McKinsey Technology

Designing data governance that delivers value

Follow these principles to shift from a data-governance model of loosely followed guidelines to one that makes the most of digital and analytics.

by Bryan Petzold, Matthias Roggendorf, Kayvaun Rowshankish, and Christoph Sporleder

June 2020

Executives in every industry know that data is important. Without it, there can be no digital transformation to propel the organization past competitors. There are no analytics driving new sources of revenue. Even running the basic business well isn't possible. But for data to fuel these initiatives, it must be readily available, of high quality, and relevant. Good data governance ensures data has these attributes, which enable it to create value.

The problem is that most governance programs today are ineffective. The issue frequently starts at the top, with a C-suite that doesn't recognize the value-creation potential in data governance. As a result, it becomes a set of policies and guidance relegated to a support function executed by IT and not widely followed—rendering the initiatives that data powers equally ineffective. In other cases, organizations try to use technology to solve the problem. While technology solutions such as data lakes and data-governance platforms can help, they aren't a panacea.

Without quality-assuring governance, companies not only miss out on data-driven opportunities; they waste resources. Data processing and cleanup can consume more than half of an analytics team's time, including that of highly paid data scientists, which limits scalability and frustrates employees. Indeed, the productivity of employees across the organization can suffer: respondents to our 2019 Global Data Transformation Survey reported that an average of 30 percent of their total enterprise time was spent on non-value-added tasks because of poor data quality and availability (Exhibit 1).

While it's challenging to directly attribute value to data governance, there are multiple examples of its significant indirect value. Leading firms have eliminated millions of dollars in cost from their data ecosystems and enabled digital and analytics use cases worth millions or even billions of dollars. Data governance is one of the top three differences between firms that capture this value and firms that don't. In addition, firms that have underinvested in governance have exposed their organizations to real regulatory risk, which can be costly.

Building the foundation for effective governance

While many organizations struggle to effectively scale data governance, some have excelled. For example, a leading global retailer, whose data governance was managed within IT, struggled to capture value from data for years. Then, as part of an enterprise-wide analytics transformation, it invested in educating and involving the entire senior-executive leadership team in data governance. It assigned to each executive leader (CFO, CMO, and so on) several data domains, or business-data subject areas, some of which, such as consumer transactions and employee data, spanned multiple functions or lines of business.

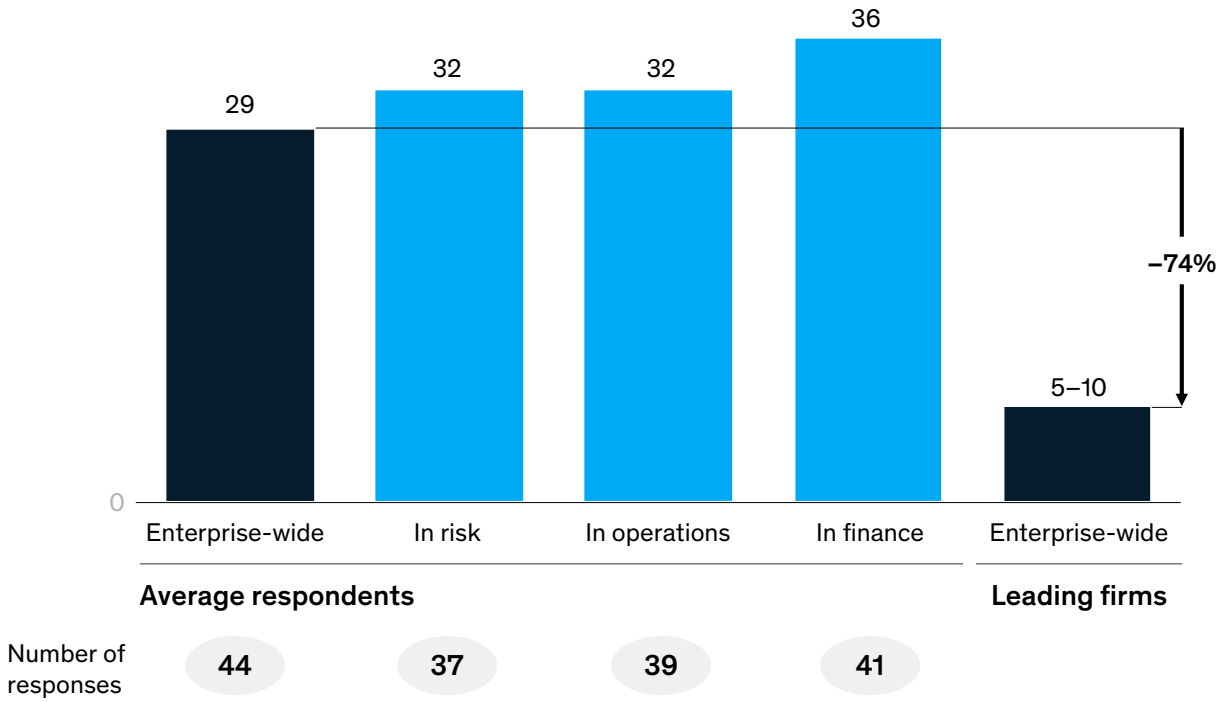
Once these leaders grasped the value of data governance, they became its champions. Within their domains, they selected representatives to act as data-domain owners and stewards and directly linked data-governance efforts to priority analytics use cases. They then worked in sprints to identify priority data based on the value they could deliver, checking in with the CEO and senior leadership team every few weeks. These efforts have begun to pay off, allowing the organization to stand up priority data domains over the course of a few months (versus years) and reduce the amount of time data scientists spend on data cleanup, accelerating analytics use-case delivery. The program continues to grow over time.

