EXPLORE AI ACADEMY

## Data sources and access

# An introduction to data governance, legislation, and ethics

# Data governance, legislation, and ethics

## Why do we care?

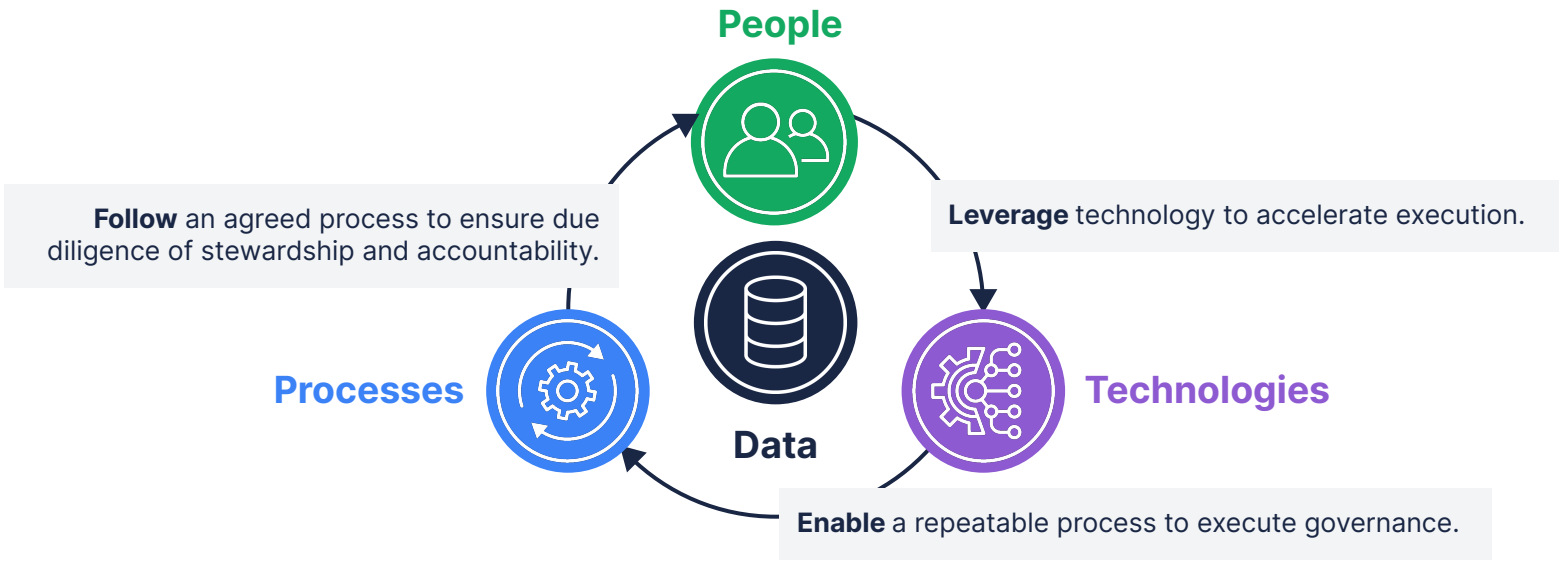
These principles ensure that individuals and organisations:

- Are **transparent** and **accountable** for how they manage their and others' data.
- Make data-driven decisions based on **relevant**, **accurate**, **quality**, **usable**, and **trusted** data.

Organisations have **legal** and **ethical responsibilities** to ensure that personal data are protected.

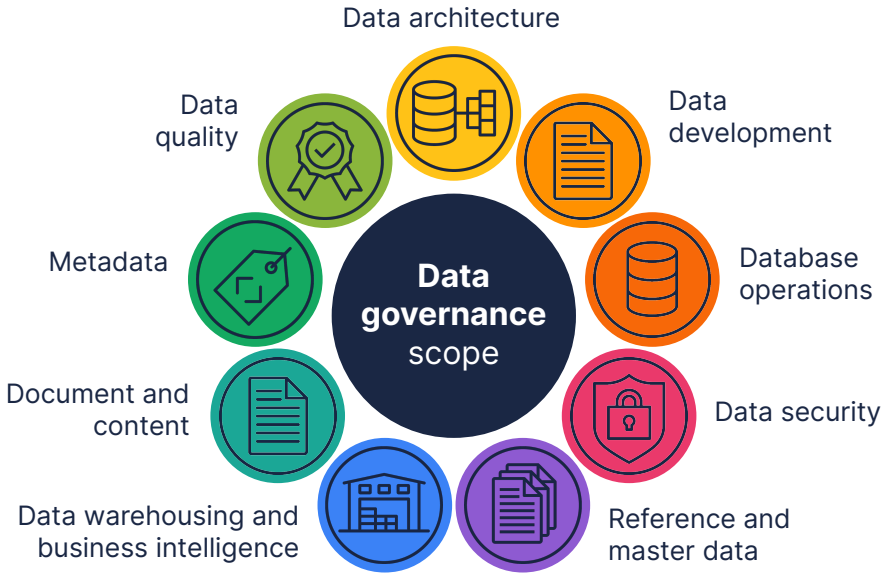
# Data governance

> The people, processes, and technologies used to **manage**, **protect**, and **use data**. It ensures **consistent** and **trustworthy data** that organisations can leverage as an organisational asset.



