DATA GOVERNANCE

Data governance is a set of principles and practices that ensure high quality through the complete lifecycle of your data.

Data governance is a collection of processes, roles, policies, standards, and metrics that ensure the effective and efficient use of information in enabling an organization to achieve its goals. It establishes the processes and responsibilities that ensure the quality and security of the data used across a business or organization.

Data governance defines who can take what action, upon what data, in what situations, using what methods.

Data governance is needed to improve data quality so that data is trusted.

Data quality if of utmost importance because when companies work with inferior data, this negatively impacts their downstream insights, analyses, and recommendations.

Data quality must entail the data is complete, unique, valid, timely, accurate, and consistent.

Data quality problems can impact on business operations causing process errors, process delays, unplanned operational costs and inaccurate decisions.

Data needs to be governed across a distributed computing environment

Data governance is the process of managing the availability, usability, integrity and security of the data in enterprise systems, based on internal data standards and policies that also control data usage. Effective data governance ensures that data is consistent and trustworthy and doesn't get misused.

It's increasingly critical as organizations face new data privacy regulations and rely more and more on data analytics to help optimize operations and drive business decision-making.

A well-designed data governance program typically includes a governance team, a steering committee that acts as the governing body, and a group of data stewards.

They work together to create the standards and policies for governing data, as well as implementation and enforcement procedures that are primarily carried out by the data stewards. Ideally, executives and other representatives from an organization's business operations take part, in addition to the IT and data management teams.

Data governance initiatives usually also include the following elements:

Data mapping and classification - Mapping the data in systems helps document data assets and how data flows through an organization. Different data sets can then be classified based on factors such as whether they contain personal information or other sensitive data. The classifications influence how data governance policies are applied to individual data sets.

Business glossary - A business glossary contains definitions of business terms and concepts used in an organization -- for example, what constitutes an active customer. By helping to establish a common vocabulary for business data, business glossaries can aid governance efforts.

Data catalog - Data catalogs collect metadata from systems and use it to create an indexed inventory of available data assets that includes information on data lineage, search functions and collaboration tools. Information about data governance policies and automated mechanisms for enforcing them can also be built into catalogs.

*What Data Governance is Not!*

~Data Governance is frequently confused with other closely related terms and concepts, including data management and master data management.

*Data Governance is Not Data Management!*

~Data management refers to the management of the full data lifecycle needs of an organization. Data governance is the core component of data management, tying together nine other disciplines, such as data quality, reference and master data management, data security, database operations, metadata management, and data warehousing.

*Data Governance is Not Master Data Management!*

~Master data management (MDM) focuses on identifying an organization's key entities and then improving the quality of this data. It ensures you have the most complete and accurate information available about key entities like customers, suppliers, medical providers, etc. Because those entities are shared across the organization, master data management is about reconciling fragmented views of those entities into a single view—a discipline that gets beyond data governance.

However, there is no successful MDM without proper governance. For example, a data governance program will define the master data models (what is the definition of a customer, a product, etc.), detail the retention policies for data, and define roles and responsibilities for data authoring, data curation, and access.

*Data Governance is Not Data Stewardship!*

~Data governance ensures that the right people are assigned the right data responsibilities. Data stewardship refers to the activities necessary to make sure that the data is accurate, in control, and easy to discover and process by the appropriate parties. Data governance is mostly about strategy, roles, organization, and policies, while data stewardship is all about execution and operationalization.

~Data stewards take care of data assets, making certain that the actual data is consistent with the data governance plan, linked with other data assets, and in control in terms of data quality, compliance, or security.

Benefits of Data Governance

An effective data governance strategy provides many benefits to an organization, including:

*A common understanding of data* — Data governance provides a consistent view of, and common terminology for, data, while individual business units retain appropriate flexibility.

*Improved quality of data* — Data governance creates a plan that ensures data accuracy, completeness, and consistency.

*Data map* — Data governance provides an advanced ability to understand the location of all data related to key entities, which is necessary for. Like a GPS that can represent a physical landscape and help people find their way in unknown landscapes, data governance makes data assets useable and easier to connect with business outcomes.

*A 360-degree view of each customer and other business entities* — Data governance establishes a framework so an organization can agree on “a single version of the truth” for critical business entities and create an appropriate level of consistency across entities and business activities.

*Consistent compliance* — Data governance provides a platform for meeting the demands of government regulations, such as the EU General Data Protection Regulation (GDPR), the US HIPAA (Health Insurance Portability and Accountability Act), and industry requirements such as PCI DSS (Payment Card Industry Data Security Standards).

