



**Data  
Regulations &  
Compliance**

---

Data lineage within data regulations and compliance

by: Shane James





# Data Governance & Compliance

---



# What is data governance?

The practice of overseeing the accessibility, usability, integrity, and security of the data utilized in an organization is known as data governance. It is a collection of practices created to guarantee the effective and efficient use of data as a resource for organizations. Establishing guidelines and rules that specify how data is produced, kept, accessed, and used is another aspect of data governance.

Data governance cases include:

1. Setting standards for accuracy, comprehensiveness, and consistency in data quality.
2. Creating policies for data access and use: Outlining who has access to and uses data and for what purposes.
3. Designing roles for data stewardship: Assigning management and oversight tasks and responsibilities.
4. Creating data retention policies, which specify how long information should be retained and when it should be deleted.
5. Setting up data security protocols: Creating rules and guidelines to guard against unwanted access to data.
6. Creating mechanisms for monitoring and auditing data usage and access to make sure that data governance policies are being followed

# What is the difference between compliance and governance?

---

Although compliance and governance are similar ideas, they signify different things.

The process of ensuring that an organization abides by laws, rules, standards, and policies is known as compliance. Compliance is frequently linked to specific sectors or domains, such as data protection, banking, and healthcare (HIPAA) (GDPR). Common compliance tasks include assessing an organization's compliance posture, identifying and mitigating compliance risks, and informing pertinent stakeholders of compliance status.

On the other hand, governance refers to the comprehensive set of guidelines, procedures, and procedures that a company implements to guarantee that its operations are moral, open, and responsible. Compliance falls under the better idea of governance, which also covers risk management, performance management, and strategic planning. Establishing rules and processes, ensuring they are followed, and reporting on the organization's performance and compliance status are all part of governance tasks.

# What are the risks in data governance?

---

First, there's the problem of data leakage, which refers to any time-sensitive information that leaves a company without proper authorization. This can happen when there need to be more safeguards in place or when bad actors are trying to access sensitive information.

Problems with data integrity occur when information needs to be more accurate or consistent. Because of this, conclusions and statistics may need to be corrected. Thirdly, problems with data privacy occur when a firm fails to safeguard its customers and workers' personal information effectively.

Issues with data security occur when an organization needs to take adequate precautions to prevent the loss, theft, or illegal use of its data. Data is misused when information is applied to activities aside from those for which it was initially collected, such as advertising or marketing. Data privacy and security laws necessitate that businesses adhere to specific guidelines. Not doing so may result in financial and legal consequences.

## **How is data lineage used within data governance?**


---

Data lineage is a critical component of data governance because it allows organizations to understand the quality and origin of their data. It can be used to help with data governance in a variety of ways:

1. **Data Quality:** By understanding the data's history, companies may detect poor-quality data sources and decide which data sources to trust and which to ignore.
2. **Data Compliance:** Data lineage can be used to verify compliance with various requirements, including GDPR, HIPAA, and SOX. Organizations can use data lineage information to demonstrate how data was gathered, used, and safeguarded.
3. **Data Governance Workflows:** Data lineage information can help with data governance workflows. When a business user seeks access to a specific data source, for example, data stewards can utilize data lineage information to determine whether the request is suitable and whether additional controls or rules are required to protect data security and privacy.
4. Data lineage can be used to enable automated data governance activities such as data quality checks, data lineage mapping, data lineage traceability, and data lineage impact analysis.
5. **Auditing and Monitoring:** Data lineage provides a record of data's path across various systems, which may be used to identify errors and concerns and track the effectiveness of data governance processes over time.
6. **Understanding the Business Context:** Data lineage gives context for the data and how it is utilized. It may be used to understand the business impact of data changes and make data-related choices accordingly.

Overall, data lineage assists companies in better understanding their data and making better decisions about how to use it. As a result, data-driven decision-making, risk management, and compliance are improved.

**How can data lineage be used to demonstrate compliance with various regulations such as GDPR, HIPAA, and SOX?**




Data lineage can be used to verify compliance with various requirements such as GDPR, HIPAA, and SOX by providing a clear record of how data is gathered, utilized, and secured.

