ACUATE
The Data Quality Company

Seven Dimensions of Data Quality

How to improve business performance by improving data quality

ACUATE
The Data Quality Company

Improving Data Quality, Improving Business Performance

On the corporate agenda, data quality is rising in importance, but often it's only when something goes wrong that management get involved and a review of data quality is initiated. Customer complaints, poor lead generation, ineffective account management are a few of the many reasons for highlighting data quality concerns. So, if data quality can have such a clear impact on business performance, why is data quality not a higher priority?

Data quality can be measured for sales & marketing data, product data, financial data and supply chain data.

The link between data quality and business performance is not always clear. Business performance is measured in numbers. Revenues, profits, cashflow, inventory turn, shareholder return, earnings per share, etc. have specific values and percentages that can be compared, subsequently decisions on performance and future actions are taken. For data quality to be linked to business performance, data quality must be measured to quantify the impact on the business.

Let's look at an example, we know that sending duplicate invoices can result in customer queries which could delay payments and impact real cashflow. If this occurred frequently, cashflow performance would be affected negatively and possible interest earnings from cash in bank accounts would be reduced. Incorrectly addressed invoices also have the same result, namely, delays in payments. There are also indirect costs, for example, people costs can be attributed in wasted administration - what other tasks could the finance executive be performing instead of resolving invoice queries? Intangible costs are incurred as a result of the organisation's reputation being affected. A waterfall effect of increased costs materialises as a result of poor quality data.

If we can proactively measure the quality of our data, then many scenarios can be avoided. The above example is one of thousands that can arise, with all of them affecting the business performance metrics of revenues, profits, etc. in same way or another. ROI cases can be completed to link the quality of poor data to business metrics (request Acuate's whitepaper on Measuring Data Quality ROI.)

Creating ROI for specific scenarios can be time consuming and with large quantity of data quality problems affecting the business, it is impractical to calculate for all these scenarios. Acuate have created seven metrics that underpin the most common data quality problems found today. These are called the Seven Dimensions of Data Quality.

ACUATE
The Data Quality Company

The Seven Dimensions

Why are there seven dimensions? There are more attributes one can look at, for example, relevancy, timeliness, reputation. These are either subsets of the main seven or are less prevalent in every day use. The seven outlined here are sufficient to provide a well-formed set of metrics to measure data quality – each one is mutually exclusive.

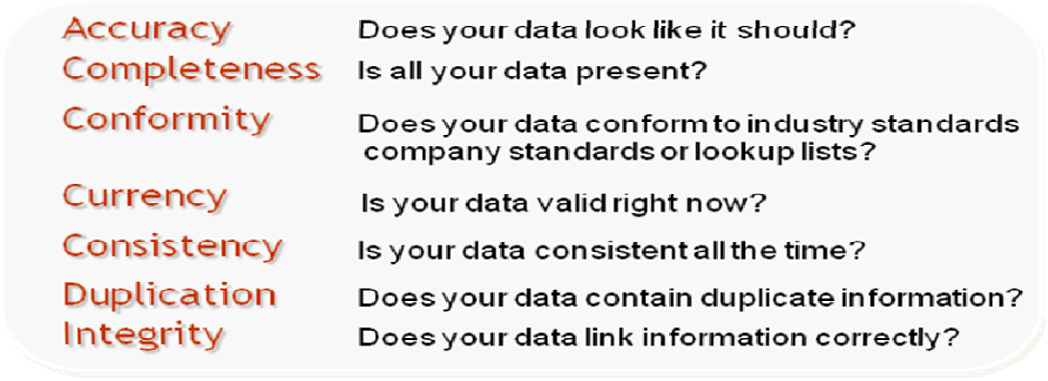

Fig. 1 Seven Dimensions of Data Quality

These definitions are best explained through an example. Below are four records with four fields from a table in a CRM application. The fields are Company Name, Date of Incorporation, Number of Employees and Main Phone Number:

