

**Capscan**



# **Data Quality: Perception Versus Reality**

**Graham Rhind**

Sponsored by Capscan Ltd



## TABLE OF CONTENTS

### Page

Executive Summary	2
Introduction	3
Key Findings	4
- Responsibility	4
- Perception and action	6
- Return on investment	7
- Eternal optimism	7
- Non customer-centric	11
- Blame the customer	11
Conclusion	13



“Although, many organisations recognise the importance of data quality management to a business, its penetration amongst all organisations as a strategic agenda is still low.”

## EXECUTIVE SUMMARY

A survey, commissioned from MCC International Ltd by Capscan and conducted in August 2008, provides a mixed picture of data quality management in the organisations surveyed.

Most organisations surveyed saw data quality management as primarily the responsibility of those involved in marketing and information technology, and of sufficient importance to merit the attention of senior management, reflecting the increasing importance that data quality is given in many organisations. They do, however, continue to see data quality as the responsibility of an individual or department, rather than of all staff within a business, to be tackled operationally or technologically, rather than through improved business practices and processes.

Although, many organisations recognise the importance of data quality management to a business, its penetration amongst all organisations as a strategic agenda is still low. Respondents over-estimate the quality of their data and are less aware of the damage that poor quality data can do to their business. There is a distinct gap between words and intentions, and actions. Few have the necessary processes, tools or strategic will to take action to improve data quality. Data is still not measured or financially valued at a strategic level and contact information collected by organisations in many cases is still not checked or validated, despite its value and importance.

However, some of the more forward thinking organisations surveyed have invested in various appropriate data management technologies. The most popular technologies used amongst surveyed organisations includes, for example, address, identity and data quality management solutions and banking validation software. In contrast, the penetration, use and awareness of data screening services amongst those surveyed were still fairly low. Further, the gap between the organisations stating that they had a data quality management strategy and those able to realise it is still fairly large.

### *Reality still lags considerably behind the rhetoric.*

Today, many organisations still face huge challenges in terms of managing data. Practical operational problems like keeping data up-to-date and accurate, making sure it is secure and well managed and compliant remains the primary concerns for these businesses. For these organisations, successful handling of such problematic data issues would bring the desired benefits of improved operational efficiency and revenue opportunities, reduced costs, and the resulting greater customer satisfaction.



## INTRODUCTION

Data quality is not a purely operational or technological issue. A significant amount of improvement in data quality can be achieved through a change in the way companies and their staff perceive their data and its quality. To what extent does senior management now buy in to the need to improve data quality, and are they providing the resources to allow this to be achieved? How do companies perceive the quality of their data, and who is responsible for that quality? Are policies in place to manage data quality, or is there a piecemeal approach? Do people’s perceptions of their data match actual reality, and are words and intentions being matched with actions?

To help answer these and other questions, Capscan commissioned MCC International Ltd to perform an independent data quality management research project in August 2008 amongst the readers of CBR Online, analysed and written up by Graham Rhind. 189 completed questionnaires were received. The readers are skewed towards higher management, particularly in information technology, within large organisations and mostly within the United Kingdom.

**Fig 1: Readership Breakdown**

	CATEGORY	% CIRCULATION
JOB FUNCTION	Head of IT/IT Director/CIO	49%
	MIS/IT Manager	19%
	Head of Finance/Finance Director/CFO	8%
	Managing Director/CEO	10%
	Business Development Manager	5%
	Systems Development Manager	4%
	Investment Banker/Financial Analyst	1%
	Consultant	3%
	Others	1%
NO. OF EMPLOYEES IN ORGANISATION	<250	5%
	250-500	23%
	500-1,000	22%
	1,000-5,000	30%
	5,000-10,000	17%
	>10,000	3%