Under the General Data Protection Regulation (GDPR), for example, companies must be able to establish that they have a legal basis for collecting and processing personal data. Data lineage information can be used to determine where personal data originated, how it was used, and who accessed it.

Similarly, corporations are obligated to employ protections to preserve the privacy and security of personal health information under the Health Insurance Portability and Accountability Act (HIPAA) (PHI). By providing a clear record of where PHI came from, who has access to it, and how it is being used, data lineage information can be used to establish that PHI is being managed in conformity with HIPAA regulations.

The Sarbanes-Oxley Act (SOX) requires firms to keep accurate financial records and have strong internal controls to avoid fraud. Data lineage information can be used to establish compliance with SOX regulations by providing a clear record of where financial data comes from, who has access to it, and how it is used.

Overall, data lineage assists firms in demonstrating regulatory compliance by giving a transparent and auditable record of how data is gathered, used, and safeguarded. This record can be used to demonstrate regulatory compliance to both regulators and internal auditors. It also enables enterprises to take proactive steps to address any concerns discovered, such as identifying and removing sensitive data or establishing access controls for sensitive data, before data breaches occur.



# What is a data transformation?

Data transformation is the process of converting data from one format or structure to another format or structure. Data transformation's goal is typically to prepare data for further analysis or use or to make it more compatible with other systems or applications. There are several different types of data transformations, such as:

1. **Data Mapping:** Mapping data from one format or structure to another. For example, data might be mapped from a flat-file format to a database format.
2. **Data Cleaning:** Identifying and removing inaccuracies, inconsistencies, or missing data from a dataset. This step is essential to ensure the quality of data and to remove any outliers, mistyped values, or irrelevant data
3. **Data Normalization:** This transforms data into a consistent format to be easily compared, analyzed, or aggregated.
4. **Data Aggregation:** This combines data from multiple sources into a single dataset or report. This is done to provide a holistic view of the data and to make it more useful for analysis or reporting.
5. **Data Enrichment:** Adding new data, such as demographic information or new columns, to existing data.
6. **Data Encoding:** This converts data into a different format, such as text, to numerical or categorical.
7. **Data Masking/Anonymization:** Data masking/anonymization is obscuring sensitive data in a database by replacing it with realistic but not real data. This protects individuals' privacy and prevents unauthorized access to confidential information. Data masking/anonymization can protect data in databases, files, and applications. It can also protect data in transit, such as when sent over the internet.



# How is data lineage used with data quality?

Data lineage, a critical measure for assessing data quality, is the capacity of an organization to trace the history of its data back to its sources, which is a significant aspect in establishing the correctness and completeness of that data. The following are some examples of data lineage applications that improve data quality:

- Data can be traced back to its source, and data lineage helps identify issues in the data flow. Therefore, identifying and resolving data quality concerns early is a significant benefit to organizations.
- By tracing a piece of data's history, or "data lineage," it is possible to examine the impact of poor data quality on succeeding systems and processes. As a result, an organization may prioritize its efforts to fix the problem by using data lineage information to identify which systems and processes are affected by a data source's subpar output.
- To demonstrate compliance with data quality laws and standards, data lineage information can be utilized to provide an auditable record of data quality activities.
- Data lineage offers a mechanism to trace the development of data quality through time. Businesses may track the quality of their data as it moves through their systems and procedures using data lineage information. They can look for patterns that will help them better maintain high standards for data in the future.

A few examples of automated data governance processes that may be enabled by data lineage include data quality checks, data lineage mapping, traceability, and data lineage impact analysis. This enhances overall data quality by allowing organizations to evaluate and fix concerns with data quality regularly.

By utilizing data lineage to comprehend their data's origins, flows, and quality, organizations may improve the effectiveness of their data quality programs, resulting in improved decision-making and compliance.

## How is data used within provenance tracking?

Following the sources and movements of data is referred to as provenance tracking. This technique may be utilized to understand better the reliability and validity of the data being tracked. Because it offers a transparent record of the beginnings and destinations of data as it travels through various systems and processes, data lineage is an essential piece of equipment for provenance tracking.

**Traceability:** Data lineage records how data has moved through various systems and processes. This record may be used to track the data sources and identify any issues that may have happened along the way. 2. **Data provenance:** Data provenance provides a clear record of how data has migrated through various systems and processes. Establishing the data's dependability and validity may be accomplished with the help of this.

