets.

Through researching perceptions, Quocirca uncovers the real hurdles to technology adoption – the personal and political aspects of an organisation's environment and the pressures of the need for demonstrable business value in any implementation. This capability to uncover and report back on the end-user perceptions in the market enables Quocirca to advise on the realities of technology adoption, not the promises.

Quocirca research is always pragmatic, business orientated and conducted in the context of the bigger picture. ITC has the ability to transform businesses and the processes that drive them, but often fails to do so. Quocirca's mission is to help organisations improve their success rate in process enablement through better levels of understanding and the adoption of the correct technologies at the correct time.

Quocirca has a pro-active primary research programme, regularly surveying users, purchasers and resellers of ITC products and services on emerging, evolving and maturing technologies. Over time, Quocirca has built a picture of long term investment trends, providing invaluable information for the whole of the ITC community.

Quocirca works with global and local providers of ITC products and services to help them deliver on the promise that ITC holds for business. Quocirca's clients include Oracle, Microsoft, IBM, Dell, T-Mobile, Vodafone, EMC, Symantec and Cisco, along with other large and medium sized vendors, service providers and more specialist firms.

Details of Quocirca's work and the services it offers can be found at

<http://www.quocirca.com>