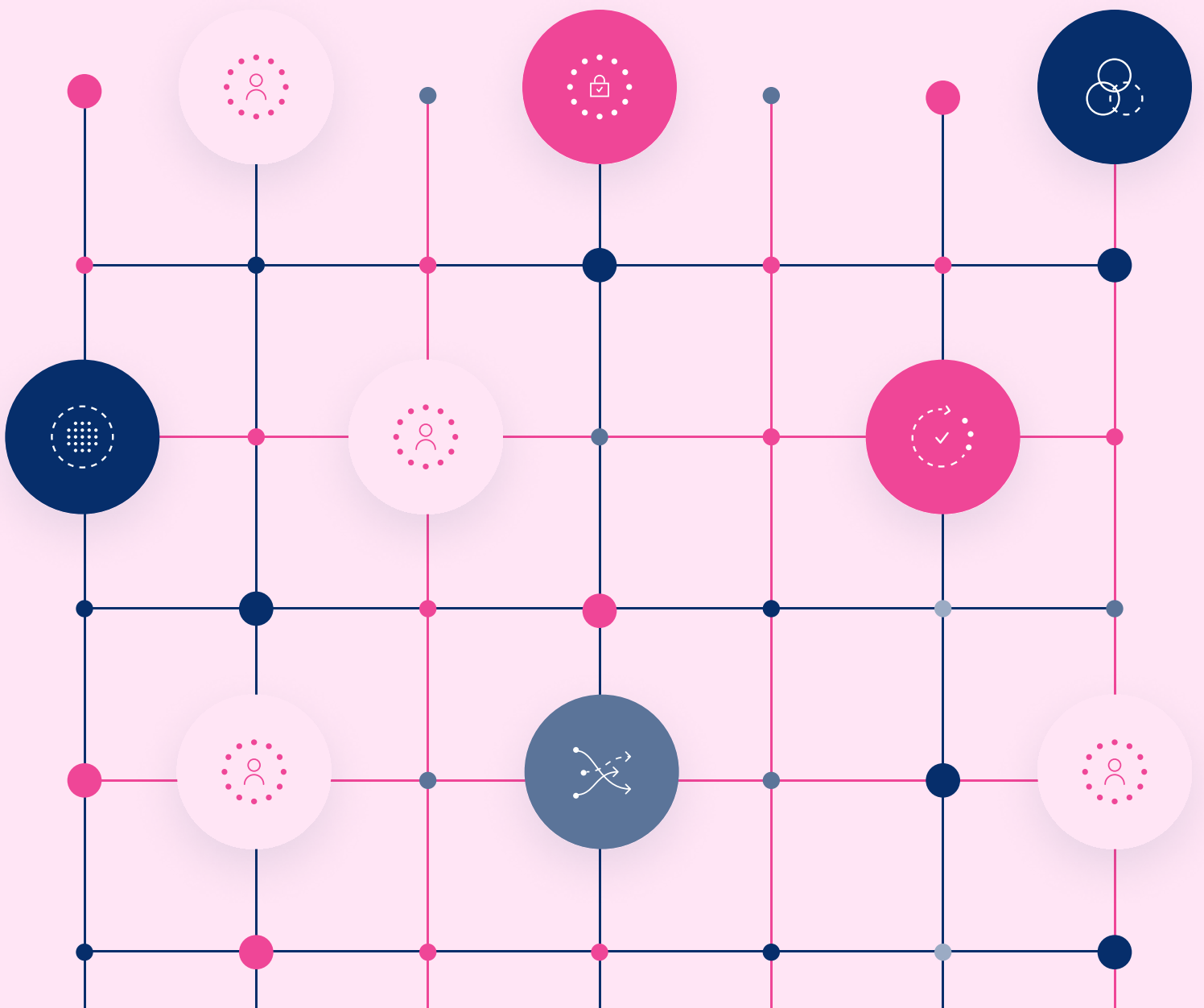GUIDE

The Ultimate Guide to Data Governance

Executive Summary

Data governance is not a standalone idea, but a strategy that enterprises can depend on to become data-driven. Good governance is good business, and it ensures regulatory compliance, enhances data discovery and collaboration, provides mechanisms for auditing and monitoring, and enriches data through transformations and linking. Through transparency, accountability, and engagement, organizations can deliver data governance strategies that bring technology platforms, people, and processes together around the central pillar of digital transformation: data.