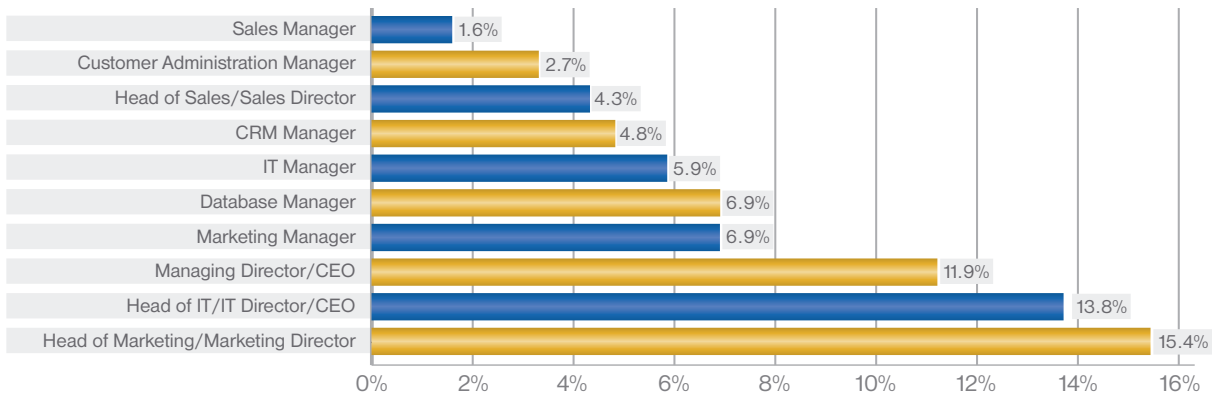
Questions were asked about the respondent’s contact database: 78% of respondents managed a contact database internally.

## KEY FINDINGS

### Responsibility

Asking respondents about those within their organisation responsible for data quality produced some interesting results and highlighted immediately one of the main issues impeding increased data quality within any organisation.

**Fig 2: Who is responsible for data quality within an organisation**



11.1% pointed to upper management as responsible; 30% to members of the IT departments; 22.2% to leaders within the marketing department; 5.8% to leaders within the sales department. Yet only a single person gave the response which shows true understanding of data quality: "All staff". Almost every person within an organisation, and certainly all those who contribute to, and consume from, data resources, must understand the importance of maintaining quality to the future of their organisation, and to be responsible for their part in that process. Allowing a false sense of security to allowing staff to point at somebody else as being responsible – an "it's their problem" attitude – will always lead to a lower level of data quality than when all staff are part of the data quality process.

Though only 40.7% of companies surveyed had an enterprise-wide data quality management policy in place, 87.3% named a staff position as responsible for data quality. Though it may be the case that a person has that responsibility named within their job description, without a data quality management policy in place,

little effective data quality process improvement is likely to take place.

The preponderance of information technology (IT) staff regarded by respondents as being responsible for data quality is also indicative of an issue within data quality. Traditionally, IT staff have been responsible for the data container: the programming, the architecture, the hardware. They concern themselves with speeds, capacity, binary processes. They rarely appreciate that the importance lies not with the container but with its contents: the data. It is, naturally, not universally correct to say that IT staff concentrate only on the data container and not on the data, but it generally holds true and is an issue which has proved to be an obstacle in many companies for improving data quality.

A significant positive result, though, is the increased representation of upper management within those named as responsible for data quality, reflecting the increased weight that some organisations are giving to data quality issues. This is reflected by 88% of respondents stating that their organisations view data quality as important or fairly important.

