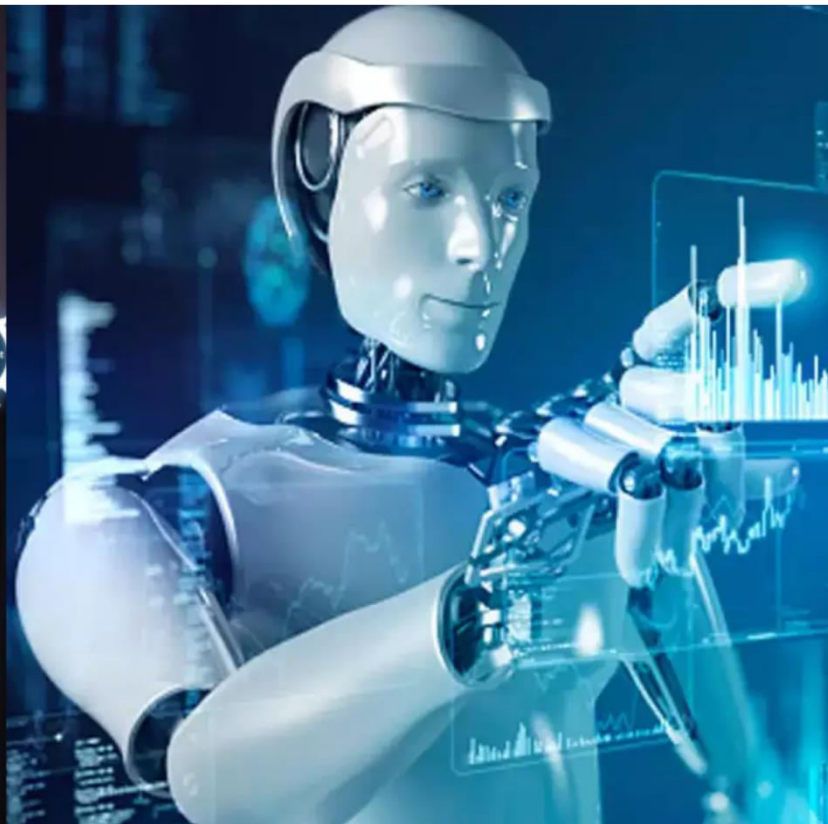




International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 2, February 2025

Achieving High Customer Data Quality: A Comprehensive Guide

Bihag Karnani

Google, New York, USA

ABSTRACT: The quality of customer data is critical for organizations in the information age. In this paper, we highlight the need for the right data quality, its attributes, and reasons for poor customer data quality in sales and revenue. It also covers the causes of poor customer data quality and offers tips on how to get and keep quality data. This paper highlights the importance of a structured approach that includes building a data quality team, data profiling, and data quality remediation. The size of the business case is such that investing in customer data quality initiatives means less waste, more sales and better customer loyalty.

KEYWORDS: Customer Data Quality , Data Quality Attributes, CRM Data, Data Governance, Data Profiling, Data Remediation

I. INTRODUCTION

As we move forward from the last decade of the 20th century, where the focus was on deploying IT solutions to the 21st century, which is the Information age, we realize how data both enables and cripples companies and industries. The focus of ERPs, CRMs and various other systems deployed in the 20th century was to integrate systems and to collect data. The focus of the current century professionals is to gather insights from the data collected in various databases and warehouses. As the focus on Data Analytics and Data Science increases, enterprise companies are increasingly realizing the nuisance created by poor data quality. The Data Warehousing Institute estimates that data quality problems cost US businesses \$600 billion a year¹. Organizations can potentially lose and alienate customers by inaccurately addressing emails or failing to recognizing them when customer call. This loss of customer is a potential loss of future revenue. It is time for companies to view data as a critical resource and make a corporate commitment to manage data quality leading to establishing a program with dedicated people, process, systems and tools to achieve high data quality. Although, data quality problems will exist in every business unit of a company, this white paper will focus specifically on Customer Data Quality, which has a direct impact on sales and thus, revenue.

