Hitachi Solutions

Data Governance:

Creating a Foundation for a Data-driven Future

Table of Contents

INTRODUCTION

Today, organizations have an ever-increasing appetite to leverage data for business advantage. But how can you successfully do that if you can't trust your data? The old-school adage of "garbage in, garbage out" is a quick way to summarize why insight-driven businesses are leaning in on **data governance** strategies — if there is concern or confusion about the data going "in," there is little value to the data being reported "out." And data confusion is a data governance problem.

Ultimately, data needs to work for the enterprise in a sustainable way. Most companies aren't sure how to accomplish this effectively with data governance, so doubt is removed as to the validity of data and clarity is established for where data lives, how to get to it, and how to use it.

Data governance is a broad topic, for this eBook, we'll focus on the key foundational components that can help build and sustain a successful data governance effort.

What is Data Governance?

Data governance is about strategy, policies, processes, and technology. It provides a framework that connects people to processes and technology, by assigning responsibilities and accountabilities to specific people for specific data domains and data types. When you create a data governance program, you define the standards, processes, and documentation structures for how the organization will collect and manage data.

In the simplest terms, data governance is about managing data as a strategic asset. It involves ensuring that there are controls in place around data, its content, structure, use, and safety. To provide effective data governance, you need to know what data exists and where, whether the data is of good quality, whether the data is usable, who's accessing it, who's using it, and what are they using it for.

So what does that mean exactly? It means that if you need to ensure data quality and information security, and enable master data management, you need to have a solid foundation tying all these practices together and defining and supporting the processes, tools, and resources needed to make these practices successful.

Business Benefits

- Provides holistic view of your data assets across your organization
- Helps control data sprawl and reduces infrastructure costs
- Increases productivity through coordination and reduced operational friction
- Improves quality of data
- Responds to evolving regulatory requirements

Data governance is frequently confused with other closely related terms and concepts, including data management and master data management. They are different but complementary, building on one another, but are all based on a foundation to ensure data quality and control.

Here's a quick primer on the concepts:

Data Management

Data governance is frequently confused with other closely related terms and concepts, including data management and master data management. They are different but complementary, building on one another, but are all based on a foundation to ensure data quality and control.

Master Data Management

Master data management (MDM) goes beyond data governance, by identifying and improving an organization's key entities, like customers. Because those entities are shared across the organization, MDM reconciles fragmented views of

Data Stewardship

Data stewardship refers to the overseeing the activities necessary to make sure data is accurate, maintained, and easily discoverable. Data stewardship is about execution and operationalization.

Why is Data Governance Important?

As a part of the day-to-day usage of any application, users generate significant amounts of process-related and customer-related data — structured and unstructured. For example, customers generate data by visiting a website or using a specific product. In turn, businesses generate data about their products or services, as well as how customers interact with those offerings.

IDC predicts that the data we generate, collect, and consume will increase from about 50 zettabytes in 2020 to 175 zettabytes in 2025. Businesses are finding it more and more difficult to store and manage and use all this data effectively.

- Statista.com

The pace of this data growth is increasing dramatically. New data sources are being created and connected every day, and the associated data needs to be captured, rationalized, and secured. In addition, the new technologies such as internet of things (IoT) and artificial intelligence (AI) bring great promise for intelligent insight, but also underscore the need for capturing, rationalizing, and securing data through effective data governance. What else can data governance help with? It's a long list; here are a few of the major problem-solving benefits:

Breaks down departmental silos

Traditionally, departmental, or regional data silos build up when individual business units deploy separate transaction processing systems without centralized coordination or an enterprise data architecture. This silo structure makes it difficult for people to locate the data they need to answer their business questions. A data governance program, built on the right intelligent technologies, makes siloed data a thing of the past.

For example, data on the profitability of a particular customer can be used in multiple business contexts. It can drive day-to-day decisions regarding how to prioritize a customer's call in a customer service center. It can feed analyses of product or customer segment profitability. It can drive how marketing designs loyalty and retention programs. And it can feed into the finance department's reconciliation of margins and earnings. That's a lot of different departments that can benefit from having the right access to the right information.

Fortifies trust in a data-driven organization

Business decision makers are hesitant to make decisions based on data they don't fully trust, and traditional data quality rules have proven insufficient to support the swelling amount of data across the organization. Leaders will continue to "go with their gut" and make decisions based on instinct until a unified approach to data, built with fundamental governance practices is in place. When you can qualify data, decision makers will trust it and use it.

Supports innovation and growth

Having trust in your data can also foster innovation and growth. Companies across industries can uncover data patterns and use them to drive a huge variety of initiatives, for example, a bank could use permissioned data to drive sales of products or services, such as wealth management. Once your organization has taken on data governance, it can start to move forward and make critical business decisions on what to do with that information and how to drive further insights from it.