Measuring data quality at migration provides baseline for future measures

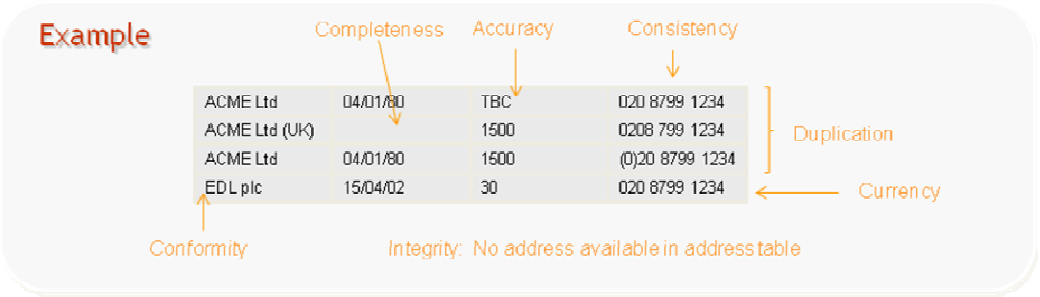

Fig. 2 Examples of Data Dimension Problems

Accuracy

Accuracy is defined as: does your data look like it should? This will depend on the type of data type that is in question. Most common data types are text, numbers and dates. In fig. 2 the first field is a company name (a text type), an incorporation date for the company (a date type), an employee count (a number) and a main phone number (a text type). Let's look at how accuracy can be measured for each field:

ACUATE
The Data Quality Company

Field	Accuracy Dimension Issues
Company Name	Does this text field have the right characters that one would expect for a company name? Does the text string look like a company name?
Incorporation Date	Is this a valid date?
Employee Count	Is this a valid number above 0? Does the field contain only characters that represent a number?
Main Phone Number	Does this have the right number of digits for a UK phone number (assuming a UK record)? Does the field have the right characters for a phone number? Are there extra pieces of information, such as extension number embedded in the text?

Completeness

Is the information your require present? Missing information can lead to many problems, for example:

(i) missing emails reduce email campaign effectiveness reducing the number of potential opportunities and ultimately sales figures
(ii) missing telephone numbers result in ineffective telemarketing
(iii) missing product quantities affect stock count and product performance reports

Completeness is important when data is being captured. In a call centre or a direct sales force, capturing contact information is vital to subsequent marketing campaigns, up-sell & cross-sell opportunities, renewals, account management, etc. At the point of data capture you should capture all the information you require. It is likely the best source of the information, customer or prospect, is on the phone at that moment.

Conformity

Conformity requires data to match a particular set of values that are known to be true. The set of values can be either internal to the organisation or external to it. For example, a Job Title or Product Name can be a set of values that is applicable to the organisation (internal) or an business name can be matched to an industry reference standard (external, for UK this will be Companies House data.) Both internal and external conformity can be measured.

Good quality data can significantly improve business performance

ACUATE
The Data Quality Company

Improve CRM data, improve sales

Non-conforming internal data is a nuisance when segmenting data for marketing purposes. If job titles are incorrectly assigned or do not match a standard set of values then segmentation errors cause a reduction in marketing effectiveness.

Non-conforming external data can result in the information not being current. In a B2C environment, matching an individual to industry reference data (e.g. Experian's Consumer database) can verify whether the individual details are correct.

Currency

Currency is a measure of whether the data or record is valid today. For, example, if a product catalogue contains a product name, but the product is now obsolete, then the record is no longer current and should be archived. Another example, a contact record with an address may be perfect according to the other six dimensions, but that individual moved house one month ago, the record is no longer current. In the B2C market, reference data is available to determine currency of individuals.

Consistency

Consistency measures the uniformity of the data across a table of information. In the above example, the phone number field contains data that is accurate, complete and conforms to industry standards, but is not consistent across the table.

Consistency is important when reporting, cleansing data or analysis. If the information is consistent - reports will be displayed correctly, cleansing is improved because of standardisation and analysis is more accurate because like-for-like data is being compared.

Deduplication

Deduplication is a common problem. In the above example, the first three records are the same. These should ideally be merged together.

