Prioritising and deploying data quality improvement activity

Iain Henderson

is the Founder and a director of Information Answers Ltd. He has 15 years of client-side experience in direct and database marketing in the retail, computing and distribution sectors followed by several years as a customer management consultant. His primary research activity at present is in the area of privacy and the role of informediaries in resolving privacy conflicts.

Don Murray

is a director of Information Answers Ltd. He has nearly 20 years experience in the design of very large customer databases and has worked in, or advised, over fifty companies in the UK, Ireland, France, Germany and Sweden, as well as in the US and Japan. Don has research interests in the areas of data warehousing and business intelligence.

Abstract   Data quality is an issue that is currently attracting much attention in customer management circles. A number of significant customer relationship management (CRM) project failures have been attributed to weaknesses in the data quality. Industry commentators are picking up on this trend: Gartner recently expressed the view that some 25 per cent of data within major corporations is inaccurate and that this situation is likely to remain the case for some time.[1]

That data quality is a problem is not in question; this paper seeks to take a step back and review the context in which data quality exists, and to use that context to provide guidance on how improvement activity might be prioritised and then delivered.

BACKGROUND

Over recent years, Information Answers has developed a methodology for assessing organisational capabilities around the sourcing, management and use of customer and related information. This methodology, built from the combined experiences of a range of data specialists across many sectors, starts with what the authors term 'the vicious or virtuous circle' of customer data. Figure 1 illustrates the circular relationship between the data available to an organisation, the quality of that data, and the use that is made of that data.

Information Answers' experience is that changes made in each of these areas have a 'cascade' effect into the others — and those changes can be either positive or negative in impact. For example, an organisation that operates in a business-to-business environment would benefit from understanding organisational structures and linkages between individual sites that it may have identified as customers or prospects. If data on corporate structures either does not exist at all within the selling organisation, or is of poor quality, then various usages of that data — such as key account management, or credit management — will either not happen at all or will be incomplete or ineffectual. In turn, this non-existent or sub-optimal usage of data will limit the frequency with which records are refreshed or updated, further reducing the content or the quality and further negatively

Iain Henderson
Information Answers Ltd,
33 Macready House,
London, W1H 5LR, UK.

Tel: +44(0) 7803 929 603;
e-mail:
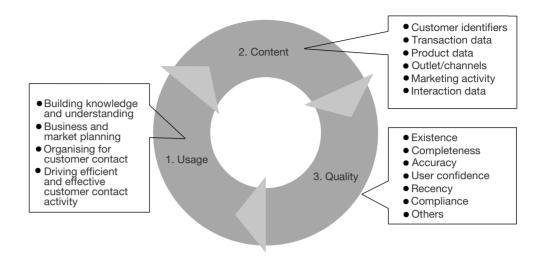iain.henderson@information-answers.com

Figure 1: The circular relationship between content, quality and use of company data

impacting on the next usage of the data. Conversely, if an organisation initiates a significant improvement project in any of the three areas then that will have a knock-on effect and, in all likelihood, will trigger a further series of changes in a positive direction.

Of course, much of the above is only common sense applied to a customer data scenario. The complexity lies in managing those same relationships over the multiple types of data, the different dimensions of data quality and the many and varied ways in which data can be used.

In the view of the authors, the route to managing this complexity — and thus the route to prioritising data quality improvements — is to focus on the uses to which data will be put within the organisation. After all, using data is the area that can make money for the organisation: the others are cost centres. Indeed, where an organisation has good metrics on its customer base it is possible to quantify the impact of improved data content, quality or usage — by extrapolating the impact of what could be done that is not currently possible.

HOW DATA IS USED IN ORGANISATIONS

Data usage across an organisation comes in many flavours, differing in emphasis by sector and size of business. Figure 2 illustrates the high level breakdown Information Answers uses as a generic starting point in assessing an organisation's capabilities or defining its requirements.

Each data usage area breaks down into specific practices (eg customer lifetime value analysis), and at this level of detail it is possible to articulate the data content required to enable this particular usage of data. In this example, both a minimum and a desirable set of data elements required to drive customer lifetime value analysis can be defined.

Once the data content required for each usage of data is understood, the next step is to prioritise these data usages according to their importance to the organisation. Typically one would expect anything from ten to 30 specific data usages to be selected as high priority for an organisation. These priorities would vary by sector and by business situation: A start-up company will be focused on
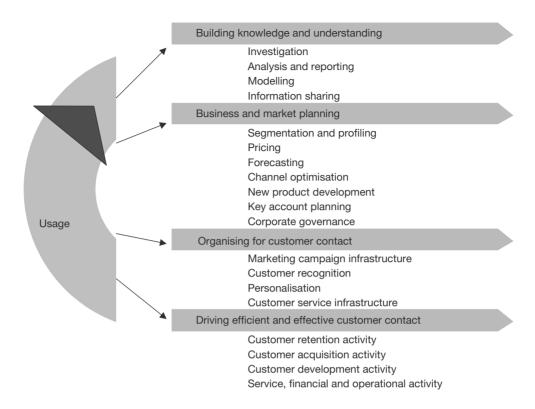
Figure 2: High level breakdown of organisations' capabilities/requirements

customer acquisition–related usages whilst an established business may be more focused on retention or on reducing operational costs.

