What Is Data Governance? Understanding the Business Impact

Andy Petrella

What Is Data Governance? Understanding the Business Impact

Andy Petrella

What Is Data Governance? Understanding the Business Impact

by Andy Petrella

Table of Contents

What Is Data Governance?

Data isn't just a commodity; it's a capital asset. It drives businesses, creates value, and inspires new ventures. If your enterprise has conscientiously collected, ingested, stored, maintained, and used data, now is the time to attach value to it and make it work for you. If you govern it properly and instill confidence in its quality, your data can give your business a competitive advantage. This report will walk you through that process.

Like money, data must be treated with a serious dose of respect, and a project using data should generate value with a reasonable maintenance cost over time. Everyone needs to understand why data is so important to the business and how serious the consequences of a data breach can be. Instilling that understanding and respect requires a cultural shift that affects the entire enterprise. That's the purpose of *data governance*, the discipline of creating and enforcing policies and standards to ensure that data is always treated with care and respect for safety and the law.

This report will help you understand what data governance is and why it's crucial to your business.

This chapter looks at the basic ideas involved in data governance: what it is, who does it, and how it fits into an organization. You'll learn why data must be treated and valued as an asset and what sets it apart from other kinds of assets. We'll discuss the fundamental principles that underlie data governance. You'll also come to understand why you need to think of data governance as an ongoing

program that's integrated into your company culture—not just a project that begins and ends.

In Chapter 2, we'll use real-world examples to look at the business advantages of a well-run data governance program, including risk management, efficiency, and brand image. You'll see how a culture of data governance can reduce your time to market, help you realize value faster, build confidence, and protect your brand.

Chapter 3 covers how a company's data governance program evolves and matures over time. You'll learn how to assess the maturity level of your data governance program (even if you don't yet have a formal program!) and what's involved in moving your program to the next level.

Data governance is often misunderstood; many people think of it as a bureaucratic obstacle or a hoop to jump through. Nothing could be further from the truth—in fact, data governance is crucial to reaching your business objectives. You can't possibly understand the full value of your data without it. We'll begin, then, with a look at data asset valuation.

Data Asset Valuation

Money has a life cycle, and the better you understand it, the better you can use it to increase value without increasing risk. In its simplest form, this life cycle describes how revenue is generated and how budgets are spent—in short, cash flow. Every enterprise knows that, which is why no one disputes the importance of accounting and oversight units. They manage cash flow, conduct audits, create budgets, optimize investments, and ensure that the principles of accounting are followed in every financial transaction, among other important functions.

Data asset valuation is the process of assigning economic value to data so that you can understand its worth just like you would real estate, machinery, or any other financial asset. However, there are a few important distinctions that set data apart from other kinds of assets. Briefly, these include the following:

- Data can be reused indefinitely.
- Maintaining data has a cost.

- The more data you maintain, the higher your risk. (We'll discuss this in Chapter 2.)

- Data can be sold multiple times.

- To acquire value from data, you need to use it; unused data is a cost until it is used to generate value.

Understanding that data has a life cycle, just like money does, is key to turning data into a competitive advantage.

The data life cycle is driven by two forces, shown in Figure 1-1: *execution* (or management) and *oversight* (or governance). It's important to understand the difference. An *enterprise data management (EDM)* system's functions include data risk management, data quality management, and metadata management. The enterprise's *data governance* arm, by contrast, acts as a control body, overseeing and ensuring the effectiveness of the data management system.

These two forces are handled by separate organizations within the enterprise in order to maintain impartiality. (This is similar to the "separation of powers" in governments; you can think of governance as the judicial/legislative branch and management as the executive branch.) People often conflate the two because data governance is part of the data management framework, but they are different functions performed by different people. Together, they comprise the *enterprise data management framework*.

This structure is known as the *Governance V*, adapted from John Ladley and represented in Figure 1-1.[1]
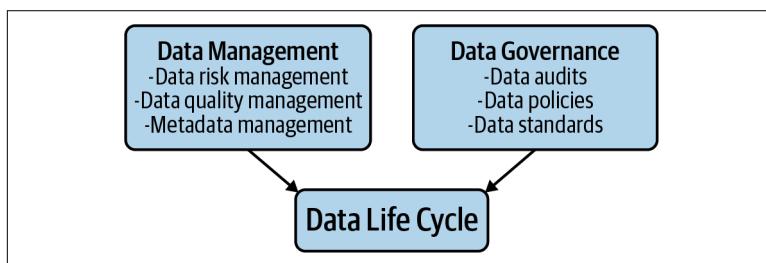

Figure 1-1. The "Governance V"

---

1 John Ladley, *Data Governance: How to Design, Deploy, and Sustain an Effective Data Governance Program* (Waltham, MA: Morgan Kaufmann, 2012), 10.

Data governance can't be separated from the enterprise itself—it has to become part of the culture of the organization. The resulting mindset embeds good data-management habits into every level of the organization, not just IT. An example of good habits might be setting up processes so that an administrative assistant who creates a new file is automatically prompted to add metadata tags and register the file in the appropriate catalog. The idea is to make the correct processes as routine and automatic as possible for every role.

Key Concepts

Data governance and data management come with a rich vocabulary. This report will focus on only the most essential terms and concepts, but for a deep dive, see the 2017 second edition of the *Guide to the Data Management Body of Knowledge (DAMA-DMBOK2)*.[2] The DAMA-DMBOK2 is published by DAMA International (formerly the Data Management Association). This important reference, created by DAMA's ecosystem of experts, is the gold standard of the field, providing data-management and data-governance professionals with shared definitions of important concepts, activities, tasks, roles, and responsibilities. It's important to have precise definitions because a common vocabulary is fundamental to aligning everyone across teams, units, and divisions on the most important responsibilities, priorities, and activities.

