



THE BOSTON CONSULTING GROUP

METADATA ISN'T JUST FOR TECHIES ANYMORE

By Rash Gandhi, Antoine Gourévitch, and Nic Gordon

COMPANIES THAT PURSUE DATA-DRIVEN transformations often overlook a key ally: metadata. And little wonder. Metadata—essentially “data about data”—isn’t sexy, is often poorly understood, and tends to be treated as a purely technical topic, a sort of secret language for the folks in IT. But pigeonholing metadata is a mistake. Metadata, when integrated with business and operational perspectives, becomes far more powerful, helping companies extract the *full* value of their data.

An integrated approach to metadata helps companies classify information and ensure its consistency and trustworthiness. It lets them better understand the lineage of data—how a piece of information may have been manipulated or changed over time. It gives them a more complete—and more immediate—view of what data exists within the organization, what it means, and where it can be found.

This can boost efficiency and speed immensely. Data scientists can devote themselves to creating models that generate in-

sights, for example, instead of spending 70% to 80% of their time searching for and massaging data so that it can be used in a new way. When metadata is managed with both business and technical views in mind, companies no longer have to reinvent the wheel with each new data initiative. Transformations become more sustainable and more potent. Companies unlock value they were previously leaving on the table.

Pitfalls and Unrealized Potential

Today, virtually every company is looking to data to drive its business forward. Information is at the heart of so many transformations. Ideally, a single piece of data could be used in different ways by different stakeholders across the company. But in practice, two main challenges hinder such multipurpose use:

- **A Proliferation of Data That Hasn’t Been Harmonized.** In this scenario, every business unit or function—such as marketing and customer support—is generating its own data and using its

own terminology to describe what the information means.

- **A Lack of Knowledge and Understanding of the Data That Exists Within the Organization.** The data that a business unit or function needs may indeed exist, but it is effectively invisible or unusable. Prospective users may not know that the information they seek has already been captured. Or they may not know its lineage—and as a result, they have little confidence in the data.

In many cases, these challenges come as an unpleasant surprise. Companies, after all, tend to put a lot of thought and effort into data-related issues: how they deploy technologies that use or generate data and how they address the concerns of customers and regulators. To be sure, these are important areas that warrant a company's attention. But often, the data management piece gets short shrift—and inefficiencies and added complexity result.

This is a real problem that plays out across all industries. The examples of organizations that end up paying some kind of price—a lost opportunity, additional costs, or disappointing results from a new initiative—are numerous. Consider just a couple of real-world cases:

- Seeking to optimize its balance sheet, a financial company took a close look at how it performed its risk-weighted asset (RWA) calculation. The company discovered that it was holding far too much regulatory capital—funds that had to be set aside to cover potential losses. The reason: low confidence in the numbers produced by the RWA process. Why the lack of faith? Each business unit was performing its own RWA calculation in its own way using its own data sources—and then manually adjusting the resulting number to what was considered “correct.” All the nuances and manipulation made it hard, at a company level, to see the data lineage—how, exactly, each business unit came up with its figure. As a result, there was

uncertainty regarding the enterprise-wide RWA number; to be safe, the company overallocated capital. The additional capital was now sitting on the sidelines instead of being used to generate revenue.

- A consumer goods company decided to replace its outdated CRM platform with a more modern one that could provide a comprehensive view of every customer (including all transactions the customer has had with the company). But because customer data was spread across multiple legacy systems, and often duplicated and inconsistent, the company had a poor understanding of where specific information resided and what data should be integrated with the new platform and what data should be discarded. Consequently, the company was unable to tie together its data—and while the new CRM platform won raves for usability, it was unable to meet its prime objective of giving a 360-degree view of the customer.

Data management is no simple thing. It requires companies to govern data from cradle to grave: acquisition, modification, retention, and deletion. (See “How to Avoid the Big Bad Data Trap,” BCG article, June 2015.) It means managing access to the data—and ensuring that security and privacy are handled responsibly. Metadata management is just one piece of a holistic approach to data management, but it is especially important for two reasons: it is the piece that is most often overlooked, and it is the piece that specifically addresses the data challenges discussed above.

