

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,400

Open access books available

174,000

International authors and editors

190M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



# Introductory Chapter: Data Integrity and Data Governance

*B. Santhosh Kumar*

## 1. Introduction

The maintenance and assurance of data accuracy and consistency throughout a system's entire lifecycle are referred to as data integrity. It is an essential component of the design, implementation, and utilization of any system that stores, processes, or retrieves data. Data integrity can be compromised in a number of ways. The word covers a large swath of ground and may signify a lot of various things depending on the precise context in which it is used, even when discussing topics that fall within the overarching category of computers. It is often used as a stand-in for the phrase data quality, but data validation is an absolute need for ensuring that data is complete and accurate. On the other hand, data integrity is the absence of any corruption in the data. Any data integrity method should ultimately strive to accomplish the same overarching goal, which is to guarantee that data are captured precisely as desired (such as a database correctly rejecting mutually exclusive possibilities). In addition, when the data is retrieved at a later time, you need to make sure that they are the same as they were when they were first recorded. In a nutshell, the goal of maintaining the integrity of data is to avoid making unauthorized modifications to the data itself.

The concept of "data integrity" should not be confused with "data security," which refers to the practice of safeguarding information against access by unauthorized individuals. The term "data integrity" refers to the quality and consistency of the data (also known as "validity") throughout the data's lifespan. At the end of the day, compromised data is of little value to businesses, and that is before we even consider the risks that come with losing sensitive data. For this reason, ensuring the data's integrity is one of the primary focuses of the majority of business security solutions. There are a number of different ways in which the integrity of data might be undermined. When data is copied or transmitted, it must be preserved in its original form and not be changed in the intervals between updates. It is common practice to rely on error-checking techniques and validation processes in order to safeguard the authenticity of data that is copied or distributed without the aim of modifying it in any way. Another source of misunderstanding is the word "data integrity," which may mean either a condition or a process depending on how it is used. A data set is said to have data integrity if it is both valid and accurate, which are two very different things. On the other hand, data integrity as a process explains the methods that are used to assure the validity and correctness of a data set or all of the data that is stored in a database or other construct. This may be done by comparing the data against a standard. For instance, procedures for checking for errors and validating data may be referred to as "data integrity processes."

“Data governance,” sometimes known as “DG” for its shortened form, refers to the process of managing the availability, usefulness, integrity, and security of the data that is stored in organizational systems. This management is carried out based on the company’s internal data standards and rules, which also serve the purpose of controlling the method by which data is used. When it comes to data governance, doing it right means making sure that the data is reliable, consistent, and not utilized in any way that may be considered exploitative. A further benefit of data governance is that it assists in preventing the improper use of data. It is becoming an increasingly important factor as businesses are being forced to comply with new regulations concerning the protection of customer data and as businesses are becoming more dependent on data analytics to assist in the optimization of operations and the driving of business decisions. This is one of the reasons why it is becoming an increasingly important factor. One of the reasons why it is becoming an increasingly significant element is because of this particular aspect. A programme for the governance of data that has been established with care often consists of three distinct components: a governance team, a steering committee that serves in the function of the governing body, and a group of data stewards. Each of these components will be broken down into even more specifics in the following paragraphs. They collaborate in order to determine the rules and standards for the governance of data, as well as the techniques for its implementation and enforcement. This is largely the job of the data stewards, but it is also a part of their responsibilities. If all goes according to plan, participants in the dialog will include not only members of the IT and data management teams but also executives and other representatives of an organization’s business activities [1].

Nicola Askham, an independent consultant, wrote in a blog post that she authored in January 2022 that in order for the governance programme of an organization to be successful, the organization in question must place primary emphasis on the anticipated business benefits of the programme. This was stated by Askham in a statement that was included in the post. This information was provided inside the framework of the blog post that Askham had written. This is still the case in spite of the fact that data governance is a crucial component of any all-encompassing data management strategy. During a session that took place during the 2022 Enterprise Data World Digital conference, Eric Hirschhorn, the chief data officer of The Bank of New York Mellon Corp., made a comment that was quite similar to the one that was just mentioned. The remarks made by Eric Hirschhorn may be found at this link. According to what he had to say, “excellent governance” on its own could not be considered a sufficient achievement and of itself. The final result must be an improvement in the way the organizations are handled in order to be considered successful. This extensive reference on data governance provides more clarity on what data governance is, how it operates, the benefits it brings to businesses, best practices, as well as the challenges that come with managing data. During the course of the governance process, you will not only find a discussion of numerous other pertinent technologies that may be of service to you, but you will also find an overview of data governance software. This will be included alongside the conversation. You will find hyperlinks on almost every page of this guide. These hyperlinks, when clicked, will take you to relevant sites that cover the same or similar subjects as the ones that are now being discussed. These links are scattered throughout the manual in a number of different places where they may be located [2].