In addition to the DAMA-DMBOK2, I also recommend reading John Ladley's *Data Governance* (Morgan Kaufmann, 2012) for an in-depth explanation of these concepts (as well as several useful tools).

Data Management

Data management is a global framework that organizations use to ensure that their data is under control. The need for data management accompanies data from its origins to all of its destinations.

DAMA's data governance wheel (Figure 1-2) puts data governance at the center of its framework.[3] Radiating out from it are the various pillars of data management. The purpose of data governance is to

---

2 DAMA International, *DAMA-DMBOK: Data Management Body of Knowledge*, 2nd ed. (Basking Ridge, NJ: Technics Publications, 2017).

3 Ibid., 35–39.

---

guarantee that each pillar is executed properly and that data is treated as an asset. I have added arrows to DAMA's graphic to represent how the data governance body monitors each pillar.
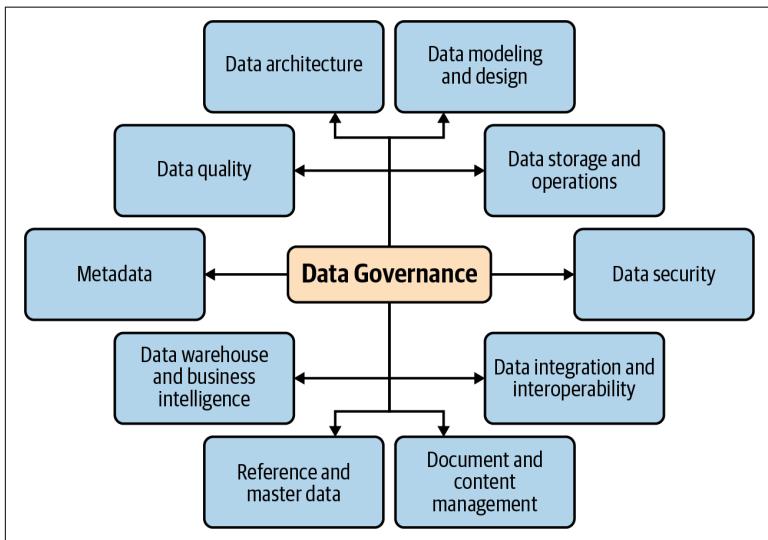
Figure 1-2. A slightly modified version of DAMA International's data governance wheel, which depicts all major components of the enterprise data management framework

Covering all of the pillars is outside the scope of this report (indeed, that's the purview of the entire DMBOK), but I do want to place special emphasis on two of them: metadata management and data quality management.

Metadata is information about data, and metadata management incorporates information about how the data is used and the systems that store it. Metadata is the data used for data governance. Managing it conscientiously is an important part of implementing a sustainable and efficient data governance program that is woven into the very fabric of the business.

For an enterprise to consider its data as an asset with value, it must have great confidence in its quality. Creating that confidence requires significant work to ensure the data's quality as well as to demonstrate to the rest of the organization that the data it relies on meets the standards defined by the data governance body. The processes and policies used to do this comprise *data quality management.*

Data governance can only be achieved through a cultural change that spans the whole organization, across functions, units, domains, and roles. This requires a holistic approach based on a set of principles.

Principles

At the core of any data governance program lies a set of values or principles. Just as a government's constitution attempts to apply a set of principles to create rules, data governance programs apply these principles to set policies for data management across the enterprise. It's important to understand that they are philosophical values, not tasks. They operate at a high level and are applied universally, across the board and across time.

DAMA's starting point for this set of values is the Generally Accepted Information Principles (GAIP), which in turn is based on the Generally Accepted Accounting Principles practiced by accountants around the world.[4]

Let's look at a few of those key principles, which need to be understood by everyone in an enterprise:

Asset principle
> Data needs to be treated and intuitively understood as one of the company's capital assets, just like personnel, cash, hardware, or buildings. This is also called the content as asset principle.

Value principle
> Data contains value. The value of its usage, loss, or acquisition can be estimated and reported. This is also called the real value principle.

Going concern principle
> Data is critical for ongoing business management. It is alive and needs to be handled accordingly: as a continual part of the business, not as a project with an endpoint.

---

4 DAMA-DMBOK2, 78–79. See also John Ladley, *Data Governance*, 17; and Accounting.com, "What is GAAP?".

Risk principle

Data is an asset that presents internal and external risks (to be clear: internal and external to the *data itself*, not to the company). Internal risks are related to the content itself; for example, it might become outdated or corrupted. External risks are associated with the data's usage, such as being misused, stolen, or sold improperly (such as in the infamous Cambridge Analytica scandal). Because data usage must meet regulatory standards for protection—such as Basel II, Solvency II, or the EU's General Data Protection Regulation (GDPR, covered later in this report)—misuses that violate these regulations also carry the risk of heavy legal penalties.

Due diligence principle

Humans can introduce bias at different stages of data processing, as well as other kinds of risk. Every person and group in the enterprise, from entry-level employees to the very top, should take responsibility for reporting any risk or anticipated risk associated with data. Internal departments and/or the data governance body should confirm any reports and raise issues to the next level as appropriate, up to and including the relevant government agencies or committees.

Quality principle

The quality of data should be monitored according to a high standard and the results reported up to the highest level. Even seemingly small quality issues can allow for incorrect interpretations, which can lead executives to make the wrong strategy decisions.

Audit principle

Proper documentation of data content and usage, including its provenance, freshness, and treatment, must be reliably maintained so that it can be audited by an independent party.

Accountability principle

Data has value not only for the organization managing it, but also (for personal data) for the individuals behind it. Therefore, throughout the data life cycle, specific people must be officially accountable for that data. This helps avoid problems stemming from data misuse.

Liability principle

> Any organization that manages data is liable for how it stores and uses that data and must do so ethically and in compliance with all relevant laws.