**Data Quality:** The quality of the data Organizations can detect the sources of data that are of low quality and make judgments on which data sources to trust and which data sources to throw away if they are aware of the data's lineage.

**Data Governance Processes:** Information about the data lineage may be utilized to assist data governance workflows such as data quality checks, mapping of the data lineage, tracing the data lineage, and impact analysis of the data lineage. Continuously identifying and resolving data compliance concerns is made more accessible for enterprises as a result of this functionality.

**Auditing and Monitoring:** Data lineage offers a record of the route data takes through multiple systems. This record may be used to assist in the identification of faults and concerns, as well as to assess the efficacy of data governance procedures over time. In addition, the quality of the data may be monitored over time using this, and choices can be made on the genuineness and dependability of the data based on the findings.

**Compliance and Regulations:** Data lineage can be used to demonstrate compliance with various regulations such as GDPR, HIPAA, and SOX by providing a transparent and auditable record of how data is collected, used, and protected. This is possible because data lineage can be traced back to the original source of the data.

Organizations can guarantee that they are utilizing data of high quality and that they can trust by identifying the origins of the data and its movements. This helps to improve the accuracy and efficacy of the companies' data-driven decision-making. In addition, the data lineage helps understand the various biases and difficulties associated with the data, which contribute to the high quality and reliability of the data.

## What is GDPR?

---

The General Data Protection Regulation (GDPR) sets out several key principles and rights for individuals regarding their data. Some of the primary directives include:

1. Lawfulness, fairness, and transparency: Personal data must be processed fairly, transparent, and lawfully.
2. Purpose limitation: Personal data must be collected for specified, explicit, and legitimate purposes and not further processed in a way that is incompatible with those purposes.
3. Data minimization: Personal data should be adequate, relevant, and limited to what is necessary for the purposes for which it is processed.
4. Accuracy: Personal data must be accurate and, where necessary, kept up to date.
5. Storage limitation: Personal data must only be kept for as long as is necessary for the purposes for which it is processed.
6. Integrity and confidentiality: Personal data must be protected by appropriate technical and organizational measures against unauthorized or unlawful processing, accidental loss, destruction, or damage.
7. Accountability: Organizations that process personal data must be able to demonstrate compliance with the GDPR.
8. Right to access, right to be forgotten: Individuals can access the personal data that organizations hold about them and have their data deleted in certain circumstances.
9. Data breach notification: Organizations must report certain types of data breaches to the relevant authorities and, in some cases, to the individuals affected by the breaches.
10. Data protection by design and by default: Organizations must implement appropriate technical and organizational measures to ensure that personal data is protected by default.

11. These are some of the core principles of GDPR, And also GDPR sets out specific requirements for obtaining valid consent, the appointment of a data protection officer, and additional obligations to protect sensitive data.

## How do organizations demonstrate compliance with GDPR?

---

Organizations must show GDPR compliance by establishing adequate technological and organizational measures to protect personal data and being able to produce evidence of such compliance upon request.

Organizations may demonstrate GDPR compliance in a variety of methods, including:

1. Conducting frequent internal audits: Organizations should regularly assess their data processing operations to guarantee compliance with the GDPR.
2. Appointing a Data Protection Officer (DPO): Organizations that conduct large-scale regular and systematic surveillance of data subjects or that process sensitive personal data on a broad scale must employ a DPO to guarantee GDPR compliance.
3. Organizations must retain records of their data processing operations, including the processing aims, the data and data subjects categories, and the security measures.
4. Privacy notifications: Organizations must provide individuals with clear and comprehensive privacy notices alerting them of their rights and how their data will be used.

5. Conducting Data Protection Impact Assessments (DPIA): Organizations must conduct a DPIA where there is a severe risk to natural people's rights and freedoms.

6. Notifying data breaches: Organizations must notify the competent supervisory authority of certain data breaches within 72 hours of becoming aware of the breach.

7. Employee training: Organizations should give their workers data protection and GDPR training to ensure they understand their duties and can detect and report any breaches.

8. Keeping consent records: Organizations must preserve records of when and how they got consent from individuals to process their data.