# What you need to know about data governance

What is data governance?

The answer depends on a few factors. The current state of your data ecosystem, the amount of data you ingest, the size of your team, and your operational priorities all factor into how important data governance is for your organization. Ultimately, however, the simplest answer is that data governance should be as important to your organization as data is.

If you don't use data and never will, good data governance isn't something that you need to think about. If, however, data is — or is becoming — an operational priority, you must **consider data governance the foundation of your data strategy**.

42% of results produced by data teams are not used by decision makers[1]
SAS
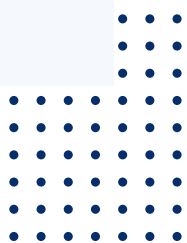
Some think that the idea of data governance is inherently restrictive. Governance implies a belt-tightening and application of rules to an otherwise free system. For others, governance means data collection and organization; a systematized approach that effectively curates and manages the information powering your decisions.

The truth of the matter is that data governance includes both of these use cases and more. For an unruly system that needs protection, data governance can create boundaries and healthy restrictions that make the organization more compliant. For a modernizing organization that wants to operationalize the data they have access to, data governance may help develop an information management system.

Although the end result may look different, at its core, **data governance is a methodology backed by both a system of rules and an infrastructure** that, combined, increase an organization's ability to find, use, and trust the data that's powering their decision science.
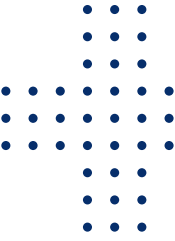
What does data governance mean to me?

For any organization looking to maximize the impact of data, good governance is a prerequisite. What form this governance takes may, however, be different from one organization to the next.

A successful data governance strategy is not rigid — the changing nature of data regulations, the evolution of data-enabled teams, and the emergence of external data as a pivotal resource all require a flexible framework designed to accommodate new opportunities and requirements.

Key reasons an organization may choose to develop a data governance strategy:

Data Compliance

Between GDPR and CCPA, the changing face of regulatory compliance and the rise of large fines for non-compliant data use has made data governance top of mind for most organizations.

Data Visibility

Organizations that control or gather data from many sources need to break down silos between data users, and are applying data governance to their environment to securely increase visibility for every team.

Data Management

Organizing data, managing its metadata, and generating insights from a database requires a data governance solution that lets organizations arrange their data catalog according to a defined and unique system of rules.

Data Monitoring

Increased use and interest in external data has uncovered a problem for organizations. Monitoring data for changes and guaranteeing its quality over time is now a prerequisite for organizations aiming to develop trustworthy applications.

Data Auditing

As organizations use more data, it becomes increasingly important to be able to audit who is using the data, for what purpose, and ensure they are connecting to it in a secure way. Auditing data access establishes data trust within an organization.

Data Enrichment

Data governance is also about modifying data to work better together. Linking data, developing a master data management strategy, and appending unique identifiers to data to combine, analyze, and enrich a database are all features of a good data governance strategy.

85% **of organizations whose data practices inhibit their business & operating strategies will empower CIOs to lead cross-enterprise investment in data governance, quality, and compliance by 2026[2]**

IDC
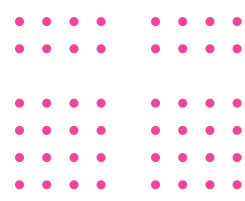
What is flexible data governance?

As its name suggests, flexible, or adaptive, data governance is the process of implementing a data governance framework that can evolve over time, change to meet new needs and regulations, and accommodate the specific data governance requirements for any organization, regardless of its size and data maturity.

As with any data governance strategy, flexible data governance is supported by both policy and infrastructure.

Designing Flexible Data Governance

Who is responsible for defining your organization's data governance policy? For many, this will be a task for the CIO or CDO, but hopefully they're not doing it alone.

Good data governance policy isn't an out-of-the-box solution, and should be tailored to your organization, its goals, and its people. Conducting a data audit with key stakeholders within your organization is a good way to understand how data is currently being used (and not used) in your environment.