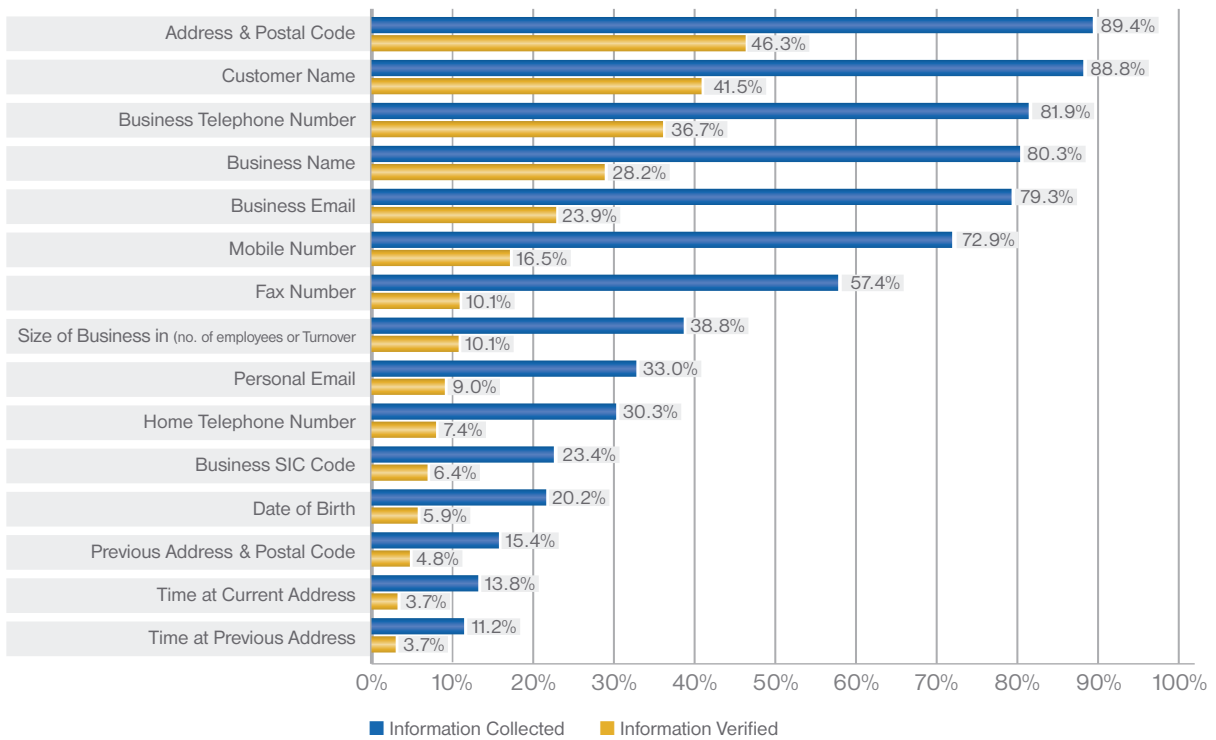
“Only 40.7% of companies surveyed had an enterprise-wide data quality management policy in place.”



## Perception and action

There is a marked contrast between words and actions shown in the survey results. Whilst 88% of the organisations gave some degree of importance to data quality issues, a very much lower 41% states that they have an enterprise-wide strategy. Furthermore, the use of software and processes to actually improve and protect data quality is relatively low.

**Fig 3: The information collected vs validation by the organisation**



Software with the highest implementation degree is de-duplication software at 32%. This would suggest that, without stringent checks and controls everywhere in the company, the other 68% of companies will have duplicate-ridden databases, leading potentially to massive costs and lower business intelligence value. Address validation software was present in only 29% of the companies for UK data and in only 17% for international data.

The presence of software and validation levels showed discrepancies. For example, whilst only 38.6% of respondents had



address validation software installed (UK and/or international) 48% claimed to be validating that data using software. In large companies especially, there is a tendency to over-estimate the amount of validation taking place. There are many gateways through which data enters most organisations, and rarely do all those gateways implement controls over the traffic. Companies may, for example, implement address validation within a call centre, but fail to attempt any validation from data entering from their web forms.

Whilst almost 90% of companies surveyed collected company name and address information, only 48% validate this information using software, illustrating a major source of poor data quality and a large potential market remaining for data improving software and procedures. There is also a clear and significant gulf between the appreciation of the importance of data quality and steps having been taken to put into practice processes to ensure that high data quality is achieved. This may in some part be due to the continuing necessity for those working with the data to persuade upper management of the need to take steps to assure data quality.

## Return on investment

The continuing necessity to persuade management of the necessity of data quality, and to prove return on investment on something intrinsically obviously advantageous to any company, is shown in the survey's results. 27.5% of respondents stated that "establishing a link between data and ROI", 31.2% "Getting senior management buy-in for a data strategy" and a massive 39.2% "Getting organisations to understand the impact of poor data" were one of their biggest data challenges today.

Whilst it is dispiriting that staff who understand and care about data quality must spend so much of their time trying to bring this message to executives who should already understand the importance of data quality, there are some small signs of improvement. Gartner reported in November 2008 an increase in spending on master data management (MDM) software of 24% in the past year, as companies begin to understand that one can never avoid the need to maintain data, even during periods of economic uncertainty.

## Eternal optimism

A major barrier to better data quality is an eternal optimism that many companies and their staff have about their data and its quality. Companies will be unable to overcome the inertia in improving data

“Whilst almost 90% of companies surveyed collected company name and address information, only 48% validate this information using software.”

“A major barrier to better data quality is an eternal optimism that many companies and their staff have about their data and its quality.”

quality if they cannot recognise that their data has quality issues, nor the advantages that they will accrue from improving that quality.

