



GOOD DATA STARTS WITH GREAT GOVERNANCE

By Elias Baltassis, Antoine Gourévitch, and Lucas Quarta

IT WAS NOT THAT long ago that pundits were ringing in a new era of big data in which all of a company's information, together with the abundance of data available in the world, would come together in a glorious engine of growth for companies everywhere. To their credit, many organizations sat up and took notice. Today, it's accepted wisdom that data and analytics provide an essential tool in competitive differentiation.

Still, too many companies embrace data and leave it at just that. To be sure, leading companies in every industry—including retail, telecommunications, pharmaceuticals, and banking and insurance—have adopted advanced analytical methods and high-performance data-handling capabilities to improve cost performance and increase revenues.

Indeed, many have achieved small, early wins with pockets of analytics applications, but scaling those wins requires internal development of a full set of data capabilities.

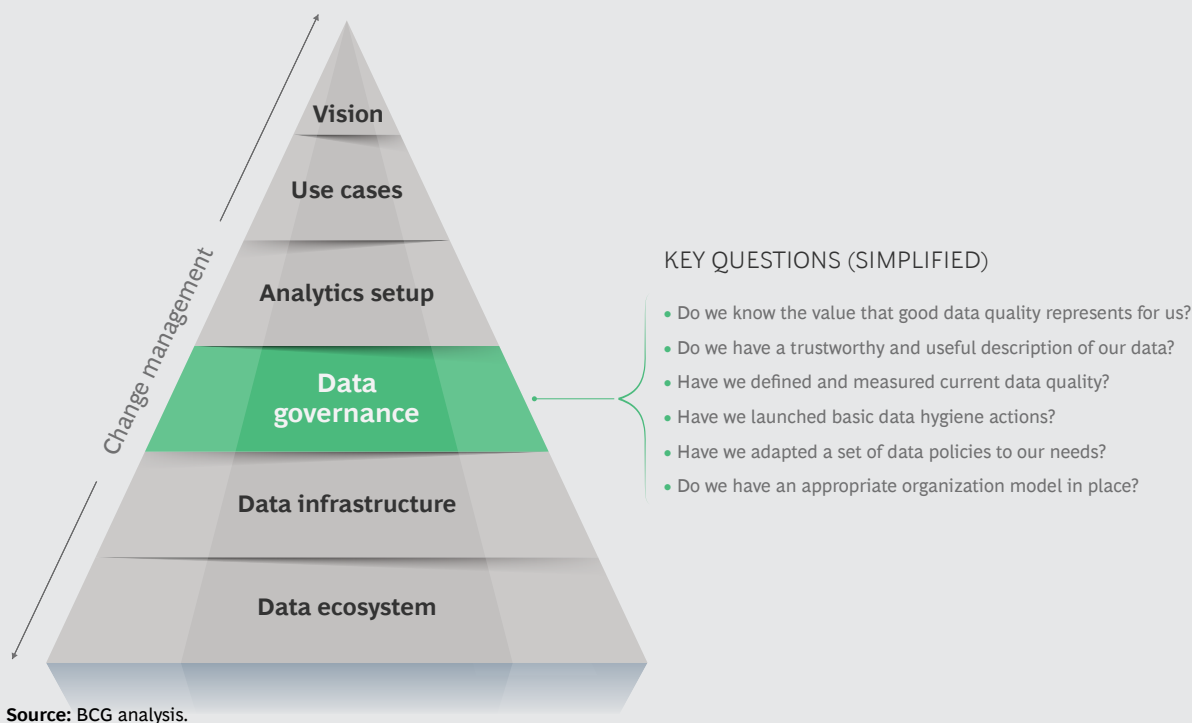
In the coming years, the companies with the best data capabilities—and best data quality—will dominate. On the basis of our client work, BCG has built a structured framework that defines the requisite capabilities for the transformation of a company's operating model and for achieving success in today's data race. (See Exhibit 1.) At the heart of the matter, data governance comprises four building blocks: data structures, data policies, data tools, and the organization's participants and target operating model (TOM).