A good data governance policy is one that takes into consideration both the prevailing regulatory requirements for data use and the specific needs of the end users in your organization.

System of Rules

Organization Design

Technology

> " Walking the tightrope of good data governance means balancing the operational goals of your organization with prevailing regulatory requirements."

A data governance policy designed with only security in mind will potentially overlook the end users within your organization who need to access the data, leading to workarounds that undermine the policy in the first place. At the same time, a policy that only considers operational objectives will not be compliant, and increases risk of breaches or fines.

Walking the tightrope of a good data governance policy means balancing the operational goals of your organization with prevailing regulatory requirements without compromising either. Exceptional data governance enhances operational outcomes while increasing compliance.

Deploying data governance infrastructure

Data governance policy is a blueprint. To construct the building, you need the right tools. For many organizations, their current digital infrastructure is designed to support data ingestion and analysis. ETLs and data lakes are critical elements in an organization's data environment. And while they're ideal for increasing the amount of data companies can ingest, they lack the management component that catalogs the contents of a data lake. These tools don't make data more visible, and don't provide the mechanisms to audit the usage of data.
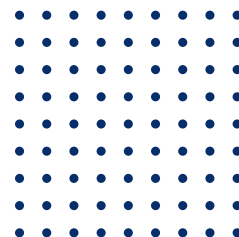
The rise of data catalogs has reflected a pressing need for organizations of every size and scope to gain insight into the status, use, and health of their data. This is especially important as organizations leverage more data from more sources. As organizations become more reliant on data they didn't themselves create, it becomes critical to monitor this data, ensure its quality, and easily govern and understand who has access to it.

A good data governance infrastructure provides a single-pane view into the data flowing through your environment. It also provides role-based controls to ensure that data administrators have complete access to your data ecosystem.

Access usage

Identify redundancy

AUDIT

Replace paid with free

Standardize contracting

Create visibility

Monitor flow

ORGANIZE

Centralize access

Quantify ROI

Identify gaps

Build your roadmap

GROW

Purchase with purpose

Leverage economies of scale

A system of visibility will increase organizational transparency, which results in getting more out of the data assets you already have.

Data auditing through lineage tracking, data access tracking, and dataset monitoring increases trust in the data that moves through your environment and provides the functional backbone of a compliant system.

Finally, deploying a familiar and intuitive user interface is more important than it seems. In order to increase data governance across your entire organization, it is important to have your entire organization feel comfortable accessing and using the data. Good governance may limit who has access, but it also democratizes the process of using data, ensuring that those who have access also have the means to use and understand the data.

The building blocks of good data governance

Data governance is not a fixed set of principles, and therefore does not have a fixed implementation. Rather, good data governance is the successful adherence to organizational priorities and policies through a set of principles decided upon by the organization and delivered through a collection of purpose-built tools.

Internal priorities, a shifting regulatory landscape, and an economy derailed by a global pandemic will all shape how organizations design and deliver their data governance strategy in the coming years. A need to bolster key performance indicators and deliver on new and exciting strategic initiatives such as AI will continue to be a priority, as will reducing operational overhead and mitigating risk. These two pillars of data strategy, innovation and optimization, will shape the conversation surrounding governance in the near and long term.

Whether your organization is primarily focused on innovation or optimization will undoubtedly affect how you design your data governance strategy. The key tenets of a good data governance policy are **Accountability**, **Engagement**, & **Transparency**.

Accountability

Good data governance should not merely restrict data use to a small group of people, and neither should it open data access to the entire organization. Flexible data governance must provide an opportunity for your organization to create configurable, role-based data access across the entire organization and then enforce accountability for how the data is being used.

Accountability is a carrot and a stick. Yes, it is important to know how data is being used in order to protect against data misuse, but accountability also elevates data users, establishes clear lines of ownership, and reduces confusion around who may use what data for what purpose. By establishing good governance, some data use will be limited, but other data use cases may emerge from unexpected places.