This has been well illustrated in this study. 52% of respondents viewed their organisation’s data as being of excellent/good quality, and a further 40% considered their data to be “alright – but could be better”, leaving only 8% to admit that their data was poor. Given the relatively low number of respondents utilising validation and cleansing software such as de-duplication and address validation, it is clear that the quality of data is being grossly overestimated. This may partially be due to the data being hidden behind complex programs and interfaces, reducing the ability of the users to actually view the data outside reports where data is coalesced into a set of figures; and also due to most data professionals not being able to recognise problems in their data even when they are able to browse through it.

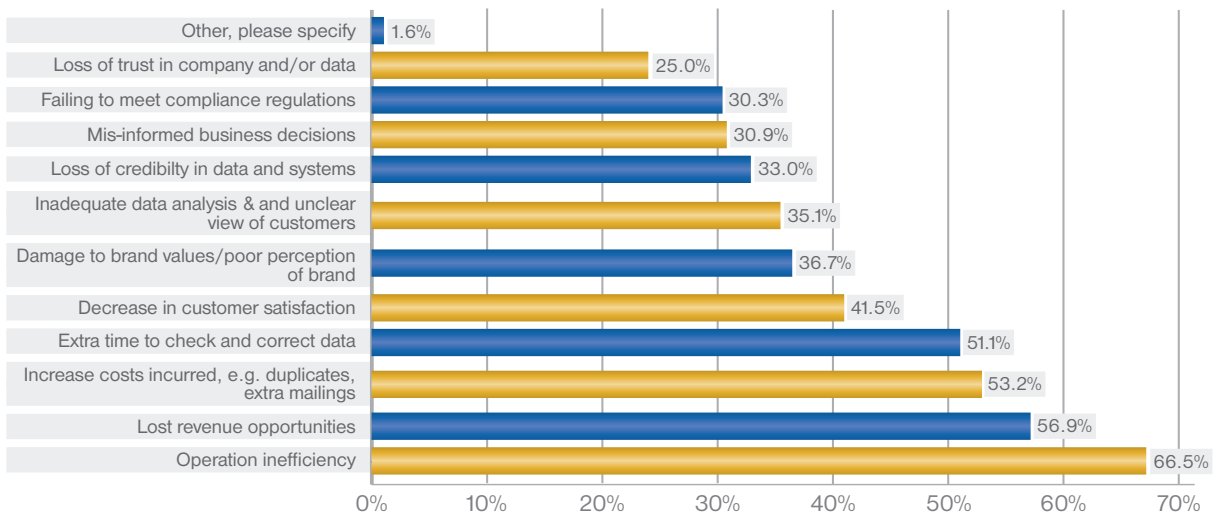
Data needs to be collected in as clean a way as possible, with validation at source, and it will start to decay even before it enters the database. Many of the 92% of customers assuming that their data is “alright” or better are usually oblivious to the reality of data quality issues. This optimism is further illustrated elsewhere in these results. It is clear that if good quality data increases, for example, the ability to provide a good service to your customers, then poor quality data will decrease that same ability. In this survey the respondents were asked how negative data quality would affect their business, and later how good data quality would affect their business, with a mirrored set of answer possibilities.

Each respondent noted an average of 4 effects of bad data quality, but 5.4 effects of good data quality, with anything up to 15% higher responses for positive effects than for the mirrored negative