Level of valuation principle

> Data contributes to the enterprise's goodwill valuation. Like any assets, the valuation varies over time and includes the cost of acquisition; unlike most other assets, data can be sold and resold more than once. (Think of this as a bit like selling a digital ebook file: unlike a physical book, it can be sold again and again without incurring additional production costs.) The possibility of multiple resale prices and kinds of revenue must be accounted for in any valuation.

It's important to understand the difference between policies and standards, both of which put these principles into practice.

Data policies formalize principles, translating them into rules that guide how to implement data management. These rules mandate the use of *standards*, which outline measurable units of work or actionable tasks at a very granular level, which can also include how and when those tasks are performed and by whom. (If, like a national constitution, policies put principles to work, then standards function like laws—translating high-level rules into specific ones.)

Data processes are the procedures used to implement and maintain those data standards. Together, policies, standards, and processes form the operational model of a data governance program. The operational model outlines the roles and responsibilities involved.

Roles

Data governance can be organized and implemented in many different ways, with varying levels of hierarchy as well as potential roles. This section will focus on roles. Please note that a role is a function, not necessarily a job position; a role can be filled by a dedicated person, by a team, or by someone who also fills other roles. The only role we'll discuss that is clearly a *job* is that of a chief data officer:

Chief data officer (CDO)

The leadership position of CDO is, as Sunil Soares defines it, "the C-level executive with overall accountability for Enterprise Data Management."[5] The CDO reports to one or more other executives, such as the chief risk officer, chief information officer, or chief executive officer. They often set up a *steering committee* composed of leaders (such as vice presidents, product owners, and even vendors) to oversee the data governance program.

Data owner

A *data owner* is generally a senior person in the company who is accountable for all management activities around a specific set of data, including quality, liability, and compliance. This role is outlined in the GAIP's accountability principle. In most cases, data ownership is shared by several people and can be organized by data domain, business function, or other metrics.[6]

Data stewards

Data stewards, as the name implies, take care of the data, carrying out the daily operations for which the data owner is responsible. Data stewardship can take several forms.

For example, a *business data steward* is generally a person with deep domain and context knowledge who understands how the data generates value for the organization. They manage how the data is used for business purposes and identify misuses.

A *technical data steward* ensures that information about the data itself is gathered and maintained and can be shared with others. They identify quality and process issues, such as sloppy entry procedures that lead to problems later, and go deep into the details to form descriptive knowledge about how the systems work.

An *operational data steward* operates at a more granular level, taking responsibility for the ingestion of data, fixing problems, and administration.

---

5 Sunil Soares, *The Chief Data Officer Handbook for Data Governance* (Boise: MC Press, 2015), 6.

6 Jill Dyché and Analise Polsky, "5 Models for Data Stewardship", SAS Institute.

Summary

Now that you've learned the basic principles of data governance and the key roles involved, you're no doubt noticing that this is a big undertaking. It can involve hiring, reassigning roles, training employees, reorganizing systems, documenting processes, and creating a shift in company culture. Why make such a serious investment? In the next chapter, you'll learn just how much value data governance adds—and what kind of return on investment you can expect when you implement data governance and transform into a truly data-driven business.

The Benefits of Good Governance

A data-driven business is one that relies on facts—specifically, the facts found in its data—to support its decision-making processes. These include *strategic* decisions, such as acquiring a competitor or restructuring the business; *tactical* decisions, such as embedding data into business processes by using it in recommendation systems; and *operational* decisions, such as performing prescriptive maintenance on machines or making internal business processes more efficient.

The data-driven business sees data as an asset and a strategic resource and has a company culture that values data and sees maintaining its quality as crucial, as discussed in Chapter 1. All of these decisions rely on creating such a culture.

All of these decisions are only as good as the data that drives them, which means that data governance is at the core of the business. This takes us back to the quality principle from Chapter 1, often summarized as "garbage in, garbage out." Good quality management and good processing lead to good data. Good data might not automatically result in good decisions, but it is a necessary condition for them; bad data will result in bad decisions. And good decisions lead to better performance, a better reputation, and better revenue.

A data-driven business culture is therefore one that values and invests in its data governance body. Let's look at how this plays out in practice.

Risk Management

While data governance is important throughout an enterprise, as a control body, perhaps its biggest business impact is on risk management, through processes and internal audits.

The emerging literature on data governance is still in its early stages (and commonly blurs the distinction between data governance and data management), but most of it focuses on presenting business cases for the importance of a data governance body. Increasingly strict national and international regulations have raised the stakes with stiff penalties for violations. The intuitive response from business leaders is to strengthen controls (like audits) and reassure stakeholders that they are managing risk appropriately—this is data governance.

The Role of Regulations

Thanks to the introduction of regulations like the EU's General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA), formal structures of data ownership are now mandatory for most businesses. This provides a high level of accountability about how data containing personally identifiable information (PII) is ingested, stored, and used. The regulations apply to third-party vendors that work with the data as well as to the entities that own it.

These laws, and other similar frameworks, outline and protect the rights of *data subjects* (the people behind the data). They restrict how different types of data can be used, how they can be stored, and for how long, and prescribe methods for transparency throughout the process. Failing to abide by data protection laws can result in fines heavy enough to cause financial distress even to very large companies. GDPR fines, for example, can go as high as 4% of a company's revenue.

The role of an internal audit of any kind is "to provide independent assurance that an organization's risk management, governance and internal control processes are operating effectively." Internal *data* audits give executive management a window into the effectiveness of their data-driven business processes and how to improve them. Internal auditors usually report directly to the CEO and operate

with a great deal of autonomy. They are separate from the data governance body but collaborate closely with it—like in governments, where police forces are separate from and yet crucial to court systems.

