



THE INFLUENCE OF DATA WAREHOUSE: DATA QUALITY

**Ahemad
Noormahamad
Nadaf***

Bachelor of Pharmacy, Student, ClinoSol Research Pvt. Ltd.
*Corresponding Author

**Adinarayanamma
Lakkireddy**

Bachelor of Pharmacy, Student, ClinoSol Research Pvt.

Sankalp Arvind Bali

Bachelor of Pharmacy, Student, ClinoSol Research Pvt.

ABSTRACT

Nowadays, if you want to run a company then you must want one thing which is data. If you have huge, neat, and sorted data then your organization can beat the market marathon for a long while. A huge amount of data can be stored in data warehouses. If you have the data and knowledge then you have the greatest weapon of all time because the organization needs the data and the decision-taker wants the knowledge. Data collection is a challenging and continuous process, everyday data gets generated from somewhere and somehow, but the challenging part is its collection as well as storage. In the market, various databases are available for the storage of data. It does not end when data is collected or captured, the management of data throughout its lifecycle is also important. This data is precious because it can lead the organization in an upward or downward direction.

KEYWORDS : Data, organization, marathon, knowledge, storage, data collection.

INTRODUCTION

If you have high-quality data then there are high chances of getting success in a project. So, the data quality brings either the success or failure of your project. Also, the quality of data can influence the growth of a private or a government sector and to avoid malfunctioning in the sector there are laws which are Data Quality Act, 2002 and Data Quality Initiative Framework, 2004 by US and Wales respectively. That's why; data quality is an influential factor in data management. Since dirty data often leads to inaccurate data analytics results in poor data quality which leads to poor patient care further creating problems.

Data quality management (DQM) has developed from a tool-oriented set of IT activities to a managed set of best practices for managing the data in private, public as well as governmental use of data warehouses to facilitate data processing at various levels including, operative, administrative and strategic[1]. Data should be stored in databases that comprise several tables or views and data warehouses are also composed of several instances of databases, which further facilitates decision support at all organizational levels. Data needs to be collected and processed efficiently and effectively to support the data consumers (people, groups, or systems that use data). Data quality may differ.

Due to the collection of data that is often incumbent on humans (systems), in a study by H. Muller [2] the reasons for poor data are lazy input habits, erroneous measurements, omissions while collecting and maintaining the data, and misinterpretations in data analysis.

There are several types of databases, some are elegant and some are gauche but what brings it, is the assessment of data quality, expenditure of data quality analysis, and quality processes. We can come across this information by screening the database catalogs, if you don't want the screening you can go with the scanning of database contents.

Data quality is a stage at which the need of the client connects with a dataset. Data quality should be errorless and clean so that any type of user can access it. Data is a variable aspect so it should be evaluated and re-evaluated in a regular timeline, resulting in sustainable data in the dataset. We have various

databases, one of it the Canadian Primary Care Sentinel Surveillance Network Database [3]. From the study analysis of the Chinese National AIDS Information System, it is found that several aspects influence the database. The factors are capturing of data, management of data, information system, and environment of data capturing [4]. One report from Scotland claims that indicators of Quality and Outcome Framework and GP software were tested with T-test and Chi-squared tests regarding National data. And surprisingly, the outcome was 77% performance of the Quality and Outcome Framework (QOF) [5].

Impact of Poor Data Quality

Nowadays, the data is used by both beings, i.e. human beings as well as system beings, so the quality of data plays a vital role in bringing the quality of care. The impact of data quality can bring you the number of clients; if you bring high-quality data it going to increase the number of clients for your firm. There are proof reports which enlighten the low-quality data that devastated the firm at various business levels.

The four considerable areas where the impact of data quality is major are as a. Confidence and satisfaction-based vitality: Referring to satisfaction of clients, and suppliers, results in decreased trust level towards the firm. b. Productivity: Having high-quality data considerably brings you the number of clients [6]. On the other side, if you have low data quality brings you lots of problems for your firm. It can directly affect the workload of the firm, reflecting the image of the firm. c. Risk and Compliance: Here identifiable risks are funding, position in the market, stalk holders, and policies given by the government or firm itself. d. Financial vitality: Low-profit results in expensive operational costs. This cascade flashes a loss of opportunities. A high level costs a huge debt to the firm [6].