These are a few examples of how firms might show GDPR compliance. Because GDPR provides flexibility in demonstrating compliance, organizations should take a risk-based approach and adjust their procedures to their requirements.


Organizations may use data lineage to comply with GDPR in several different ways:

1. Organizations may more easily adopt the necessary protections to secure the data by identifying where personal data is kept, processed, and accessed in their systems using data lineage.
2. Organizations may identify and track personal data flows and spot any unwanted access or use of personal data by using data lineage.
3. Data lineage may help companies identify and record the legitimate grounds for personal processing data, which is necessary for GDPR compliance.
4. A clear and comprehensive history of personal data, which is a requirement of GDPR, is provided by data lineage, which can assist enterprises in demonstrating compliance with the regulation.
5. Organizations can comply with the GDPR's "Right to be Forgotten" rules by identifying personal data that is no longer needed and deleting it. This is made possible via data lineage.
6. Organizations can follow data subjects' requests for access to their personal data, which is required by GDPR, using data lineage.

By giving enterprises visibility into how personal data is used and safeguarded throughout its life cycle, data lineage can assist organizations in complying with the GDPR's data protection by design and default principle.

## What is PHI?

The acronym PHI refers to "protected health information." It means that any information linked to an individual that pertains to their health condition, the provision of healthcare, or payment for healthcare is considered PHI information.



Individuals' health records, test results, diagnoses, treatment plans, and insurance information are all examples of PHI. Additional criteria include treatment plans and insurance information, written records, spoken communications, or electronic data, which might all be different forms of protected health information (PHI).

In the United States, Protected Health Information (PHI) is safeguarded by the Health Insurance Portability and Accountability Act (HIPAA), which sets the standards for ensuring the confidentiality and safety of PHI records. The law requires healthcare providers and insurers to implement safeguards to protect the confidentiality, integrity, and availability of protected health information (PHI). Additionally, the law restricts PHI's use and disclosure to the minimum amount required to accomplish its intended purpose.

To maintain compliance with the HIPAA regulations, healthcare providers, insurers, and other organizations that handle PHI must implement administrative, physical, and technical safeguards to protect PHI from being accessed, used, disclosed, altered, or destroyed without authorization.

Data lineage can give visibility and traceability into the movement of PHI throughout the company, assisting healthcare providers, insurers, and other businesses that handle PHI in maintaining compliance with HIPAA laws.

Tracking a piece of data's beginnings and journey via numerous systems and applications is known as data lineage. Data lineage in the context of PHI can offer a comprehensive history of PHI data, revealing its origin, previous users, processors, and storage locations.

In a few different ways, data lineage may assist firms in keeping up their HIPAA compliance:





1. Organizations may use data lineage to determine where PHI is processed, accessed, and stored inside their systems, simplifying implementing security measures.
2. Organizations can identify and monitor PHI data flows with the use of data lineage, and they can also spot any unwanted access or use of PHI.
3. Data lineage, which provides a comprehensive history of PHI data access, including who accessed it, when, and where, can assist companies in identifying and correcting any PHI data breaches.
4. By providing a comprehensive and transparent history of PHI data, which is a requirement of HIPAA, data lineage can assist companies in demonstrating compliance with HIPAA laws.
5. To maintain compliance with data retention laws, data lineage can help companies identify PHI that is no longer required and can be removed.

Data mapping, discovery, profiling, and governance tools and technologies are just a few of the tools and technologies that may be used to execute data lineage.

## What is a Data Protection Impact Assessment (DPIA)?

A Data Protection Impact Assessment (DPIA) is a process organization must conduct, per the General Data Protection Regulation (GDPR) and other data protection laws, to identify and mitigate the risks of personal processing data.

A DPIA is a risk management process designed to help organizations identify and assess the potential risks to the rights and freedoms of individuals resulting from data processing activities and implement measures to mitigate those risks. DPIA should be conducted before any processing of personal data that is likely to result in a high risk to the rights and freedoms of individuals.

DPIA generally consists of the following steps:

1. Identify the need for a DPIA and the type of data processing activities that will be included in the assessment.
2. Identify and assess the risks associated with the data processing activities.
3. Identify the measures that can be implemented to mitigate or eliminate the identified risks.
4. Consult with the relevant stakeholders, including data protection authorities and other interested parties.
5. Review the effectiveness of the measures implemented.

