

Data quality is the foundation of any data intensive business like yours, whether you are a bank, an insurance company, a mutual fund or a telecom service provider - and assuming that you have good quality data, is like assuming that the foundation of your office building is good!

Where's Your Data Quality Scorecard?

Hurry Today? Or Worry Tomorrow?

If you are a senior manager at a bank, insurance company or a telecom service provider, and do not have a Data Quality scorecard, or worse still, do not know what it holds, you will soon join the worried and troubled lot.

If you are the CEO / CFO / CMO / Compliance Head / CIO of a bank or insurance company or a telecom service provider - you should be really worried!

Our senses receive stimuli and transmit emerging data to the brain almost instantly, after some fairly complicated processing. From childhood, due to the proven reliability of these processes, we have learnt to totally trust our senses.

Do we trust our computerised data acquisition systems as much? Do we always have the correct customer name? Or address? Or demographic, psychographic, economic details? More often than not, the answer will be NO. Why not? Because, most organisations often don't know the level of quality of their data. We generally assume that we have good quality data. Assumption is not assurance, however! Data quality is critically essential for good decision making.

Data Quality is the foundation of any data intensive business like yours. Assuming that you own data of high quality, is like presuming the foundation of your office building is strong! For the building foundation, an official agency is responsible - how they do their job being completely irrelevant to this discussion! However, for Data Quality, absent any specific individual responsible

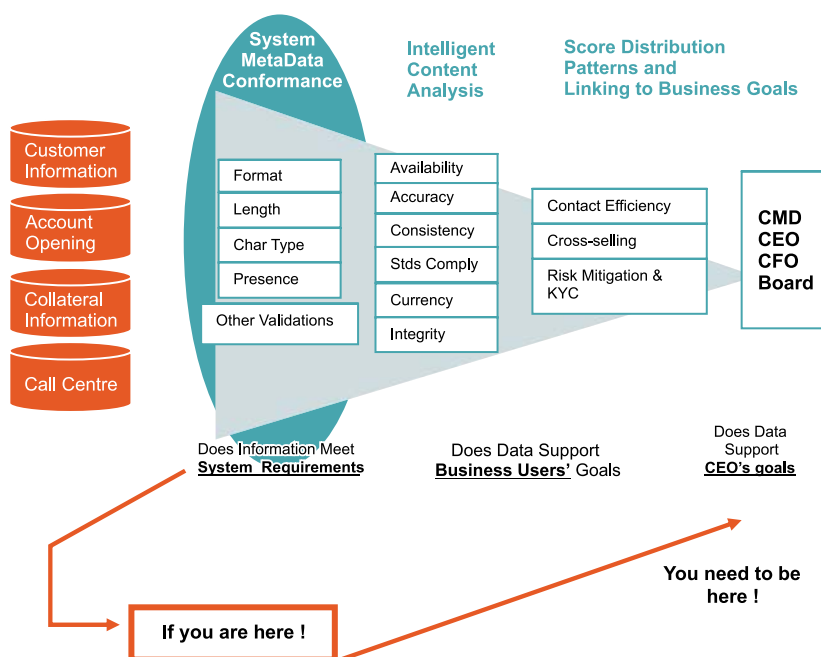
for it, it is a case of 'everybody's responsibility being nobody's'. As a result, you really have no clue about the foundation on which the current MIS reports, that you receive daily /weekly /monthly /quarterly are based. Are you getting fine looking reports that mean very little, since the underlying data is suspect? Not inconceivable! Worried?!! **It would be wise to get worried. But, wiser still to get moving to get your data "scored"!!**

Data Deficiency - What is measured gets done! What is not measured is not understood!!

Studies across the globe clearly state that data deficiency is the biggest nightmare for most business users. What is worse is that the level of data non-availability is never calculated. Data deficiency assumes alarming proportions in various forms.

