

# WHERE IS DATA SHARING HEADED?

By Massimo Russo and Tian Feng

*This is the ninth article in a multipart series.*

**O**N ITS OWN, MOST data has limited value. But when aggregated with other data and forged into insights and applications, data can transform businesses, create enormous value, and help solve some of our biggest societal problems. But all these powerful use cases depend on the sharing of data, which holds risks and has ramifications that can deter many companies from

acting on potential opportunities. (For more on these and other topics related to data sharing, see the sidebar “Previous Articles in This Series.”)

The challenges of data aggregation have led to the emergence of platforms and ecosystems that facilitate data sharing. These are growing in size and number (think smart cities), and more and more of them are enabling integrated, citizen- and

## PREVIOUS ARTICLES IN THIS SERIES

[Innovation, Data, and the Cautionary Tale of Henrietta Lacks](#)

[Contact Tracing Accelerates IoT Opportunities and Risks](#)

[The Risks and Rewards of Data Sharing for Smart Cities](#)

[What B2B Can Learn from B2C About Data Privacy and Sharing](#)

[How Far Can Your Data Go?](#)

[Simple Governance for Data Ecosystems](#)

[Europe Needs a Smarter, Simpler Data Strategy](#)

[Sharing Data to Address Our Biggest Societal Challenges](#)

business-centric solutions that use data from disparate sources. Tech giants, being data-centric businesses, have established themselves as early movers in shaping the data-sharing marketplace.

As the benefits of data aggregation increase, four models appear most likely to facilitate broad data sharing within and across industries. Companies that are thinking about participating in data ecosystems need to consider the opportunities—and the risks—that each one presents.

The first, vertical platforms, is already established. Two others—super platforms and shared infrastructure—are beginning to take shape. These are all models in which data is centralized, which has many benefits. But because these approaches can concentrate data in just a few repositories, they potentially raise issues regarding control over the market, access to the data, and the value the data generates. We believe that on the horizon is a viable fourth, decentralized model, in which syndicated data resides at its source and is accessible to others. (See the exhibit.)

### Vertical Platforms

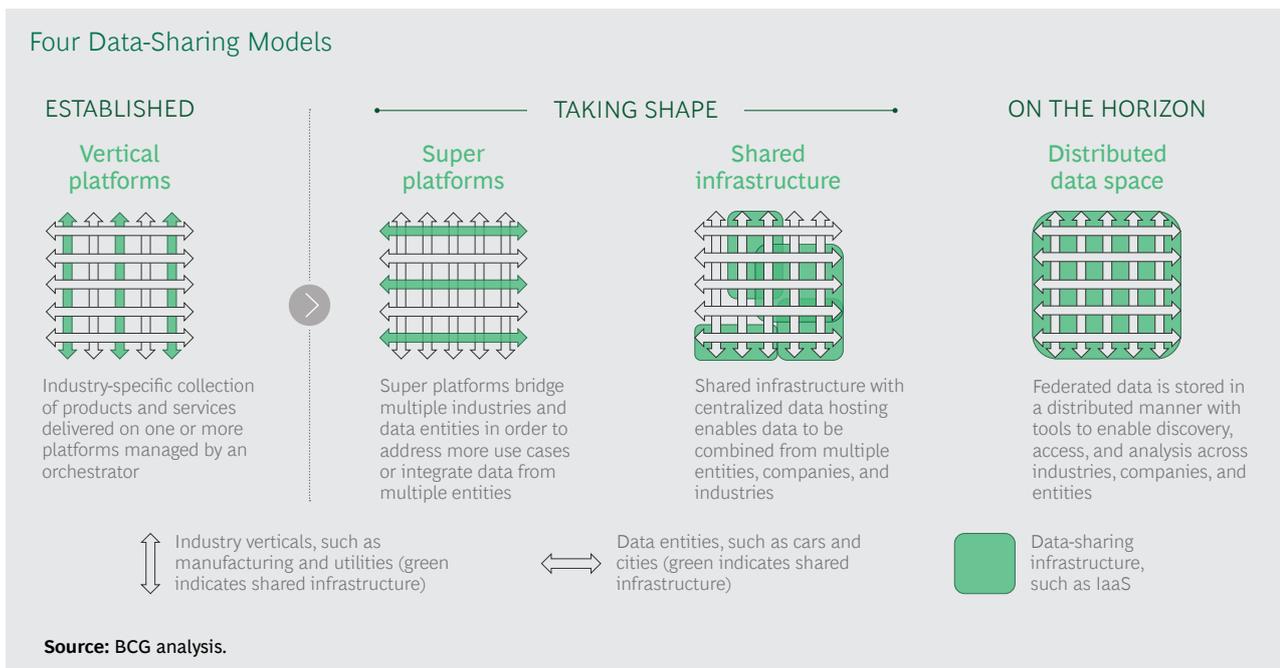
Within individual industries, vertical platforms have formed in order to share data

