DECEMBER 2016

© Heizfrosch/Getty Images

BUSINESS TECHNOLOGY OFFICE

Using agile to accelerate your data transformation

Data-management issues can undermine companies' ability to create value from analytics. Some businesses are using agile principles to make sure data are available when and where they are needed.

Chiara Brocchi, Brad Brown, Jorge Machado, and Mariano Neiman

Data have become core strategic assets in most organizations, and data management has become a top priority for most C-suite leaders. The structured and unstructured information companies collect about people and processes has the power to spur cutting-edge customer acquisition and retention strategies. It can reveal areas where processes could be made more efficient, and it can help business leaders make better decisions that reduce overall organizational risk. Data are the cornerstone upon which companies are launching their digital transformations—investing in analytics capabilities, machine learning, robotics, and other technologies to boost their odds of success.[1]

Indeed, companies are spending hundreds of millions of dollars to transform their data-related IT infrastructures and processes. But for most, the benefits of doing so have been limited to discrete areas of operation. Data-migration solutions created for specific business units or functional areas—such as the ad hoc repositories built for financial or administrative data—have been difficult to replicate company-wide because there is no end-to-end logic or central governance associated with them. Critical business information remains trapped in isolated systems.

Additionally, most companies face a big talent gap when it comes to data management. There is typically limited expertise within IT and business groups about newer data-migration technologies, capabilities, and architectures, as well as approaches to data delivery. And the relevant subject-matter

Takeaways

Companies collect more data than ever before, but most are still trying to determine how best to turn all that information into relevant business insights. An agile approach to data management can help companies gain this capability.

Agile data is similar to agile software development: it relies heavily on joint collaboration between business and IT. Cross-functional teams work in "data labs" that are focused on generating reliable insights from data sources. These labs develop minimally viable products and systems for transmitting data rather than waiting for "complete" solutions to emerge.

To get started, companies should use pilot projects to gain buy-in for agile data. They should empower the teams that work in the data labs to make some decisions unilaterally. They should make the necessary updates to technology infrastructure. And they should introduce new forms of measurement and communication to ensure that data-transformation projects stay on track.

experts, who could help define best practices in data migration and transformation, tend to be as siloed as the data flows they oversee.

In too many companies, the benefits of data remain undefined. About 60 percent of banks, for instance, say they have never quantified the potential value to be gained from investments in data-migration tools and capabilities.[2] Without a clear vision and outcomes-based metrics to guide executives' strategies and decisions, data-transformation projects can drag on for years.

Businesses must generate analytics-based insights much faster than that. They need a coordinated data-management program that explicitly involves the business and can be deployed across multiple functions and business units.

Some leading-edge companies are using an agile approach to run their data programs. Agile is a time-tested methodology used in IT organizations to build software or manage processes more effectively. Broadly, it is a collaborative approach in which cross-functional teams design and build minimally viable products (MVPs) and features quickly, test them with customers, and refine and enhance them in rapid iterations.[3] Agile data similarly relies on a joint approach to development and delivery: cross-functional teams comprising members of business and IT work in "data labs" that are focused on generating reliable insights that allow the company

to address its highest business priorities and realize positive outcomes quickly.

In this article, we explore the principles of agile data and the steps companies can take to introduce it in their organizations. The companies that deploy agile data can realize significant process and product improvements in the near term and set the stage for future advances and experimentation in big data infrastructure.

Understanding agile data

An approach to agile data necessarily relies on several core principles and organizational capabilities. The first is a *business-driven approach* to digital transformation and, hence, to data migration and management. Under this approach, companies create a master list of possible business use cases for advanced analytics, as well as opportunities for new or enhanced products and processes. They take inventory of the different types of data associated with those use cases and opportunities. During this process, they identify the most important customer characteristics and activities across a range of business domains. An insurer facing disruption from digital entrants, for example, may consider ways to conduct more detailed analyses of factors such as customer purchase behaviors or time to serve customers. In this way, it could improve its underwriting processes, reduce costs, and increase quality of service.