As the example demonstrates, effective data governance requires rethinking its organizational design. A typical governance structure includes three components:

- a central data management office (DMO), typically led by a chief data officer (CDO), with a targeted data strategy and governance leaders who set the overall direction and standards
- governance roles organized by data domain where the day-to-day work is done
- a data council that brings domain leaders and the DMO together to connect the data strategy and priorities to the corporate strategy, approve funding, and address issues

Exhibit 1

Lack of data quality and availability can cause employees to spend a significant amount of time on non-value-added tasks.



[1] Data sourcing, data aggregation, data reconciliation, data cleansing, manual reporting, etc.
Source: McKinsey Global Data Transformation Survey, 2019

This structure serves as the foundation for data governance, balancing central oversight, proper prioritization, and consistency while ensuring that the employees creating and using data are the ones leading its management (Exhibit 2).

Six ways to drive data-governance excellence

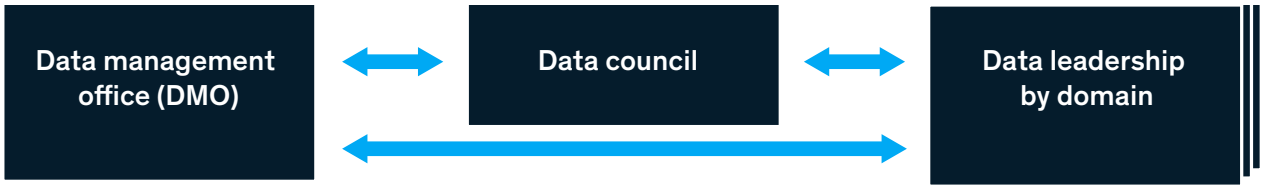
The organizational foundation alone, however, is not enough. Six critical practices are needed to ensure data governance creates value.

1. Secure top management's attention

As the aforementioned example highlights, success with data governance requires buy-in from business leadership. The first step is for the DMO to engage with the C-suite to understand their needs, highlight the current data challenges and limitations, and explain the role of data governance. The next step is to form a data-governance council within senior management (including, in some organizations, leaders from the C-suite itself), which will steer the governance strategy toward business

Exhibit 2

A best-practice data-governance model typically includes three organizational components.



- Defines policies and standards
- Empowers data leaders in the domains with tools, playbooks, training, etc
- Ensures coordination and consistency across key roles in the data life cycle
- Links data-governance efforts to technology and other related functions
- Provides targeted support for issues

- Owns strategic direction and principles
- Proposes DMO structure, defines domains, and assigns leaders
- Confirms adherence to standards and policies
- Reviews progress on initiatives and serves as funding gateway
- Resolves and/or escalates issues
- Chaired by head of DMO

- Sets and executes strategy of the domain—initiatives, external sources, etc
- Understands and meets needs of data consumers
- Owns and manages the data—assesses and improves data quality, defines definitions and data model, etc
- Participates in the data council

needs and oversee and approve initiatives to drive improvement—for example, the appropriate design and deployment of an enterprise data lake—in concert with the DMO.