Metadata Is the Who, What, When, and Where of Your Data

Many companies approach metadata from a technical perspective—and only from a technical perspective.

To be sure, technical metadata is essential stuff. It describes how data is structured, including its format (relational, hierarchical, or columnar, for example), its physical location, table and field names, and allowable

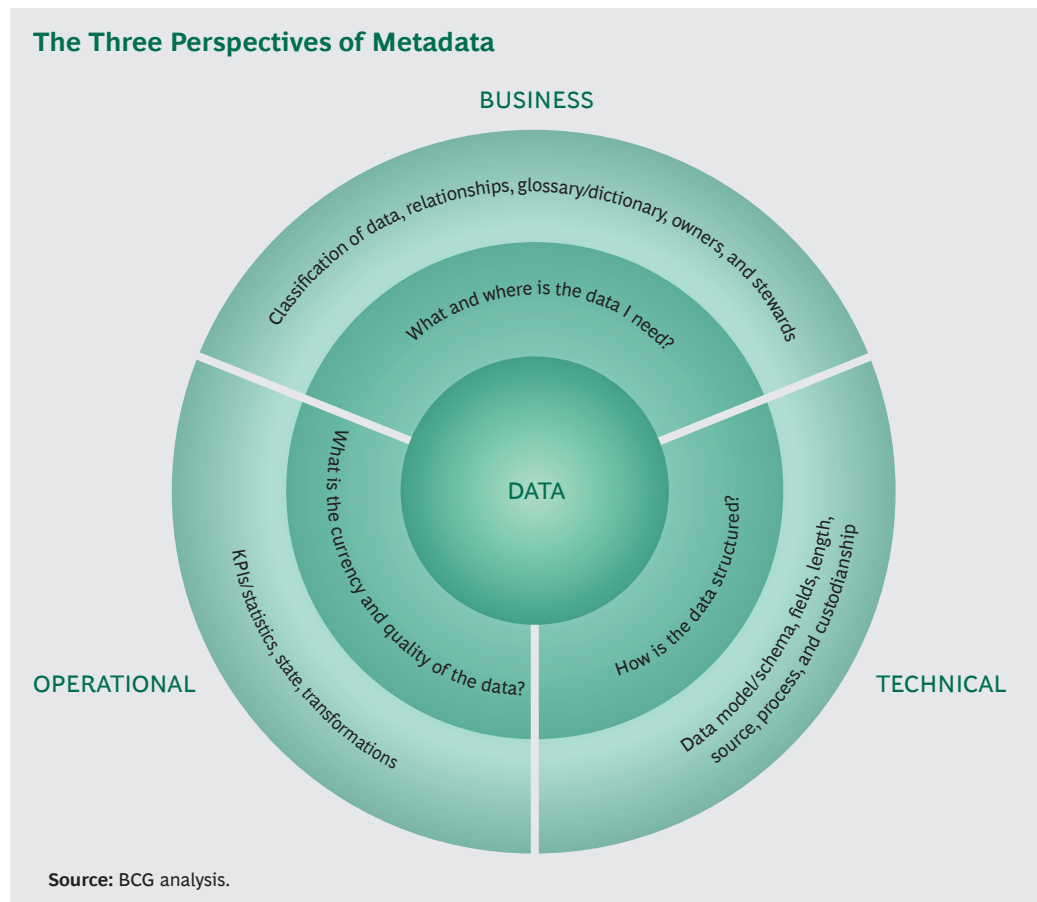
values. It enables companies to create databases that have the right data inserted in the right fields in the right tables. But it only goes so far in unleashing the value of data.