The aspect of capitalization is not just related to your organization, it's come to your nation also. It means, can influence your market of the nation either in the positive or negative side of the graph. One of the developed countries has issued the classification of cost impact as well as came to know that there are two types:

a. Direct cost: This comprises poor data capturing in the process and the required cost to enhance that poor data.

Managing the data requires less cost than the other.

- b. Indirect cost: This is accompanied by a loss of credibility and client satisfaction. Client happiness is vital while keeping your firm evergreen, so you won't go with poor data or poor data management. You have to keep the data of clients at a consumable level so that you can maintain your position in the market.

Data quality expenditures are reported of three types, which are

- a. Entry of Data: Now this stage is a vital stage where the magic happens, if you collect data in a high range then you can maintain the pillars of your firm. If data collection is of poor type then you have problems. The firm has to face two problems, no. of clients and further correction of data or call it data management. In the initial stage, if you spent lots of liquid, you haven't seen further steps.
- b. Processing of Data: Processing of data is divided into further subtypes like reworking dirty data and process improvement costs. In cases of improvement, you have to go with steps like Detection, analyzing, and reporting the dirty data.
- c. Use of Data: Data use costs account for compensation costs, lost revenue, opportunity, liability costs as well as what you fear is a lower reputation.

Table 1 Impact of poor data quality

No.	Administrative Impact	Operational Impact	Strategic impact
1	Impact on decision making	Reduce efficiency and productivity	More difficult to set a strategy
2	Harder to implement a data warehouse	Lower customer satisfaction	More difficult to execute a strategy
3	Missed opportunities	Lost in revenue	Contribute to the issue of data ownership
4	Misanalysis	Lower employee satisfaction	Divert management attention
5	Difficult to perform root cause analysis	Increased financial cost	Compromise ability to align the organization

The Efforts to Improve the Data Quality

Improvement of data quality is very important for getting accurate and errorless data as well as more effective research. There are six methods to improve data quality in any organization [8].

a. Conduct routine audits for health data quality or assess the data

Every improvement cycle starts with an assessment. A detailed audit helps list the strengths, weaknesses, threats, and opportunities that exist in your data. Audits typically end with an exchange of recommendations and suggestions for improving data quality.

b. Implement systematic data quality management in healthcare institutions

On-demand execution of data quality techniques yields results today but does not guarantee consistent data quality results in the future. This is where you should implement a data quality management system, preferably via automated workflows, so that new and upcoming data is batch processed for data quality checking and correction before being stored in the destination source.

c. Involve healthcare leadership and management

Making changes across the organization is possible only with the consent or approval of the organization's leadership and senior management. Many healthcare facilities hire on-site

data quality managers - these roles are responsible for applying data management best practices to minimize data loss and maximize data quality. They are considered gatekeepers or health data caretakers.

d. Perform root-cause analysis

A root-cause analysis for data quality errors may entail obtaining the most recent data profile report and working with your team to find answers to questions such as:

- What data quality issues were discovered?
- Where did they come from?
- When did they first appear?
- Why did they get into the system despite all data quality checks? Did we overlook something?
- How can we keep such errors from recurring in the system?

e. Train and educate healthcare teams:

The ability to properly and accurately read, understand and analyze data at all levels enables medical and clinical staff to make the right decisions. It also ensures operational efficiency and reduces communication errors in data-related content.

f. Maintain data quality by using technology

The key to increasing data quality in healthcare facilities is to use technology to establish a long-term data quality management lifecycle.

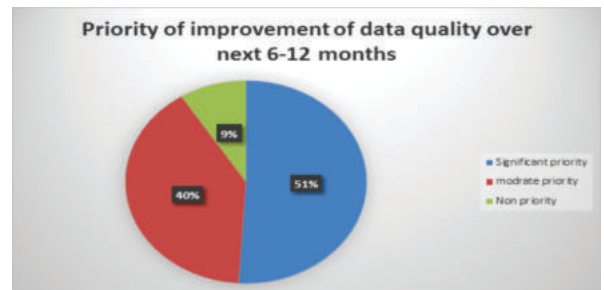


Figure 1 Priority of improvement of data quality over the next 6-12 months

Best Practices for Data Quality in Data Warehouses

Data quality is a crucial element of any successful data warehouse solution as the complexity of data warehouses increases, so does the need for data quality processes when an organization achieves consistent high-quality data. They are better positioned in the market, get buy-in, and make data quality an enterprise-wide priority. The quality of data can't be taken for granted as it is a very influential factor for the organization [9].