The DMO and the governance council should then work to define a set of data domains and select the business executives to lead them. These leaders drive governance efforts day-to-day by defining data elements and establishing quality standards. Companies need to invest the time to introduce these leaders to their new roles, which are typically added to their primary responsibilities. They should understand the value they will generate in these roles and be armed with the skills they need, including an understanding of the relevant regulations and core elements of the data architecture.

Critically, having top-down business-leadership buy-in will avoid the usual challenges around role clarity and empowerment. Data stewards on the business side will understand that the effort is an enterprise priority and make time to address it (which might be facilitated by a shift in their performance metrics or an adjustment in their other responsibilities). Top-down mandates also make it possible to immediately address conflicts over data ownership.

Leading organizations also create tangible ways to track progress and value creation. For example, they can measure the amount of time data scientists spend finding, curating, or enabling data for priority use cases, or the dollar losses associated with poor-quality data and associated business errors. Tracking impact metrics like these helps ensure the attention and continuing support of top management.

2. Integrate with primary transformation themes

To ensure that governance efforts create value, link them directly to continuing transformation efforts that already have CEO attention, such as digitization, omnichannel enablement, or enterprise-resource-planning modernization. These efforts typically depend on data availability and quality.

Linking governance to transformation themes simplifies senior leadership buy-in and changes the organizational construct. Rather than governance running on its own, such initiatives shift data responsibility and governance toward product teams, integrating it at the point of production and consumption.

For example, a European retailer embarked on a digital transformation of its core business and a rapid extension of its online business, which required significant redevelopment of the e-commerce stack, including back-end platforms. Data was identified as a critical enabler, and a DMO and a data council were set up to develop the core framing on the future ecosystem, as well as the structure of data domains, including the strategic goals on managing data in the future.

Lead product owners, who were heading several digital-transformation squads in dedicated functional areas, became data leaders within their area of responsibility. Product owners became data-domain owners. For example, the product owner working to drive process improvements around in-store checkout owned the sales and payment domains. This structure ensured that governance efforts were oriented primarily to enabling business needs and that the leaders creating and consuming data were actively shepherding it.

3. Prioritize data assets and focus data leadership accordingly

Many organizations approach data governance in a holistic manner, looking at all data assets at once. But such a large scope means slow relative progress in any given area and a risk that efforts aren't linked directly to business needs. To succeed, data assets should be prioritized in two ways: by domains and by data within each domain.

The data council, supported by the DMO, should prioritize domains based on transformational efforts, regulatory requirements, and other inputs to create a road map for domain deployment. Then the organization should rapidly roll out priority domains, starting with two to three initially, and aim for each domain to be fully functional in several months.

For example, a North American retailer set a bold aspiration to transform the company over three years with advanced analytics. The company quickly realized that its current data would hold it back and established a DMO and data domains to scale governance. It identified ten domains across the enterprise and prioritized deployment of the first two—transactional data (logging in-store purchases) and product data (establishing a clear hierarchy of products and their details). This helped accelerate priority use cases around in-store assortment and inventory.

In addition to prioritizing domains, prioritize data assets within each domain by defining a level of criticality (and associated care) for each data element. Critical data typically represents no more than 10 to 20 percent of total data in most organizations. Critical elements, such as customer name or address, should receive a high level of care, including ongoing quality monitoring and clear tracking of flow across the organization, whereas for elements that are used less often in analytics, reporting, or business operations (such as a customer's academic degree), ad hoc quality monitoring without tracking may suffice. This significantly narrows the scope of governance efforts and ensures that they are focused on the most important data.

4. Apply the right level of governance

Data-governance programs can vary dramatically across organizations and industries. Leading organizations take a "needs-based" approach, adopting the level of governance sophistication appropriate to their organization and then adjusting the level of rigor by data set.

