What is Data Governance?

Table of Contents

Sources

Data Governance The Definitive Guide (ISBN: 9781492063490)

DAMA DM-BOK (ISBN: 9781634622349)

What is Data Governance?

A data management function to ensure the quality, integrity, security, and usability of data collected by an organization.

What is Data Governance?

This is done by ensuring that:

- the data is available to the right people
- the data can be used for its intended purpose
- the data conforms to regulations and standards
- the data is correct, up to date and consistent
- the data is accessed only by permitted users in permitted ways
- all accesses, including changes, are logged

What is Data Governance?

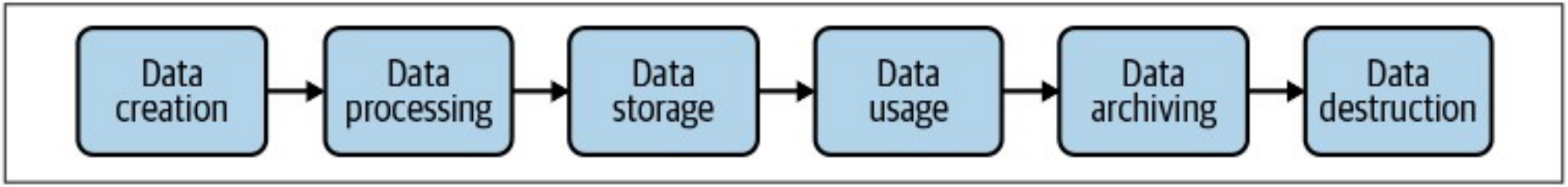
- No two organizations are exactly alike, and each organization should adopt a governance model that supports its business strategy and organizational culture.
- However, creating a data governance program might lead an organization to think differently about its data, to become data-centric, by managing its data as a corporate asset rather than a by-product of business processes.

What is Data Governance?

- A Data Governance Program is an ongoing process, not a project with a defined end. This requires organizational commitment.
- The activities related to the Data Governance program must be embedded in the day-to-day practices of developing software, analyzing data, and managing risks.

The Data Life Cycle

When establishing a data governance policy in an organization, processes pertaining to the whole *data life cycle* must be considered:



Phases of a data life cycle (DG TDG p. 87)

A Data Governance Framework

Processes to be considered are:

1. Data discovery and assessment
2. Data classification and organization
3. Data cataloging and metadata management
4. Data quality management
5. Data access management
6. Auditing
7. Data protection

1. Data discovery and assessment

As the quantity of data available to an organization grows, there is an increased risk of *ungoverned migration of data assets*. This is the potential loss of knowledge of what data assets are actually available.

The data discovery and assessment process is used to identify and organize data assets, to trace and record the data origin and lineage, transformations and metadata.

2. Data classification and organization

This process can help categorize the data, and also infer whether the data contains sensitive information so that it can be classified by level of sensitivity - personal, private, confidential, etc.

3. Data cataloging and metadata management

During this process a *data catalog* is created which contains metadata and information about sensitivity levels.

Handled correctly, the data catalog can be the way for data consumers to find the data they need for a certain task.

4. Data quality management

This process provides a means of documenting the expectations of data quality, as well as the tools and techniques for ensuring that the data quality is as expected.

Examples of this is creating controls for validation, enabling quality monitoring and reporting, and tracking data incidents.

5. Data access management

By defining identities, groups and roles, and assigning these to users based on their access rights, ensures that only authorized individuals and systems have access to certain data assets.

6. Auditing

Regular audits, checking for quality, access and transformation history, ensures that issues of data security can be identified early.

6. Auditing - an example

An example of finding a security issue through regular auditing could be found in the news recently. An employee of the district court in Malmö has been detained, suspected of accessing unathorized data.[1]

The accesses were discovered during a routine check of the access logs.

1. Svdsvenskan. 2024-09-17

6. Auditing - an example

– Det var vi själva som upptäckte det vid våra rutinkontroller och gjorde en polisanmälan. Vi kollar regelmässigt i våra loggar över vad våra medarbetare varit inne i. Då hittade vi något som vi misstänkte var fel, säger Fredrik Lassen, kommunikatör på Malmö tingsrätt.

Screenshot of sydsvenskan.se 2024-09-17

> We discovered it ourselves during a routine check and filed a police report. We continously check our logs of what our employees have accessed, and we found something we suspected was wrong, says Fredrik Lassen, communicator at Malmö district court.

7. Data protection

In the event of a data leak, additional layers of protection - encryption at rest and in flight, data masking and permanent deletion - are all tools to ensure that leaked data can’t be read.

Business drivers

The most common driver for data governance is regulatory compliance. The explosion in machine learning and data science has created an additional driving force.

The two main focus areas behind implementing a data governance program is

- **Reducing Risk**: data security and privacy
- **Improving Processes**: regulatory compliance, data quality improvement, metadata management