Data Governance Affects the Entire Organization

Data governance may include organizational and technological elements that facilitate the sustainable improvement of a company's data quality. Our relatively broad definition of data quality includes data completeness, accuracy, consistency, accessibility, and the qualities that are important to the particular business and are ultimately determined by that individual company.

EXHIBIT 1 | To Fully Leverage Their Data, Companies Require Capabilities That Work Together in a Cohesive Operating Model

BCG'S DATA CAPABILITIES MODEL



Good data matters—not least for compliance and operational excellence. Some industries, such as banking and pharmaceuticals, are subject to regulations that compel them to improve their data in a sustainable way. But in addition to complying with regulatory obligations, clean data allows an organization to optimize efficiency, offer modern and streamlined customer journeys, anticipate continually evolving customer needs and desires through effective advanced analytics and artificial intelligence (AI), and even create new businesses.

Good data also offers innumerable opportunities for unlocking value. A major European bank freed up several billion dollars of capital after improved exposure data decreased its regulatory capital buffers. A leading consumer goods company generated hundreds of millions of dollars annually after implementing a hyperpersonalized approach that leveraged customer data by integrating it with external data in AI models. A global steel manufacturer reduced costs by hundreds of millions of dollars by

streamlining its supply chain data, integrating internal and external data flows, and using the data to optimize its operations. Properly implemented, data governance that is focused on data quality can help companies reap significant quantifiable benefits.

We hear too many complaints about the quality, consistency, and accuracy of company data, as well as the difficulty of accessing it. In the same breath, many blame problematic data for their inability to leverage advanced analytics and AI. In a recent benchmarking study that covered more than 600 companies, BCG found that more than 60% of those companies assessed their data governance capabilities at various levels of underdevelopment.¹

Organizational Obstacles to Overcome

What's keeping companies from developing and embedding data governance that can improve the quality of their data?

There are three primary culprits:

- **Expectations That Run High.** As companies absorb the positive impact of analytical and AI applications, they develop a new appetite for higher data quality and faster delivery speeds.
- **Resources and Capabilities That Run Low.** Most companies have relatively limited experience using a company-wide data governance function. As such, there are very few executives—either inside or outside the organization—with the necessary experience and expertise to drive a data governance operation.
- **The Lack of a Shared Understanding of Data “Ownership.”** Failing to recognize that data is a critical *business* asset may be the biggest problem of all. Many companies still believe that anything that contains the word “data” belongs exclusively to IT. As such, they adopt an IT-driven approach to data governance that misses the point, leading to a disconnect from the potential business value it can create.

The failure to implement data governance has very real and long-lasting ramifications. We’ve seen companies embark on multi-year projects, struggle with regulations, institute governance rules that aren’t needed, and spend millions of dollars on tools with-

out moving the needle much on improving data quality or creating value.

The Four Building Blocks of Data Governance

Best practices create the optimal environment for successful data projects. (See Exhibit 2.)

Data Structures. This is the starting point. Data structures help create an inventory and shared language around data. They include descriptions of the company’s data, defined and classified in glossaries, domains, families, models, dictionaries, and flows. This isn’t exactly new: companies have been using such tools for decades. More often than not, however, they are outdated, incomplete, and not fit for purpose.

The most important of these structures are data glossaries and data domains, which help define, organize, and assign the management of company data.

A *data glossary* serves as an important exercise in semantics. The glossary is a list of the terms by which the company’s data is categorized, so the selection of the terms is crucial to the way the business is conducted and aligns the organization on their meaning and use. For example, a leading retail bank spent several months moving

EXHIBIT 2 | The Building Blocks of Data Governance

1. DATA STRUCTURES

- MECE data domains (and their owners), families, business data objects
- Prioritization of data domains and families on the basis of the company’s projects
- Data glossaries, models, dictionaries, flow maps

4. DATA ORGANIZATION PARTICIPANTS AND THE TOM

- Key stakeholders, mandates, R&Rs for each participant
- Decision framework, specifying different levels of decision-making authority

2. DATA POLICIES

- Rulebooks on, for example, data quality, data documentation, data accessibility, master data management
- Policies that correspond to specific issues and root causes; no bureaucratic “Ministry of Data”

3. DATA TOOLS

- Basic data hygiene tools supporting the data dictionaries and flow maps
- Advanced data management tools adapted to the company’s needs: MDM, lineage, KQI automation



Source: BCG analysis.