Teams then rank-order the opportunities identified and consider, for each, the levels of data governance, architecture, and quality required—customers' preferred channels of consumption, for instance, or the "golden" (or irrefutable) sources of data required, or the latency of data. The result will be two detailed, aligned road maps—one highlighting digital business objectives, budgets, and time frames and milestones; the other defining the data requirements to build an effective big data architecture and provide seamless analytics support.
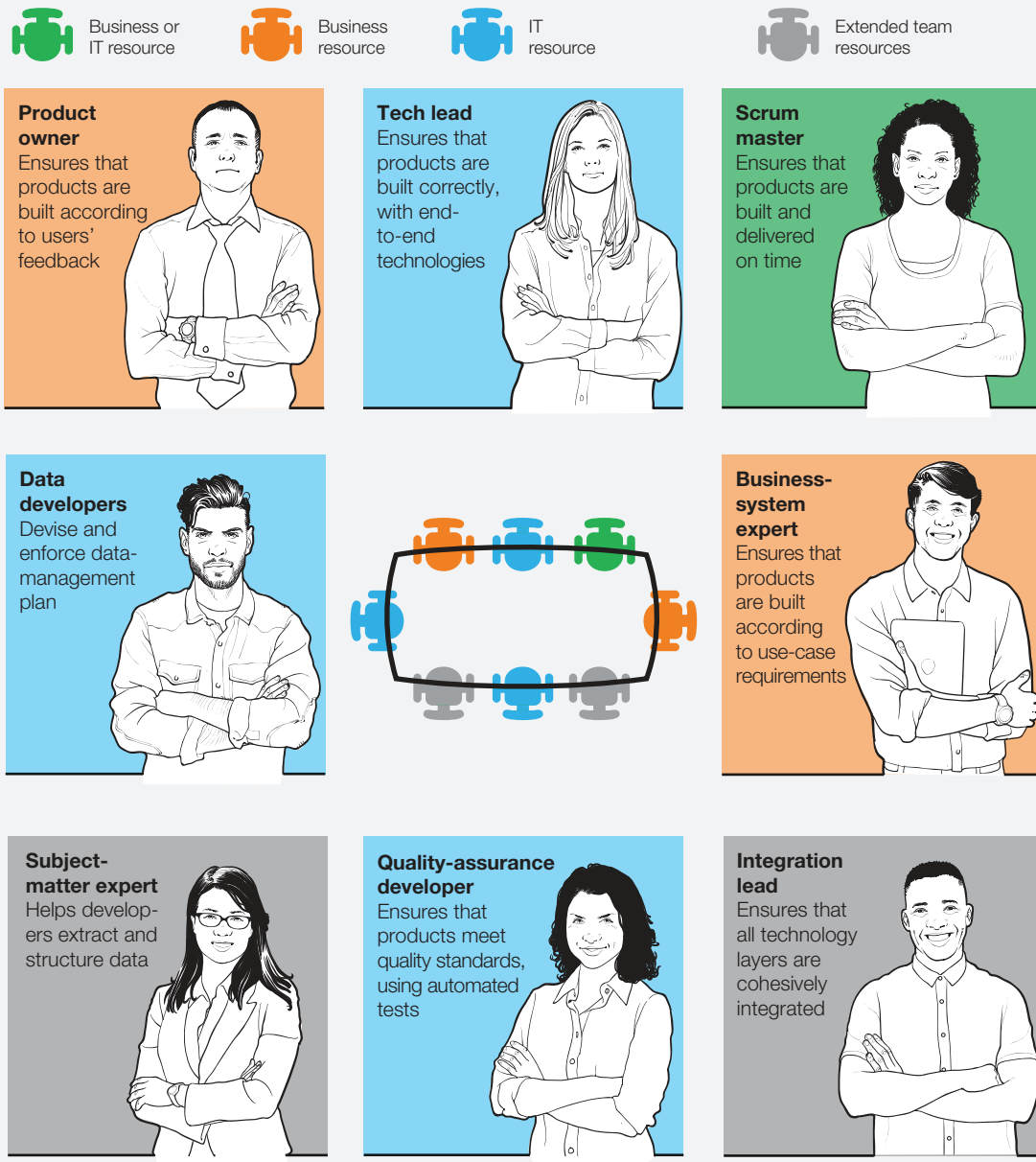
Another critical principle of agile data is *joint ownership*. Interactions between business units and the IT organization in most companies are typically limited to business requests and IT solutions thrown back and forth over a wall separating the two. To successfully deploy an agile data approach, representatives from the business side need to physically sit with members of the IT organization. Having them in the same room can help break down cultural barriers—businesspeople can learn more about technology, and IT managers may be exposed to more elements of the business. Perhaps more important, joint ownership of data-migration and data-management protocols can help the organization define just-in-time data requirements, quickly validate the business case of proposed solutions (rather than waiting for approvals to cascade through traditional channels), and assure the quality of the solution (through constant joint monitoring). As the agile approach takes hold within the company, dispersed teams may be able to use social-networking and collaboration technologies to achieve the same effects as they would from co-location.