First Things First: Understand the Lifecycle of Data

Good governance provides the rules that manage data throughout its lifecycle — or stages that data moves through from its initial generation or capture to eventual archival or deletion. Different policies and processes can be applied to different types of data at the various stages, so you'll always understand how data is being captured, used, saved, and shared at each stage.

Let's look at each stage of the data lifecycle and the concerns for governance:

Create or capture

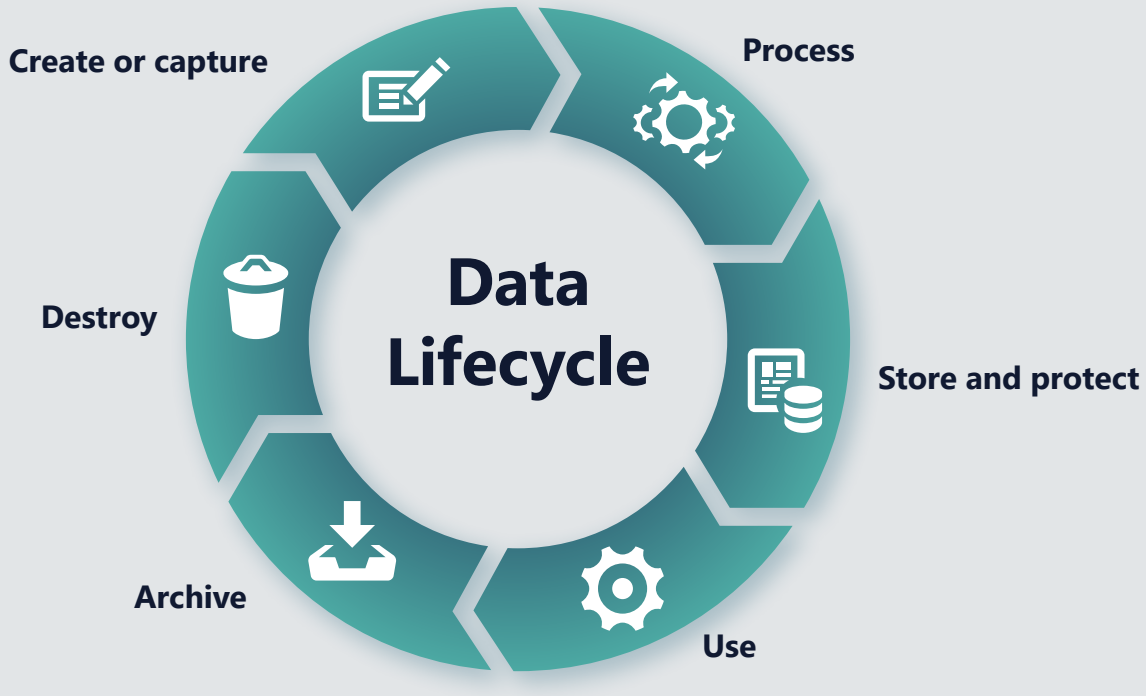
Identifying all the sources of data coming into your organization is a first step to understanding your data estate and determining what specifically needs to be governed. As we discussed, data is generated from multiple sources, in different formats, both structured and unstructured, and in different frequencies (batch or stream). Data sources and connection mechanisms need to be documented, understood, and maintained.

Process

Once data has been captured, it is processed. This is done prior to use and before the organization uses it to derive any value. Data processing includes cleaning, scrubbing, or extract-transform-load (ETL) to prepare the data for storage and eventual analysis. It's at the heart of the effort that will ensure a sustainable data governance program.

Store and protect

Once processed, data and metadata are stored on systems and devices with the appropriate levels of protection whether that's a SQL server, a data warehouse, or a data lake. While data is "at rest," it should be encrypted to protect it from intrusions and attacks. In addition, it needs to be backed up regularly in the event of a data loss, accidental deletion, or disaster.



Use

It's crucial to understand how data is consumed within an organization. As data is used by multiple internal and external stakeholders during this phase, proper access management and audits are key. In addition, regulatory or contractual constraints may dictate how data can be used; part of the role of data governance is to ensure that these constraints are observed accordingly.

Archive

Over time, the volume of data grows as does the cost of maintaining and storing it. A data archival and retention strategy allows regular shifting of data from one storage location to another. Data is often retained for potential audits, and data sets in different jurisdictions will have different retention schedules. A good data governance plan should guide the retention of data and define the length of time it will be stored, including the different controls that will be applied.