If only half of the company is committed to ensuring data quality. Then we can expect no better than 50% data quality. All the stakeholders must understand and take responsibility. We need a way to measure data quality by establishing metrics that apply to the goals and business targets to achieve the best quality data and understand how accurate our data is. If we fail to investigate data quality, it stands to reason that errors will continue, correcting errors in data can be a difficult time-consuming task [9].

These all require a deep understanding of data quality principles, processes, and technologies. Establish and implement data governance guidelines that go beyond rules and data protection. It engages business users in best practices and as members of the data team and also establishes a data auditing process.

Audits on data, within data, are the best way to build trust in data, and also the frequency of audits is important to the acceptance and success of audit data. Good data quality is both essential and difficult to achieve so we have to plan how

to get started on the best pathway to improve the best quality data. Auditing is beneficial for the sponsor organization to reduce any type of bias in the data captured.

For the outcome elevation of the project, the best method should involve all participants like the primary health network, researcher group, digital developers in the healthcare sector, and the group of consumers. Through the digital platform, you can safely utilize real-time general data. Primary health network has a great chain working system around the world.

Framework to improve data quality

The framework provides the standards for accessing and reporting statistical information and also provides an industry-developed best practices guide for the improvement of data quality and allows companies to better cover their data quality programs first we have to desire goals and objectives regarding data quality. It is always best, to begin with, the end in mind. what are the casts that can be attributed to poor quality this is all should be designed a data quality framework must begin with a clear definition of desired outcomes then identify areas which area leads to poor quality and fewer data and then we should apply our implements for improvements? An effective data quality improvement is also required through documentation [10]. The project leads should establish a methodology for establishing data, also monitor changes and plan further better improvements for measuring best quality data and compare data before and after implementation of improvements.

Nowadays, handling the huge amount of data in electronic health records is easy. It can be done in the following steps,

- a. Determination of answers regarding the research questions concerning data source.
- b. Detecting the presence of any other variability of data.
- c. Use the approach to search variability and assess the data quality.
- d. Create a systematic dataset and perform scanning.

With the use of the above framework, it is easy to report quantified and provide novel and sturdy quality data. It can develop and validate risk management models for many projects. This framework can contribute to achieving project completion [10].

Data Quality Management

First of all, we have to understand that poor data quality is nothing but unwanted trash that you can't dumb for an organization, because it neither fully helps you nor you can leave it. There is a need for data quality management for ensuring consistent data in our databases. As the year passes, the increase in demand for data quality management is seen in every industry, aiming for the detection and correction of errors in data, resulting in the addition of accuracy and value to business processes. Improper data quality management, even small errors can results in huge revenue loss, inefficiency in the process as well as non-fulfillment to comply with industry and government regulations. The various stages of data quality management comprise finding rules, inspecting inconsistency, and restoring, necessary to be done acceptably [11].

Data management generally involves handling data throughout its lifecycle. This means managing how they are collected in one or more systems and how they are represented and arranged in database systems. It also means controlling how information is fully documented, securely backed up, and monitored over time. It is protected from unauthorized access and changes shared with other people and systems and updated with new information. Checked quality and fixed errors. How it is transformed for different purposes. And finally, how it is destroyed. Data management is relevant to many areas such as business, marketing,

customer service, hospital billing, and arguably clinical research [11].

For enhancing the knowledge of the medic field the satisfactory way is to conduct a clinical trial. After the conduction of a trial, it provides the data and for assurance of that quality, data there are so many national and international guidelines.

Some project demands internet-based entry of data system and real-time data access allowance [16]. And side by side incoming data should be monitored and cleaned in an accessible way. Managing the data quality in a proper way results in a positive impact on the organization.

CONCLUSION

Data is everywhere and everything for the organization and business; And in the market, various databases are also present. It can affect the growth of the organization either in a positive or negative direction. Data quality can bring the number of clients to the organization.