It is also critical to build *cross-functional, or scrum, teams* (Exhibit 1). Companies can benefit from assigning dedicated teams to their data-transformation efforts. These scrum teams could include representatives from the business units and IT—for instance, data scientists, data engineers, business-information owners, IT developers, and quality-control specialists. In our experience, these teams work best when they are ring-fenced from the rest of the organization in data labs and when 100 percent of their time is allocated to agile data management. These scrum teams would be focused on developing and delivering minimally viable data-migration products and processes that could be released, tested, and refined quickly—thereby accelerating companies' ability to gain insights and business value from the data at their disposal. As data-transformation programs encompass more and more business units, multiple data labs could coalesce into a data factory. Additional support systems, such as a project-management office, could be helpful in monitoring scrum teams' activities and ensuring that data-migration efforts proceed as planned (according to predefined milestones) and that any data-related issues are resolved quickly.

Finally, companies will need to adopt *emerging technologies*. Data lakes are particularly promising (Exhibit 2). A data lake is a repository for all structured and unstructured business information collected from the myriad systems located in different business units and functions in a company. It would include current and archived data and, through the use of APIs, could be enriched with information from external providers—social data from Facebook or Twitter, for instance—or with open data, such as GPS coordinates. Unlike a traditional data warehouse, data lakes do not store data in files and folders; the data remain in their original formats. The cost to store these data therefore remains low, and the ability to configure or reconfigure data on the fly remains high. Beyond efficient storage capabilities, the data lake would be compatible with standard data-discovery tools, making it easy for IT and business users to find the information they need. The data lake can initially be built separately from core IT systems (in a discovery environment) so that updates

Exhibit 1   **An agile data lab should comprise a mix of business and IT expertise.**

Business or
IT resource

Business
resource

IT
resource

Extended team
resources

Product owner
Ensures that products are built according to users' feedback

Tech lead
Ensures that products are built correctly, with end-to-end technologies

Scrum master
Ensures that products are built and delivered on time

Data developers
Devise and enforce data-management plan

Business-system expert
Ensures that products are built according to use-case requirements

Subject-matter expert
Helps developers extract and structure data

Quality-assurance developer
Ensures that products meet quality standards, using automated tests

Integration lead
Ensures that all technology layers are cohesively integrated
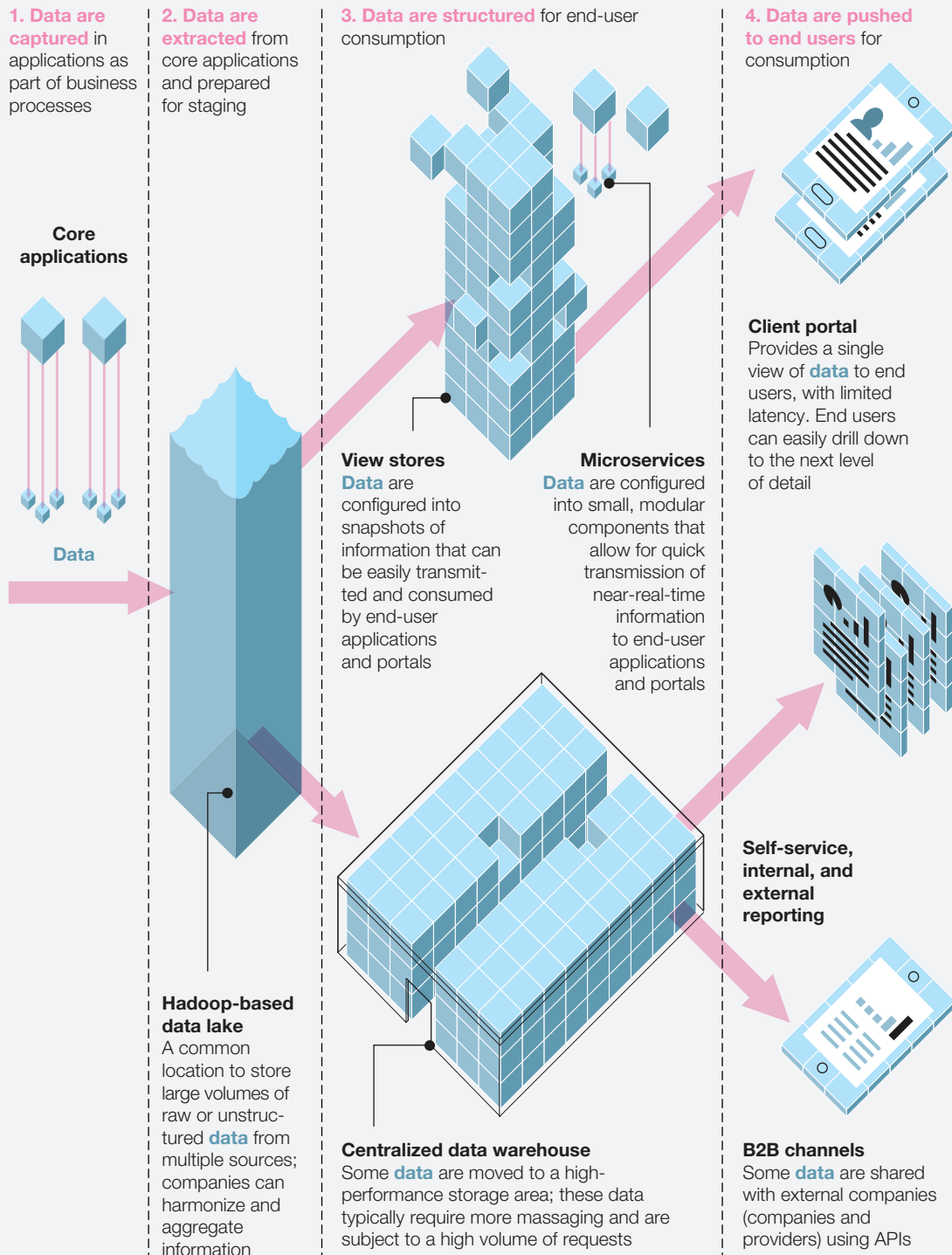
Source: McKinsey analysis