Duplicate records waste costs in direct marketing, give an incomplete view of a customer leading to poor account management and lead to poor matching across multiple data sources.

Integrity

This is the most complex measurement category. Integrity identifies if real world objects are linked to correctly to other real world objects. For example, are account records linked to their correct contacts, are order lines connected to the correct order header record, are activities connected to the correct contact, etc.

ACUATE
The Data Quality Company

Improve financial data, improve cash flow

These are important when deploying your CRM application. The migration must ensure the right business objects (accounts, contacts, products, etc) are correctly connected to others. If connections are not present, then you can lose information through orphaned records or you may not be able to rely on reports. Incorrect links will lead to unexpected results which only damage the organisation's reputation.

Continual Improvement

Proactively measuring data quality according to the Seven Dimensions will highlight critical business issues. As a result, corrective action can then take place and the measurements can than be applied again to determine the level of improvement. Continually measuring and improving data will lead to greater confidence by all concerned.

More importantly, there is now a baseline for your data quality. A set of measurements exists that can be built upon. The data can be measured weekly/monthly/quarterly depending on quantity and frequency of changes. Trends can be identified and questions can be raised to aid continual improvement.

Data degrades continually – here are some reasons why:

1. A piece information is no longer true, e.g. a contact at an organiation leaves – email campaigns are likely to less effective immediately
2. A new list is imported into the database – duplicates arise or incomplete records are imported leading to incorrect customer views
3. Humans entering data into a system daily – many data entry issues occur because of poor data entry
4. A new CRM system is deployed (or any other application) – migrating non-cleaned data will further cause problems

Degradation is continuous and given the large impact on of poor data, it is imperative that continual measurements of data quality are performed to sustain quality levels.

ACUATE
The Data Quality Company

Seven Dimensions and Business Performance

Linking data quality metrics to business performance is now simpler with the Seven Dimensions of Data Quality metrics. For example, measuring direct marketing effectiveness we need to add some business results to see how effective campaigns have been. If we know the number of returned mailings, the conversion rates to sales, cost of sending one item (e.g. brochure cost and postage) and the average order value then we can work out the cost wastage and lost opportunity:

Item	Value
Number of Mailings	400,000
Cost of Item	£5
Returned Mailings %	4%
Returned Mailings	16,000
Cost of Return Mailings	£80,000
Conversion Rates	1%
Lost Conversions	160
Average Order Value	£2,000
Lost Opportunities	£320,000
Total Lost Costs & Sales	£400,000

Note: Figures in Red indicate calculated figures

In this example we have a potential £320,000 lost sales and a further £80,000 in direct costs.

Looking at the data before the mailing, we can determine how many duplicates there are, how complete the data is or did we have conformity problems with addresses. These can then be directly attributable to the £400,000 cost to the business.

Dimension	%	Cost
Accuracy of Name	10%	£40,000
Accuracy of Address	50%	£200,000
Completeness of Addresses	20%	£80,000
Duplicate Records	20%	£80,000
	Total:	£400,000

The Seven Dimensions provide a solid set of data quality metrics that can be linked to business metrics. From this example, budgeting internal resources, services or technology for improving data quality has a crystal clear return.

ACUATE
The Data Quality Company

Conclusion

Acuate's Seven Dimensions of Data Quality allow organisations to:

1. Define a standard of data quality metrics
2. Provide objective measurement for data quality projects
3. Provide a method for continual improvement
4. Link data quality to business metrics
5. Improve business performance through proactive improvement in data quality

If you want to improve any business process then measurement is required, Acuate's Seven Dimensions of Data Quality enable organisations to create a base line for data quality. This baseline can trigger process improvement programmes and training programmes. Post-project measurement determines your success criteria. The circular sequence of measure, create action plan, execute and measure again matures the organisation's data quality and its performance.

Improve Data Quality, Improve Business Performance

Contact Us

Telephone:	+44 (0) 20 8799 0360
Fax:	+44 (0) 20 8799 0301
Email:	info@acuate.com
Web:	www.acuate.com

Acuate Ltd
Level 7, Westgate House
West Gate,
London W5 1YY
United Kingdom