Integrated metadata—the kind that helps companies continually and fully extract value from information—includes business and operational perspectives as well. (See the exhibit.) It doesn't just tell you where you can find a certain piece of data, it also helps you trace the path the data has taken since it was first created—and see how it was transformed as it flowed across systems. This lineage is a key characteristic—and benefit—of good metadata management. It allows companies to know what calculations and manipulations were applied to the data and to determine whether the final value makes sense and can be trusted.

And by swinging the metadata pendulum from a purely technical to an integrated form, companies will mobilize their business and operational arms in generating value from data.

About Business Metadata. Business metadata provides a meaningful business context to the data: what the data is, who “owns” it, who can access it, and so on. It enables the various departments and business units of a company (and the data analysts, business analysts, and data stewards working with or within them) to understand, find, and use any piece of information stored anywhere in the organization. For example, one piece of data might simply be a number: say, 229215941. Business metadata would reveal that this number is, in fact, a customer ID.

The key to using business metadata is consistency. This can be harder than it might seem—within a single company, different stakeholders will tend to use different terms for a common concept. For example, within a bank, one business unit might refer to a “customer” while another might label the same individual a “consumer” and still other units might use terms like “client,” “beneficiary,” “counterparty,” or “obligor.” So a starting point for metadata



management is to gain agreement on the meaning of business data (here, for instance, all of the different labels refer to an individual with a product or service relationship with the bank). Companies can then create a master glossary to map the different terms to a common definition.

This sort of “stitching” is no modest task, but the payoff can far outweigh the effort. Individual business units or departments can work the way they always have, using the terms they always have. Yet at the same time, their data becomes far more useful—and more readily available—to other stakeholders who might be able to extract value from it.

About Operational Metadata. Operational metadata describes characteristics such as data quality, lineage, and currency (incorporating time stamps, for example). This kind of metadata is of particular value to the data stewards and custodians who are playing an increasingly important role within businesses. It can enable notification methods to alert interested parties as data elements are debated, agreed to, and potentially modified. It can also help identify—and correct—errors.

Consider, for example, a user who wants to know why only 50 customers were selected for a marketing campaign intended for 100 customers. Operational metadata would help guide that user through the underlying process—and discover where and why things went wrong. For example, say customers were to be selected if their age fell within a certain range. Tracing back through the metadata, one might see that in 50 instances, the data that was used contained an incorrect age—or was missing an age altogether. As a result, 50 customers mistakenly fell out of the targeted range. Quickly revealed, this error can be quickly addressed.

Strategies and Best Practices

Metadata management is a complex and demanding undertaking. Indeed, that’s one reason why companies avoid it. But there are ways to smooth the path. We’ve found

the following practices to be particularly helpful.

Don’t boil the ocean—instead, take an incremental approach. A sure-fire route to frustration—and often failure—is to try to capture metadata in a big-bang sort of way, looking at every piece of data throughout the organization. For most businesses, that’s bound to be an onerous undertaking that will never reach a conclusion, with momentum lost somewhere along the way.

A better way to do it: take a more focused approach. Start small with a new initiative that creates or leverages data. Capture metadata for all the information within the scope of the initiative. Project by project, you’ll build momentum in a disciplined way.

Embed responsibilities into existing roles, instead of creating new roles (and hiring new employees). Most organizations already have employees who specialize in data-related work—data stewards, data architects, and data administrators, for example. Instead of hiring new full-time employees to oversee metadata management and enforce policies, companies should integrate these new responsibilities into existing roles. This not only saves money, it puts metadata management into the hands of those who already have significant insight and knowledge regarding company data. It is often far more efficient to have the person who is already capturing data capture the metadata as well than to have a second individual tackle the job from scratch.

Embed clearly defined processes. Metadata needs to be managed throughout the life cycle of data. So, processes should be clear, and they should cover the capture, approval, registration, publishing, and use of metadata. They should ensure that data held across multiple sources is described completely, consistently, and unambiguously—and make it easy for users to rapidly identify and locate the data relevant to their needs. KPIs should be defined and employed not only to measure performance but also to create a culture where metadata capture is embedded across the organization.

