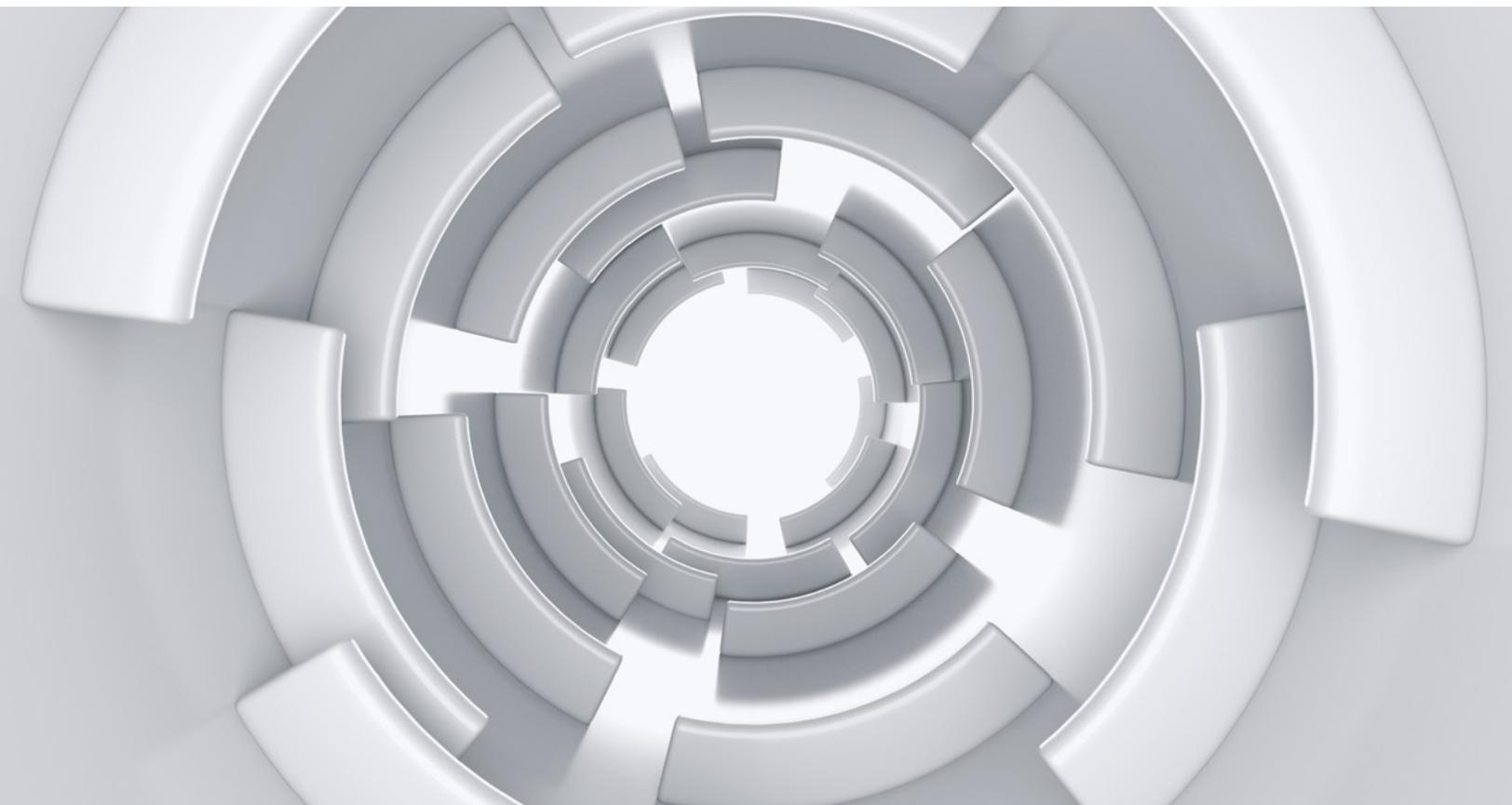


# Four ways to accelerate the creation of data ecosystems

Public and private organizations rapidly set up data ecosystems to help fight COVID-19. Even in the absence of a crisis, companies can fast-track ecosystem development.