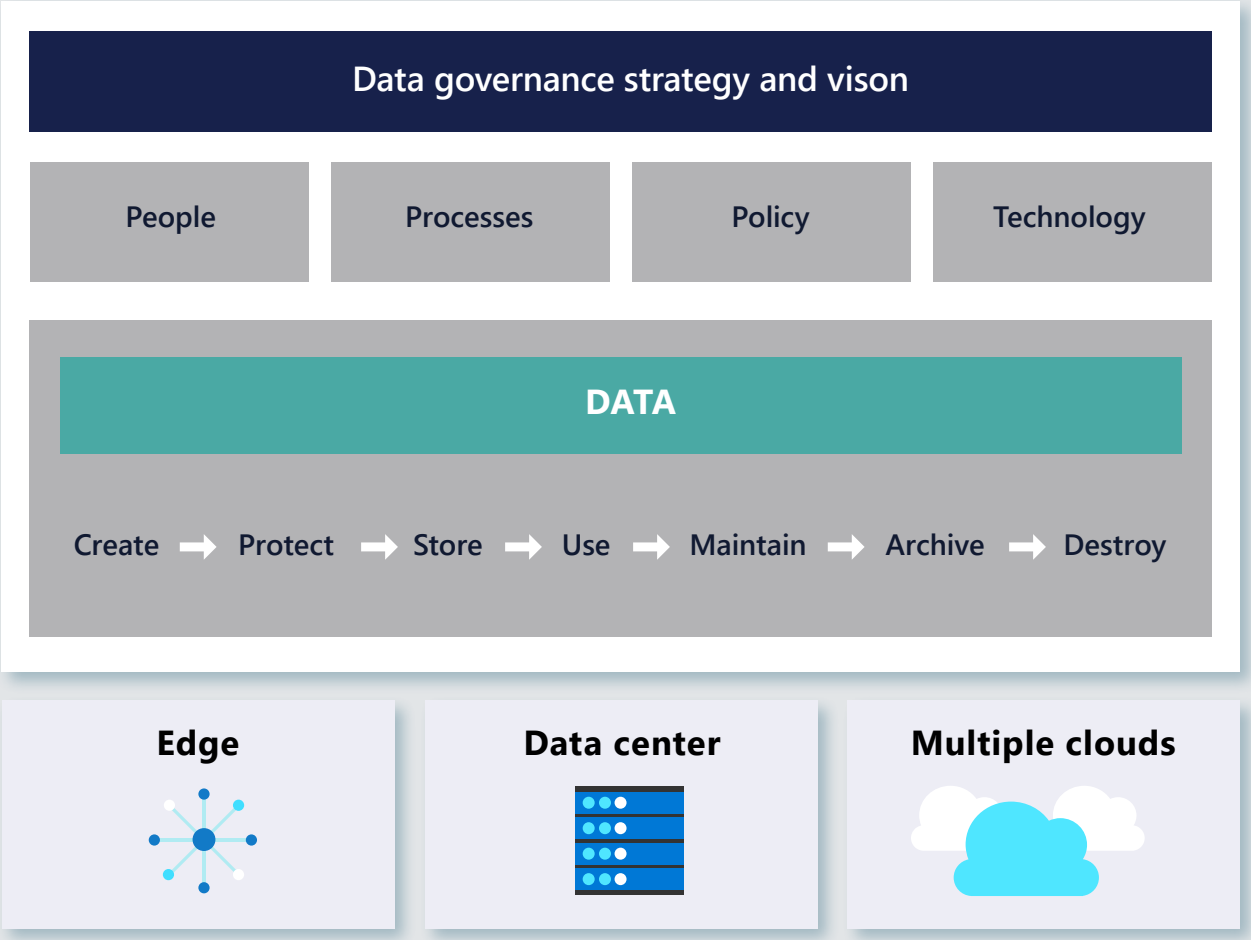
Destroy

Before destroying any data, it is critical to confirm whether there are any policies or regulations that would require you to retain the data for a certain period. Coming up with the right timeline for this cycle means understanding state and federal regulations, industry standards, and compliance policies to ensure that the right steps are taken.

Building the Foundation: A Data Governance

As data moves through the lifecycle stages, a well-defined structure will determine how and who handles the data. This is defined as the data governance framework. Let's dive into each of the framework components: people, policies, processes, and technology, all driven by a clear strategy.

What is in a data governance framework?



A vision and strategy

Creating a data governance vision and strategy is a matter for the entire business. The business side of the organization is closest to the data, the ways it has been created, and how it is (or should be) used.

Your team can help by answering these questions:

- What problems do you want to solve?
- What opportunities do you want to pursue?
- How do these translate into specific data needs and what does good-quality data look like?