DPIA is a continuous process, and organizations must monitor and review the DPIA regularly to ensure it remains current and effective. By conducting a DPIA, organizations can demonstrate that they have taken a proactive and risk-based approach to data protection and that they have considered the potential impacts of their data processing activities on individuals' rights and freedoms.

## **What is the potential evolution of data lineage?**

Data lineage is a relatively new field and is likely to evolve in several ways in the future. However, here are a few potential developments that may shape the future of data lineage:

1. Greater automation: As organizations increasingly rely on automation to manage and analyze their data, data lineage systems will likely become more automated. This may include the ability to automatically discover, track and map data lineage without human intervention and automate the compliance checks on lineage.
2. Greater integration: Data lineage will likely become more integrated with other data management and analytics tools, such as data catalogs, data quality tools, and data governance platforms. This will allow organizations to more efficiently view and analyze data lineage information in the context of other data management activities.
3. Increased real-time tracking: With the growing use of real-time data processing, data lineage will likely evolve to include real-time monitoring and visualization of data movements and transformations. This will help organizations quickly identify and address data quality and compliance issues as they arise.
4. Greater collaboration: Data lineage will likely become more of a collaborative effort, involving not only IT and data professionals but also business stakeholders, data stewards, and data governance teams. This will help organizations govern and use their data more effectively.
5. Greater use of AI/ML: The use of AI and Machine Learning will grow in the field of data lineage; it can be used for automating the discovery, mapping, and visualizing lineage, identifying lineage issues, and also can be used to predict potential lineage issues.

Overall, the future of data lineage will likely be shaped by the growing importance of data governance and the need for organizations to understand better, trust and use their data. This will likely lead to new technologies.

## **Regulations in the insurance industry**

---

1. In the United States, the Health Insurance Portability and Accountability Act (HIPAA) sets standards for protecting the privacy and security of personal health information.
2. The General Data Protection Regulation (GDPR) in the European Union (EU) regulates the handling of personal data, including how it is collected, stored, and shared.
3. The EU's Insurance Distribution Directive (IDD) regulates how insurance products are sold, including rules for disclosing information to customers.
4. In Canada, the Personal Information Protection and Electronic Documents Act (PIPEDA) regulates how personal information is handled by organizations in the private sector, including insurance companies.

Some additional regulations in insurance are below:

1. The Gramm-Leach-Bliley Act (GLBA) requires financial institutions, including insurance companies, to protect the privacy of personal financial information.
2. The Fair Credit Reporting Act (FCRA) regulates how insurance companies and other organizations collect, use, and share consumer credit information.
3. The Children's Online Privacy Protection Act (COPPA) regulates the collection of personal information from children to 13 by insurance companies and other organizations.
4. The Affordable Care Act (ACA) includes provisions related to the privacy and security of personal health information, including rules for how insurance companies must handle this information.
5. The New York Department of Financial Services (NYDFS) Cybersecurity Regulation (23 NYCRR 500) applies to insurance companies and other financial institutions operating in New York State. It requires them to implement a cybersecurity program to protect sensitive data.

# Regulations in the healthcare industry

The healthcare industry is subject to a wide range of regulations, both at the federal and state level in the United States. Some of the most important regulations include:

1. The Health Insurance Portability and Accountability Act (HIPAA) sets standards for protecting the privacy and security of personal health information. HIPAA applies to healthcare providers, health plans, and healthcare clearinghouses.
2. The Affordable Care Act (ACA) includes provisions related to healthcare reform, including rules for the individual mandate, coverage of pre-existing conditions, and Medicaid expansion.
3. The Clinical Laboratory Improvement Amendments (CLIA) sets standards for laboratory testing and quality control, and applies to all facilities that test human specimens for health assessment or to diagnose, prevent, or treat disease.
4. The Food, Drug, and Cosmetic Act (FD&C) regulate the development, testing, and marketing of drugs and medical devices, including pre-market clearance or approval by the FDA.
5. The Center for Medicare and Medicaid Services (CMS) sets standards for quality and safety in healthcare facilities participating in Medicare and Medicaid programs.
6. The Occupational Safety and Health Administration (OSHA) sets standards for workplace safety, including in healthcare settings.
7. The Drug Enforcement Administration (DEA) regulates controlled substances and the handling of drugs by healthcare providers.
8. The Clinical Laboratory Improvement Amendments (CLIA) is a federal regulation in the United States that sets laboratory testing and quality control standards. It applies to all facilities that test human specimens for health assessment or to diagnose, prevent, or treat disease.