Prioritised data usages, ie those that are expected to generate the highest return (current or future) for the organisation can be arrived at by interviewing stakeholders across an organisation. These usages will each have required and desirable data content — to which must then be added an interpretation of data quality. This can be done qualitatively or quantitatively; often both routes are deployed in tandem. A qualitative assessment will typically be carried out with power users to allocate a score for each data element based on its completeness, appropriateness, accuracy, currency, compliance and its suitability for grouping or meta-linking. A

quantitative assessment will most likely involve SQL querying of the data set or matching against external reference files.

This assessment of data quality is where the underlying problems become apparent. In isolation, Information Answers would typically expect to see reasonably good data quality on customer names and addresses — especially where these are required for billing. Similarly, with data quality on transactional data — although this is most often an incomplete picture due to the fact that, in many organisations, this data is held in separate transactional systems or 'silos'. External data brought into the organisation will be as good, bad or indifferent as the source, with currency often being the weakest link. Data on customer interactions with the organisation often vary according to the sophistication of

existing CRM deployment, but comparatively few organisations in the authors' experience can claim complete and accurate coverage of customer interactions. This problem comes into sharp focus when one uses the lens of data usage to assess how data quality is limiting the organisation. The majority of the ways in which data can be used require data from across the spectrum of data content — and not just those parts of the customer record that are easiest to maintain.

Findings from assessments

The need for tools and processes that specifically help organisations to address data-related issues becomes apparent on review of Information Answers partner QCi's benchmark database of some 650 organisational assessments worldwide. Findings to date include that:

— 75 per cent of organisations have no explicit incentives or sanctions in place to drive data quality improvement
— 56 per cent of organisations do not have an historic measure of data quality against which to measure improvement
— 39 per cent of organisations do not have any data quality standards in place
— 41 per cent of organisations fully understand the implications of privacy/data protection legislation in all of the countries in which they do business

So, if organisations find that data quality is a significant limiting factor on their ability to manage customers, what might they do to address this problem?

PRACTICAL STEPS TO ENSURING DATA QUALITY

In order to tackle this difficult issue, organisations need both a top-down approach to data quality — sponsored by the most senior levels of management — and a comprehensive bottom-up analysis of data sourcing, usage and content — including an assessment of the enterprise's capabilities in terms of data management, relevant tools and people skills.

The following practical steps will make a difference — both strategically and tactically — to data quality within the organisation:

1 Establish a formal body — accountable at board level — responsible for the setting of quality goals, for the management of improvement activity and for the provision of training. Data quality issues are best addressed through formal bodies with sufficient time and resources to undertake the tasks involved. As data quality issues are often organisation-wide, this body should operate with the full backing of senior levels of management.
2 Have this body clearly define what is meant by 'data quality'. Defining the term and confirming which data content, which systems and which processes are to be within scope for improvement, are the first steps towards building a platform for measurement of quality, for root cause analysis of faults and for repair of both processes and data. The components of organisational data quality listed below provide a good starting point (NB precise definitions will differ across organisations or sectors.)
— Completeness: versus the total of all possible relevant sources of this data;

— Appropriateness: how closely the data meets business requirements;

— Accuracy: the overall accuracy of the data record;

— Grouping accuracy: the accuracy when consolidated and matched from disparate sources;

— Accessibility: the speed of data provision to point of use;

— Confidence: a measure of the users' levels of confidence in the data;

— Currency: time since data capture, verification or update;

— Regulatory and compliance: the degree to which the data meets regulatory and compliance requirements; and

— Meta-linking: the degree to which the data can be combined to derive higher-level variables.

3 Ensure that data is seen as, and treated as, a distinct and separate asset within the organisation — and not just part of IT. Data on customers, transactions, products, outlets, interactions and the like are the life blood of an organisation. Data should be regarded as a distinct asset in their own right, and not as a subset of a technology platform or process that stores, creates or amends data.

4 Strive to keep ownership of data and their quality management within the business community. Ownership and strategic direction of the information resources and assets should reside firmly with a senior business user — typically in one of the big usage functions such as Marketing, CRM or Finance.

5 Recognise that poor data quality has an adverse impact on customer satisfaction. Improving quality of data that impact directly on the customer sales and service interaction must be a major priority for an enterprise. It is equally important that active measures are in place to monitor and refine improvement.

