datacamp

Data Governance Fundamentals Cheat Sheet

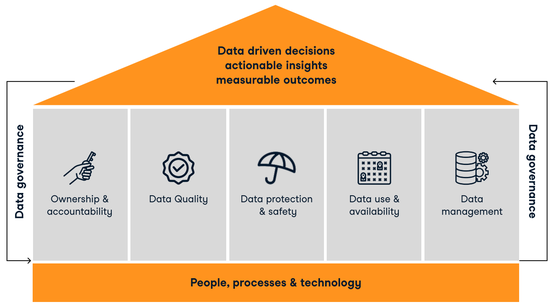
Learn online at www.DataCamp.com

> Definitions

Data governance: is a set of principles and processes for data collection, management, and use. The goal is to ensure that data is accurate, consistent, and available for use, while protecting data privacy and security.

A data governance framework: Is a set of policies, procedures, and standards that implements data governance for an organization. That is, where data governance describes what the organization needs to do, the data governance framework describes how to do it.

> The pillars of data governance



Ownership and Accountability

In order to ensure that there are no gaps in data governance, and that work is not duplicated, a clear structure of accountability needs to be defined.

- Every data asset should have an owner.
- For each data governance task, the person or role responsible for working on that task should be clear.
- For each data governance task, the people or roles accountable for the completion of the task should be clear.

Data quality

Data quality means whether or not data is fit for purpose. There are several factors that affect data quality.

- **Accuracy:** Is the data free from errors? Does the data capture the thing it is meant to represent?
- **Completeness:** Is all the data that is needed present? Have missing values been dealt with appropriately?
- **Timeliness:** Is the data available when it is needed?
- **Consistency:** Is the data the same, regardless of where it is accessed from?
- **Integrity:** Can you guarantee that the data has not been corrupted or tampered with?

Data Protection & Safety

Data protection and safety refers to measures to prevent misuse of data. There are several aspects to it.

- Preventing unauthorized access of data through access management.
- Preventing personal or commercially sensitive data being leaked publicly through data masking and encryption.
- Preventing data being deleted or corrupted through disaster recovery.

Data Use & Availability

Data use and availability refers to users being able to consistently access the data when they need it.

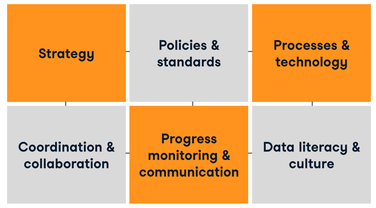
- Can legitimate users get access to the data assets they need?
- Are the data assets consistently available when they are needed?
- Are the data assets in a state that they can be used?

Data Management

Are the processes and technologies in place to ensure other aspects of data governance?

- Are the data assets stored securely?
- Are the data assets kept up to date?
- Are the data assets protected from problems?

> Data Governance Framework Components



- **Strategy:** Strategy defines how data should be used, treated, and managed safely, efficiently, and effectively to solve business problems and meet business goals to ensure that data is consistently recognized and leveraged as a valuable asset.
- **Policies and Standards:** Policies are documents of data management principles that outline decision-making rights, goals, expectations, and responsibilities. Standards are guidelines of best practices to comply with the policies.
- **Processes and Technology:** Processes involving data should describe procedures for monitoring data quality, handling issues, permissions for sharing data, managing metadata, and master data management. The technology needed to implement, monitor, and maintain these processes also needs to be described.
- **Coordination and Collaboration:** Data governance is a cross-team effort, and the responsibilities of each role must be documented, as well as the processes for how each role interacts with the others.
- **Progress Monitoring and Communication:** The progress of data governance at the organization must be monitored using metrics for data quality, risk exposure, policy compliance, and return on investment.
- **Data Literacy and Culture:** Workforces should be able to understand the importance of data governance, as well as how to get value from data assets. Good data governance requires a culture in which data governance is respected and encouraged.