Note: MECE = mutually exclusive, collectively exhaustive; MDM = master data management; KQI = key quality indicator; TOM = target organization model; R&R = roles and responsibilities.

from “individual” customers to “household” customers and deciding which data should be associated with them. A global automaker needed a similar effort to define and categorize “spare parts.” A leading luxury goods company had trouble with “points of sale.” Fully owned boutiques are, of course, points of sale, but what about multiple stands in a department store? Does each serve as one or as one of several points of sale? Should they be described by the same data? And how are “sales” defined? Nominal revenues? Nominal revenues minus discounts? Minus commissions? These are some of the concerns that data glossaries address.

Data domains focus on where data resides and specify ownership of the data defined in the glossary. This high-level, unapologetic classification allows for no gaps and no overlaps. Achieving this is easier said than done. Some critical data is used in varying use cases and with different understandings of its meaning and purpose. It thus seems owned by many people or corporate entities. For example, who owns the customer data? The sales department? Marketing? Finance?

Once the critical data has been classified, data owners are assigned to each domain and given ultimate responsibility for all data quality decisions, taking into account the specific needs for the data quality of their business area. It’s worth noting how important it is to untie the overlapping and conflicting uses of data and assign all data to well-defined domains. When data is spread across domains, responsibility is diluted and decisions and actions necessary for the domain’s quality are easily overlooked.

Data domains also play useful roles in identifying areas for improving data quality. Matching domains—and their subdivisions and families—to data used by various ongoing and near-future projects and use cases is an easy and very effective way to prioritize areas in which data governance actions should be focused. (See Exhibit 3.) This approach has, therefore, a broad impact on the efficiency of an organization.

A *data dictionary* (not to be confused with a data glossary) focuses on data about the data—metadata. Metadata can be descriptive and related to data models, security restrictions, quality indicators, and data governance. A data dictionary is either passive (maintained separately from the dataset it describes) or active (updated automatically when the dataset structure changes).