# Data governance

Data governance means different things to different people because the scope is broad.




However, it can be summarised as:

A collection of **data management principles and practices** that help a company manage internal and external data flows.

The Data Management Body of Knowledge (DAMA-DMBOK2)

# Data governance frameworks

> A **data governance framework** ensures that data governance is actionable by directing how data will be managed.

## Rules

**and rules of engagement**

The policies, requirements, standards, and controls, and the rules that govern how different people will enforce it.

## People

**and organisational bodies**

The stakeholders and stewards that make and enforce the rules.

## Processes

Governs the data while creating value, managing complexity and cost, and ensuring compliance.

The specifics of this framework are particular to an organisation. In general, the framework has three focus areas.

# Data governance frameworks

> The **Data Governance Institute (DGI)** data governance framework outlines ten components that address the **why-what-who-how** of data governance.

## Rules

**and rules of engagement**

1. Missions and vision
2. Goals, governance metrics and success measures, and funding strategies
3. Data rules and definitions
4. Decision rights
5. Accountabilities
6. Controls

## People

**and organisational bodies**

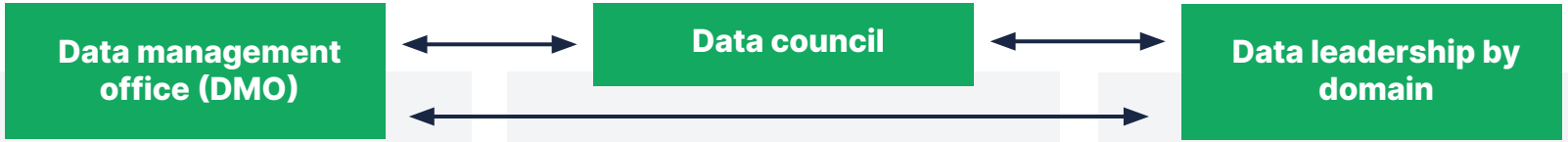
7. Data stakeholders
8. Data (governance or management) office
9. Data stewards

## Processes

10. Proactive, reactive, and ongoing data governance processes.

# Data governance frameworks

> The McKinsey framework places **people** at the foundation of **effective data governance**.




**Defines** the policies and standards of the data governance framework, and:

- Empowers data leaders through tools and training.
- Ensures **coordination** and consistency across the data lifecycle.
- Provides targeted **support** on issues.

**Owns** the data governance principles and decides the strategy, and:

- Defines domains, proposes DMO structure, and assigns leaders.
- **Manages compliance** to standards and policies and reviews progress.
- Resolves and/or escalates issues.
- Chaired by the head of the DMO.

**Owns** and **manages** the **data,** including data quality, models, etc., and:

- Sets and executes strategies of the specific domain.
- Understands and manages the needs of data consumers.
- Participates in the data council.

Data sources and access

# Data governance

Regardless of the specific framework, an **effective framework provides**:

01. Better **decision-making** on issues related to data and data management.

02. Reduced **operational friction** and increased **effectiveness**.

03. Established **best data practices** and **processes** and the **training** thereof.

04. Process and practice **transparency** and **regulatory compliance**.

# Absence of data governance

> In the **absence** of data governance, there could be a **lack** of **clear ownership**, **accountability**, and **oversight** of data.

**This leads to:**

## Data quality problems

Without **clear ownership and oversight**, different stakeholders in the organisation may unknowingly, for example, **create duplicate data** for the same entity.

Duplicate data could lead to **inconsistencies** and **inaccuracies** in the data, making it **untrustworthy**.

## Data security breaches

Without **clear roles and responsibilities**, an organisation's employees, for example, may **not know** which data is **sensitive** or how to protect it.

This could lead to **sensitive data being shared** with unauthorised parties or **stored in unsecured locations**.

# Absence of data governance

## Compliance violations

Without **clear policies and procedures** for handling data, an organisation **may not be aware** of their **obligations** under laws such as the GDPR (General Data Protection Regulation).

This could lead to an organisation **collecting, storing, and sharing personal data** without obtaining the required **consent**.

It could also lead to compliance violations under laws such as the Payment Card Industry Data Security Standards (PCI DSS) if the organisation **fails to maintain the integrity and confidentiality** of financial information.