When accountability is done correctly, users will become less afraid of using data because they will understand precisely what it may be used for. Data users who understand their access level will be empowered to use the data because they know their access is intentionally provided.
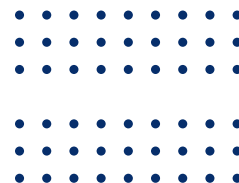
Data compliance

GDPR has resulted in over €1.67B (Aug 2022)[3] in fines since Jan 2020, the majority of which have been for "insufficient legal basis for data processing" and "insufficient technical and organizational measures to ensure information security." The first of these types of fines highlights the need for data governance policy; the second, the need for data governance infrastructure.

In order to become compliant, organizations of every size will need to consider not only why they are connecting to and using data, but how. The benefit gained from becoming compliant is not just avoiding ramifications, but adopting a technological framework to use data better across your entire organization.

Data auditing

All data governance frameworks need to provide a picture of how the data is being used. Organizations will have to provide auditable data access logs and a clear line of sight into who is connecting to the data and what they're using it for.

Data access auditing gives organization administrators insight into what's being built, but may also provide ancillary benefits to the organization. Visibility into the way data is used by one group or user, when exposed, lets other users and groups understand and interpret the data from different vantage points, enabling cross-functional learning and collaboration.
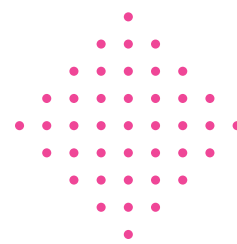
Engagement

The goal of data governance is to protect data use. A good data governance framework provides an environment where stakeholders feel confident using and reusing data, have insight into the data that belongs to their organizations, and can find, access, monitor, and deliver the data they need.

Engagement from a policy perspective may mean establishing a data governance committee, understanding how stakeholders use data and how they would like to through surveys and feedback, and establishing an iterative framework.

From a data perspective, engagement means not only data use, but metadata management, data cataloging, and data enrichment. Rather than set up a governance policy that puts data into a heavily-guarded warehouse, good data governance encourages proper engagement with the data through appropriate channels.

Data Management

At a foundational level, data management involves cataloging data. A sophisticated

solution, though, also involves augmenting the catalog with metadata and providing the ability to query and analyze data at both dataset and catalog levels. In order to manage data effectively, it is important to set up role-based access controls to ensure that data stewards may control the flow of data and set custom permission levels on a dataset-by-dataset basis.
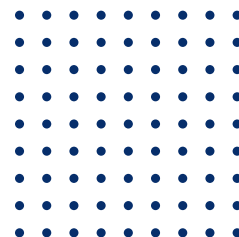
For many organizations, the past decade has been about filling their data lakes, with a notion that more data means more insight. The coming years will focus on management: organizing data into a manageable structure, adding metadata to increase exposure, and applying access rules to data. The library is filled with books; it's time to organize them.

Data Enrichment

For many, data enrichment is the primary goal of their data strategy. Enrichment is about maximizing the impact of your data, whether that's by linking data, enhancing it with additional features, or priming it for AI. If the goal of a good data governance strategy is to use data more effectively, data enrichment is part of the marketable result.

Deploying a data enrichment strategy relies on many of the elements of good data governance strategy outlined above. In order to modify and consolidate data, it must first be exposed, accessible, and integrated. Furthermore, with a data auditing and lineage platform, data users will be able to track how the enriched dataset has been modified over time, generating valuable insight into how data changes, & what the result is.

Transparency

Transparency does not mean that everyone gets access to everything. Rather, transparency as a general principle means that your governance framework provides controlled exposure, an open dialogue, and a strong, iterative feedback loop.

Transparency may mean that the process of developing your data governance policy is collaborative. It may also mean that the data in your organization is more exposed to stakeholders. Ideally, adopting transparency as a feature of your governance policy includes both ideological and technological elements.

Institutional transparency is a cultural decision. Engagement with key stakeholders in your organization to make decisions is probably already a strategic priority, and how open you want this process to be is unique to the way your organization operates. That said, collaborative approaches to data governance often include multiple perspectives and goals, and are stronger for these additional dimensions.