## **2. Data integrity and data governance**

In data integrity, the use of developed information evaluation techniques is required in order to investigate typically unknown, legal prototypes, and relationships in massive data sets. Mathematical prototypes, numerical procedures, and machine learning strategies might all be included in these tools. As a result, it entails the collection, organization, and storage of data, which includes evaluation and prognosis. It was possible to do this on information that was represented in quantifiable, text-based, visual, picture, or hypermedia patterns. For the purpose of doing an evaluation of the data, the apps could use certain metrics. They include things like relationship ordering or route assessments, classification, grouping, and estimations. Various businesses amass immeasurably vast amounts of information. The strategies were easily adaptable for use on traditional software and hardware platforms, which allowed for an increase in the value of already existing resources. Additionally, the strategies were adaptable enough to be combined with newly developed products and systems, which were readily available online. The repositories and information repositories are getting more and more appealing, and they are making use of the enormous number of data that needs to be evaluated effectively. It is possible that the process of data exploration in repositories entails the examination of data that is alluring, concealed, and in the normal course of events unknown from the vast repositories [3].

If the information depository has database management systems that may assist with the additional supply needs of information mining, using data integrity repositories may be more rational than using a physical subgroup of the information depository. It is recommended that a separate repository is maintained whenever it is feasible to do so. In common parlance, the phrases data integrity and data quality are often used synonymously with one another. On the other hand, they often share very few distinguishing characteristics with one another. The validation of the data and the maintenance of its unaltered state throughout its life cycle are two aspects of data integrity. On data, a wide variety of actions, including storing, retrieving, updating, and others, are carried out on a frequent basis. The procedures guarantee that the data will remain in the same format in which they were entered, regardless of the number of activities that are carried out. A few procedures, such as encrypting and backing up data, controlling who may access it, and validating it, help to keep data integrity intact. On the other hand, data is said to be of high quality if it is relevant and comprehensive as well as if it is acceptable for the purpose for which it was collected. According to the standards, data quality may be defined from three distinct vantage points, including that of the customer, that of the company, and that of the standards themselves [3].

In the case that appropriate data governance is not put into place, there is a good chance that data discrepancies that exist in a variety of systems that are spread out throughout an organization will not be handled. For instance, the systems that are used for sales, logistics, and customer support might, each in their own right, list the names of customers in a different manner from the other two systems. This might make the process of data integration more complicated and lead to problems with the integrity of the data. This, in turn, would have an effect on the precision of applications such as business intelligence, corporate reporting, and analytics. Additionally, there is a possibility that data inconsistencies will not be identified and corrected, which will have an additional negative influence on the accuracy

of business intelligence and analytics. If this occurs, there will be an additional negative influence on the accuracy of business intelligence and analytics. Inadequate data governance may also make it difficult to comply with regulatory norms, which may be very irritating for everyone engaged in the process. Businesses that are required to comply with the ever-increasing number of regulations regarding the privacy and protection of data, such as the General Data Protection Regulation of the European Union and the California Consumer Privacy Act, could find themselves in a precarious position as a direct result of this. It is generally important for an organization to establish both standard data formats and common data definitions as a part of its overall data governance programme. This is because of the interdependence between the two. In the end, increasing the quantity of data consistency, which is advantageous for the purposes of both business and compliance, is the outcome of implementing these standards and formats across all business systems [3].