The most difficult aspects of internal data audits are getting access to the right data, ensuring its quality, and doing so in a way that doesn't disrupt operations. Audits are often performed using interviews, surveys, and questionnaires, all of which are extremely intrusive and take great time and energy from both auditors and staff.[1] If the data governance program has good data management in place, however, the necessary data can follow the same life cycle as any other data. High-quality and good metadata and search capabilities make audits easier, more precise, and better able to flag issues for quick action.

When the culture of an entire organization shifts toward new strategies, like operating in a data-driven fashion, the positive effects extend to every department and functional body in the enterprise. Although there is also disruption, the improvement can be profound—and it certainly isn't limited to regulatory compliance and audits. The remainder of this chapter will look at other areas where data governance makes a big impact. For simplicity's sake, I've organized them into two categories that reflect the challenges and rewards you'll likely encounter as you implement your data governance strategy: improvements to efficiency and improvements to brand image.

Efficiency Improvements

Efficiency is pragmatic, tangible, and easy to measure relative to other impacts; it requires collecting only some fundamental metrics from business processes. Here, I'll focus on three areas where a solid data governance program can particularly improve efficiency: responding to changes in the market, reducing a product's time to market, and realizing value from mergers and acquisitions.

---

1 S. Rao Vallabhaneni, "Domain 3," in *Wiley CIAexcel Exam Review 2014: Part 1, Internal Audit Basics* (Wiley, 2014), *https://oreil.ly/TX0m9*.

Time to Awareness: Responding to Changes in the Market

Markets change. That's what they do. And change is becoming even faster, broader, more complex, and more drastic than ever.

In such troubled times, organizations have to be smart about customer behaviors, segment adjustments, and budget allocations. Collecting the right data, internally and externally, is a crucial first step in tracking the viability of any business and making business decisions. This data needs to get into the right hands quickly—the goal is to decrease the *time to awareness*, or how long it takes to make sure that problems come to the attention of people who can act on them.

When you put appropriate data governance and a complementary data-driven culture in place, you ensure that relevant and accurate information flows to decision makers consistently, continuously, and promptly. This helps them to make better decisions faster, even when change is swirling all around them.

Time to Market: Smoothing Your Processes

Reducing *time to market* is a key business driver: when you have an innovative product, you need to get it into the market before your competitors do. But before deploying a new product or marketing campaign, you need to ensure that the underlying data is reliably reported and accessible.

Did you know that data scientists, on average, spend 80% of their time identifying, cleaning, validating, and organizing data and just 20% on analyzing it?[2] Making these processes more efficient can help get products to market much faster.

One aspect of achieving that efficiency is putting the right policies in place to ensure well-documented metadata about ingested data and generated intermediate results. This makes it easier to share results and methodologies. Here, again, we see the distinction between governance and management: the data governance body creates those

---

2 Armand Ruiz, "The 80/20 Data Science Dilemma," *InfoWorld*, September 26, 2017, *https://www.infoworld.com/article/3228245/the-80-20-data-science-dilemma.html*.

policies and oversees their use, while the data managers (in this case, metadata managers) implement them.

Time to Value: Efficient Mergers and Acquisitions

During mergers and acquisitions (M&A), the spotlight will shine directly on your data. Are you ready for that? This happens at the highest levels during the due diligence process, but it continues at a smaller scale well into the integration period. Projects migrate, teams and units learn to work together, and everyone looks to the data to find their footing.

In addition to integrating two companies and their products, management, processes, and systems at every level, the M&A process means integrating those companies' data. It all amounts to a daunting job under the best of circumstances; even two fully data-driven businesses might well have different principles, policies, and standards that must be compared, rationalized, and brought into harmony. Good data governance can, however, significantly simplify many aspects of this undertaking.

At the highest level, the challenge of M&A is about realizing value. We do M&A to increase value, but we only see that increase once the integration is complete and updated processes and policies are in place across the newly larger enterprise. A data governance program can make business processes more transparent and ensure that clear, well-organized metadata eases the process of combining data. The more efficiently that happens, the faster you can get to that increase in value—that's *time to value*.

Brand Image Improvements

You've seen some examples of the many ways that data governance improves efficiency across the enterprise. I'll turn now to a closely related concept: brand image. I noted previously that strong data governance and a strong data culture are ultimately felt even outside of the organization, which directly affects the company's brand image. Let's look at why.

A well-governed data management system embeds quality management and accountability into its processes and provides users with facts about the reliability of their data. Put simply: *you can trust it.*

When people know that they can trust the data—that it is well maintained, accurate, and reliable—they feel more comfortable putting it to work. They can use it to back up innovative strategies, proposals, and projects. Taking a risk on a new idea becomes less risky because they know that the data will hold up to scrutiny. Stakeholders are less reluctant to drive strategic evolutions. Without confidence that appropriate data governance and culture are in place, however, controlling risk takes precedence over data-driven innovation.

Figure 2-1 shows several ways that data governance can play a role—large or small—in improving brand image.
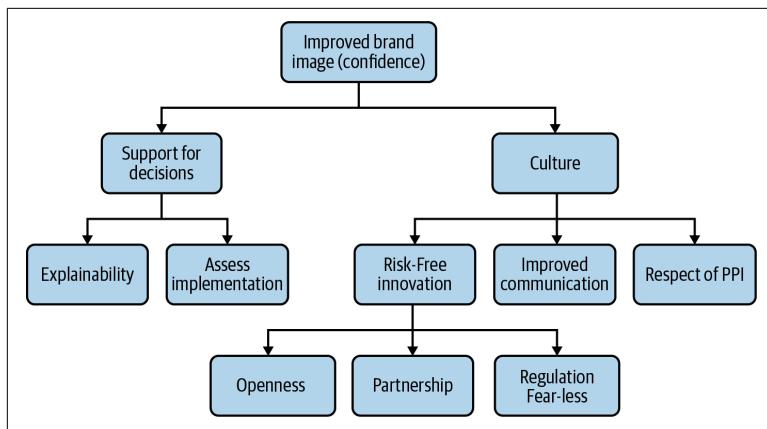
Figure 2-1. Data governance programs impact brand image and consumer confidence in a variety of ways

We'll go through each of these, beginning with something truly fundamental: culture.