Example 1: A consumer durables company realised that they could never locate a customer on their ERP system when s/he called in for the first service. Worried about displeasing the customer, they simply added his/her data to the system and created an additional entry for the same customer - without verifying whether the customer was a bonafide first-time buyer and therefore eligible for first service or not.

Result: Burgeoning Data with duplicates, money lost in providing free-service to buyers in their second year of purchase.



Example 2: A bank wanted to reduce the risk of lending by choosing less risky borrowers. This bank had default values for telephone and date of birth entered in a large portion of the existing records. Most new loan applicants' match against existing customer data resulting in wrong matches, loan rejections, visibly agitated applicants and unpleasantness.

Result: The bank suffered badly in market reputation and it's managers lacked confidence in the quality of its own database to help manage risk.

Over a period of time, such incidents can add up to a total decision-making paralysis and huge monetary losses.

It is critically important, therefore, to know whether the data meets the requirements of people at various levels from operations to the 'C' Level and for various purposes from routine decision-making to strategy formulation, including facilitating legal and statutory compliance. Business users need to decide whether their data is showing a green, red or yellow signal indicating a "Go", "Caution" or "Stop". This can only be done by assessing the current level of information quality - in line with business goals, making improvements and measuring them again. The process needs to be done on an ongoing basis.

The standard way of measurement is the Data Quality Scorecard.

Measuring key factors that impact Data Quality and inform the Scorecard -

Available & Complete

The first and primary consideration is whether data is available - establish the presence or absence of information. However, this alone can be misleading. Information may be present in certain fields but inadequate or totally nonsensical. That would be equivalent to non-availability. So the second metric is completeness.

Data Quality is the foundation of any data intensive business.

Example: If the 'City' field is not populated - this is missing data. If the field has abbreviated content, then too the data is as good as missing, assuming those abbreviations do not stand for any city codes. If the address field contains - AB Road - there is content but it is again as good as missing, since information is not sufficient or complete.

Correct = Accurate and Current

The third measure is correctness, i.e. accurate and current. For example, if the Date of Birth field is to contain data in dd/mm/yyyy format and it contains "22/01/1898" the data is available but most likely inaccurate - since the person is dead long back. In an extreme case, it was found that residents of a particular village all had the same dates of birth for their insurance policies. This is a case where date of birth is valid and recent but demonstrating a fraudulent pattern.

Consistent

The fourth measure is data consistency - if data is available and correct, it is necessary to establish whether or not it is consistent with other elements in the data.

For example - the city field may contain "Mumbai" which is correct, the Pincode field may contain 110021 which too is correct on a stand-alone basis, but taken together the data is inconsistent. Should the record figure in the MIS for Mumbai or Delhi? The answer may emerge from other components of the available data.

Standard Compliant

The fifth measure is data standardisation. If data is accurate, available and consistent - it may still not figure in the decision making process because it is not represented in standard form.

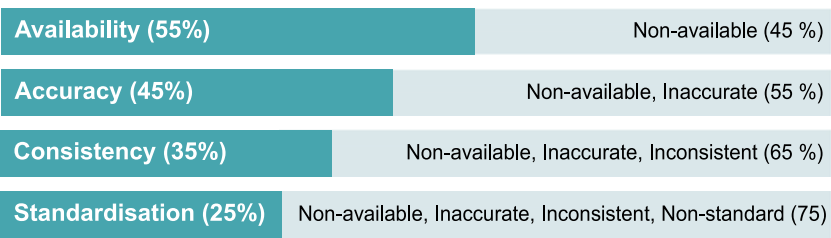
In the above example: If Mumbai was written as Mmbai and the Pincode field contains "400021" -the conditions of availability, accuracy and consistency are met. However, they fail on the data standardisation measure.

A query on Mumbai will not result in this record being fetched, specially if "Mumbai" is not present in the city field but in some other field.

There are several other measures - but this article is focussed on some of the primary considerations in creating a Data Quality Scorecard.