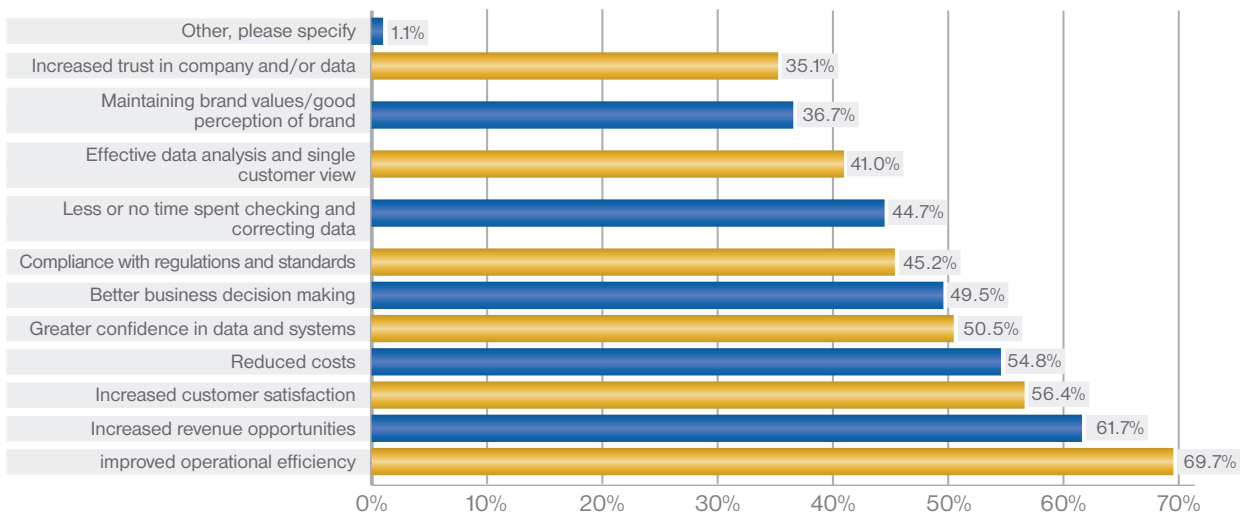


effects. It is clear that whilst respondents see the positive effects of improving their data quality, they do not always identify the associated negative effects, allowing a “do nothing” attitude to data quality until the effects of poor data quality begin to seriously affect the business’ bottom line.