### **3. Data governance goals and benefits**

The elimination of data silos inside an organization is one of the primary aims of data governance. Common causes of the formation of such silos include the deployment of independent transaction processing systems by distinct business units in the absence of either centralized coordination or an enterprise data architecture. The goal of the collaborative process that is data governance is to harmonize the data contained inside those systems. Participants from across all of the different business units take part in this process. A further objective of data governance is to guarantee that data is utilized appropriately. This is done for two reasons: first, to prevent the introduction of data mistakes into systems, and second, to prevent the possible abuse of sensitive information and personal data concerning consumers. This may be achieved by establishing consistent guidelines for the use of data, as well as processes that can be used to keep track of how the guidelines are being followed and ensure that they are consistently adhered to. In addition, data governance may assist in striking a balance between the practices of data gathering and the laws pertaining to privacy. Improved data quality, reduced costs associated with data management, and increased access to necessary data for data scientists, other analysts, and business users are some of the benefits that come as a result of better data governance. Other benefits include increased accuracy in analytics and enhanced regulatory compliance. By ultimately providing executives with more accurate information, data governance may ultimately aid in the improvement of company decision-making. In a perfect world, this would result in competitive advantages, more revenue, and increased profits [4].

### **4. Components of a data governance framework**

In the context of a governance programme, the policies, rules, procedures, organizational structures, and technology that are established as part of the framework for data governance are referred to collectively as “governance framework.” Additionally, it outlines things like a mission statement for the programme, its objectives, and the manner in which its success will be assessed. Furthermore, it specifies who is responsible for making decisions and who is accountable for the

different duties that will be a part of the programme. Documenting and disseminating an organization's governance structure should be done on the company's intranet for the purpose of making it immediately apparent to all parties involved how the programme will function. On the technological front, data governance software may be used to automate many parts of maintaining a governance programme. This saves a great deal of time. Even though data governance tools are not required components of the framework, having them may help with managing programmes and workflows, collaborating on the design of governance rules and process documentation, and more, in addition to supporting the building of data catalogs. Tools for master data management (MDM), data quality management, and metadata management are some examples of complementary applications that may be used with these [5].

## **5. Recommended procedures for the administration of data governance projects**

As a result of the fact that data governance often imposes restrictions on the way in which data is handled and exploited, the technique has the potential to spark controversy inside enterprises. When it comes to information technology and data management teams, one of the most common concerns is that business users would see them as the “data police” if they take the lead on data governance initiatives. This is one of the reasons why this concern is so popular. Data governance managers who have years of experience and industry consultants both advocate that programmes be business-driven, that data owners be consulted, and that the choices on standards, policies, and procedures be made by the data governance committee. This will help eliminate pushback to governance initiatives while also promoting buy-in from businesses. Training and education on data governance is a necessary component of initiatives, particularly for the purpose of acquainting business users and data analysts with rules governing the usage of data, privacy mandates, and their responsibility for contributing to the maintenance of consistent data sets. This is particularly important for the purpose of acquainting business users and data analysts with rules governing the usage of data. In order to keep in continual communication with corporate executives, business managers, and end users about the creation of a data governance programme, a variety of outreach tools, including but not limited to reports, email newsletters, seminars, and other types of events, are necessary. This exchange of information is a precondition that must be met. A second piece of writing by Farmer presents a rundown of seven recommendations for effective data governance. Two of these seven recommended practices are training and communication, and they are both included here. Some of the others include enforcing data security and privacy standards at a location that is as close to the source system as is practically practicable, implementing suitable governance policies at every level of a business, and frequently reviewing governance policies. This proximity to the source system is important because it helps ensure that sensitive information is protected [6].

## **6. Data governance challenges**

Due various areas of an organization sometimes have different perspectives on essential data entities, such as customers or goods, the first steps in attempts

to manage data may frequently be the most challenging. This is often the case because of the complexity of the situation. It is necessary to find a solution to these disparities as part of the process of data governance—for instance, by reaching a consensus on standard data definitions and formats. Because this may be a difficult and contentious endeavor, the committee in charge of data governance has to have a well-defined process in place for resolving disputes. The following are some more typical difficulties that businesses have while attempting to regulate their data.

Providing evidence of its worth to the company. It may be difficult to get approval, funding, and support for a data governance programme if there is no proof of the anticipated business advantages provided by the initiative up front. Askham said in a blog post that she published in January 2022 that corporate leaders want to know what is in it for them right from the beginning of a governance initiative. According to what she wrote, “if you can’t answer it in a manner that they are interested in and that helps them, then they’re simply not going to be interested.” To demonstrate the worth of an investment to a firm on an ongoing basis demands the establishment of quantitative indicators, especially for the enhancement of data quality. This might include the amount of data inaccuracies that are fixed on a quarterly basis as well as any revenue increases or expense reductions that arise from these fixes. Other frequent metrics for measuring data quality include accuracy and error rates in data sets, in addition to associated characteristics like the completeness and consistency of the data. Learn more about the strong relationships that exist between data governance and data quality, in addition to the many types of metrics that may also be used to illustrate the efficacy of a governance programme, by reading more about these topics.