It is difficult to acquire new customers but far easier to generate more revenue from existing customers. Invest in your customer data so that you can reach out to them and build a bond which will reap rich dividends.

Only 25% Data Typically gets Reflected in the Output of most MIS or Business Intelligence Systems



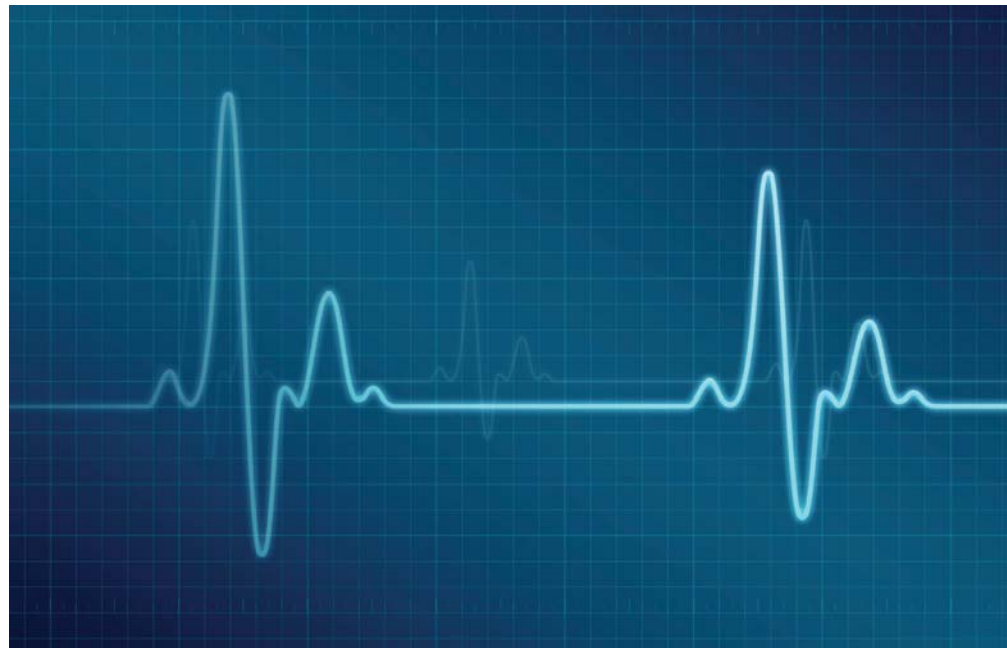
Building a Data Quality Scorecard

To make a Data Quality Scorecard truly representative of the data, with metrics that are aligned to business goals, the following steps are critical:

1. Defining needs of various business users through exploratory methods.
2. Understanding the currency, geography spread, product spread of data and evolving sampling methodology such that each sampling cell behaves homogeneously. For example: housing loan and personal loan data may have to be separately audited since the data capture methods of both business are different.
3. Defining the business rules as well as data quality rules for analysis.
4. Exploratory analysis of data to unearth anomalies not part of regular rule set.
5. Preparing the low level summary, high level summary and composite scores through a system of weights and thresholds.
6. Comparing this with industry benchmarks if available.
7. Linking score values to other insights from the business and providing recommendations.
8. Repeating process at regular intervals to ensure data quality stays at required levels.

Data Quality Scores and Profits

Data Quality Scores and Profits are very tightly linked. It may not be as obvious as linking customer additions to topline but increasingly companies are measuring the health of their data - as a key indicator in measuring