can be made more easily as data technologies evolve. Eventually, however, the data lake should be integrated within a company's existing data architecture and serve as the primary source of information. Data would be fed into the lake for each use case; but over time the company would have access to a trove of reliable data that could support a range of applications.

Exhibit 2   **Data lakes are allowing companies to transmit data to end users faster.**

1. Data are captured in applications as part of business processes

2. Data are extracted from core applications and prepared for staging

3. Data are structured for end-user consumption

4. Data are pushed to end users for consumption

Core applications

Data

View stores
Data are configured into snapshots of information that can be easily transmitted and consumed by end-user applications and portals

Microservices
Data are configured into small, modular components that allow for quick transmission of near-real-time information to end-user applications and portals

Hadoop-based data lake
A common location to store large volumes of raw or unstructured data from multiple sources; companies can harmonize and aggregate information

Centralized data warehouse
Some data are moved to a high-performance storage area; these data typically require more massaging and are subject to a high volume of requests

Client portal
Provides a single view of data to end users, with limited latency. End users can easily drill down to the next level of detail

Self-service, internal, and external reporting

B2B channels
Some data are shared with external companies (companies and providers) using APIs
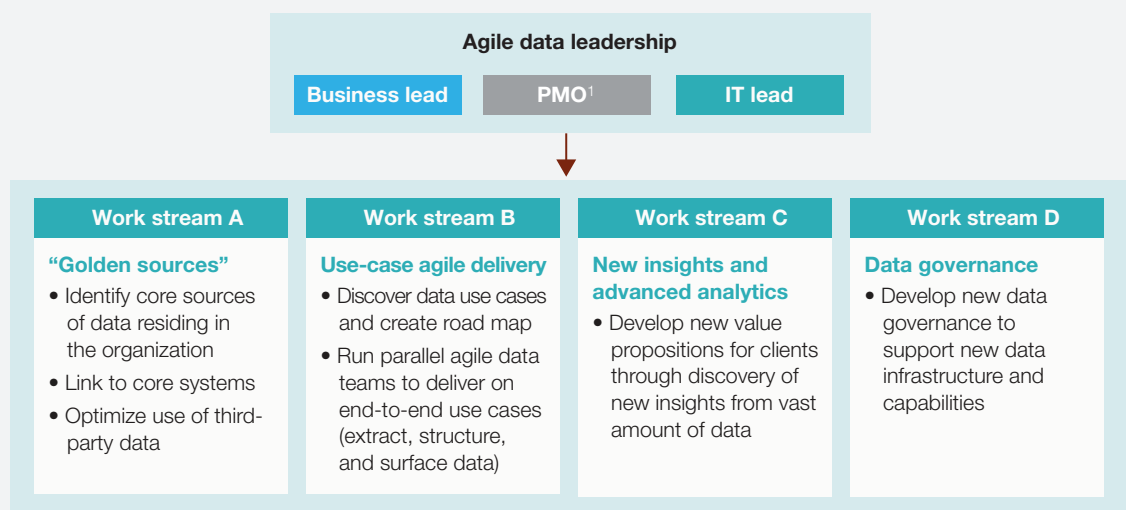
Demonstrating impact with agile data