and provide solutions to targeted needs, such as predictive maintenance, supply chain optimization, operational efficiency, and network optimization. Airbus’s Skywise and Penske’s Fleet Insight, for example, provide benchmarking and other services using aggregated data from airlines and logistics providers, respectively. Volkswagen recently opened its Industrial Cloud to external companies, inviting platform partners to both contribute software applications that increase the carmaker’s production efficiency and improve their own operations by scaling their applications.

According to a recent BCG global survey, three-quarters of manufacturing managers are considering data sharing to improve operations. Vertical platforms are redefining competition and developing new opportunities, particularly in B2B industries, but again, their scope tends to be limited to direct customers, suppliers, or partners within a single industry or sector.

### Super Platforms

Big companies, both tech giants and traditional industry incumbents, have recognized the value in data sharing and are positioning themselves to capture a significant share by expanding their existing vertical platforms to become super plat-



## ABOUT DATA ENTITIES

People or things that generate data for which there are numerous potentially valuable use cases—such as mobility services, electric-grid efficiency, and supply chain optimization, among many other examples—are called data entities. They include cities, homes, cars, factories, farms, and, of course, individuals. As users as well as generators of data, data entities can also benefit from the additional use cases unlocked by

aggregating data from multiple, seemingly unrelated entities. Optimizing plant operations, for example, requires combining supply chain, customer, staffing, and machine operating data. Measuring an individual's carbon footprint requires data on home energy consumption, car use, personal and business travel, and food consumption, among other factors.

forms. Super platforms aggregate data across both verticals and data entities to support the development of applications that address new sets of use cases. (See the sidebar “About Data Entities.”) Smart-city open data platforms are nascent examples of this trend.

Consider, for example, the moves by Amazon, Apple, and Google to aggregate data from smartphones, home assistants, and (eventually) cars. In 2015, Google launched Sidewalk Labs, which has worked with the US Department of Transportation to develop a mobility data platform and services that improve road, parking, and public transportation use by employing a variety of data sources. Alexa, Siri, and Google Home are battling to build the biggest base of users—because it's users who generate the data that give these systems value.

Most super platforms so far have been consumer focused, but there are also early examples that aggregate data for industrial and B2B uses. Siemens' MindSphere, ABB Ability, and Honeywell Forge, among others, are competing to be the digital entry point for the factory. Similarly, John Deere Operations Center, AGCO's Fuse, Monsanto's Climate FieldView, and startup Farmers Business Network are positioning themselves as the go-to agriculture super platform by aggregating farm data and applying it to such tasks as operations management, equipment management, and land preparation and planting.

Super platforms can also address big societal issues, such as energy efficiency. Schneider Exchange connects Schneider Electric customers with an open ecosystem of analytics and solution providers that can develop applications by tapping directly into data from the entire electrical system, from generation to transmission to commercial and residential use. The Exchange ecosystem enables innovations that can improve energy efficiency and reduce greenhouse gas emissions.

A big issue for super platforms is assuring data entities that they will have some say in where and how their data is used—and that it will not be misused. Apple has been strict about not sharing customer data with third parties. Many platforms, such as the John Deere Operations Center for precision-agriculture data and Otonomo's permission management systems for smart-car data, have adopted a policy of transparency and advertise their security, privacy, and use procedures.

## Shared Infrastructure

Most large companies are migrating at least some of their data and IT infrastructure to cloud services provided by so-called hyperscalers, such as Amazon Web Services (AWS) and Microsoft Azure, some of which also provide super platforms. Hyperscalers facilitate data sharing by providing both cloud storage infrastructure and the applications that put data to use for consumers and businesses. They already host

massive amounts of data from all kinds of industries, and they are in a natural position to aggregate data by building connections across companies.

Microsoft has partnered with numerous industry incumbents to power their platforms and ecosystems with cloud storage, tools, and processing capacity from its Azure product suite. AWS pursues a similar strategy, and other tech players are competing to become preferred cloud and platform providers.

Cloud hyperscalers are also moving into data sharing. In 2019, Microsoft launched Azure Data Share, which enables Azure users to share their data sets. Microsoft recently announced its acquisition of ADRM Software, a developer of large-scale industry-specific data models. Microsoft believes that the combination of ADRM's industry modeling capability and Azure's storage and computing power will enable it to create an "intelligent data lake, where data from multiple lines of business can be harmonized together more quickly."

