



A data quality solution? I can't even see a problem.

“If you think education is expensive, try ignorance” (Derek Bok)

One of the issues faced by any data quality practitioner whose field is at all specialized, is how to persuade data owners that they have a problem with their data.

Many data owners realize that they have a problem with data quality, but the form that that problem takes is often less distinct in their minds. Some forms of data quality issues are more easily recognised than others. A sales system which suggests sales of 1 million while the accounts system indicates sales of half a million is likely to show a clear data quality problem. Engineers, building a bridge from both banks that fails to meet at the middle, know they have a data quality problem. But when the data quality issue is based on a subject area which requires a deep knowledge of that topic, the data quality issues are much more difficult to recognise – and are therefore not given priority by the data owners.

My own specialization is international personal name and address data management, and this is one of those topics. With over 6000 languages written in one of tens of different scripts, over a hundred different address formats and around forty different personal names formats, there's a lot to know and learn. In a world where an alarming number of people cannot even locate the country that they live in on a map, the general knowledge level on this topic is shallow.

Often, when a name and address is output and printed on the correct part of an envelope, the data owner is happy. The fact that the name and address are incorrectly formatted, or that the wrong information has been printed, cannot be recognised without a deep understanding of that subject. If the envelope can be sent, and is not returned, there is often no indicator that there is a data quality problem.

This is a very common problem. When I speak at conferences, I know that I am usually speaking to those few who recognise a problem with their name and address data. The vast majority is in blissful ignorance of their data quality problems. How can we persuade these people that some data quality issues cannot immediately be seen, and that the effects of these issues are rather more insidious, so that they may not be able to easily measure their effects?

For a few of us, it is still the case that one treats one's customer, and their data, with respect as a matter of course. We are in the minority. Most companies are only looking at the bottom line – what is this costing us, how much can this save us? When a data quality issue is not obvious, it is easy to overlook. Whereas a customer will be very quick to complain when their invoice overcharges them, when they receive mail that does

not correctly use their name and address data, most will sigh, show their irritation in any number of ways, but rarely will they contact your company to complain.

“Ignorance is never out of style. It was in fashion yesterday, it is the rage today and it will set the pace tomorrow” (Frank Dane)

Before most conferences I get mail from the sponsors extolling the virtues of the products that they will be displaying on their stands, and afterwards I get mails thanking me for visiting their stands, even when I didn't. Whenever the conference is held in any country other than the one in which I live, my address details on those mailings will always be incorrectly formatted. Given that most of these companies claim to be practitioners of data quality in one field or another, this is not a positive sign.

Even when there is recognition that the data quality is not good, how do companies handle this information? As an experiment, I have tried contacting each company that had sent me mail with incorrectly formatted data to point out the problem and to point them to a cheap and effective solution. The reactions have ranged from polite indifference to total silence. As a result, I have had to practice my right as a consumer to vote with my wallet. This is a hidden cost of poor data quality that most companies are unaware of because it cannot be easily measured.

Given the immense importance that data quality has in the success or failure of any company, it seems surprising that practical data quality is all too often entrusted to lower-level employees for whom recognizing, and acting upon, new data quality issues will only increase their workload. A friend of mine works in a logistics company, and he spends more time than he should trying to persuade colleagues of such simple truths as Monaco is not part of France and that the Channel Islands are not part of the United Kingdom. His efforts, supported with documentation and impassioned argumentation, are usually greeted with a grunt or, if he is lucky, an “oh, that's interesting”, but the extra effort needed to correct this is too much for the administrative staff concerned, who do not benefit directly from it and who can rarely be persuaded that attention to such, apparently unimportant, details is beneficial for their companies and are part of their job description. This company continues to make itself look rather silly, their customers continue to be irritated, and the financial effects on the company remain unaltered and unchallenged.

“In the Valley of the Blind, the one-eyed man is king” (H.G. Wells)

Though keeping one's customers happy should be reason enough to give attention to data quality, lack of knowledge in subject areas leaves companies at the mercy of some of the less scrupulous companies operating in the data management field. There are those that offer address management for more countries than exist, those that offer postal code validation in countries that have no postal codes, and so on. Failure to acquire knowledge can lead to poor business decisions.

Given these facts it is inevitable that many of us involved in data management will be more involved in pointing out that problems are there, than being asked to provide the solutions. We shall continue to plug away, evangelizing wherever we can and however we can.

About the author:

Graham Rhind is an acknowledged expert in the field of data management. He runs his own data consultancy company, GRC Database Information, in The Netherlands, (<http://www.grcdi.nl>) where he researches postal code and addressing systems, collates international data, runs a busy postal link website and writes data management software. Graham also regularly speaks on the subject and is the author of *Building and Maintaining a European Direct Marketing Database*, *The Global Sourcebook of Address Data Management* and *Practical International Data Management - a guide to working with global names and addresses*.

End

Graham Rhind
GRC Database Information
Nieuwe Prinsengracht 80-hs
NL-1018 VV AMSTERDAM
020-5286120 (telephone)
020-422 93 11 ('fax)
graham@grcdi.nl
<http://www.grcdi.nl>