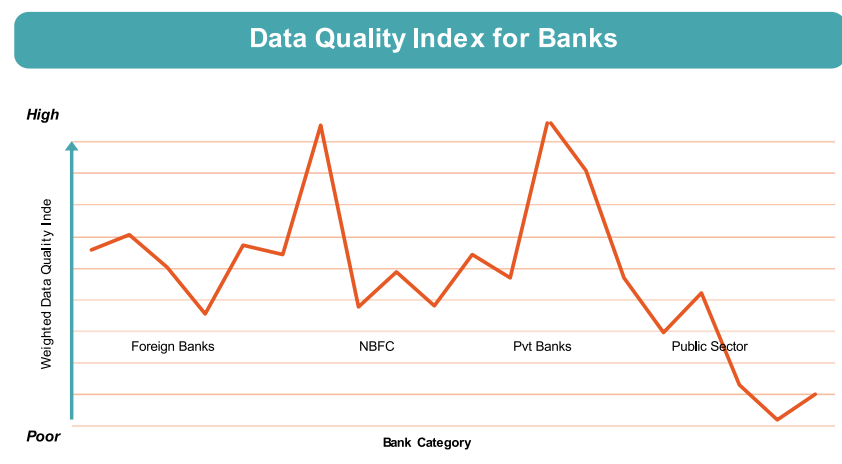
intrinsic business worth and ability to generate new business.

Almost 20-25 % of an organisation's revenue is lost because of poor data quality. An equivalent amount could be lost through opportunity gains foregone.

In order to beat competition and grow inorganically, mergers are the order of the day. How do you know whether the organisation you hope to acquire or merge with has intrinsic business value?

Customer data will reveal a major part of the story. It is difficult to acquire new customers but far easier to generate more revenue from existing customers. Invest in your customer data so that you can reach out to them and build a bond which will reap rich dividends.

Given below is a sample of a data quality Index created for the loan data across various banks and financial institutions



Source: Ixight Benchmarking Study done in 2008 for CIBIL

News

Ixsight Technologies to Accelerate Growth Rate

Having built a reputation for being a reliable, competent and qualitatively superior provider by solving tough data quality problems for corporates, Ixsight Technologies has recently received early stage funding from a leading Mumbai-based venture capital company.

The funding took place after a detailed due diligence of the market space, Ixsight's progress and its future potential. The funding will enable Ixsight to accelerate the growth of the company's proprietary and competitive solutions while expanding its market reach through intensified sales and marketing efforts, and serving existing customers more comprehensively.

Ixsight, the only Indian technology company to be listed as an upcoming promising entity for the second year in a row in Gartner's Data Quality Magic Quadrant, is now moving up the value chain from providing Data Quality Management to helping clients gain significant business insights from their data.

The company provides solutions across industry verticals and helps its customers meet specific business goals - reducing credit risks - operationally and strategically, providing superior customer service, and incisive customer insights at a geographic and demographic level, undertaking informed market expansion while significantly reducing communication costs. ●●●

What do Data Quality Scores do for You?

Data Quality Scorecard can be the key to meeting various business objectives. **Composite scores generated periodically for various key result areas form a direct indicator of potential success.**

As an indicator, a 50% composite data quality score for HNI (High Net-worth Individuals). 50% data quality score for HNI (High Net-worth Individuals) data may indicate that you can target no more than 30-40% business from existing HNIs. A 30% score on customer profile and identification data may mean that the organisation's risk exposure is very high, and drastic measures to improve knowledge about customer need to be taken up.

The Ixsight DQINDEX - is a measure of how well you can expand the business using existing customer footprint, customer value and therefore intrinsic business worth.

The DQINDEX indicates:

- How much revenue can be generated using existing customer goodwill?
- What is the level of operational risk that the organisation carries?
- What is the spend leakage because of bad data?
- What is the reputational risk because of bad data?
- What is the statutory non-compliance proportion, if any, due to bad/missing data?

Even if the graph is showing a positive movement, it may be a matter of concern to see how it compares with the competition, and if it is improving rapidly enough to make a difference to your business. If the index is low, it is time for action on a war-footing.

Like all other audits - Security or Accounting - since it deals with your most precious asset - Data - it therefore deserves to be addressed twice a year! ●●●

About the Author

Vivek Likhite is Chief of Credit and Operations at Bajaj Auto Finance Ltd. The views expressed are his own and do not necessarily represent those of his company.