II. WHAT IS DATA QUALITY?

While we continue to discuss the causes, impacts and solutions for poor data quality, it remains pertinent to define what data quality is and what its constituent attributes are. Data quality is not just data that is devoid of errors. Incorrect data is just one of the characteristics of bad data. A complete definition of Data Quality is made up of 6 attributes mentioned below.

Volume 14, Issue 2, February 2025

[DOI: 10.15680/IJIRSET.2025.1402009]

III. ATTRIBUTES OF DATA QUALITY

Data quality is not just limited to the accuracy of raw data, but a list of attributes that characterize the quality of data:

Image 1: Data Quality Attribute Framework



1. **Accurate:** Data should accurately represent reality or a verifiable source
2. **Consistent:** Data elements should be consistently defined and understood
3. **Complete:** All the necessary data should be present
4. **Valid:** All data values should fall within acceptable ranges defined by business.
5. **Real time:** Data should be available when needed
6. **Accessible:** Data should be easily accessible, understandable and usable.

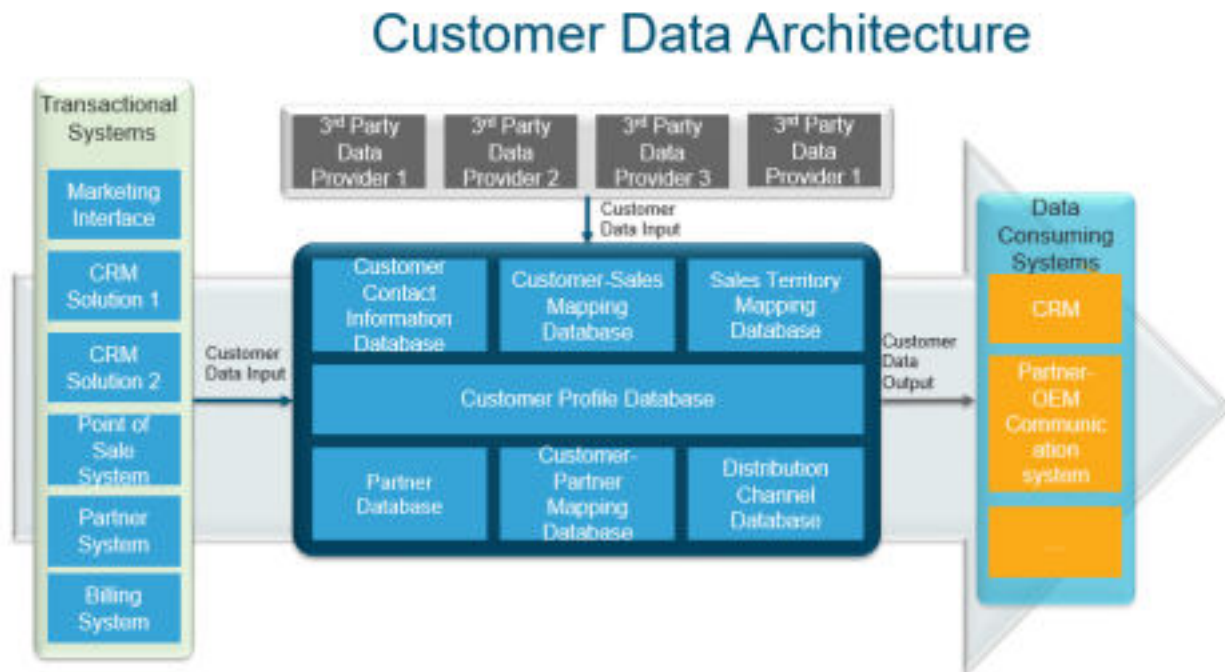
IV. CUSTOMER DATA QUALITY

According to Gartner, the amount of prospect and customer data in a business-to-business organization typically doubles every 12 to 18 months². Further, between 10 to 25 percent of customer and prospect data has critical data errors ranging from inaccurate demographic data to lack of current disposition². Organizations need to maintain high data quality in order to ensure efficient operation of CRM systems on which they have spent millions of dollars. While CRM systems act as integrators, their output is only as good as the quality of the data that they use as input. Maintenance of high customer data quality also ensures maintenance of a meaningful and accurate set of customer metrics. Achieving and maintaining a level of highly integrated and consistent customer data, necessary for CRM success is not an easy task. Lack of customer data quality may also lead to inaccurate sales forecasting which sets inaccurate goals for the entire organization, lost sales opportunities for new sales, upsell, and cross sell opportunities and incorrect sales data compensation allocation which further leads to loss of employee morale. In order to understand customer data better, it is impertinent to understand the architecture for customer data in an enterprise organization.

V. CUSTOMER DATA ARCHITECTURE

The Customer Data Architecture provides a base for collecting, storing and consuming customer information. It defines the systems, databases, data flow and data consumption systems used in a typical enterprise organization. The self-explanatory Customer Data Architecture shown below would be used as a baseline to identify and to detail Customer Data Quality problems and then to identify solutions to said Data Quality problems.

Image 2: Customer Data Architecture



VI. SOURCES OF POOR CUSTOMER DATA QUALITY

Now that we have drawn our focus to customer data quality and briefly on the impacts that poor customer data quality might have, let us look at some common sources of poor customer data quality:

Human error factor

As is mostly the case, most data pollution issues start with inaccurate data entry. The more manual the process of data entry is, the more inconsistent, inaccurate and incomplete is the data. Since sales is a relationship driven activity, most of the data entry at CRM or ERP systems is manual. The added complexity is the resistance of the sales agents to new creating new processes and systems.

Multi-channel data inconsistency

Since the focus of sales organizations is to make it easier for customers to purchase products and services, B2B organization sales business units often create multiple touch points to serve the customer. Customer data usually resides in multiple systems like CRM, Point of Sale systems, Billing Systems etc, coupled with an organization's efforts to purchase data from 3rd Party providers or validate customer information from 3rd Party providers, it makes achieving high quality customer data a nightmare.

Multiple data stores

Supporting the multiple data source systems are multiple data stores, which have their own integration limitations along with multiple data formats, attribute definitions and table structures.

Data definition and Validation

In all probability, each of the system identified in the Customer Data Architecture are used by a different team, which implies that each team would have their own definition for data and a different set of validation rules that govern data integration. It can become challenging for any CRM system and sales team to create a consistent customer definition from this chaotic data architecture.

Business Process Breakdowns

While data migration and conversion projects are performed with the good intent of reducing data discrepancies, ETL tools that pull data from one system to another also generate defects. Although systems integrators may convert databases from one format to another, they may fail to integrate different business processes. This breakdown now becomes the source of creating data pollution.

Dynamically Changing Dimension

Mergers and acquisitions, which are quite commonplace in the enterprise IT world, pose another challenge for data consistency. The intent of an M&A might be to acquire product or to acquire a new demographic, but the challenge in integrating the acquiring company systems with that of the acquired company systems, creates a maze of data that is difficult to navigate. Another case of M&A that significantly impacts customer data is that of M&A in the client structures⁴. If, for example, company A merges with company B, and both company A and B were customers of firm C, it is a huge task for company C to collate company A and B sales and customer data to ensure ongoing opportunity mapping and maintaining a clear view of future upsell, cross sell opportunities.

In summary, making it easier for the customer to purchase product in turn poses a risk of a diminishing customer data quality for the organization, which increases operational costs and jeopardizes future sales opportunities.