Every person responsible for creating, managing, or using any data should understand and support the governance vision and strategy. It's the baseline, or north star, you will follow.

People

In most cases, the business assumes ownership of the data and takes the lead in driving data governance. This makes sense because data ultimately exists to serve the business, and the business is the primary beneficiary of effective data governance. IT then partners with the business to implement the technology aspects of the data governance program.

A well-designed data governance program requires the direct attention of the CEO, with a structure that is fully supported by other executives and sponsors. This program should also include:

- A data governance team to oversee and approve all governance measures
- A steering committee to monitor data governance decisions and implementation of the framework
- A group of data stewards to drive data management projects forward
- Representatives from each line of business to take part in the decision-making processes
- Clearly defined roles, responsibilities, and accountabilities for resources

Simply put, you need to get the right people and organize them appropriately. Be sure to include other data owners and data users. Joining producers and data consumers in conversation will only improve data quality and access policies. When the entire team understands how data is used holistically, the easier it is to identify gaps and security and compliance needs.

Policies

Well-crafted policies with clear compliance measures are the foundation for a successful data governance program. Policies need to address quality, access, security, privacy, and usage, at each stage of the data lifecycle. While the number of policies can vary from organization to organization, here are four that are critical:

Program structure policy

A program structure policy defines how data governance will be practiced at the organization and includes formal guidelines to manage company data and information resources, and the requirements for employees to follow them. It also defines who should manage data governance at the organization. By documenting who is responsible for each task, you can help ensure nothing slips through the cracks and all necessary steps are taken to ensure your new policies are effectively implemented and followed.

The program structure should also define how new policies will be communicated to the organization, how the information will be stored, and who will be responsible for educating data consumers and answering their questions.

Data usage policy

Data usage activity is divided into several categories: reading data, creating data, updating it, and distributing it. A data usage policy outlines ethical use for each of these categories and should be in accordance with all applicable laws and standards for each security level. To maintain standards, employees should only access and use data for business purposes.

Data access policy

A data access policy ensures data is accessed only by authorized individuals and is used correctly. It should be applied across the company and all business units. Make sure the policy outlines the circumstances under which an employee will be granted temporary access and how you ensure that access is revoked once it is no longer needed. Also include a formalized process for revoking the credentials of former employees. Even if a former employee is unlikely to act maliciously, unattended credentials could be used by hackers or other bad actors to gain access to your system and your data.

Data integrity policy

Data integrity relates to the validity, reliability, and accuracy of data. It should include a clear understanding of the business process on which the data is based and the routine review of individual data points to guarantee validity. Those responsible for data quality should conduct regular audits based on the defined usage and access policies, as well as risk assessments to outline potential threats and the steps for mitigation.

Processes

While policies provide the guardrails for your data governance program, processes describe how data is handled at each point in the data lifecycle. The key to designing processes starts with the data catalog, which can serve as the technical underpinning of a successful governance effort.

Ensuring quality and discoverability

To provide effective governance you need a full view of all your data assets and the ability to discover those assets regardless of where they exist. Here's where a data catalog comes in. As a central hub for enterprise information, the data catalog acts as an information map that lists all data points, their originating source, and where they currently reside. Typically, data catalogs include a well-developed structure for searching using metadata.

Here's a breakdown of the automated elements in an ideal data catalog:

- A **common glossary** defines data entities, allowing users to easily browse and search the data catalog using familiar terms. For example, a customer could also be referred to as client, purchaser, or buyer, but all those terms would be mapped to the same assets.
- Data **classifications** are the metadata or labels you assign to your data to organize it, such as sensitivity, compliance, industry, data type, etc. The value of metadata is immense in the world of data governance. For example, by classifying data as sensitive, you can kick off policy enforcement processes to restrict search ability once the data is in the catalog.
- **Data lineage** traces the flow of data and the transformations it undergoes throughout its lifecycle. This data flow is usually displayed in a graphical format for easier understanding and traceability. Without it, users can spend a lot of time unsuccessfully searching for the root cause of a data issue.
- **Profiling** examines the data available in an existing information source like a database and collects statistics or summaries about that data. These summaries can be used to ensure that the data meets the standard statistical measures and business rules of the organization.
- **Cleansing** fixes or removes incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset. When combining multiple data sources, data is often duplicated or mislabeled. Cleansing can fix these errors and standardize how data will be combined.

Data catalogs also make data easily discoverable and understandable. Once a search is entered, the return list of data assets is matched to the relevant keywords, collections, and classifications you've selected for a holistic view across your data estate.