It's important to realize that data governance was largely first championed by banks under pressure from BCBS 239[1] and other regulations that required sophisticated governance models. Most other industries and organizations don't

[1] Basel Committee on Banking Supervision's standard number 239: "Principles for effective risk data aggregation and risk reporting."

face the same level of regulatory pressure, so the design of their programs should align with the level of regulation they uniquely face and the level of their data complexity. Organizations with multiple, distinct businesses spanning many geographies have more complex needs than those with a business in only one geography; similarly, a high pace of data change or low level of technology automation increases data complexity (Exhibit 3).

The most comprehensive governance model—say, for a global bank—will have a robust data-governance council (often with C-suite leaders involved) to drive it; a high degree of automation with metadata recorded in an enterprise dictionary or data catalog; data lineage tracked back to the source for many data elements; and a broader domain scope with ongoing prioritization as enterprise needs shift.

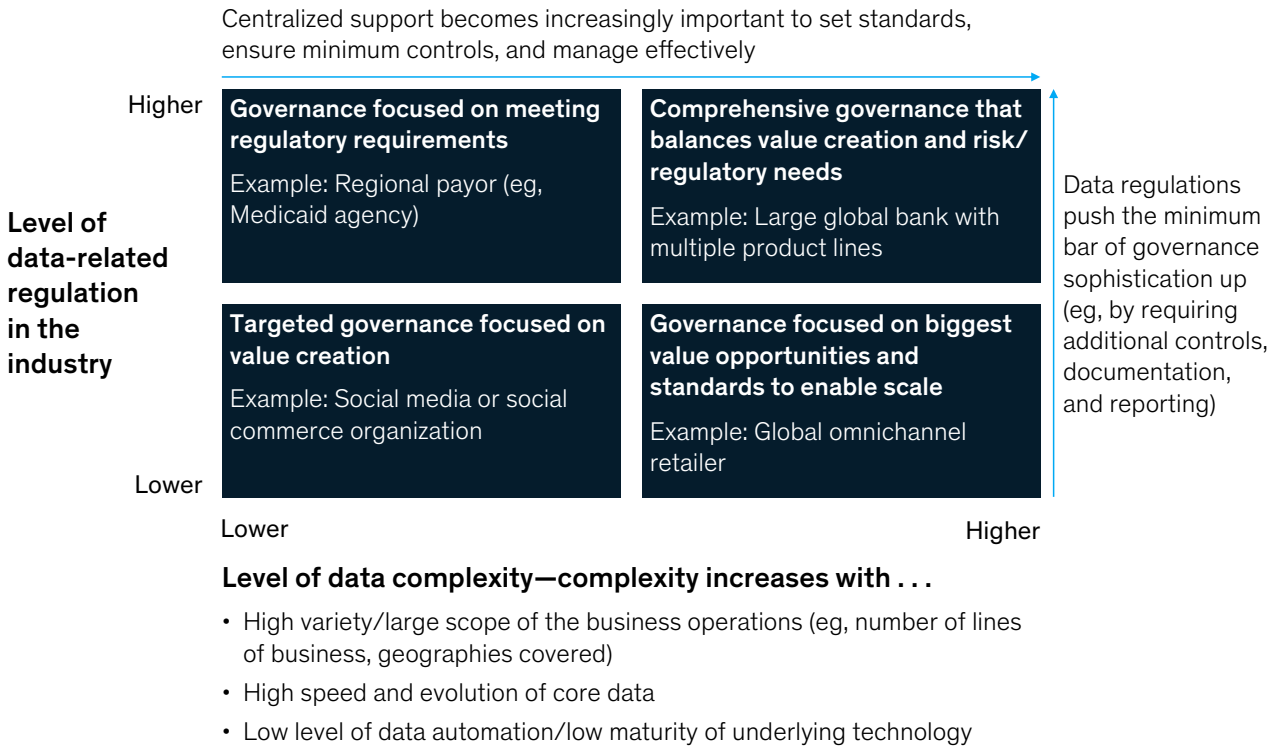
In contrast, targeted data governance for a regional technology company might have a data council that meets less frequently and includes C-suite leaders only periodically; metadata tracking that could even start in Excel; limited lineage tracking; and narrower domain scope, at least initially, to enable priority use cases.