*Improved data management* — Data governance brings the human dimension into a highly automated, data-driven world. It establishes codes of conduct and best practices in data management, making certain that the concerns and needs beyond traditional data and technology areas — including areas such as legal, security, and compliance — are addressed consistently.

If you’ve managed to get this far, the benefits are probably obvious. Data governance means better, leaner, cleaner data, which means better analytics, which means better business decisions, which means better business results. Better market positioning. Mindshare in your space. Reputation. Better profit margin (everybody likes this one).

Cloud Data Governance

As more and more businesses and organizations realize the benefits of moving some or all of their data storage and processes to cloud integration strategies and iPaaS, the need for effective data governance increases at scale.

Moving to the cloud is all about delegating certain tasks to third parties, such as infrastructure management, application development, security, etc. Cloud is also about virtualization of technical resources, which can create data sovereignty challenges—such as with regulations that mandate that data resides in a certain place or country. In addition, cloud-first strategies generally encourage decentralization, allowing lines of business or workgroup to roll out their own system independently, which could result in a uncontrolled data sprawl.

That’s where governance finds its place. First, a strategic data governance plan is crucial for migrating content to the cloud. Whether an organization is moving to a hybrid or completely cloud data model, the data migration process will enjoy all the same benefits of an overall data governance plan, and the migration itself will be more efficient and secure.

Additionally, moving data processes to the cloud adds a layer of complexity regarding security and access. While a completely on-premises data solution still needs a robust data governance strategy, stakeholders especially appreciate the value of data governance when that data is moving through the cloud.

Data Governance Tools

In order to find the right data governance approach for your organization, look for open source, scalable tools that can be quickly and economically integrated with the organization’s existing environment.

Additionally, a cloud-based platform will allow you to quickly plug into robust capabilities that are cost-efficient and easy to use. Cloud-based solutions also avoid the overhead required for on-premises servers.

As you start comparing and selecting data governance tools, focus on selecting tools that will help you realize the business benefits laid out in your data governance strategy.

These tools should help you:

Capture and understand your data through discovery, profiling, and benchmarking tools and capabilities. For example, the right tools can automatically detect a piece of personal data, like a social security number, in a new data set and trigger an alert.

Improve the quality of your data with validation, data cleansing, and data enrichment.

Manage your data with metadata-driven ETL and ELT, and data integration applications, so data pipelines can be tracked and traced with end-to end data lineage.

Control your data with tools that actively review and monitor.

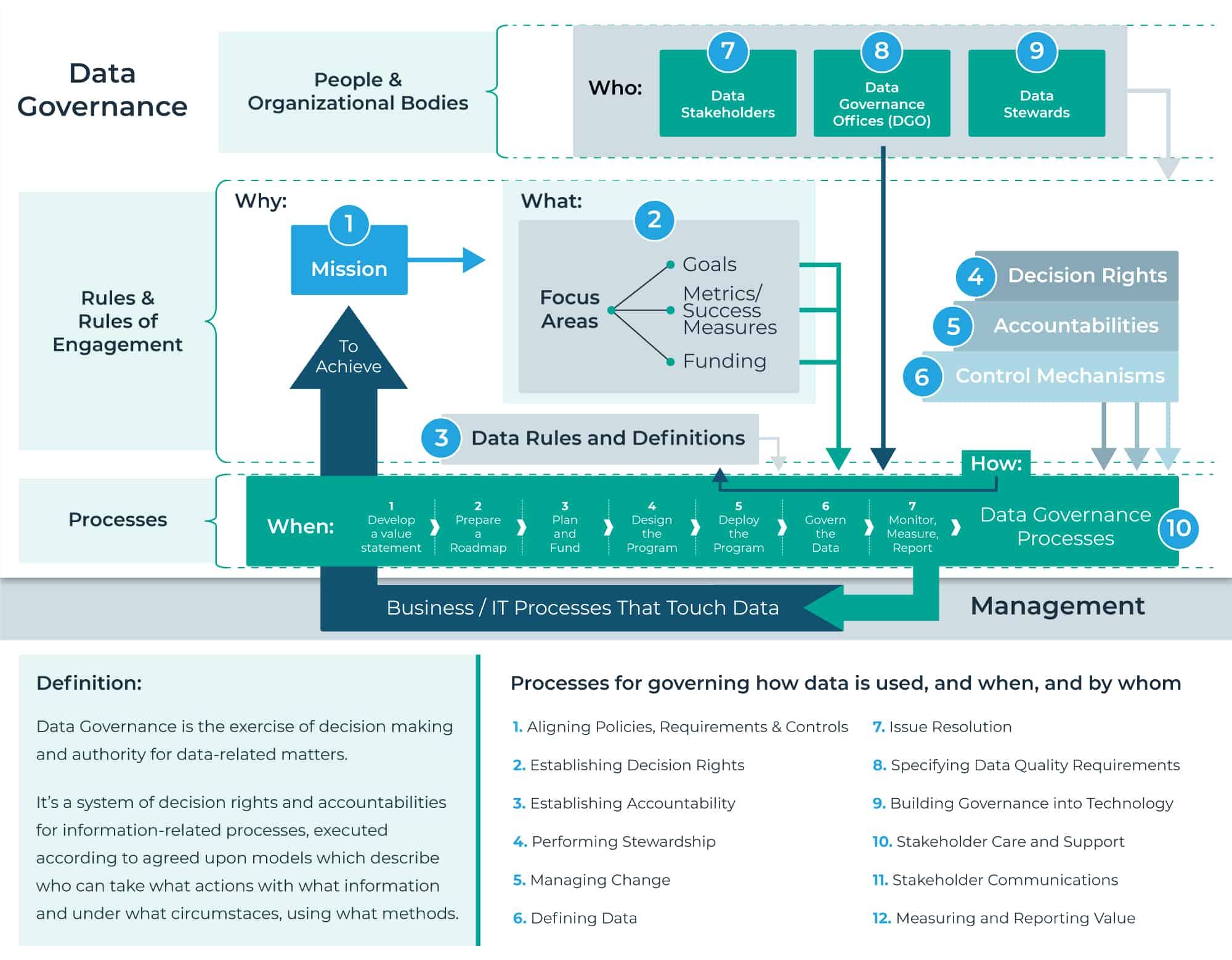
Document your data so that it can be augmented by metadata to increase its relevance, search ability, accessibility, linkability, and compliance.