Microsoft has clearly made "open data" one of its strategic priorities. Along with two other enterprise software powerhouses, SAP and Adobe, it has launched the Open Data Initiative, which offers customers "a platform for a single, comprehensive view" of their data. Others recognize the potential: AWS, IBM Cloud, and Google Cloud are all facilitating the streaming of common data such as weather and financial market data.

As data sharing generates more value, and as more data migrates to the cloud, providers can differentiate themselves by offering data connectivity services that help clients capture and retain business. Down the road, cloud providers can offer additional features that allow companies to control access to their data, trace it as it is being shared across ecosystems, and monitor—and potentially charge for—its use. Ecosystem orchestrators looking to bolster data sharing can shift the entire ecosystem to the cloud platform with the best sharing functionality. Equally, companies with plat-

forms focused on a single purpose, such as supply chain optimization, can tap the capabilities of shared infrastructure, as Volkswagen has done with AWS in establishing its "automotive industrial cloud."

## A Potential Alternative: Distributed Data Space

There are two drivers of innovation from data sharing: aggregation and access. Aggregation of data from disparate sources can lead to more innovation as hidden relationships are revealed. Aggregation of more and more data from the same source across time and space can facilitate benchmark comparisons and generate insights into trends. Likewise, greater access and more open platforms unlock innovation by allowing a broader base of talent to solve problems. Innovation contests such as Kaggle and DrivenData can help connect data sets and problems with analytical talent.

But data concentrated in a few companies' hands can also hamper innovation if those companies aggregate only limited data types or seek to control access. As tech giants and others build out infrastructure and services to consolidate data, the impact of network effects propels them into powerful positions in the market.

Governments are pursuing different ways of overseeing this concentration of information and capability. A technological solution could emerge in the form of a fourth data-sharing model that we call distributed data space, a framework in which data is generated, stored, and shared in a decentralized manner.

For example, as the EU seeks to bolster data sharing in order to maintain AI competitiveness in the face of US and Chinese tech giants, the need for such a decentralized model is implied in the European Commission's [European Data Strategy](#), issued early in 2020 with proposed regulations following late in the year. While many details remain to be clarified, the initial proposal states that "the aim is to create a single European data space—a genuine single market for data, open to data

from across the world—where personal as well as nonpersonal data, including sensitive business data, are secure and businesses also have easy access to an almost infinite amount of high-quality industrial data, boosting growth and creating value.”

There are big questions, however, not least of which are where all this data will be hosted and what the viable European alternatives are to super platforms and hyperscalers. A number of initiatives and startups may provide some answers and contribute to versions of a distributed data space model coming to market. Here are five examples that address different stages of the data journey.

**Collection.** “Citizen sensing” is an emerging way to crowd source data by allowing individuals to collect data for the purpose of fact-finding and policymaking. Pilots in Europe and the US, such as the Making Sense project, have captured data on pollution, noise, and radiation, among other data types. Several sessions in the 2020 UN World Data Forum touched on the value of distributed sensors, both for scaling up data capture and for encouraging participatory policymaking.

**Computation.** Distributed computing creates a virtual network of machines that provides an alternative to cloud computing. DFINITY, a foundation based in Zurich, is building the “Internet Computer,” which uses independent data centers to combine the power of individual computers in an attempt to “reverse Big Tech’s monopolization of the internet.”

**Learning.** Federated learning is an approach to data sharing in which the underlying data doesn’t leave the control of its owners. Instead, algorithms are remotely trained with their cooperation, and insights are then centralized for the owners’ use.

**Storage and Sharing.** Distributed ledger technology can enable trusted, distributed data storage and sharing. IOTA, a not-for-profit organization based in Germany, is building a blockchain-based data platform for IoT applications. It has the support of

large industrial players, such as Bosch, Volkswagen, and Schneider Electric.

**Ownership.** MIT professor Tim Berners-Lee, the inventor of the World Wide Web, is trying to reinvent the internet by storing consumer data in highly decentralized “pods” controlled by the individuals who generate the data. This concept, called the “solid framework,” is being pursued by a startup, Inrupt. Another MIT professor, Sandy Pentland, creator of the MIT Media Lab, is exploring citizens’ data cooperatives, which help people keep control of their data while maintaining the benefits of pooling and aggregation.