In parallel with establishing the right level of governance for the organization as a whole, adjust the level of governance rigor across data sets. Many organizations' legacy data standards set conservative restrictions on quality and access across the board. This minimizes risk but can stifle innovation. Leading organizations consciously balance opportunities and risks and differentiate governance by data set.

For example, organizations can apply light governance for data that is used only in

Exhibit 3

Data-governance archetypes can be used to inform the level of sophistication needed.



an exploration setting and not beyond the boundaries of the science team. The team may also not need perfectly prepared and integrated data with full metadata available. Data masking may be appropriate to ensure privacy, together with strict internal non-disclosure agreements (NDAs). However, as soon as such data is used in a broader setting, such as in interactions with customers, stronger governance principles need to be applied.

An Asian financial institution took an aggressive approach to "free the data" using these principles. It agreed on the sensitivity level for each data set and was able to free the roughly 60 percent of enterprise data that was low risk, giving all employees access to use and explore it. On the other hand, highly sensitive data, such as personally identifiable information, was highly restricted both in terms of who could access it and how.

As organizations mature and their governance capabilities and technology continue to advance, scope becomes less important. A suite of tools is beginning to automate data-governance activities, and its coverage and cost-effectiveness will only improve over time. Both newer platforms, such as Octopai and erwin, and established organizations, such as Informatica and Collibra, are rolling out capabilities for automated metadata harvesting, lineage creation, data-quality management, and other governance functions.

5. Choose iterative and focused implementation

To ensure that data governance creates value fast, tailor governance priorities to the domain, and use iteration to adapt quickly. This goes beyond integrating governance with business needs, prioritizing use cases and domains, and applying needs-based governance; the key is to adopt iterative principles in day-to-day governance. For example, if there is a backlog of known data-quality issues, review and reprioritize daily, working to maximize the benefit to the business as priorities shift.

Push to enable priority use cases quickly even if the solution isn't perfect. Longer-term development to make use cases production ready (by integrating with the core customer-relationship-management and operational customer master data) can occur once value has been demonstrated. For example, enhancing customer campaigns may not require a fully integrated set of data across the entire enterprise, but rather a tailored approach in a dedicated platform. Data governance should support and accelerate this tailored approach, focusing on solving issues around availability and quality in addition to establishing strong master-data management.

6. Generate excitement for data

When people are excited and committed to the vision of data enablement, they're more likely to help ensure that data is high quality and safe. Leading organizations invest in change management to build data supporters and convert the skeptics. This can be the most difficult part of the program, as it requires motivating employees to use data and encouraging producers to share it (and ideally improve its quality at the source).

Successful organizations use a combination of interventions to drive the right behavior. These can include role modeling from the CEO and other senior leaders, recognition for high quality, responsive sources, and new demonstrated-use cases. Some organizations also offer training and qualifications, often as part of a larger academy approach, together with communicating about career opportunities in data jobs. Others have used successes in data and analytics to create excitement in the form of events, publications, or even data art.[2] What works is highly dependent on the culture of the organization.

[2] TED compiled a series of talks on data art: ted.com/playlists/201/art_from_data.

Getting started

Companies should begin their new data-governance approach by asking these six questions:

1. What is the opportunity cost of not getting data governance right in terms of missed upside, extensive time lost in manually cleaning data, or incorrect and suboptimal business decisions?

2. Who is leading governance efforts today, and what would it look like to elevate the conversation to the C-suite? Who should be involved?

3. Where is governance most important? What domains and parts of domains does the organization most need right now?

4. What governance archetype best fits the organization, and are current efforts aligned to that level of need?

5. How can governance be accelerated by adjusting its focus and injecting iterative working concepts?

6. Do you have the in-house capabilities to manage such a shift?

Data governance is critical to capturing value through analytics, digital, and other transformative opportunities. While many companies struggle to get it right, every company can succeed by shifting its mindset from thinking of data governance as frameworks and policies to embedding it strategically into the way the organization works every day.

Bryan Petzold is an associate partner in McKinsey's Silicon Valley office, **Matthias Roggendorf** is a partner in the Berlin office, **Kayvaun Rowshankish** is a partner in the New York office, and **Christoph Sporleder** is a partner in the Frankfurt office.