Empower the people that know the data best, to contribute to the data stewardship tasks with self-service tools.

THE DATA GOVERNANCE FRAMEWORK

A data governance framework is a set of data rules, organizational role delegations and processes aimed at bringing everyone in the organization on the same page.

There are many data governance frameworks out there. As an example, we will use the one from The Data Governance Institute. This framework has 10 components; let’s discuss in detail:



Data governance involves the whole organization to a greater or lesser degree, but let’s break down the most commonly involved stakeholders:

Data Owners: First, you will need to appoint data owners (or data sponsors if you like) in the business. These must be people that are able to make decisions and enforce these decisions throughout the organization. Data owners can be appointed at the entity level (e.g. customer records, product records, employee records and so forth) and supplementary on the attribute level (e.g. customer address, customer status, product name, product classification and so forth). Data owners are ultimately accountable for the state of the data as an asset.

Data Stewards: Next, you will need data stewards (or data champions if you like) who are the people making sure that the data policies and data standards are adhered to in daily business. These people will often be the subject matter experts for a data entity and/or a set of data attributes. Data stewards are either the ones responsible for taking care of the data as an asset or the ones consulted on how to do that.

Data Custodians: Furthermore, you may use data custodians (or data operators if you like) to make the business and technical on boarding, maintenance and end-of-life updates to your data assets.

Data Governance Committee: Typically, a data governance committee will be established as the main forum for approving data policies and data standards and handling escalated issues. Depending on the size and structure of your organization, there may be a subcategory for each data domain (e.g. customer, vendor, product, employee).

These roles highlighted above should optionally be supported by a Data Governance Office with a Data Governance Team. In a typical enterprise, here are some folks who might make up a Data Governance Team:

Manager, Master Data Governance: Leads the design, implementation and continued maintenance of Master Data Control and governance across the corporation.

Solution and Data Governance Architect: Provides oversight for solution designs and implementations.

Data Analyst: Uses analytics to determine trends and review information

Data Strategist: Develops and executes trend-pattern analytics plans

Compliance specialist: Ensure adherence to required standards (legal, defense, medical, privacy)

One of the most important aspects of assigning and fulfilling the roles is having a well-documented description of the roles, the expectations and how the roles interact. This will typically be outlined in a RACI matrix describing who is responsible and accountable to be consulted and be informed within certain enforcement, a processor for a certain artifact as a policy or standard.

Resources -

<https://profisee.com/data-governance-what-why-how-who/>

<https://www.techtarget.com/searchdatamanagement/definition/data-governance#:~:text=Data%20governance%20(DG)%20is%20the,and%20doesn't%20get%20misused>.

<https://www.talend.com/resources/what-is-data-governance/>