> Data Governance Framework Principles

- **Integrity:** Data stakeholders should be honest and transparent with each other to ensure the success of the data governance program, as well as promoting a culture of trust, teamwork, and collaboration.
- **Ownership and Accountability:** The responsibilities of each data governance role must be clearly defined in order to prevent gaps in ownership or duplication of work.
- **Standardization and Consistency:** Data definitions need to be standardized to make it easier to use data across multiple teams and projects. Consistency ensures that processes are repeatable throughout the organization and over time.
- **Change Management:** The impact of new data governance policies or processes against existing projects should be considered, and the framework needs to be robust against changing business needs and new employees.
- **Risk Management and Compliance:** The data governance framework should comply with any relevant laws and regulatory frameworks. This includes having auditable processes, and controls to ensure compliance.
- **Strategy Alignment:** In order to ensure the continued existence of a data governance framework, it needs to be clear how that framework supports business goals and drives value.

> Data Governance Roles

Role	Description	Governance responsibilities
Executive Sponsor	A senior employee acting as a conduit between the C-suite and the data governance lead or council.	• Coordinate data governance activities across teams. • Ensure alignment between corporate goals and the data governance program.
Data Governance Lead	Usually the CIO, CTO, or CDAO. Oversees the data governance program.	• Develop and implement a data governance framework. • Communicate with stakeholders.
Data Governance Council	Group of individuals setting the strategic direction for the data governance program.	• Define what the data governance program needs to accomplish by what date. • Prioritize data governance initiatives.
Data Owner/Admin	Someone with authority to make decisions about a dataset.	• Define standards for data use, including who has access. • Define data quality properties such as accuracy, completeness, reliability, relevance, and timeliness.
Data Steward	Provides enforcement of the rules set by the data owner.	• Implement the data use and data quality rules set by the data owner. • Understand how the data is used by the organization.
Data Custodian	Manages and protects the data assets. Usually from an IT team.	• Manage and monitor data access and usage. • Provide backup and recovery of data. • Respond to data breaches.
Data Stakeholder	Anyone affected by data governance decisions.	• Provide feedback to the Data Governance Lead or Council
Data User	Someone who makes use of the data assets. Can include employees or external users.	• Generate insights or value from the data

> Common Regulatory Frameworks

Regulation	Region	Description	Who this applies to	Key highlights
GDPR	European Union	Designed to provide greater control over personal data for EU residents.	Businesses with customers in EU	• Need to obtain explicit consent from individuals before collecting and processing their personal data. • Data breaches must be reported within 72 hours.
CCPA	California state	Designed to provide greater control over personal data for California residents.	Businesses with customers in California that have gross revenues over US$25M	• Consumers have the right to know what information is collected about them, the right to have that information deleted, and the right to opt out of the sale of their personal data.
NY SHIELD	New York state	Designed to provide greater control over personal data for New York residents.	Businesses with customers in NY	• Provides broader definitions of "private information" and "breach" than federal law. • Businesses must implement a data security program including employee training, vendor contracts, risk assessments, and timely data disposal.
PIPL	China	Designed to provide greater control over personal data for China residents.	Businesses with customers in China	• Similar to GDPR, but for China. • Has stricter rules around data storage and international transfer.
Sarbanes –Oxley	United States	Designed to protect investors by improving the accuracy and reliability of disclosures by publicly traded companies	Publicly traded companies in USA	• Financial statements must be checked by independent auditors. • Controls and procedures must be in place to ensure financial statements are in line with Generally Accepted Accounting Principles (GAAP).
CCAR	United States	Designed to ensure resiliency of large banks against severe economic situations	Banks and bank holding companies with US$50B assets	• Banks must submit a capital plan to regulators. • Capital reserve data and forecasts under economic scenarios are reviewed by regulators.
HIPAA	United States	Designed to protect individuals personal health information	Organizations dealing with health data	• Sensitive information cannot be shared without patient knowledge of consent • Patients must be educated on their data privacy rights, and given access to their medical records