Growing a Culture: Social Attitudes Toward Data

Your organizational culture isn't just about ping-pong tables in the break room or the tone of chatter in the office—it's about how every person in the enterprise approaches their work.

Organizational culture starts with the CEO, who plays a fundamental role in creating, shaping, and fostering it. But it extends to every level of the enterprise. No one is untouched by it. Yet a culture can't be taught; it must be set up and nurtured. It lives in the interactions of people: the interdependent day-to-day behavior and group dynamics of every single person working in or with the

organization. In this way, it's just like any other kind of culture in the world, from nations to religions to youth subcultures.

Culture determines how employees see themselves and their place within the company. It shapes what they do and how they do it. An unhealthy culture can damage a business, but a healthy, positive, data-driven culture can do wonders—improving not only the quality of the work or product, but also the company's brand image, the happiness of its customers and employees, and its resilience in the face of change. It shapes what the company builds (or rebuilds). Just as a national culture, for example, might include a food culture, a musical culture, and an artistic culture, *data culture* is an important component of a company's culture.

The principles of data governance (Chapter 1) are the soil in which your data culture will grow. The company's mission and values plant the seeds, determining what kind of culture you'll cultivate. And what waters and weeds and shines sunlight to make it grow? Social behavior.

Human beings demonstrate what they value by how they behave. Their principles show up in what they do and how they do it, how they treat one another, and—yes—how they treat their data. When people care about the quality of the data and respect it as a fundamental part of the business (following the asset principle), they're willing to adopt habits that support this value. They do the right thing because they understand that the data is an important business asset; they would no more harm the data than a farmer would trample their field. (This brings us back to the due diligence principle.)

You've seen that policies and standards are tangible, operational, measurable artifacts that embody principles. They are introduced and adjusted over time. Data culture is less tangible, but very much present. It becomes an intrinsic characteristic of the group, so people come to respect the principles in a natural way, without cumbersome controls and tedious processes.

When values become habits, there is social pressure to demonstrate the values by performing the habits. When someone breaks a principle—for example, by treating data carelessly and violating the audit principle—it makes everyone noticeably uncomfortable: *that's just not how we do things around here.* The offender is quickly set straight; their colleagues are eager to help them understand why the habit in question is supported by a fundamental principle. Such

interactions allow knowledge about principles, policies, and standards to spread through the organization organically. That kind of environment makes good data habits sustainable.

Growing Accountability: Data-Driven Decisions

More than ever, as you've seen throughout this report, businesses and consumers expect decisions to be based on data. Indeed, it sounds almost silly to talk of basing decisions on anything else, yet the history of business is full of decisions based on hype, fear, gut feelings, personal relationships, inflated optimism, and many other powerful forces. Furthermore, even when you think you're working from evidence, it can be easy to see what you want to see or to present facts selectively. With data, anyone can dig in and examine the facts.

Data-driven decision making offers two considerable advantages. First, it's easier to explain decisions. Rather than simply issuing an edict, leaders can provide the data behind their actions, showing clearly why the action is important. A company's decision to pivot toward a new product, for example, will get a great deal more buy-in if everyone can see that consumer interest in that type of product has risen strongly and consistently for the last few years.

Sometimes turning to the data helps to hone the approach. In a case study, Stanford professors David Larcker and Brian Tayan cite the example of a fast-food chain that began an effort to reduce employee turnover as a way to improve customer satisfaction. But when they turned to the data, they found a different problem. Journalist Michael Mauboussin summarizes:

> As the turnover data rolled in, the executives were surprised to discover that they were wrong: Some stores with high turnover were extremely profitable, while others with low turnover struggled. Only through proper statistical analysis of a host of factors that could drive customer satisfaction did the company discover that turnover among store managers, not in the overall employee population, made the difference. As a result, the firm shifted its focus to retaining managers, a tactic that ultimately boosted satisfaction and profits.[3]

---

3 Michael J. Mauboussin, "The True Measures of Success," *Harvard Business Review* (October 2012); David Larcker and Brian Tayan, *Corporate Governance Matters*, 2nd ed. (Upper Saddle River, NJ: Pearson FT Press, 2015).

The second advantage is that data-driven decision making provides accountability. From decision to execution to results, having good data helps us follow the trail and assess what we've done well and how we can improve. Let's say that sales of a new product are much softer than anticipated. What happened? Why is there such a gap between the company's initial expectations and the actual results?

If the company based its decision on data, a postmortem review of the product launch can easily show why those expectations were set. If it was based on other factors, examining the data might help explain why the expectations were unrealistic. Data about how, where, and when the product was launched and who did (or didn't) buy it can inform decisions about where the missteps happened and how to do better next time. Data about who did what, and clear ownership of data, are also important drivers of accountability.

Here's my take on accountability: it is largely composed of being able to explain what is happening, why it's happening, and how things got there, with the goal of determining the next best actions to take. Accountability is *not* the same thing as blame. It's about facts. When this is clear to everyone, the accountability principle becomes an attitude, a duty to self—not a sword of Damocles hanging over everyone's heads.

Can you name three companies that have recently suffered major data breaches? I bet you can. Companies that fail to maintain ownership and accountability for their data risk serious damage to their brand image. They face more than hefty fines and legal issues—they also face public anger and become the butt of jokes and memes on social media and television. That's embarrassing, but far worse is the loss of trust in the eyes of customers, partners, and shareholders, not to mention the advantage their competitors gain.

Growing Respect: Protecting PPI

One of the first signs that an organization's data governance is maturing (a process we'll discuss in the next chapter) is a culture of strong respect for personal private information (PPI) and personally identifiable information (PII).