## Not Yet Ready for Primetime

These five examples are interesting early steps, but to take hold, such initiatives will need to provide more than just data collection and storage. Data from disparate sources usually requires additional context (in the form of data definitions, data models, and metadata, for example) in order to be gathered, understood, and used effectively. Much of this context is use-case-dependent and under a distributed model—with no centralized control or curation—can be difficult to achieve. Distributed data spaces also require strong governance, standards, and “wrangling” tools to render the data fit for use. For example, in the Making Sense project, participating citizens were provided not only with sensors but also with the necessary training on calibration, data collection, and record keeping.

The distributed data space concept is capturing the attention of think tanks, academics, researchers, and investors. For example, the UK’s Open Data Institute is exploring the value of data sharing as well as models such as data trusts and other data institutions. New York University’s GovLab is cataloguing different forms of data use for the public good, including open-access data collaboratives. Venture capital fund Outlier Ventures is supporting technologies that enable an open-data economy.

While the early initiatives are encouraging, it will take a combination of government

policy, technical and governance innovation, research, and investment for distributed data platforms to emerge as viable alternatives to super platforms and shared infrastructure.

## Implications for Ecosystem Participants

The various data-sharing models present different choices and challenges, depending on the company's desired role in a given ecosystem. We have previously defined three main data-sharing ecosystem participants—orchestrators, contributors, and enablers—and each faces its own set of issues.

**Orchestrators.** These are the organizers and managers of the ecosystem. They set the governance and value capture rules, provide the platform for data sharing and innovation, coordinate the activities of participants, and provide a channel for the products and services of contributors. Orchestrators typically leverage platform- and infrastructure-as-a-service solutions from hyperscalers. A few orchestrators will try to establish super platforms. In a distributed data space, orchestrators will need to rethink their priorities.

From the point of view of technology, orchestrators need to select partners with sophisticated data-sharing capabilities or explore emerging distributed data space solutions in order to avoid supplier lock-in. From the strategic point of view, they need to evaluate whether their business models will work if data ownership and control become distributed and how they can continue to add value under this model. Under all models, orchestrators will need to encourage trust and participation by protecting contributor data in a transparent manner.

**Contributors.** Contributors provide data and services to the ecosystem and build products and services using the data that it makes available to them. Contributors have a choice of ecosystems in which to participate and are likely to work with a few simultaneously. They should consider the

data-sharing model underlying each ecosystem, how that model will support their own data strategy, and how power will be shared with the ecosystem orchestrator. The super platform and shared-infrastructure models can address the need for contributors to easily syndicate data across multiple ecosystems. Contributors can then concentrate on developing value-added solutions instead of managing data. If distributed data spaces mature, contributors can turn to these solutions to reduce their dependence on hyperscalers and avoid vendor lock-in.

**Enablers.** These companies provide the ecosystem with infrastructure and tools, such as connectivity, security, or computing resources. Multiple enablers will be needed to support the different models. The hyperscaler tech giants clearly have a critical role to play in vertical platforms and super platforms. Several large cloud and platform providers are making major investments, and the market is recognizing the importance of data sharing, with cloud data platform companies, such as Snowflake, going public at high valuations. Enablers of secure, distributed data spaces are being pursued by startups and technology incumbents alike. IBM, for example, is building a distributed data platform based on blockchain technology. The acceleration of data sharing creates opportunities for enablers to build more sophisticated data-sharing solutions based on either centralized or distributed models.

**A**S THE VALUE realized from sharing data grows, the models discussed here will evolve and others may emerge to unlock the value potential of the data economy. Understanding which models are available and how they align with specific data-sharing strategies will help companies decide what role they want to play in the fast-growing world of data-sharing ecosystems.

## About the Authors

**Massimo Russo** is a managing director and senior partner in the Boston office of Boston Consulting Group. You may contact him by email at [russo.massimo@bcg.com](mailto:russo.massimo@bcg.com).

**Tian Feng** is a project leader in BCG's Boston office and a BCG Henderson Institute ambassador. You may contact her by email at [feng.tian@bcg.com](mailto:feng.tian@bcg.com).

## About BCG

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.

## About BCG Henderson Institute

The BCG Henderson Institute is Boston Consulting Group's strategy think tank, dedicated to exploring and developing valuable new insights from business, technology, and science by embracing the powerful technology of ideas. The Institute engages leaders in provocative discussion and experimentation to expand the boundaries of business theory and practice and to translate innovative ideas from within and beyond business. For more ideas and inspiration from the Institute, please visit <https://www.bcg.com/featured-insights/thought-leadership-ideas.aspx>.

© Boston Consulting Group 2021. All rights reserved. 2/21

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