*by Mohammed Aaser, Kumar Kanagasabai, Marcus Roth, and Asin Tavakoli*



**When organizations come together** to share and manage data, they can create value far beyond what would be available to the individual partners. These powerful data ecosystems can be harnessed to help solve public health problems or provide greater insights to governments as they form policy. During the COVID-19 crisis, for example, public–private partnerships were able to quickly create data ecosystems that did everything from tracking infection rates to providing job exchanges for displaced workers. (See sidebar, “Putting data ecosystems to work to battle COVID-19.”)

Data ecosystems also offer tremendous potential value for businesses, whether several come together to share research with one another or a single company creates a data ecosystem that enables customers and other stakeholders to share and access data. Credit bureaus are a long-standing form of data ecosystem that provide shared value for their bank partners by providing data that can help them achieve lower default rates.

Tapping the value of data ecosystems can be a challenge for companies. We have seen many organizations get stuck in the initial stages, particularly in deciding legal issues (What data

are we able to share?) and questions of value-add (Will we get the fair share of value?). If not managed, such pitfalls can doom a commercial data ecosystem before it ever gets implemented. Even companies that successfully navigate these early questions tend to experience slowdowns (or full deal breakdowns) when implementation begins.

While crises like COVID-19 can spur faster ecosystem development, we have seen accelerators emerge in non-crisis situations that can offer a path forward for business. By starting small and scaling later, building on proven technology and evolving the data model as they progress, and involving partners early on, organizations can dramatically lower risk and accelerate implementation.

### **The value of sharing data across organizational boundaries**

In our experience working with companies across sectors, we see three main ways data ecosystems provide value to companies:

- *Growth*: They enable companies to pursue new business opportunities by extending their core business or even enabling completely new products or lines of business. Credit-card

## **Putting data ecosystems to work to battle COVID-19**

---

With the urgency of the COVID-19 crisis as a spur, public–private partnerships were able to stand up data ecosystems rapidly, providing data on mobility, infection rates, ICU availability, test capacity, and other issues for the public good. The following are some examples:

- *Opportunity Insights Economic Tracker*: A high-frequency data hub with a number of data partners that is available to assist governments in making health policy and economic decisions.
- *COVID-19 Mobility Data Network*: Data providers, universities, and communities came together to exchange relevant mobility data across the globe in a privacy-preserving approach to identify potential high-risk COVID-19 zones.
- *Carnegie Mellon COVIDcast*: A partnership with social networks and healthcare institutions to create a view on emerging hot spots, frequently consumed by both public and private institutions.

processors, for example, have developed strategic insights on consumer shopping journeys and purchasing behaviors that they share with retailers and brands.

- *Productivity:* They can help companies improve operations. Online travel portals that offer insights into consumer behavior can help airlines and hotels plan for demand and set pricing. Other data ecosystems enable automotive suppliers and OEMs to share performance and usage data, generating insights that can improve product design and processes.
- *Risk reduction:* Data ecosystems are important in reducing risk, especially for industry consortia in which every member contributes data. Banks, for example, pool data to identify fraudulent transactions and accounts. Trucking fleet operators share data with insurance companies on frequency of hard-braking incidents by geography for use in risk analysis.

Before organizations can access this value, they must overcome a number of challenges as they set out to build data ecosystems, which broadly take one of three forms (Exhibit 1). For some companies, building the ecosystem may be their

first experience in sharing data, and leaders may be hesitant due to potential risks from sharing sensitive data with competitors or having that data disclosed to the media. Working out governance and sharing protocols so that all members receive fair value is another hurdle.

### How to fast-track data ecosystems

The starting point of building a data ecosystem is, of course, having a clear definition of the business problem and value to be generated. For example, some ecosystems create business value by removing high-friction steps in a flow or supply chain, some provide insights that drive value for specific business segments, and some aim to truly disrupt a market by assembling a variety of partners that solve for all aspects of the business need. The essential point in many cases is to bring in a partner with a valuable data set that will enable the partners, by working together, to create a unique offering. We have observed four actions that are critical to getting data ecosystems off the ground fast.