Corporations generate and collect a monumental amount of personal data about individuals, much of it quite sensitive, including home addresses, government ID numbers, credit card information, GPS data, medical records, marital status, criminal record, and

internet history, just to name a few. The consequences of a failure to steward such data are very serious, sometimes life and death. In the wrong hands, PPI can be used to find someone's location, blackmail them, gain a competitive advantage, conduct credit card fraud, or discriminate against members of protected groups.

The data governance body monitors and enforces compliance with data protection laws. As you've learned, it creates the policies and standards that govern how aspects of each person's job can be done in ways that maximize respect for and protection of data. That means helping everyone understand how even their small actions can affect the data, the data subjects, and the company's overall compliance. That understanding allows for a healthy data culture to grow.

Holding PPI data is a responsibility that carries significant risk. Companies that hold PPI are obligated, legally and ethically, to protect it. Sometimes that obligation can get in the way of an innovative or potentially valuable idea—but the need for protection against widespread misuse of data has imposed this constraint. That sense of obligation and responsibility needs to suffuse the company's data culture. When everyone, from the CEO on down, understands that protecting PPI means protecting the people behind the data and acts accordingly, the ensuing sense of trust becomes attached to the brand's image.

Growing Profits: Realizing Value from Data Assets

The rights associated with PPI and PII are not the only constraints to innovating with data. One of the biggest challenges in data innovation is the legacy of *silos*. Older methods of data storage were localized and hard to share, and over the decades, most organizations dealt with this by not sharing it. Instead, each department or division found its own way to handle data. The eventual result was a set of silos, or separate, walled-off parts of the organization, each with different storage, organizational, and processing methods, database software, and quality levels. As cloud architect Mike Kavis notes, "This model served its purpose well when software was built as large, monolithic applications that were deployed on physical infrastructure and planned in quarterly or biannual release cycles."[4]

---

4 Mike Kavis, Chapter 1 in *Accelerating Cloud Adoption* (Boston: O'Reilly, 2020).

The advent of cloud computing has removed those technological limitations, but in most enterprises, the silos remain in place. Today, when software might deploy multiple times a day, silos slow things down—and, as you learned in "Time to Value: Efficient Mergers and Acquisitions" on page 15, no company can afford to lose any time in leveraging the value of its data assets.

Data silos don't shut innovation down entirely—certainly the last few decades saw enormous innovation despite them. However, even aside from their impact on speed, they're harmful to company culture. The existence of silos fosters a culture of separation, where people focus on how to optimize their use of data within their own small domains in competition with others, instead of working systematically to improve the value the company overall can generate from its data. It's hard to monetize your data quickly when everyone handles the data differently and no one talks to one another about it.

It might seem strange to talk about the need for openness and data sharing right after discussing the dire risks of losing control of PPI. But the openness I'm talking about happens *within* an enterprise, and *open* doesn't mean irresponsible, or uncontrolled, or delaying the risk-management process to a later stage. What it means is lowering the cost of introducing new data into the business, finding cost-effective and efficient ways to store and work with data, and improving knowledge sharing and collaboration across the organization. All of those things have a direct effect on the bottom line.

In Chapter 1, you learned that data's ability to be sold more than once makes it unique as a type of asset. Data monetization—selling lists of data to third parties or sharing them with partners—is a major source of value. A data governance program can ensure that this is done safely, legally, and ethically. In addition, as we've seen, proper data governance increases the *quality* of the data, which in turn makes it more valuable. In the data market, a reputation for well-managed, high-quality data is a competitive advantage.

Summary

This short chapter could not possibly list all the positive effects of properly governing and managing data; there are too many. Instead, I've covered several of the most important benefits of good data governance, with a focus on risk management, efficiency, brand image, and company culture. You've seen how serious the consequences can

be when data governance isn't up to snuff and how rich the rewards can be when it is. In the final chapter of this report, you'll learn how companies get from here to there.

Maturing Your Data Governance Program

Good data governance doesn't happen overnight.

In order to succeed, your data governance program must be well planned, with a clear understanding of its impact on every part of the business. It must be carefully deployed, without skipping any steps. What's more, it must also align closely with the needs, goals, values, and context of the business so that objectives are clear and reachable.

Once you've deployed it, your data governance program still needs to mature. You'll have to make adjustments, correct misalignments, and learn to adapt to changes as they come. As your employees continue to work within the new system, they'll develop habits, as you learned in Chapter 1; it's up to you to ensure that the habits they develop are good ones. They'll also learn how it all works, becoming more confident and knowledgeable and developing new processes that deepen efficiency and minimize room for error. As your company culture changes to incorporate a strong and healthy data culture, this knowledge and confidence reinforces itself and spreads through the organization. That, in broad strokes, is what it means for a data governance program to mature. We can define *data governance maturity* as the development and deployment status of the data governance processes, functions, principles, and standards necessary for a company to use data as an asset successfully.

Data Governance Maturity Models

Any parent of a teenager can tell you that maturity is a difficult thing to measure. It doesn't happen all at once, or on an even increase; there are leaps forward, steps backward, long pauses. In a data-driven business, though, you need a fairly precise way to measure maturity. How does that work when every organization and every data governance body, much like every teenager, is unique?

That's where *data governance maturity models* come in. A maturity model is a tool for guiding strategy that incorporates a rating scale that measures how well a data governance program is operating. As it evolves and grows and becomes embedded in company culture, the program moves up through the levels, which are determined using sets of specific indicators (more on those in a moment). These levels are discrete and provide a good understanding of the situation but give little visibility into the continuous progress made to move to the next level. They are performance indicators that tell you how far the program has progressed and what it can do next to improve.