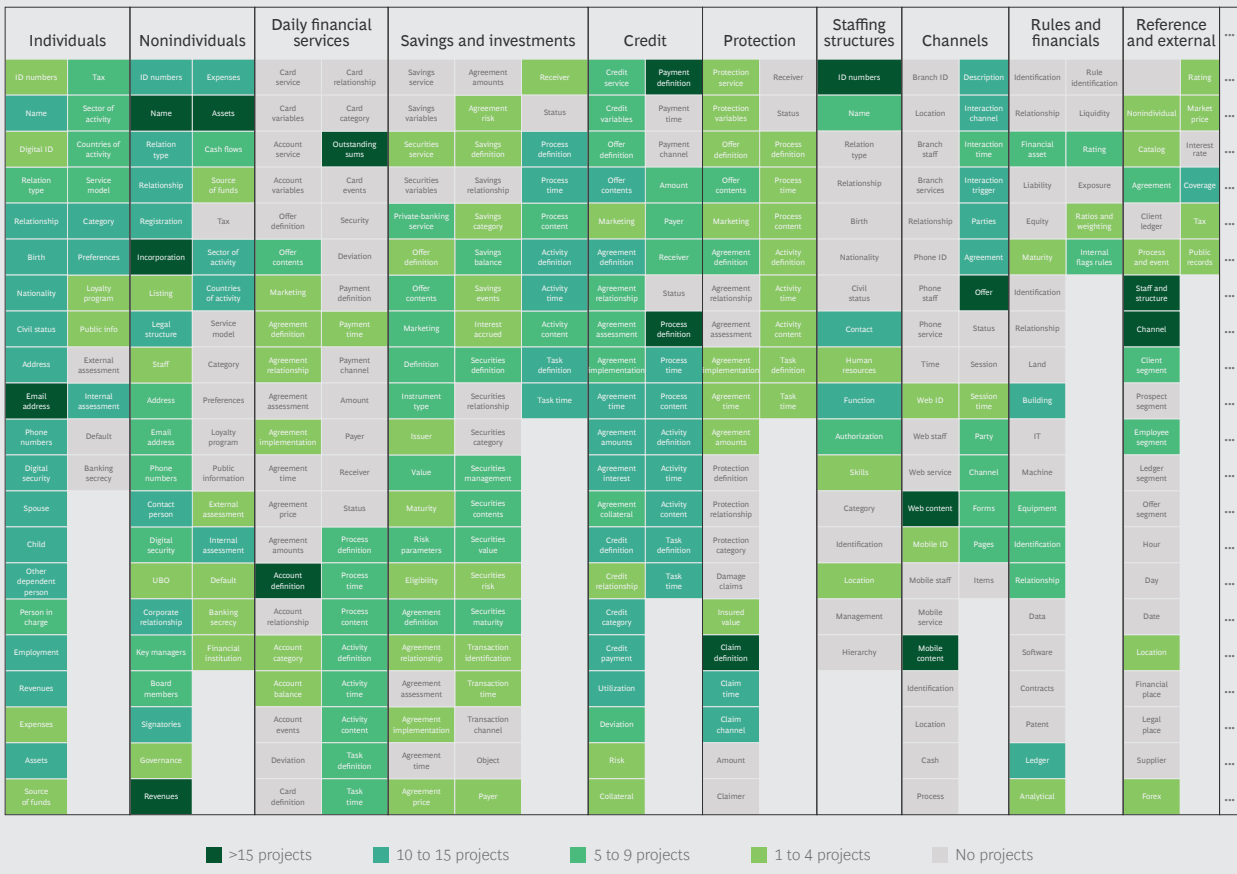
Data flows help stakeholders understand the paths data takes in an organization. They provide visibility on data location and status, simplifying error tracing back to the error source. Data flows have become increasingly important for the data governance of companies subject to the European Union’s General Data Protection Regulation.² Several of the rights guaranteed by the EU regulation are offered only with up-to-date data flows.

Data Policies. Once a company has organized and clarified its data, it should develop rules related to, for example, its processes, actions, roles, and budget allocation principles. We have seen several companies take the wrong approach to this, launching multiyear efforts to develop policies about everything and anything. This invariably leads to a bureaucratic and inefficient “Ministry of Data” effort. Because such initiatives generate no value, top management generally discontinues them.

Our position on this point is clear: all companies need clear and unambiguous definitions of data quality, measurement guidelines, and the key roles and responsibilities that govern improvements.

At least two dozen sets of criteria are available to describe data quality. They range from the very basic that are easy to measure with simple key quality indicators (KQIs), such as “completeness,” to the very complex that can be measured only by applying sophisticated business rules, such as “accuracy.” Which and how many should a company use? Opt for simplicity. On the basis of the needs and uses of data, companies should select a small number of criteria and enshrine them as the data quality definition.

EXHIBIT 3 | A Banking Prioritization Heat Map: Data Domains and Data Families in Current Projects



Source: BCG case experience: sanitized extract.

A company should next specify the ways these criteria will be applied and measured and establish a baseline for the current situation. It's important to note that criteria are not equally important across domains. For example, the quality of customer email addresses is likely less important than the quality of sales data. While the criteria remain the same, the KQI objectives may be different, reflecting the various levels of quality needed for each type of data. It is the role of all data owners to define the quality objectives of their specific domains.

The data quality policy must also lay out the processes by which KQIs are defined and measured, the roles and responsibilities of the people participating in their improvement, the budget allocation process for these activities, and, more generally, all organizational and technical elements necessary for continuously monitoring and sustainably improving the quality of the data.