## A summary of the consequences:

- It could be **difficult to leverage data effectively** for data-driven decision-making and business operations.
- An organisation may struggle to have a comprehensive understanding of their data and may **not be able to use their data to drive business growth**.
- If the organisation fails to comply (even unknowingly) it could result in **penalties** and **reputational damage**.

# Data legislation

> The **rules** and **regulations** that govern the collection, storage, and use of data on a legal and political level. Data legislation is mostly about data privacy and information privacy.

**Data legislation** protects **individuals** and **organisations** from the misuse of data and ensures that personal data are handled responsibly and transparently..

In practice, **data governance** and **data legislation** work together to ensure that organisations **comply** with laws around **data privacy**, **data security**, and **sensitive data**.

**Data privacy:** protecting personal information of individuals from being collected, used, and shared with others without consent.

**Data security:** the practice of protecting the availability, integrity, and confidentiality of organisational data from unauthorised access, use, modification, disclosure, and destruction.

**Sensitive data:** data that require additional protection due to the sensitive nature thereof; this includes government and personal data, such as financial and health data.

# Data legislation

**Data privacy** is the practical measures and principles of data legislation that individuals and organisations need to implement and follow.

Data privacy laws include:

**GDPR** — **General Data Protection Regulation** in the European Union (EU)

**CCPA** — **California Consumer Privacy Act** in the United States

**POPIA** — **Protection of Personal Information Act** in South Africa

Data is subject to the laws of the location where the data is collected and processed – this is called **data sovereignty**.

Only **15%** of all countries have no data legislation as of 2022.

Due to the worldwide increase in internet usage and data generated by humans and machines, **data legislation** has become an **integral part of society**.

# Data legislation

> Many data privacy acts are modelled after the **GDPR** (General Data Protection Regulation), but the GDPR has the **widest reach**.

The **GDPR** sets out **seven key principles** for processing personal data.

**1. Lawfulness, fairness, and transparency**

Obtain data in a lawful manner, inform the individual, and keep your word.

**2. Purpose limitation**

Be specific on what the data will be used for.

**3. Data minimisation**

Collect only the required data.

**4. Accuracy**

Data must be accurate and kept up to date, as necessary.

**5. Storage limitation**

Only keep data for the necessary limited time period, and then destroy it.

**6. Integrity and confidentiality**

Ensure the data is protected against unauthorised and unlawful access.

**7. Accountability**

Record and be able to prove compliance to the rules and regulations.

# Data ethics

> The **moral obligations** of collecting, storing, and using data.

Although we have **data legislation** to guide what we can and cannot do with data, there are still grey areas that individuals and organisations use to their advantage and often the disadvantage of others.

This is where **data ethics** comes into play.

Just as we have ethics to guide our behaviour when it comes to other people, money, etc., we must **keep each other accountable** for how we use data, models, and insights.

Data ethics compels us to ask:

**"Are we doing the right thing?"**

# Data ethics

> Some of the ways we can **keep each other accountable** when collecting, storing, and using data:

Take the **responsibility** to use, collect, and process data in a way that is **ethical** and **respects** the **interests** of individuals.

**Transparency** on how data are being collected and used for individuals to make **informed decisions** about sharing their data.

Personal data should be collected, stored, and used in a way that **respects an individual's privacy**.

Ensure data are **protected** from access and use by **unauthorised** individuals and systems.

Individuals should give **informed consent**, **understanding** how their data will be **used,** and have the opportunity to **opt out** of data collection.

**Fairness** in data collection and usage by **not discriminating** or **disadvantaging** individuals based on the collected data.

Ensure that collected and used data are **accurate and reliable**.

Protect the **confidentiality** of data by **not sharing personal information** without authorisation.

# The rise of data ethics

Data ethics is **increasingly important** for the following reasons:

The **increasing amount of personal data** being collected, stored, and processed has prompted for more responsible and transparent data practices, which has led to a growing awareness of data ethics.

**Data breaches and cyber-attacks** are also increasing. This has brought data security to the forefront of public and political attention, which has led to increased expectations on organisations to ensure that data are handled responsibly.

The **increasing use of AI** (Artificial Intelligence) and the recognition that AI can perpetuate or even amplify biases, which could lead to discrimination, has brought ethical considerations into the world of data.

The **introduction of data legislation** and laws such as the GDPR has also played a significant role in the rise of data ethics. Organisations are now required to be ethical when handling data.

# Data governance, legislation, and ethics

The concepts of **data governance**, **data legislation**, and **data ethics** include similar principles and are closely related in that they ensure that we are **transparent** and **accountable** for how we manage our and others' data.

This also means that any **data analysis** we do or **models** we build should be **accurate** and **represent the truth** as far as possible.