#### Accelerator 1: Build a bold long-term vision, but start small

In our experience, companies that are able to build ecosystems quickly start small and scale, focusing initially on just a few partners and a limited number

Exhibit 1

### Data ecosystems take three main forms.

#### 1. Closed data ecosystems

Organizations share data in a closed environment



#### 2. Strategic partnerships

Small number of organizations share data for a dedicated purpose



#### 3. Open data ecosystems

Organizations share data for the public good openly



of data sets to reduce complexity. Often, a founding company seeks out a single lead partner. Together, they start by defining a five-year vision and detailing the first six months, with a clear focus on value creation for both partners. The vision articulates the objectives and strategic cornerstones of the data ecosystem as well as a strong partner and stakeholder governance framework that clearly states the benefit that each organization hopes to achieve.

At this early stage, they also discuss important design characteristics, including the resources each provides, value-sharing models, privacy requirements, technology, the talent needed, and a view on how the ecosystem could scale. These specifications and goals are often codified and provide a framework as the ecosystem starts to grow.

Consider a recent effort in which a digital services provider and a supermarket chain partnered to solve a perennial challenge for consumer-packaged-goods (CPG) companies: understanding how advertising affects purchases. Each partner gathers large amounts of data—the digital services provider collects advertising and audience data from millions of accounts, and the supermarket chain collects purchasing data from millions of customers. Combined, these data sources can be used to uncover the link between advertising and purchases—a significant challenge for both CPG marketers and digital advertising platforms. The ecosystem creates benefits for both partners. For example, the digital services provider will be able to attract CPG advertisers to its platform, while the supermarket chain will be able to drive growth with its CPG partners.

By starting with just two players, they will be able to build trust and learn rapidly (for example, which data to share for highest value and how to structure the data model). They may also define new use cases to target, identify additional resource needs, adapt commercial models, or improve the technical data architecture. The insights generated can help determine which additional partners should be invited into the ecosystem to maximize value for

all participants, as well as enhancements to the ecosystem operating model.

### **Accelerator 2: Simplify—but don't compromise—on the legal and risk-management process**

It's critical to involve the legal and compliance teams at the start of the process. This step sounds simple, but in our experience, this is the number one factor in slowing implementation. Many companies do not address these important issues until late in the process. When discovered late, legal constraints often result in having to redesign parts of the business model or the data assets themselves, setting the timeline back by months.

To avoid serious setbacks and accelerate implementation, founding partners should focus on these four main areas of risk at the outset of the partnership: privacy risk, reputational risk, business risk, and data security and governance (Exhibit 2).

In one recent case, a failure to fully understand the risk implications of data sharing for customers derailed the entire ecosystem effort. Two agricultural companies announced a partnership to enable farmers to share data across platforms with the goal of helping farmers streamline their data collection and operations. Customers perceived that their data were being shared in a way that would hurt their ability to operate and maintain their land leases, and they took to social media to complain. Both organizations explained that data were never shared without farmers' permission and were not used in the way the farmers perceived. The damage, however, was already done. After the backlash, the partnership agreement was terminated.

The goal of the legal and compliance teams is to work on a common vision and to set guardrails across risk areas. This enables cross-company teams to work out the details. As the work progresses, teams should design a data model that meets all compliance and security requirements, including the ability to trace data from source to user.

Exhibit 2

## Ecosystem partners should engage legal teams early to explore four potential areas of risk.



**Privacy risk.** Are the data elements being shared in line with regulatory requirements, especially when dealing with consumer or individual data?



**Reputational risk.** Is there a risk that a data partnership could be misconstrued by channel partners or customers and lead to a negative response?



**Business risk.** Would sharing data with an external party create an operational, financial, or competitive risk to the organization?



**Data security and governance.** What platform is the data stored on, who can have access to the data, who approves the data's downstream use, and how are the data tracked?

One consumer-finance company in Latin America developed a streamlined, four-step legal and proof-of-concept (POC) process for onboarding partners to its data ecosystem:

1. Lay out general terms and conditions of collaboration in a memorandum of understanding that allows data sharing for an initial POC period, and work closely with a specialized law firm to ensure compliance of the business idea with all legal and regulatory requirements for data exchange.
2. Build trust by creating secure sandboxes or safety zones. The first sandbox sequesters raw data from partners, marking the data inaccessible until put in a secure form approved by the contributing partner. Other sandboxes act as secure sharing zones, or provide security as data enter and leave the ecosystem. Legal boundaries are maintained through technology.
3. Conduct a POC test in the market that includes key performance metrics to determine fair-value sharing conditions—for example, trial marketing

of a financial product to the customer base of a new partner, leveraging the joint data set.