Other policies may or may not be called for. Should a company have problems with its reference data (a resource for other data), it probably needs a master data management (MDM) policy; if it has no referential problem, it does not need an MDM. Companies that are dealing with data accessibility problems should create a user access rights policy.

As a general rule, data policies must correspond with and address the company's specific data issues and their root causes. They are a means to an end and play important roles in the smooth functioning, consolidating, and streamlining of data efforts. No more than that. Forgetting this can lead to unnecessary effort, the waste of resources, and excessive bureaucracy. We have seen companies developing numerous data policies, but it is quite rare that more than four or five policies are needed for effective data governance.³

Data Tools. Just as data policies require the simplicity approach, the approach to data governance has no need of myriad tool sets. Still, it is important to consider two types: basic data hygiene and advanced tools.

It's possible to build a list of data domains or a data dictionary using a spreadsheet or to represent a data flow using any kind of graphic software. However, sharing and updating such elements across a community of several hundred users in a few or many locations can be cumbersome and overly complex, if not outright impossible.

Enter data hygiene tools. These relatively basic, built-for-purpose tools help companies build and maintain their data structures, data glossaries, dictionaries, and flows efficiently.

Advanced tools perform sophisticated tasks. Some tools, for example, those focused on a specific area such as MDM or data lineage, serve a single purpose, while others that are multipurpose cover the full spectrum of data functions, such as calculating KQIs and workflow management.

Implementation of data tools is not a panacea and can consume resources that could be better employed in other data-related tasks. Data tools should be adapted to each company's needs and should help enforce its data policies.

The Data Organization's Participants and Target Operating Model. Successful data governance requires multilayer management that is focused on business but spans both business and IT. It is typically built around chief data officers and their teams—the only people specifically dedicated to data governance—plus a data governance council (DGC), data owners, data stewards, and data custodians.⁴

- **Data Governance Council.** The DGC is the overarching data-related decision-making body that delegates authority to the chief data officer (CDO) and appoints data owners. It defines the

company's data strategy and sets priorities for data governance objectives, standards, and policies and resolves issues escalated from other levels of the data organization.

The DGC typically includes all senior data stakeholders: key data owners (typically, heads of business units or corporate functions), the chief information officer, the data protection officer, the chief information security officer, and the CDO. It is usually chaired by the chief operating officer or the equivalent. The CDO is responsible for setting the agenda and executing strategies and decisions made in the council.

- **Chief Data Officer.** The CDO, with overall responsibility for data governance, drives the definition and coordination of the data governance framework and its components. The CDO proposes the agenda, helps prioritize business requirements, coordinates implementation, and helps transform data into a valuable asset. CDOs and their small teams are generally the only full-time, dedicated resources of the data governance organization.
- **Data Owners.** As noted earlier, data owners have oversight of data domains, holding ultimate responsibility for content and quality and, as such, deciding the appropriate level of quality and ensuring that data governance policies are applied. Data owners are typically C-suite executives, but some, depending on the size of the company, delegate responsibilities to data owners by business or region. This is mainly an undemanding decision-making role, requiring, in most cases, less than 5% of the C-suite executive's time.
- **Data Stewards.** Data stewards, part of each data owner organization, report to the data owner, working to achieve the data quality business objectives defined by the data owner. Data stewards are responsible for applying data policies and standards in their domain and for providing guidance to IT.

Depending on company size and the domain for which data stewards are responsible, several of them may support a data owner, dedicating a significant amount of their time (roughly 20% to 30%) to data-related tasks.

- **Data Custodians.** Unlike all other data governance participants, data custodians are part of the IT function. They work with data stewards, supporting their efforts to implement specific data quality measures. Many data custodians are attached to a specific IT system or database, so there may not be a one-to-one relationship, and a single custodian may work for many stewards or, inversely, a single steward may work with many custodians.