Digital transparency is a technological decision. Understanding where data resides, how it is structured, who manages it, how it updates, its metadata, and how it is used are not

theoretical questions, but infrastructure concerns. Traditional data lakes are often murky, making the process of seeing what is inside difficult and rarely comprehensive.
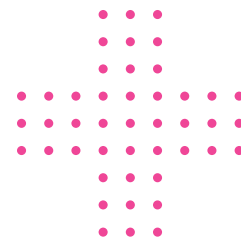
Data visibility

It is important for your organization to understand the data it is using and where it resides. Gaining insight into the entirety of your data catalog is often difficult given that, for most organizations, data is distributed across different warehouses, divisions, and, too often, localized to one individual's personal computer. Breaking down these silos may seem like an impossible process, given how instutionalized they often are.

Rather than dismantling your infrastructure to create a more transparent view of your data catalog, it is easier and cheaper – in terms of both financial and human resources – to adopt a solution or platform to pull data from wherever it resides into a central clearinghouse, a single-pane view for all data assets. Where data migration may create compliance or strategic barriers, organizations should consider options that provide data virtualization, letting users access and even transform data assets without passing them through an ETL pipeline or having them change location.
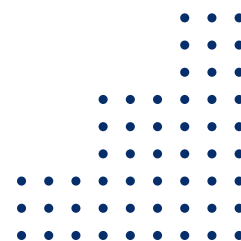
Data visibility also raises questions about data security. A good data platform will provide robust administrative controls that enable several layers of data visibility, from completely closed to completely open.

Data monitoring

Another aspect of transparency is the ability to monitor and audit data over its lifecycle. Data lifecycle management is complex, since data assets are ideally being used by many members of your organization for unique purposes. The goal of a robust data governance policy is to protect how data is being used, but it should also improve the usability of data by increasing exposure and availability to those who should have access.

It's the role of a good data governance policy to provide a framework for how organizations monitor data and data use over time. As third-party and public data grows in popularity, it is also important to adopt a monitoring solution that provides insight into how the data changes over time, when it is revised, when the schema updates, and other metainformation. Configurable alerts for specific or potentially damaging changes to the data allow end users and stakeholders to stay ahead of breaking updates – an effective data governance platform provides this functionality.

Data monitoring will let your organization gain trust in the data that's powering analysis, solutions, and models. This trust is critical in order to build products that not only demo well, but can be used in perpetuity. Monitoring is the component of data governance that makes data solutions scalable in the long term.

The building blocks of good data governance

Accountability

Data Compliance

- RBAC & ACLg models for data permissions
- administrative layer
- maintain data residency
- maintain and manage data provenance
- data deprecation workflows
- user deletion and removal

Data Auditing

- Data access auditing
- Data lineage
- Configurable inclusion of licensing and use case limitations
- Reload and restore capabilities
- Integration logging
- Addition of data artifacts
- Operational dashboards

Engagement

Data Management

- User interface for data management
- Custom scheduling
- Configurable formatting
- Intelligent file interpretation
- Metadata management
- Row- and column-level dataset sharing
- Data cataloging

Data Enrichment

- Data consolidation
- Data linkage, entity resolution
- Adding universal identifiers or anonymization
- Integration with common languages: Python, Ruby, R, Java
- Integration with common tools: Tableau, Microsoft BI, Excel

Transparency

Data Visibility

- Single-pane view of data from multiple sources
- Data virtualization
- Intelligent file interpretation
- Cloud-agnostic
- Searchable interface
- Third-party or public data discovery

Data Monitoring

- Retention of raw data
- Row- and column-level dataset monitoring
- Configurable alerts and error messaging
- Custom scheduling for data ingestion
- Revision and version tracking
- Accuracy and completeness tracking
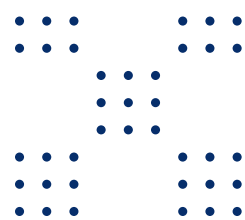
The business case for good data governance

Bad data makes bad decisions. A recent study by Gartner showed that organizations believe poor data quality accounts for ~$12.9M/year in losses[4], indicating that data quality is not merely a concern for your compliance officer, it's actively hurting your business.

If the threat of fines or missing out on the benefits of AI aren't enough motivation to develop and improve data governance, there are other important reasons to adapt your existing policies. It is important that all stakeholders **understand the benefits of good data governance as well as they understand the drawbacks of doing nothing**.