4. If the POC results are satisfactory, the parties agree on a legal partnership contract where the value and cost-sharing agreement are informed by the key performance metrics of the POC.

This streamlined process cut the time for onboarding new partners from a year to just three to four months. Most of that time was used to execute the POC in the market rather than spent on the more typical contract negotiation.

### Accelerator 3: Don't reinvent the (technology) wheel

We have seen organizations spend significant time and resources designing a data architecture blueprint and still end up in analysis paralysis. An alternative approach is to reuse existing platforms where possible, including available blueprints that combine a data-lake architecture, API-based data access, and data-management tools such as data catalogs. By repurposing wherever possible,

# We have seen organizations spend significant time and resources designing a data architecture blueprint and still end up in analysis paralysis.

they avoid overinvesting in large platforms before they actually need them.

Another approach is to leverage pre-built data-ecosystem platforms that have been built to securely manage data. These typically create a secure zone where data are shared and only approved data can be removed. In addition, they have secure capabilities so raw data are not exposed. Investigating appropriate platforms and leveraging them can remove the burden of having to design a new platform from the ground up.

Consider the example of a bank that was looking to build a data ecosystem across several partners, such as airlines and retailers, to create an ecosystem that would inform strategic marketing efforts. While excited about the opportunity, bank executives were concerned about sharing sensitive data and wanted complete control of analysis and use of its data. Moreover, they wanted to ensure that no identifiable data were accessible, to control extraction or use of data, and to establish full audit trails. The bank worked with an existing data-ecosystem platform to enable these conditions. Without the need to build a platform from the ground up, it was able to develop and share data products with partners more rapidly, eventually enabling several data partnerships and products.

#### **Accelerator 4: Build a data model that scales**

While it can be tempting to jump into creating a robust data model to combine partner data, to get a data ecosystem up and running quickly, it's best to take a staged approach.

In a first step, organizations should simply contribute their respective data "as is"—that is, in

source-data format for fast and immediate use in initial, jointly prioritized business use cases. Data teams can then begin developing an understanding of the data and how to link or join the data to other data sets.

As data ecosystems evolve, and more data sets are shared, a more sophisticated data model that can support linking of disparate data will often be required, and data teams will now be more prepared to act. They'll likely need to implement meta-models or knowledge graphs that provide a high degree of flexibility for changing and adding relationships and entities later in the process. These models can also support data governance by limiting access and use to specific data elements based on data-sharing agreements. Again, such models should be developed only in line with use cases.

While this staged approach enables organizations to start and iterate fast, it also has implications for the capabilities that are needed, requiring deep technical and data-modeling expertise at the outset so that the model is built with scale in mind. Later, self-service tools enable people who are less data savvy to join.

An organization working to build a data ecosystem around public and licensed data as its primary business model offers a good example of this approach. Initially, the organization focused on collecting data in particular domain areas, such as real-estate or consumer products. The data came from myriad sources, such as websites, government agencies, news, and licensed data providers. While the data were formatted in different ways, they still provided value "as is" to downstream users.

Next, the organization invested in creating a knowledge layer linking disparate data across the multiple data sources. For example, it defined specific concepts or entities—such as “product,” “product description,” and “product price”—and their linkages to specific data sources. Downstream users can now write a query that, for example, resolves the historical price of a product across dozens of data sources rapidly. As organizations share data with one another and data types continue to expand, these data models will be increasingly valuable and will enable a common layer for governance and access control.

---

Data ecosystems provide a powerful way for organizations to team up and solve important societal problems and deliver more value to participants and consumers. With a value-driven and iterative approach, data ecosystems can be formed quickly, delivering benefits within months and offering opportunities for expansion and more value over the long term.

**Mohammed Aaser** is McKinsey’s chief data officer, based in the Minneapolis office; **Kumar Kanagasabai** is a partner in the New York office; **Marcus Roth** is a partner in the Tokyo office; and **Asin Tavakoli** is an associate partner in the Düsseldorf office.

Copyright © 2020 McKinsey & Company. All rights reserved.