Leverage tools and structures to facilitate access to metadata. Tools are available not only to capture metadata but also to work with it, making it easier for business users to find the data they need. By deploying such tools, companies can take an important step toward creating a self-service capability for users—a worthy goal.

At the same time, organizations with legacy systems need to understand that while they may strive for a centralized metadata repository, the likely reality—at least for the short term—is that their metadata will be scattered across multiple local repositories (perhaps dozens or even hundreds of them). Instead of looking at all of these locations—an inefficient approach to working with metadata—companies should create an integrated metadata layer. This is an interface that links the different repositories and creates, in effect, a virtual centralized repository. The physical location of metadata becomes irrelevant for the end user, who can see, in one place, all the company’s metadata—even if that metadata is actually spread across the organization.

Lessons from Early Adopters

An enterprise metadata repository can be implemented using internally developed databases or by purchasing a commercial solution. When requirements are not too complex, a homegrown solution can be implemented more quickly—and, in turn, deliver benefits more quickly—than a vendor’s offering. Often, it can also prove less expensive to implement and maintain. An example: A major utility company built a bespoke solution to manage the petabytes of customer data stored in legacy systems—and believes that it achieved an 18- to 24-month lead on the industry as a result. Indeed, the effort was so successful that the company launched a subsidiary to take this technology to market.

But there’s a flip side: a vendor solution can provide substantial out-of-the-box functionality. Most products include prebuilt meta-models; scanner functionality to capture metadata from traditional sources, such as

database catalogs and generic spreadsheet-loading facilities; and front-end user interfaces (with the caveat that these will almost always require customization). Companies need to be careful, though, that as they adopt a vendor’s solution they don’t also adopt the vendor’s view of their business. No one, after all, understands your business better than you do.

How to successfully integrate a vendor’s solution? The first step is to understand the metadata you want to capture. From a practical perspective, you won’t be able to describe in metadata everything about every piece of data you possess. So don’t try. Instead, focus only on the metadata you need, such as that which describes data elements that support a key business process (for example, customer gender for use in campaign management). Once you’ve done this, you should map specific requirements to product capabilities—and land only on a package that matches.

When carefully approached and implemented, an existing solution can work well. Netflix, for example, uses the open-source tool Metacat to edit metadata in a highly effective way. The company can capture a user’s last watched movie and then automatically update the metadata across various data repositories. And because Metacat’s query interface can interact with other data platforms (such as Apache Hive and the Teradata database), it effectively creates a single integrated metadata management system.

METADATA IS A long-term ally in generating value from your data. It ensures fast access by the right people to the right data and helps companies use their information—continually—in new and profitable ways. Managing metadata can be a complex and challenging task, and too many companies have put it low on their agenda. Resist the temptation. Give metadata the attention it deserves and it will give you a smoother path to implementing—and succeeding with—the technologies and business models that are at the top of your to-do list.

About the Authors

Rash Gandhi is a principal in the London office of The Boston Consulting Group. You may contact him by email at gandhi.rash@bcg.com.

Antoine Gourévitch is a senior partner and managing director in the firm's Paris office. You may contact him by email at gourevitch.antoine@bcg.com.

Nic Gordon was formerly an associate director in BCG's London office.

The Boston Consulting Group (BCG) is a global management consulting firm and the world's leading advisor on business strategy. We partner with clients from the private, public, and not-for-profit sectors in all regions to identify their highest-value opportunities, address their most critical challenges, and transform their enterprises. Our customized approach combines deep insight into the dynamics of companies and markets with close collaboration at all levels of the client organization. This ensures that our clients achieve sustainable competitive advantage, build more capable organizations, and secure lasting results. Founded in 1963, BCG is a private company with offices in more than 90 cities in 50 countries. For more information, please visit bcg.com.

© The Boston Consulting Group, Inc. 2018. All rights reserved. 3/18