> " Data and analytics leaders need to understand the business priorities and challenges of their organization. Only then will they be in the right position to create compelling business cases that connect data quality improvement with key business priorities."
>
> **Ted Friedman**
> GARTNER

Data use is a key piece of almost every organization's operational strategy. As you consider the benefits of connecting to and using more data — artificial intelligence, automated decision-making, predictive modeling — it is important to consider that models and theories are relatively easy to offer up; without being supported by a good data governance policy, it's unrealistic to think they'll ever be put into action.

Without a system to monitor the health of data from multiple sources, share data to the teams that need it, and integrate that data seamlessly into your entire organization, your data team's efforts to advance experiments and analysis will face huge obstacles reaching production.

To successfully operationalize these advances, it is important to focus on three areas of your business that unlock the value good data governance provides. The business case for good data governance will be realized when you prioritize how your **People**, **Platform**, and **Process** come together with a common purpose to get more out of data.

People

Maintaining a common purpose and a positive culture is a difficult task at the best of times, and the new pressures of a distributed workforce — both practical and psychological — raise entirely new challenges for an organization adapting to the new normal. As the backbone of your organization, your people are the most important asset you have in implementing and perfecting a data governance strategy.

At a time when many are still figuring out the best way to work from home, balance work and life, and maintain a clear line of sight on a shifting strategic vision, the worst thing your organization can do is impose a set of new rules that alienates your team and creates more work. Good data governance will change aspects of how people work, but it should also take into account how they work currently. If someone commonly uses a process that is no longer sustainable under your data governance policy, what alternative is being provided?

The process of developing a data governance strategy may expose problems that have been lying dormant for years. Hard questions about who owns data, who uses data, and what the data is being used for will surface as your organization looks to implement good data governance. If your changing policy is clamping down on data use, it is important to develop clear lines of communication to let users request access to data, understand who owns it, and ask questions about the data in an open air environment.

Duration

Establishing good data governance will not happen overnight, so make sure your policy is implemented with a pilot-to-production model in mind. Architect plans with simple execution and measurable goals, implement them, then gather feedback and iterate.

Commitment

From top down: if data governance is a priority, it's a priority. If the people in your organization won't change their workflows to meet the changing needs of your organization, no policy will effect change. Data governance that doesn't translate from proposal to practice is not good data governance.

Effort

Studies have shown, if a new policy takes more than 10% of someone's time to implement, it will not get accomplished. By understanding the people who use data, you will be able to design a data governance framework that fits into their workflow, rather than supplanting it.
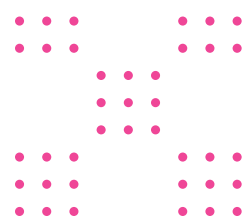
Platform

A data management platform is the interface for your data governance policy. This platform should be capable of providing the infrastructure that lets your organization carry out the requirements in your data governance policy. Data ingestion, sharing, monitoring, auditing, and integration are practical considerations underpinned by good policy.

A key component of flexible data governance is having a platform that is as adaptable as your policy. Data management solutions that require a complete digital overhaul and redistribution of all data to a central warehouse are impractical and hard to implement. As highlighted earlier, governance policies that don't become governance practices are not good policies.

It is much easier, and just as secure, to leverage a data catalog that lets organizations ingest some data and virtualize other data while maintaining current warehouses and cloud deployments.

The business case for good data governance dissolves if your organization decides to build this platform in-house. While at first it seems like tailoring a platform to your specific needs is ideal, consider: the development cycles needed; the research and specialists required; the number of unique teams that must coordinate on the project; ongoing security and reliability; and the overall financial cost.

It is easier, cheaper, and more secure to develop a data catalog or management platform with a specialized provider that will work with your organization to understand your requirements and deploy a managed infrastructure in your preferred environment.

User-friendly

The barrier to entry for a good data governance platform needs to be low. In order to democratize data use and increase adoption, your data governance platform must be usable by both the data science and business divisions of your organization. The spectrum of data use, from Excel to Jupyter Notebook, must be considered.

Secure by design

Security is the responsibility of every part of your governance policy: people, platform, and process. Whatever security policy your organization enacts will be carried out via the platform, so security certification, comprehensive penetration testing, and built-in data and access auditability are essential elements.