**Fig 4: The negative impact of bad data quality on the organisation**

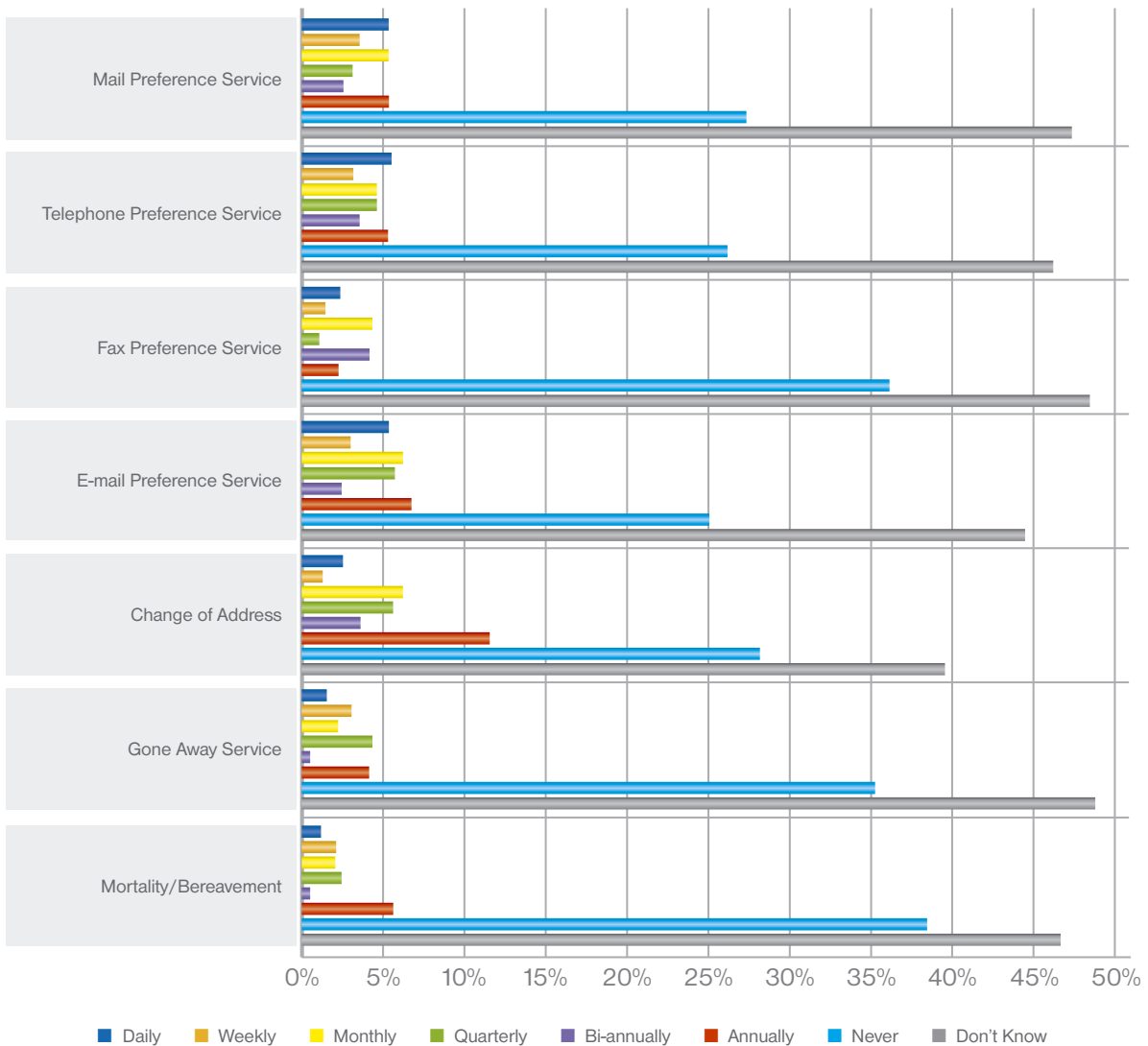


**Fig 5: The business benefits of good data quality management**



The way that data is used within an organisation will define what effects bad quality data has on a company, but in general the respondents underestimated the effects. The greatest impact was perceived as being the effect on operational efficiency (66.1%), but only 34.9% of customers perceived “inadequate data analysis and an unclear view of customers” as being an issue, whereas it is sure that a larger percentage of the respondents use their data in this way, and any poor data quality will effect these analyses.

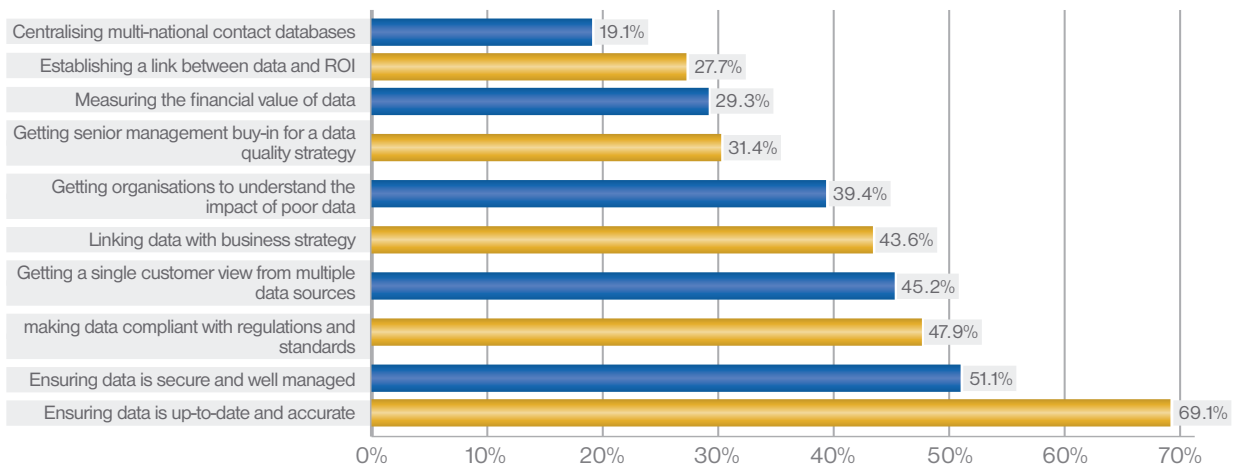
**Fig 6: How often the organisation used data screening or cleaning services**



## Non customer-centric

When asked to name their biggest data challenges today, the respondents took a largely operational line, continuing the trend of considering data quality to be a largely technical problem with largely technical solutions.

**Fig 7: The biggest challenges facing businesses today**



Perceived challenges continue to be broad, with each respondent naming an average of 4 main challenges. It is, however, promising to note that 69.3% of respondents suggested that one of their main challenges was “ensuring data is up-to-date and accurate”. Inherent data quality is when data accurately reproduces the information about the real-world entity which it represents. When the data achieves that, and only at that stage, can other challenges be successfully overcome.