VII. STEPS FOR ACHIEVING HIGH CUSTOMER DATA QUALITY

The end goal for every enterprise company is to achieve high Customer Data Quality. The more structured the approach to achieve this goal, the sooner the goal is reached. The recommendations mentioned below are a result of years of implementation and consulting efforts with methods that lead to Data Quality success:

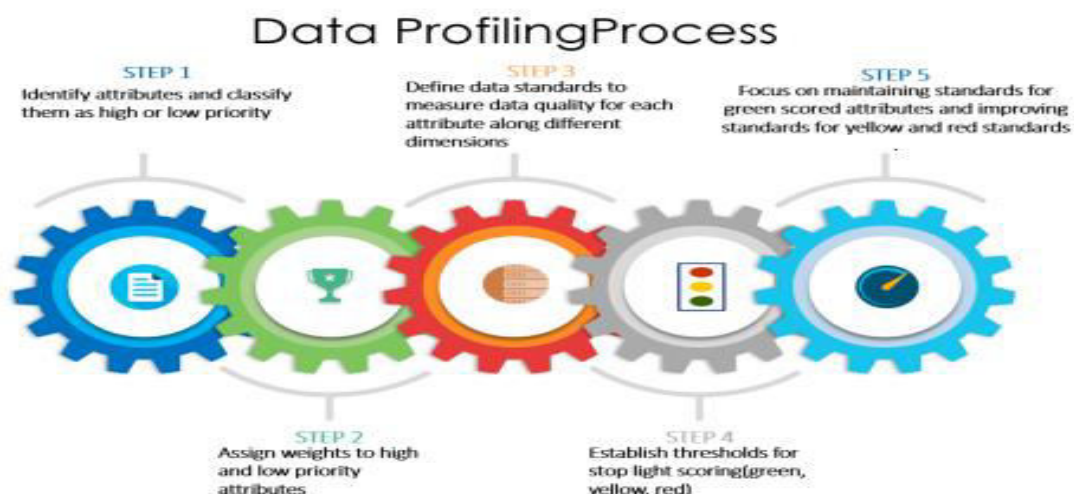
Creation of a Data Quality Team

Since data inflow and consumption is an ongoing activity with an ever-increasing scope, the Data Quality improvement and maintenance effort should also be ongoing and continuous. It is recommended to identify data owners for the customer data who manage a team of data stewards. This team of data stewards constantly comes up with process, policy and system improvements with the dual goal of improving data quality, and work efficiency. Data Custodians should implement the changes the data stewards recommend. In most enterprise firms, Data Custodians are the IT team who have to align the data architecture with the recommendations of the Data Stewards and Owners.

Profiling data

Creating a weighted average metric to measure data quality provides an organization with a perspective into what the data quality situation is and what is the short and long-term goal for data quality. The process for data profiling can be found below:

Image 3: Data Profiling Process



Data quality metrics can be formed using the attributes mentioned above. The data should be accurate, consistent, complete, valid, real-time and accessible. For each firm, the importance, availability and requirement for the data against these attributes would differ. Once the data is measured against these metrics, data attributes can be identified in red, yellow and green zones, which further helps, identify the sources of data pollution. As an example, the following metric table can be created to keep a track of the customer data quality on an ongoing basis. As more fields are added, the metric calculation would become more robust and the current situation of the quality of customer data more transparent. Overtime, the focus on the team should be to maintain the green standards for attributes and to improve the data quality levels for attributes that have yellow and red standards.

Table 1:Example of a Data Quality Table

Use Case	Rule #	Rule	DQ Weight	Confidence Level	Overall Customer Data Quality Level
Customer Contact Information	Rule 1	Single Point of Contact	30%	20%	Yellow
Customer Install Site Details	Rule 2	Multi Install Site Mapping	40%	60%	
Customer Asset View	Rule 3	All Products and Services View	30%	85%	

VIII. DATA QUALITY REMEDIATION

Remediation of poor data quality needs to have a 3 pronged approach. Improving quality at source, improving quality at storage and consumption level and improving pre-existing data pollution. All these three are explained below:

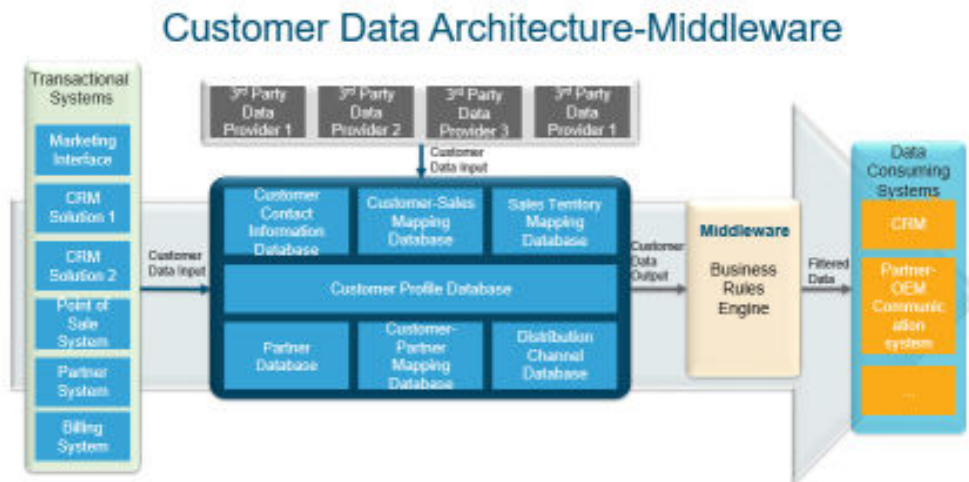
Improving data at source

Since the biggest source of data quality errors is data entry by humans, more automation and digitization is the most likely solution. With more and more automation solutions available in the market, investing in solutions that give sales agents suggestions of customer name, address, contact information etc. would help reduce manual entry and thus reduce inconsistency, and inaccuracy in the data³.

Improving data at storage

Like mentioned before, enterprise IT sales cycles would always comprise of multiple channels of sales that use multiple systems. The data from these multiple systems would inadvertently be stored in multiple databases, which may be inconsistent. In order to change this situation, there needs to be a change made to the data architecture. A middleware that acts as a data filter between data storage and data consuming systems is the need of the hour. This middleware will house the business rules applicable to the data for each consuming system depending on the need. Additionally, creating a middleware would avoid the cost of integrating multiple data sources and avoid further data mess created by ETL tools used for integrations.

Image 4: Customer Data Architecture Middleware



Improving pre-existing data pollution

While the above-mentioned 2 steps would help improve situations going forward, it is important to look back the mistakes of the past and rectify them. Creating a write-back program that determines sources of polluted data and creates methodologies to write either higher quality data back to source or at the middleware level is critical to having a long lasting effect on Data Quality. In addition to a write-back program, a continuous cleanup effort for preexisting polluted data paves the way to achieving success in achieving high quality data.

IX. CONCLUSION

In conclusion, achieving high quality customer data and then maintaining it, is not beyond any company. The more initiative an organization takes in this matter, the higher rewards it would reap by means of reduced wastage, increased sales and improved customer loyalty. Technology can help automate most of the recommendations provided. The most effective data quality improvement initiatives are created by the leaders of the organization, which ensures vision, commitment, and drive. Truly, the time to start these initiatives is now!

REFERENCES

1. Shaw, M. J., Subramaniam, C., Tan, G. W., & Welge, M. E. (2007). Customer data quality in CRM systems: An empirical study. *Journal of the Association for Information Systems*, 8(11), 583-613.
2. Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of management information systems*, 12(4), 5-33.
3. Redman, T. C. (1996). *Data quality for the information age*. Artech House.
4. Karnani, B. (2024), Subscriptions as a Revenue Model: Evolution, Comparisons and the Impact of Generative AI, *International Journal of Computer Trends and Technology*, 72(12)



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details