Diversity of Maturity Models

There is no general consensus on a single maturity model. Indeed, there are many different models, like the DAMA-DMBOK maturity model, the Stanford model, the Capability Maturity Model Integration (CMMI) (on which IBM's maturity model is based), and Gartner's Enterprise Information Management Maturity Model. Comparing these models is outside the scope of this report, but I encourage you to research them to find out which is right for your company. Here are some resources to get you started:

- "Review of Data Management Maturity Models" by Alan McSweeney
- "Data Management Maturity Models: A Comparative Analysis" by Irina Steenbeek
- "The Case for the Data Management Maturity Model" by Jennifer Zaino
- "Maturity Models for Information Systems—A State of the Art" by Diogo Proença and José Borbinha
- "Data Governance Maturity Model: How Mature Is Your Approach to Data?" by Stef Olafsdottir

---

The data governance maturity model you choose will be part of the program's implementation from the beginning and will shape its evolution through the years.

Though the models differ, they use similar methodologies and indicators to observe, assess, and measure performance, processes, and practices and recommend improvements. These indicators vary: *staged* models tend to be broad in scope, with macro-level ratings of the entire program, while *continuous* models operate at a smaller scale, measuring process capabilities in specific areas of data management and governance. They are used simultaneously, with the continuous model guiding maturation within each stage until the company "moves" to the next stage.

These two models correlate with two modes of thinking, both of which you'll need to master: strategy and tactics. *Strategy*, the big-picture plan, works from the top down: you define what your process areas are, what processes they contain, and what practices should be incorporated in them. *Tactics* deals with the logistical side of things. It starts at the bottom and works upward, looking at the tasks that make up each practice, prioritizing them, and using metrics to assess them. The staged method of assessing maturity correlates with strategy, and the continuous method with tactics.

The Staged Scale

The staged scale used in many models is a big-picture measure spanning several areas. These levels represent the *current* maturity of the *overall* organization, so there is only one rating at a time: an organization is a Level 3 organization until it moves up (or down). Most models use something like the following categories:

Initial (Level 1)
    The earliest stage. This doesn't necessary mean nothing has been done, but activity has been scattered, sparse, and/or disorganized, with no framework in place to replicate successes. There is little buy-in among staff.

Managed (Level 2)
    A data governance program is in place. It ensures that processes are planned and managed for all projects, allocates resources, and defines roles to allow the organization to react to events.

Defined (Level 3)

> The data governance program has produced full documentation of practices and precise definitions for everyone to work from. The whole organization follows standards across projects. Measurements and reporting are robust and swift, allowing the organization to move proactively based on data.

Quantitatively managed (Level 4)

> The data governance program has introduced proactive measurements into quality monitoring, with specific goals defined along the life cycle of each process. Because of the measurements and reporting defined in Level 3, the organization manages quality and performance predictably.

Optimizing (Level 5)

> The data governance program focuses on continuous improvement, monitoring the quality and performance of processes and their role in meeting business objectives. The organization efficiently and consistently improves and adapts its practices to strategic changes.

The Continuous Scale

The continuous scale focuses on achieving process goals through practices, which are the activities that make up a process. This scale is meant to be applied to a single process at a time. Unlike the staged scale, it produces many results—one for each process—and so many assessments will be going on simultaneously. Most models use something like the following categories:

Incomplete (Level 0)

> Existing practices are not well tuned or need adjustment; new practices are partially or not at all implemented or executed. Goals have not been met, and the organization may not yet be mature enough to expect achievement in the process area.

Performed (Level 1)

> Practices have been implemented and goals met. This is validated at the end of the process rather than at each step, or as part of an audit.

Managed (Level 2)

> Practices are planned and executed according to the process documentation. Operations are robust and process effectiveness

is monitored. A clear system of ownership for practices is in place. Employees are well trained and held accountable for the practices they own.

Defined (Level 3)
The process consistently and sustainably meets its goals and follows its prescribed standards. Standardization is in place, with clear, universal definitions of the requirements and objectives of the process, participants' roles and responsibilities, inputs and outputs, and validation criteria.

Next, we'll look at the assessment process.

Assessing Your Maturity Level

Now that you understand the basics of data governance maturity models, we can turn toward the first step in this data-driven journey of change: assessing your enterprise's maturity level.

I must emphasize the importance of not skipping this step. Nearly every time I've seen a data governance implementation fail, it was in a company that overlooked the initial maturity assessment. Before your journey starts, after all, you need to know in which direction to go, and to know that, you need to figure out where you are right now.

If you don't have a formal data governance program in place, don't automatically assume that your business is at Level 1 (initial). It may have reached some level of data governance maturity simply by practicing common-sense processes. For example, a well-organized company with clearly defined business processes and appropriate audits might well find that it is at Level 3 (defined), but also that it needs metrics to improve its monitoring efforts and reinforce its treatment of data as an asset. Note, too, that there is no Level 0 in the staged scale; all organizations have some degree of data governance, however unconscious or disorganized.

Each maturity model includes an assessment questionnaire and guides you through the assessment process.

Increasing Your Maturity Level

Once you discover, through the assessment process, where your starting point is, you can begin to think about how to take your organization's maturity to the next level.

This is a complicated process. I cannot tell you how to increase your organization's data governance maturity level. No one can. I will offer some general advice for each maturity level, but only you know your organization's culture, the resources available, or how urgently maturation is needed (to name a few of many factors). But the most important thing I can tell you about increasing your data governance maturity level is this: *never stop*.

Here's what I mean: it's tempting to think of increasing your data governance maturity level as a project. You propose a plan and a budget, the board of directors approves it, you start the project at a certain date and end it on the prescribed ending date, and then you're done. This is dangerous. Don't give in to that temptation.

Instead, you should think of maturing your data governance program as an *ongoing* effort. Everything you have learned in this report has emphasized embedding data governance into your organization at every level. Shifting your culture, training your people, organizing

your roles—all of this serves to weave data governance into the fabric of the organization until you couldn't remove it if you tried. That isn't a project; it's a way of life.