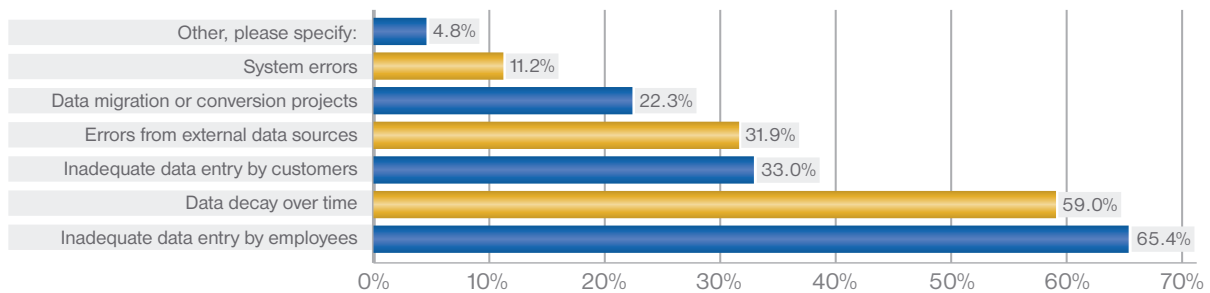
## Blame the customer

It is perhaps understandable that staff in most companies regard customers as a nuisance. If they didn't move or get married, there would be no need to update their records; if they didn't buy things the shelves at the supermarket would always be fully stocked; if they didn't call there would be no waiting times at the call centre. It remains easy to forget that most companies only exist because of their customers and that they work largely for the sake of those same customers: without them, they would fail.

It is therefore fascinating that 33.3% of respondents blamed “inadequate data entry by customers” as one of the main sources of their data quality problems.

Far more (65.4%) blamed inadequate data entry by employees. In both cases, the organisation has the power to improve results through technological implementation (of, for example, validation software) and procedural business process improvement and training. A great deal of control can be placed on data entry systems through the use of dynamic forms, validation, dialogue with the customer and a better analysis of data being gathered. These are problems which companies could begin to address today and which could have major positive effects on data quality.

**Fig 8: The main source of data problems**



Only three respondents recognised non-technological aspects for their data quality issues, noting that there was no single person responsible for their data issues; that no policies or standards were in place; and that piecemeal solutions were attempted rather than an enterprise-wide approach.

“33.3% of respondents blamed ‘inadequate data entry by customers’ as one of the main sources of their data quality problems.”



## CONCLUSION

Whilst most people are increasingly aware of the importance of increasing data quality, there remains a gulf between perception and action. Most organisations do not have enterprise-wide data quality improvement programs, and responsibility for data quality is often perceived to lie with a single individual rather than every member of staff within a company. Despite a higher appreciation of the importance of good data quality, few companies have the policies, tool and processes in place to manage data quality at all, or in the most effective manner; and most companies considerably overestimate the quality of their data.

### *A huge contrast remains between intentions and actions.*

Data quality continues to be seen as an isolated topic safely tucked away in the IT department. Higher management is increasingly becoming involved in data management issues, but often still require proof that increased data quality will positively effect their bottom line. Whereas many see the advantages to improved data quality, fewer have acted on this and even fewer have installed the necessary (technological and business) infrastructure to achieve this. A great deal of work is still required to turn around the perception of data quality issues and how they are to be tackled within companies of all types and sizes. The attitude that data quality is an operational and technical issue, divorced from business processes and procedures, remains stubbornly in place.

### *Reality still lags considerably behind the rhetoric.*



## About the author

Graham Rhind is an acknowledged expert in the field of data quality. He runs his own consultancy company, GRC Database Information, based in The Netherlands, where he researches postal code and addressing systems, collates international data, runs a busy postal link website and writes data management software. Graham speaks regularly on the subject and is the author three books on the topic of international data management.

Email: [graham@grcdi.nl](mailto:graham@grcdi.nl)

Web: [www.grcdi.nl](http://www.grcdi.nl)



## About Capscan Ltd

Capscan is a leading supplier of international address management solutions and data integrity services. The company's flagship product, Matchcode, is a fully functional address management system, available as a stand-alone programme for data capture, a web-based tool for online data capture and as a tool for batch cleansing of commercial databases. Capscan now supports addressing functionality for 240 countries or territories worldwide. In addition, Matchcode UK address data can be integrated with Ordnance Survey data sets to allow mapping and logistics rationalisation. Capscan is also expert in the integration of lifestyle and business data sets into address databases to support and rationalise canvassing or marketing activities throughout Europe.



## Capscan



Capscan Ltd  
Grand Union House  
20 Kentish Town Road  
London NW1 9BB

Tel: +44 (0)20 7428 1255

Fax: +44 (0)20 7267 2745

Email: [enquiries@capscan.com](mailto:enquiries@capscan.com)

Web: [www.capscan.com](http://www.capscan.com)