There is a considerable risk that data inconsistencies will not be managed in the several systems that are dispersed across an organization if effective data governance is not established. This is because of the widespread nature of the systems. For example, the systems that are used for sales, logistics, and customer service may, each in their own right, show the names of customers in a manner that is distinct from the manner in which the names are displayed by the other two systems. Because of this, the process of integrating the data can become more difficult, which might subsequently cause problems with the data’s integrity. As a consequence of this, the accuracy of applications such as business intelligence, corporate reporting, and analytics will be impacted. Additionally, there is a possibility that data inconsistencies will not be recognized and remedied, which will have an additional negative influence on the accuracy of business intelligence and analytics. If this occurs, there will be an additional negative influence on the accuracy of business intelligence and analytics. This is because there is a chance that discrepancies in the data will not be identified and fixed, which is the reason why there is a problem. Inadequate data governance may also inhibit attempts to comply with regulatory rules, which is likely to be frustrating for everyone involved in the scenario. Companies that are required to comply with the ever-increasing number of regulations regarding the privacy and protection of data, such as the General Data Protection Regulation of the European Union and the California Consumer Privacy Act, could find themselves in a position that is problematic as a result of this. Because of this, companies that are required to comply with these regulations could find themselves in a position that is problematic. As part of their overall data governance programme, it is often vital for an organization to define common data definitions in addition to standard data formats.

## 7. Applications

The analytical illustration provides a trade purchasing system that makes use of the majority of the items from the year before, allowing one to make an accurate prediction of the number of products that will be required during the next time period. The authentication could check conditions such as viral, but there is a possibility that the acknowledgment and withdrawal identification may be used fraudulently. This is in contrast to the fact that viral conditions could be verified by authentication. It is put to use for a wide variety of purposes in both public and private organizations. It is common practice for businesses in the banking, insurance, medical, and buying industries to make use of data integrity in order to save costs, enhance analyses, and increase trades. Consider the insurance and banking companies that have implemented data integrity tools to assist in risk assessment and the identification of fraudulent activities. The firms could design prototypes that forecast the threats prevailing to the users in terms of credits or regarding the privileges during an accident that might be false and shall be inspected more carefully if they make use of the user-related information gathered over the course of the current period. This information was gathered over the course of the current period.

### Author details

B. Santhosh Kumar

Department of Computer Science and Engineering, Guru Nanak Institute of Technology, Hyderabad, Telangana, India

\*Address all correspondence to: [bsanthosh.csegnit@gniindia.org](mailto:bsanthosh.csegnit@gniindia.org)

### IntechOpen

---

© 2023 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 



## References

[1] Tao X, Zhang H. Research on data security governance based on artificial intelligence technology. In: 2021 International Conference on Big Data, Artificial Intelligence and Risk Management (ICBAR). Shanghai, China; 2021. pp. 102-105. DOI: 10.1109/ICBAR55169.2021.00030

[2] Harwanto IM, Hidayanto AN. Data governance maturity assessment: A case study directorate general of corrections. In: 2022 International Conference on ICT for Smart Society (ICISS). Bandung, Indonesia; 2022. pp. 1-6. DOI: 10.1109/ICISS55894.2022.9915243

[3] Hongxun T et al. Data quality assessment for online monitoring and measuring system of power quality based on big data and data provenance theory. In: 2018 IEEE 3rd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA). China. 2018. pp. 248-252. DOI: 10.1109/ICCCBDA.2018.8386521

[4] Ladley J. Data Governance: How to Design, Deploy, and Sustain an Effective Data Governance Program. Second ed United States: Academic Press Inc. 2019

[5] Nayan BR. Security and governance. In: Cloud Computing. United States: MIT Press. 2016. pp. 99-126

[6] Simon AR, Shaffer SL. Chapter 9 - Data Quality and Integrity Issues. In: The Morgan Kaufmann Series in Data Management Systems, Data Warehousing, and Business Intelligence For e-Commerce. Massachusetts, United States: Morgan Kaufmann; 2002. pp. 193-208. DOI: 10.1016/B978-155860713-2/50012-7