Now, let's back up for a moment and compare the staged and continuous scales shown in Figure 3-1. The staged scale rates organizations from Level 1 to Level 5, whereas the continuous scale goes from 0 to 3. The first three maturity levels are the necessary steps to get your organization to the point where all projects, business units, and associated processes follow the data governance principles by implementing all practices. In other words, if your organization's current maturity level is Level 4, quantitatively managed, on the staged scale, all processes will need to be operating at Level 3, defined, on the continuous scale (or higher) before the organization can level up. Any new processes added will likely start at Level 0.

Staged	Level	Continuous
Optimizing	5	
Quantitatively managed	4	
Defined	3	Defined
Managed	2	Managed
Initial	1	Performed
	0	Incomplete
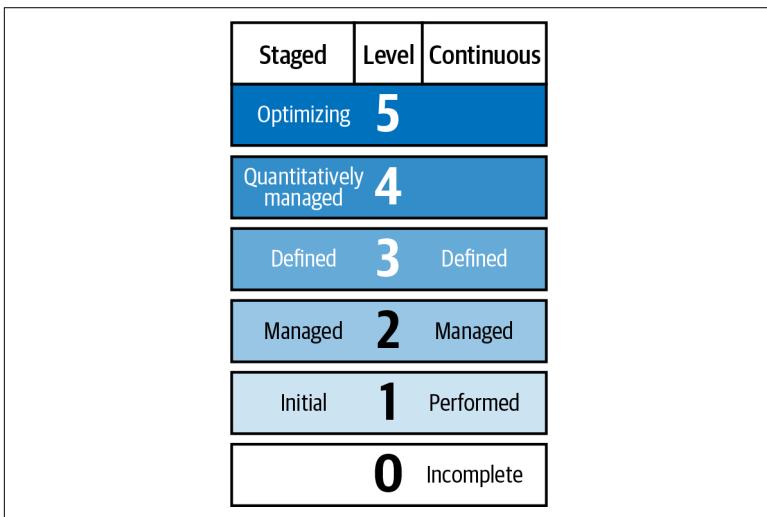
Figure 3-1. Comparing the staged and continuous scales of data governance maturity

Of course, that's not *all* you need to do. Raising the maturity level generally requires defining and implementing new processes (and, rarely, new process areas) to harmonize the overall implementation and ensure sustainability. (This makes sense when you recall that the goal of a data governance program is to create a data culture that spans across process areas, through the entire organization.) You'll need to concentrate on critical business processes, building knowledge about how the organization *really* functions, using data to fill

in the gaps. All the while, you'll be reporting processes as they are, defining metrics to measure them, monitoring their performance, and identifying opportunities to share knowledge.

Level 3 of the continuous scale, which rates a process as "defined," represents the highest level of maturity that a *process* can reach. But, as an organization, you can go higher. Once you get to Levels 4 (quantitatively managed) and 5 (optimizing), your data governance program is no longer operating process by process. By this point, it's operating as an umbrella organization, sitting above and overseeing all of the processes. At Level 3, the data governance program begins to really produce data about how the organization runs; at Level 4, the organization is leveraging this information and introducing new processes to make projections and anticipate problems. It's important to point out that these new processes need to be sustainable—if they're tedious and time consuming, they will fail. Nor can you simply throw some AI into the mix and hope for miraculous insights. You must design these processes to be intuitive and weave them carefully into the existing business processes.[1]

Once you reach the top—that is, Level 5 (optimizing)—your goal is to stay there. Sharp-eyed readers will have noticed that this level's name, unlike the others, uses a present-tense verb. This is because optimizing is *always* ongoing. For companies with the very best data governance programs, much like Olympic athletes with the very best speed and reflexes, maintaining peak performance takes continuous effort.

Summary

This report is accompanied by many references, and I encourage you to read through them and build your knowledge. As you learn, you'll begin to form a vision of how data governance could look in your organization. From there, you can assess your maturity level and make a plan to improve it.

Many people think of data governance as a boring, bureaucratic task, a trap of endless implementation that forces them to cross off

---

1 For a guide to what to do at each phase that offers the freedom to adjust to your current situation and goals, I strongly recommend the Phase Activity Task Table, a gem of a tool buried deep in the appendices of John Ladley's *Data Governance* (301–17).

every box on the compliance checklist. Indeed, it can be easy to fall into this misunderstanding, with business units viewing governance teams as working at cross purposes or slowing them down.

Properly implemented, however, the right data governance program speeds things up. Good data governance, as you saw in Chapter 2, reduces the time it takes to become aware of problems and market changes, the time it takes to bring products to market, and the time it takes to realize value from data. It builds trust among the individuals whose data you hold, the businesses you share data with, your employees, your vendors, the regulatory bodies that oversee your business, and—crucially—the public, safeguarding and polishing your brand image even as other companies' scandals erode their willingness to trust. It suffuses your company culture and guides everyone's thinking. It transforms your data from commodity into asset, from cost sink to moneymaker. It makes your business data driven—and that changes everything.

About the Author

Andy Petrella gets energy from using his expertise in mathematics, distributed computing, and data technologies to build innovative enterprise solutions and help data organizations.

In the early days of Apache Spark, Andy created the Spark Notebook, an open-source software for data scientists using Spark and Scala. Today, he runs Kensu, Inc., a confident, enthusiastic young scale-up that gives visibility on how analytics are built, maintained, and validated.

Andy regularly gives public talks, keynotes, and seminars at events like Spark Summit, Big Data Spain, and Strata (Data, AI), for which he also served on the reviewing committee. Andy teaches online through O'Reilly regularly, including courses like "Distributed Data Science Pipelines using Cassandra, Kafka, and Cassandra."