An agile approach to data migration and management conveys a number of important benefits—not the least of which is to declutter the business-information landscape (Exhibit 3). Data from multiple databases, functions, and business units can be combined and accessed more easily. Companies can realize immediate value from the frequent release of minimally viable data-management solutions. Through the data mining made possible by the development of a comprehensive data lake, companies can also identify new business opportunities. And if business units are involved in data migration from the outset, they can seize these emerging opportunities more quickly or otherwise help the IT organization prioritize data- and digital-transformation initiatives. Data and pilot projects are no longer locked in silos, invisible to all but the immediate team members.

This approach is already having a positive effect in companies across a range of industries. A large European bank, for example, has been able to launch a number of digital initiatives on time, on budget, and in accordance with regulatory requirements and deadlines, in part because of its adoption of agile data. Supported by its data-integration efforts, the bank has been able to design and launch new customer-facing services more quickly (within weeks rather than months). The bank established a data lab comprising more than 100 employees, who implement projects in six-month waves. It updated its technology infrastructure with a data lake that contains about 100,000 data elements multiple business units can analyze and use. A joint IT and business team developed a clear road map to ensure that the data underpinning the delivery of customer-facing services are always available and of highest quality. Simultaneously, the bank has established an internal "data science" service to help individual business units design analytics programs and machine-learning models. More than 200 employees across multiple departments have been involved in the data-transformation program, and they have expressed high satisfaction with the company's new "big data culture."

Exhibit 3

Agile data transformations require new organizational and leadership approaches.

Agile data leadership

| Business lead | PMO[1] | IT lead |

Work stream A	Work stream B	Work stream C	Work stream D
"Golden sources"	**Use-case agile delivery**	**New insights and advanced analytics**	**Data governance**
• Identify core sources of data residing in the organization • Link to core systems • Optimize use of third-party data	• Discover data use cases and create road map • Run parallel agile data teams to deliver on end-to-end use cases (extract, structure, and surface data)	• Develop new value propositions for clients through discovery of new insights from vast amount of data	• Develop new data governance to support new data infrastructure and capabilities

[1]Project-management office.

Source: McKinsey analysis

Additionally, the bank has been able to build a library of more than 300 uses for various categories of data by internal and external users—an activity that will eventually help it realize significant gains in efficiency.

Meanwhile, a large US-based pharmaceutical company is in the midst of a two-year data-transformation program. Its primary goal is to use advanced analytics to improve operations in multiple functions and departments; it also hopes to eventually gain competitive advantage from its use of analytics. The company has deployed agile data methodologies across different work streams (commercial practices, R&D, and supply chain, for instance) in order to build a consistent, coherent view of customer data. It established a number of small, cross-functional teams to create and launch data-integration tools and services. In a relatively short period of time, these scrum teams have been able to gather in one place critical information about products, pricing tiers, contracts, and other customer data from different systems and business domains. As one leader told us, the data have been democratized; they can now be accessed by users from different parts of the company, and the data are being updated in real time and therefore are better able to support multiple use cases. The company hopes to significantly reduce its time to market with new products while still ensuring regulatory compliance. What's more, the data can now be mined for insights relating to future mergers and acquisitions and future product launches.

Getting started with agile data

The benefits of agile data are clear. But persuading business and IT leaders to shift to an agile mind-set can be challenging, in the way that any form of organizational change can be. IT employees will likely be unfamiliar with the use of agile principles in the management of data. They may push back against the idea of breaking from the status quo—that is, no longer using waterfall methodologies[4] to build traditional data warehouses, which often require years to develop before companies can reap value from them. For their part, the business units may be leery of taking ownership of technically complex assets like data, which have

long been the domain of IT. To change these dynamics, companies should consider doing the following:

Use pilot projects to build acceptance. A pilot project is the best way to demonstrate to IT that a change in data-management methodologies is possible and to demystify data for the business units while pointing out agile data's potential value to both sides. IT and business-unit leaders should identify a single project that is cross-cutting and of very high value to the business—for instance, a data-based service that would help reduce customer churn or improve protections against fraud. Successes and failures, milestones reached, and impact realized should be clearly and frequently communicated across the company as the pilot progresses and the service is finally released. In this way, best practices can be codified, and strong project champions can be identified and encouraged.