6 Allocate and protect a separate budget to drive data quality projects through the enterprise, and ensure that this is backed by a robust and ongoing business case. Data quality improvements will either be made by addressing process issues or by conducting data cleansing and repair exercises. Both of these cost money and take up the time of internal managers. As data quality is most frequently a cross-functional issue, a separate budget of sufficient size to address the issues outside of normal departmental or project budgets is absolutely essential.

7 Establish a separate Data Quality Team, with responsibility across the enterprise, which is empowered to make significant change. Often the data that are required to drive one function have been captured by another function which may not appreciate the importance of it. The cross-functional nature of the problem dictates that a cross-functional team is best positioned to address the issue.

8 For all customer management processes and systems, establish defined data quality measures using an appropriate range of quality dimensions and implement audits to identify where action must be taken. Data quality is a complex topic; there are numerous types of data and databases will typically have many data elements. For each of these, the components of data quality (eg currency, accuracy, accessibility) must be identified and measured. Data content should be mapped against data usage in order that the most important elements can be prioritised and tackled first.

9 Make use of external research to

assess the impact of poor quality data — especially for customer data — and publish this across the enterprise. Data quality problems will result in complaints, increased customer service requests and reduction in customer satisfaction. Research should be undertaken to identify and quantify the impact of poor data quality; this will act as a feed into the business case for improvement activity.

10 Construct a formal data quality framework that sets a context and scope around data quality decision-making. A data quality framework must contain a set of policies that determines how an organisation approaches the issue. These policies should be established within an overall data quality vision, which should define what is to be achieved, by whom and within what timeframe. This vision should be endorsed by, and owned by, senior management. Aspects of a framework will include: a description of the impact of poor quality data on business effectiveness and efficiency (especially customer retention, acquisition and marketing costs); plans for data quality improvement; data definitions; data stewardship; standards for data and meta-data; *ad hoc* and continuous measurement of data; management of data suppliers; data presentation and user accessibility; and the roles and responsibilities of all staff involved in sourcing and managing data. The plan must include a prioritisation of those issues that have maximum impact on customer satisfaction and revenue, in both the short and the long term.

11 Undertake data cleansing exercises and ensure that improvement processes are in place to maintain the quality. Data cleansing exercises are often the most visible signs of data quality issues being addressed within an organisation. They may involve the use of a wide range of tools — from manual review through to the application of sophisticated, specialist data cleaning software. Cleansing exercises can be undertaken either tactically (eg in advance of direct marketing activity) without addressing the root causes of the data quality issue, or strategically as part of a wider 'process fix'.

12 Define, agree and implement common data capture and data entry standards. The definition and adoption of data capture or data entry standards across an organisation is an essential building block on the road to improved data quality. Data entry to the standards should be supported at each data capture point by appropriate software. Some flexibility to enter data outside of the defined standards must be retained. Any new data types or standards should be noted in a 'knowledge base' and the defined standards should be updated. Customer-facing and customer management staff should undergo induction and ongoing training that stresses the importance of data quality. Further training for existing staff with emphasis on advanced topics should also be available.

13 Define a feedback process to take action on behalf of customers who complain as a result of a data quality issue and measure the effect of this action on the customer retention rate. Simple mechanisms should be in place to allow customers to advise the organisation of any data errors: these should include response-paid postal options, e-mail and website access. The process must be effective:

the customer should not notice the same data quality issue recurring.

14 Specify data quality criteria when managing contracts with suppliers of information. Where a supplier generates data that enter the organisation's customer management systems, expectations around data quality standards should be set within the supplier contract: the quality of such data should be measured regularly and action taken where problems are found.

SUMMARY

All of the above will make a significant difference to data quality over time, but none more so than a review of those aspects of customer management that the organisation would wish to deploy but is unable to because of the poor content or quality of its data. Although the initial cost of setting up a project to address all of the data quality issues within an organisation will be significant, this outlay will be recovered many times over as the short-term and long-term benefits of data usage are realised.

References

1 Keizer, G. (2004) 'Gartner: Poor data quality dooms many IT projects' available at www.bizintelligencepipeline.com/datasources/20301076.

Horror stories

Gartner[2] has flagged that data quality, and failure to fully address data-related issues, has been a significant contributor to a number of CRM project failures. The details of specific failures are understandably elusive, but specific examples of problems the authors have observed after the event include:

— A £100m deployment of a leading CRM application that was written off as a failure largely due to poor quality data not being integrated/cleaned in advance of release to the sales teams

— A public sector project significantly delayed after 15 months of development due to an earlier failure to address data protection/data sharing protocols

— A mailing to 400,000 customers that had a 'returned as undeliverable' rate of 15 per cent (equivalent to wasting 60,000 times the cost of the mail piece and postage).

2 Nelson, S. and Kirkby, J. (2001) 'Seven key reasons why CRM fails', Gartner, Inc.