Enforcing policies

Policies help safeguard data and establish standards for its access, use, and integrity. But how do you enforce the rules are being followed? Typically, companies use automation to enforce policy compliance. Automated processes can be published as event-driven, or demand or timer-driven services. Let's say you have a policy that all state names be published as two-character, capitalized entries, such as MA or IA. An automated enforcement process can automatically fix the entry or surface an alert to the user during data entry.

Monitoring

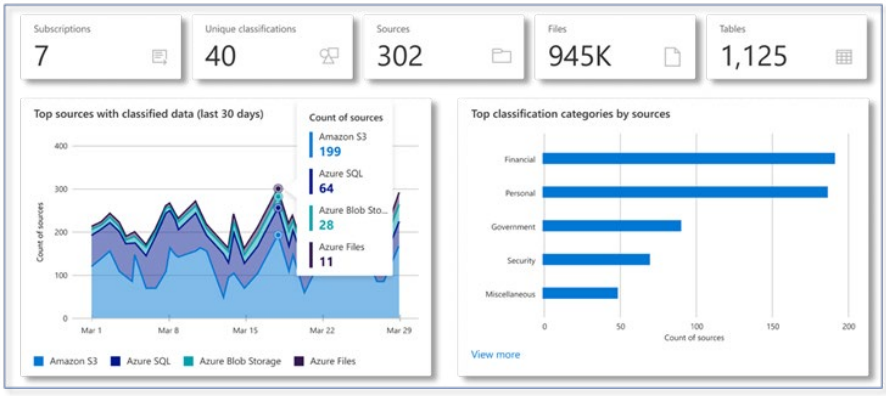
Although many data governance processes can be automated, you still need to monitor the program so you can measure its effectiveness and value. Automated processes can be set up as an early warning system for catching quality, security, or privacy compliance problems before they wreak havoc on your dependent applications, reports, and processes. The right data monitoring processes ensure checks and balances are in place to quickly react to changes as needed. They also enable transparency and the ability to better audit your assets throughout the data lifecycle.

Technology

Indeed, it is nearly impossible to achieve the ultimate goals of data governance without technology to automate and scale the development and enforcement of standards, policies, and processes.

When considering solutions, look for a platform that performs well in data acquisition, cataloging, lineage, and analysis so it can deliver **trusted data** and provide insights into the location and movement of data across the entire data estate.

Also look for a solution that includes an easy-to-use, graphical representation of the state of your data. **Microsoft Purview** has a great example of this with its Data Insights window — a consolidated view into your data assets, scans, glossary, classifications, and sensitivity labeling. It also gives you the ability to see audit trails (the lineage) categorized by sensitivity and business relevance. You can think of it as a simplified compliance risk assessment across all your data sources.



Data Insights Window in Azure Purview

What Technology Solutions Can and Need To Do:

- Improve the quality of your data with validation, data cleansing, and data enrichment
- Manage data with metadata driven ETL and ELT
- Manage data integration applications, so data pipelines can be traced with end-to end data lineage
- Control data with tools that actively review and monitor
- Document data so it can be augmented by metadata
- Empower the people with self-service tools for analysis and visualization

Get Started with **Hitachi Solutions.**

Good data governance isn't a burden, it can be a competitive advantage — you just need to know how to get there.

Chances are, you are going to need the assistance of a third party to help you assess your current data landscape, evaluate your current infrastructure, and define the policies and processes you'll use. Most companies don't have the technical and personnel resources to manage such a shift.

Hitachi Solutions has the data scientists and change advisory experts to help kickstart your data governance program with a pragmatic approach so you can truly begin using data as an asset. We look forward to joining you on your digital data transformation.

If you're still on the fence, ask yourself:

What is the cost of not getting data governance right?

- Extensive time lost manually cleaning data?
- Ill-informed business decisions?
- Lack of regulatory and privacy compliance?

Is it Time to Reimagine Data Discovery?

Hitachi Solutions can put you on the right path, starting with a **cross-industry assessment** to evaluate your data management maturity and needs. As a trained and certified member of the EDM Council, Hitachi Solutions leverages the Council's proprietary cloud data management capabilities (CDMC) and data management capability assessment model (DCAM) frameworks to provide the kind of structured, comprehensive assessments that are critical to managing, controlling, and securing data in the cloud.

The possibilities are endless when you can turn data into an asset knowing it is accurate, trusted, and secure.

Contact us to learn how we can help you!

Now is the time to talk to Hitachi Solutions about Microsoft Dynamics 365 and your digital transformation journey.

Email Us NA.Marketing@hitachisolutions.com

Visit Us www.global.hitachi-solutions.com

Follow Us

Hitachi Solutions

global.hitachi-solutions.com