Distributed

Centralizing your data operations does not necessarily mean centralizing your data. Taking advantage of distributed cloud technology, for example, will ease some of the governance requirements for your organization internally, as the burdens of operation and governance fall to the cloud service provider.

Process

Data governance process does not describe the steps towards achieving data governance, but the steps taken to achieve operational goals within your data governance framework. Process is not the blueprint for establishing data governance in the first place, but the methodology you enact for getting something out of it.

Your organization's process will relate to its overall data strategy. If, for example, your organization is focused on extracting more value from internal data, your data governance policy will aim to dismantle data silos between divisions while protecting the security of the data assets themselves. The data governance process, in this example, is the means for people to connect to this data: where they go, how they receive the data, and how they deliver it to solutions, models, and business intelligence tools.

In this way, the process is the user experience of data governance within your organization. It's important to get right, as flaws in the process may account for a lack of adoption, even if there's internal buy-in and a strong infrastructure to back it up.

Developing a good data governance process might be as straightforward as creating sufficient documentation, or developing repeatable training modules for anyone who plans to use data in your organization.

Aligned

The goal of a well-designed process is that it doesn't disrupt other processes in order to be effective. Establishing good data governance should be a priority, but it will be an uphill battle if 'getting it right' compromises the day-to-day operations of your people and your organization.

Transparent

Your policy should be a living document that can be accessed and reviewed at any time. The everyday processes in your organization's data use should be just as visible. By making data governance transparent, you expose obstacles in the process which, when revealed, lead to improvements in the overall framework.

Agile

Data governance must be flexible enough to accommodate changing regulations and needs for your organization. The process you design should wrap around your data priorities, and adjust as needed to match and enhance your strategy.

The data governance checklist

When it comes to actually implementing a technological infrastructure for data governance, there are some practical considerations about how to connect to data, how you want data to be managed, and what requirements your organization has around tracking data use, data lineage, and metadata.

The following checklist breaks good data governance into three primary components, **Platform**, **People**, and **Process**, and provides a general overview of what you should expect of each while you are implementing data governance for your organization.

PLATFORM

Ingestion
- Ingest data from multiple sources
- Handle multiple data file formats
- Ingest data on different schedules
- Wipe and reload data
- Enable data virtualization
- Log and audit integrations
- Store on any cloud for smart warehousing

Transformation
- Normalize and standardize raw data
- Preserve raw data
- Enrich data on ingest (e.g. anonymization)
- Provide APIs
- Integrate with data science toolkits
- Integrate with common programming languages
- Allow export into standard formats

Management
- Include metadata
- Provide the ability to include additional or custom metadata
- Have a queriable interface
- Provide a user-friendly interface for data access
- Allow for data classification
- Track dataset lineage

PEOPLE

Management
- Allow any number of users
- Provide role-based/alternative permission models for data access
- Allow users to only access specific datasets
- Have an administrative layer
- Restrict user access when needed
- Securely deprecate and remove user access
- Offer multi-factor authentication, SSO, or other security protocols

Activity
- Audit data access
- Configure custom alerts on datasets
- Data is accessible to and usable by both data science and business professionals
- Allow secure shares between users
- Allow row- and column-level dataset filtering
- Give users the ability to quickly correct data problems

Type
- Have a steering committee
- Operationalize data stewards
- Involve a governance team
- Integrate with IT

PROCESS

Usability
- Provide access to internal data
- Provide access to external data
- Validate data over its lifecycle
- Provide sufficient documentation
- Provide full system redundancy
- Have integrated "self-healing" mechanisms
- Exist on your organization's infrastructure

# Build a strong governance foundation that fuels growth

A successful data governance program is the key to elevating your data initiatives to new heights. Getting there requires a solution that is adaptable to the nuances and intricacies of your organization's governance requirements.

A data catalog platform can work alongside your requirements, helping to integrate governance efforts to the everyday workflows of data users. Transform your data governance program into an active organizational-wide practice.

To learn more about our platform's approach to governance, book a chat with us at **thinkdataworks.com**.