And what should data governance's TOM be? We see four main organizational archetypes:

- **Totally Decentralized Model.** Business units define data governance independently. There is no overall data strategy and very little coordination. This model allows businesses to have direct control over their data and bring agile ways of working into changes in data management. However, it can lead to inefficiencies in data operations when teams “reinvent the wheel,” generating data inconsistencies across the various business units and functions. This archetype is becoming less common, persisting only in the rare companies that have not yet entered their data transformation journey.
- **Totally Centralized Model.** At the other end of the spectrum are companies that have implemented a highly centralized organization model, in which all data-related issues are handled by the central team that includes a data “czar” and the owner of all company data—usually the CDO. This model is easy to explain to the organization and rather effective, but because it is difficult to implement and scale in large organizations, it slows the pace of change.

- **Federated Model with the CDO as “Doer.”** The CDO directly manages a relatively large team responsible for most data operations, while business units' data functions work under the strong guidance of the central team. This model is efficient and quickly improves data quality, but it sometimes faces negative reactions from business units feeling a loss of independence. It can thus be difficult to establish the CDO mandate.
- **Federated Model with the CDO as “Facilitator.”** The CDO unit is much smaller and focuses on coordinating and facilitating actions decided by business units. Business units have full ownership of their data, but their data governance staff maintains a dotted reporting line to the CDO. This model retains the responsibility for data in the business, but it also ensures good coordination and relatively rapid improvement of data quality.

It is not uncommon for companies to begin by using a federated model with a doer CDO and, as the organization matures and builds its capabilities, to evolve to a CDO-as-facilitator model.

OVER THE PAST few years, data has been established as a fundamental source of business value. Companies compete in an environment characterized by enormous—and growing—data sets, stringent data regulations, and frequent data-powered disruptions. In this context, data governance—and the resulting improvement of data quality—provides a way to achieve not just short-term results but also to embed data in the organization and succeed in its data and analytics journey.

It just requires some attention and dedication to a data transformation. Any company can get it right.

NOTES

1. BCG Data Capabilities Maturity Assessment survey, March 2019.

2. The General Data Protection Regulation, EU-2016/679, covers data protection and privacy for each individual citizen of the EU and the European Economic Area. It also addresses the transfer of personal data outside those areas.

3. We explicitly exclude the security policy and the data privacy policy: they are not always part of data governance or the responsibility of the chief data

officer. In many cases, the former is the responsibility of the chief information security officer and the latter, of the data protection officer.

4. Exact titles and granularity of roles of data governance differ from company to company, but they should result in the same thing: a stronger data foundation.

About the Authors

Elias Baltassis is a partner and director in the Paris office of Boston Consulting Group, a core member of Gamma and the Technology Advantage practice, and a global coleader of the data governance topic. You may contact him by email at baltassis.elias@bcg.com.

Antoine Gourévitch is a managing director and senior partner in the firm's Paris office, a core member of the Technology Advantage practice, the global leader of the Data & Digital platform, and the global leader of the digital and deep-tech topic. You may contact him by email at gourevitch.antoine@bcg.com.

Lucas Quarta is a partner and associate director in BCG's Paris office and a core member of the Technology Advantage practice and a global coleader of the data governance topic. You may contact him by email at quarta.lucas@bcg.com.

Boston Consulting Group partners with leaders in business and society to tackle their most important challenges and capture their greatest opportunities. BCG was the pioneer in business strategy when it was founded in 1963. Today, we help clients with total transformation—inspiring complex change, enabling organizations to grow, building competitive advantage, and driving bottom-line impact.

To succeed, organizations must blend digital and human capabilities. Our diverse, global teams bring deep industry and functional expertise and a range of perspectives to spark change. BCG delivers solutions through leading-edge management consulting along with technology and design, corporate and digital ventures—and business purpose. We work in a uniquely collaborative model across the firm and throughout all levels of the client organization, generating results that allow our clients to thrive.

© Boston Consulting Group 2019. All rights reserved. 11/19

For information or permission to reprint, please contact BCG at permissions@bcg.com. To find the latest BCG content and register to receive e-alerts on this topic or others, please visit bcg.com. Follow Boston Consulting Group on Facebook and Twitter.