Empower agile data teams. Senior leaders should give cross-functional agile data teams the leeway to make important decisions relating to data migration and architecture. Scrum teams must be fully focused on activities run in the data lab and committed to a test-and-learn approach; they cannot be 50–50 players, nor can they wait for approvals from colleagues or bosses outside the data lab. Companies therefore should set up direct channels of communication between scrum teams and seniors leaders and define escalation plans: when questions or obstacles arise, scrum teams share their concerns with the project-management office and designated senior leaders, who must respond within hours or days, not weeks.

Update the technology infrastructure. The launch of a data-transformation program provides an opportunity for companies to consider ways to improve current systems and technologies. The creation of a data lake, for instance, may require new platforms, tools, and skill sets that must be compatible with the overall enterprise IT architecture. It may also prompt IT leaders to reconsider their enterprise IT strategies, asking questions such as: How can we integrate the data lake with existing systems? How can the tools and platforms used to

support agile data be leveraged in other circumstances? Should we introduce open-source technologies?

Emphasize new forms of communication. The communication methods used in agile data labs are different from those used in the rest of the organization and must be managed that way. Outcomes should be strictly monitored and can be rendered graphically using "burn up" charts.[5] Scrum teams shouldn't share these outcomes with just their immediate supervisors; they should also share them with relevant stakeholders from the business units. Many also establish semiregular forums in which IT and business leaders can convene to discuss important industry trends and topics—in areas of technical specialization, for instance, or competitive analysis—and conduct postmortems on pilot projects. These communications channels can help ensure that everyone (inside and outside the data lab) understands how important agile data is, how to implement agile ways of working, and how to resolve any issues.

Develop and introduce key performance indicators. Metrics that are directly tied to agile ways of working should be reflected in performance-evaluation mechanisms and in performance-review discussions. They are critical for sustaining long-term commitment to agile data. In most data labs, work plans (also referred to as backlogs) are based on weekly or biweekly schedules called sprints. These schedules are strictly monitored by a project-management office. Scrum-team members have clear individual responsibilities; and certain roles are linked with specific results. The entire team's ability to complete work items, often referred to as user stories, is measured and made visible to senior leadership. Critical indicators and metrics—for example, the percentage of data mapped or the amount of new business information fed into a data lake—are collected weekly, so if there are issues, adjustments can be made immediately. Business and IT leaders receive regular reports and get the visibility they need to change course.

With better access to better pools of data, companies can deploy better segmentation strategies. They can learn more from transactional data. They can automate processes, leading to greater operational efficiencies. They can compete more effectively against rivals, and they can find new sources of growth. However, most institutions are still struggling to extract value from data. In our conversations with IT and business leaders, very few say their information-management capabilities are "advanced at scale." An agile approach to data management can help solve this problem. It can formalize collaboration and conversations between business-unit heads and IT leaders, reveal new uses for existing data, prompt the development of innovative data-driven business initiatives, and speed up the delivery of critical business information. Agile is no longer just a methodology for software development or operations management. It is becoming a critical capability for those companies that want to manage their data more strategically and deliver seamless multichannel customer experiences—which, in this era of digitization, is every company. ∎

---

[1] See "Big data: The next frontier for innovation, competition, and productivity," McKinsey Global Institute, May 2011, on McKinsey.com

[2] 2015 McKinsey survey of US global, systematically important banks on the goals and status of their BCBS 239 programs.

[3] Santiago Comella-Dorda, Swati Lohiya, and Gerard Speksnijder, "An operating model for company-wide agile development," May 2016, McKinsey.com.

[4] Under the waterfall model, development steps happen sequentially rather than iteratively; no activity can proceed until the previous one is completed and approved.

[5] A graph that compares work already completed against expected results each week.

Chiara Brocchi is a consultant in McKinsey's Milan office, **Brad Brown** is a senior partner in the Boston office, **Jorge Machado** is an associate partner in the New York office, and **Mariano Neiman** is an associate partner in the Toronto office.