Data lineage can be used to help with the regulation of the healthcare industry in several ways:

1. HIPAA: Data lineage can help healthcare organizations comply with HIPAA regulations by tracking the flow of personal health information (PHI) throughout the organization. This can help organizations identify where PHI is being stored, who has access to it, and how it is used.
2. CLIA: Data lineage can help healthcare organizations comply with CLIA regulations by tracking laboratory test results throughout the organization. This can help organizations identify where test results are being stored, who has access to them, and how they are used.
3. FD&C: Data lineage can help healthcare organizations comply with FD&C regulations by tracking data flow related to developing, testing, and marketing drugs and medical devices throughout the organization. This can help organizations identify where data is being stored, who has access to it, and how it is used.
4. CMS: Data lineage can help healthcare organizations comply with CMS regulations by tracking the flow of data related to quality and safety throughout the organization. This can help organizations identify where data is being stored, who has access to it, and how it is used.

Data lineage is a powerful tool that can help healthcare organizations navigate the complex web of regulations that apply to the industry by providing a clear and comprehensive view of how data flows through the organization and identifying potential areas of non-compliance.

## Regulations in the finance industry

Many financial regulations in the United States and Europe, Middle East and Africa (EMEA) apply to various aspects of the financial industry. Some examples of these regulations include:

## United States:

1. The Dodd-Frank Wall Street Reform and Consumer Protection Act (Dodd-Frank) is a comprehensive financial reform law enacted in response to the 2008 financial crisis. It includes consumer protection, banking supervision, and financial stability provisions.
2. The Sarbanes-Oxley Act (SOX) is a law that was enacted in 2002 in response to financial scandals at companies like Enron and WorldCom. It includes financial reporting, internal controls, and corporate governance provisions.
3. The Bank Secrecy Act (BSA) is a law that requires financial institutions to help the government fight money laundering and terrorist financing. It includes provisions related to recordkeeping, reporting suspicious activity, and customer identification.
4. The Consumer Financial Protection Act (CFPA) is a law that gives the Consumer Financial Protection Bureau (CFPB) broad authority to regulate consumer financial products and services.

## EMEA:

1. Basel III is an international regulatory framework that sets out minimum standards for bank capital adequacy, stress testing, and liquidity.
2. The Markets in Financial Instruments Directive (MiFID) is an EU regulation that governs the provision of investment services and the operation of financial markets.
3. The General Data Protection Regulation (GDPR) is an EU regulation that regulates the handling of personal data, including how it is collected, stored, and shared.
4. The Anti-Money Laundering Directive (AMLD) is an EU directive that requires financial institutions to take measures to prevent money laundering and terrorist financing.

Data lineage can be used to help organizations comply with financial regulations in several ways:

1. Dodd-Frank, SOX, and BSA: Data lineage can help organizations comply with these regulations by tracking the flow of financial data throughout the organization. This can help organizations identify where data is being stored, who has access to it, and how it is used. Data lineage can also be used to identify and document critical controls, such as data validation and reconciliation processes, that are in place to ensure the accuracy and integrity of financial data.
2. Basel III: Data lineage can help organizations comply with Basel III regulations by tracking the flow of data related to capital adequacy, stress testing, and liquidity throughout the organization. This can help organizations identify where data is being stored, who has access to it, and how it is used.
3. MiFID, GDPR, and AMLD: Data lineage can help organizations comply with these regulations by tracking personal and financial data flow throughout the organization. This can help organizations identify where data is being stored, who has access to it, and how it is used. It can also help organizations understand where sensitive data is stored, how it is used, and how it is shared with third parties, which is essential for compliance with regulations such as GDPR and AMLD.

Data lineage can be a powerful tool for organizations operating in the financial industry, providing a clear and comprehensive view of how data flows through the organization and identifying potential areas of non-compliance with regulations.