About ThinkData Works

ThinkData Works unlocks the value of your data to grow your business. Connect to any source, catalog your assets, and deliver data to the people and applications that need it while retaining visibility and control. Our unified cloud platform cuts overhead, fuels innovation, and drives revenue growth.
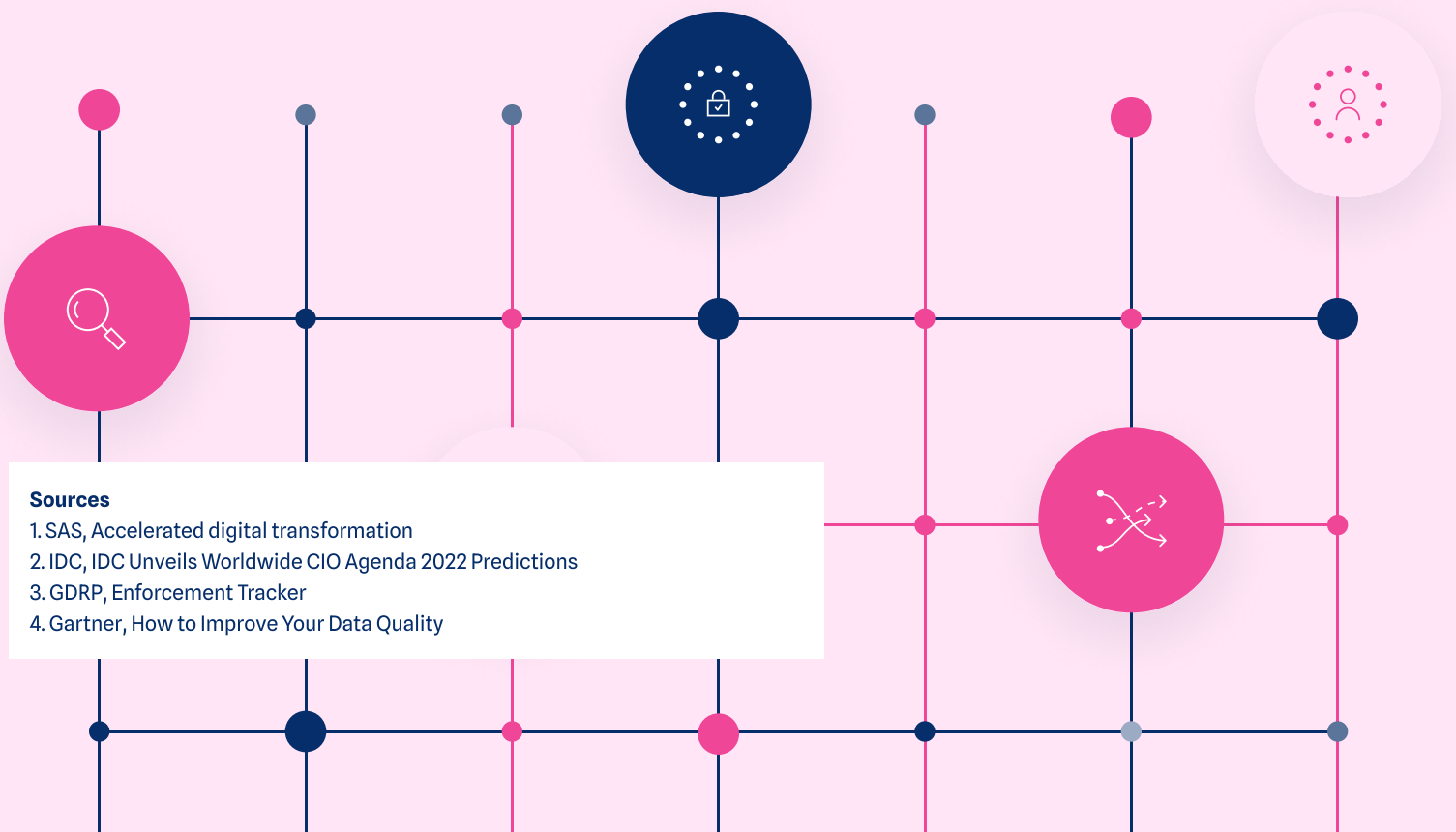
Sources
1. SAS, Accelerated digital transformation
2. IDC, IDC Unveils Worldwide CIO Agenda 2022 Predictions
3. GDRP, Enforcement Tracker
4. Gartner, How to Improve Your Data Quality