There is a huge requirement to keep the data at an acceptable scale, and the data is classified into three types, data entry, data processing, and data use. Keeping clean data cost high liquid for the organization. There is a huge scope for data quality management for keeping the data consistent in the flow. Several frameworks help to keep data in a quality manner. Based on project requirements, changes should be done in the data capturing system and working manner to reduce bias and speed up the completion of the project.

Acknowledgment

We would like to express our special thanks of gratitude to ClinoSol Research Pvt Ltd who motivated us to write a review on the topic "Influence of data warehouse: Data Quality". A special thanks to Mujeebuddin Shaik- founder and CEO of ClinoSol Research, Uma Priya- Director of ClinoSol Research, and Noorush Shifa Nizami for guiding and supporting us on this project.

REFERENCES

1. Bae, C. J., Griffith, S., Fan, Y., Dunphy, C., Thompson, N., Urchek, J., Parchman, A., & Katzan, I. L. (2015). The Challenges of Data Quality Evaluation in a Joint Data Warehouse. *EGEMS (Washington, DC)*, 3(1), 1125. <https://doi.org/10.13063/2327-9214.1125>
2. H. Müller, J.C. Freytag. Problems, Methods, and Challenges in Comprehensive Data Cleansing. In Technical Report HUB-IB-164, 2003.
3. Ehsani-Moghaddam B, Martin K, Queenan JA. Data quality in healthcare: A report of practical experience with the Canadian Primary Care Sentinel Surveillance Network data. *Health Inf Manag.* 2021;50(1-2):88-92. doi:10.1177/1833358319887743
4. Chen H, Yu P, Hailey D, Cui T. Data Quality of the Chinese National AIDS Information System: A Critical Review. *Stud Health Technol Inform.* 2017;245:1352.
5. Weatherburn CJ. Data quality in primary care, Scotland. *Scott Med J.* 2021;66(2):66-72. doi:10.1177/0036933021995965
6. N. Laranjeiro, S. N. Soydemir, and J. Bernardino, "A Survey on Data Quality: Classifying Poor Data," 2015 IEEE 21st Pacific Rim International Symposium on Dependable Computing (PRDC), 2015, pp. 179-188, doi: 10.1109/PRDC.2015.41.
7. Dr. Michael Schreffl, and Hons. Alastair Long, "Data Quality and Data Cleaning in Data Warehouses," 2007 Data Warehousing & Data Mining Seminar SS 2007, and Fachsprache Englisch Seminar SS 2007.
8. Ehsanelah, "Data quality in Healthcare - Benefits, challenges, and steps for improvement" 2022 Data Ladder.
9. Collins, "Best Practices for data quality in data warehouses" 2022 Tech Republic.
10. Stuart E. Madnick, Richard Y. Wang, Yang W. Lee, and Hongwei Zhu. 2009. Overview and Framework for Data and Information Quality Research. *J. Data and Information Quality* 1, 1, Article 2 (June 2009), 22 pages. <https://doi.org/10.1145/1515693.1516680>
11. B. Saha and D. Srivastava, "Data quality: The other face of Big Data," 2014 IEEE 30th International Conference on Data Engineering, 2014, pp. 1294-1297, doi: 10.1109/ICDE.2014.6816764.
12. Jarie Bolander, "How Poor Data Negatively Impact Your Business" 2019 THE DAILY MBA.
13. Objective, "THE IMPACT OF POOR DATA QUALITY IN 2021," 2021.
14. Redman, T. C. (1998). The impact of poor data quality on the typical enterprise. *Communications of the ACM*, 41(2), 79-82. doi:10.1145/269012.269025
15. Experian™, "2021 Global data management research," 2021.

16. Houston L, Probst Y, Yu P, Martin A. Exploring Data Quality Management within Clinical Trials. *Appl Clin Inform.* 2018;9(1):72-81. doi:10.1055/s-0037-1621702
17. Watson NL, Prosperi C, Driscoll AJ, et al. Data Management and Data Quality in PERCH, a Large International Case-Control Study of Severe Childhood Pneumonia. *Clin Infect Dis.* 2017;64(suppl_3):S238-S244. doi:10.1093/CID/cix080