The Benefits of Data Sharing Now Outweigh the Risks

APRIL 29, 2024

By François Candelon, Guillaume Sajust de Bergues, David Zuluaga Martínez, Harsha Chandra Shekar, and Marcos Aguiar

READING TIME: 12 MIN

Many of today's biggest industry challenges won't be solved by a company toiling alone, drawing only on its proprietary data. Complex issues such as fraud detection, supply chain optimization, and drug discovery can often be tackled most effectively through collaboration, pooling data from multiple industry players. In this scenario, everyone wins—both individual companies and the industry at large. And the win could be substantial: in 2019, the Organization for Economic Co-operation and Development estimated the value opportunity of data sharing at 2.5% of the global GDP.

Although most companies remain resistant to strategic data sharing, some industry heavyweights are already realizing the benefits. In the US, automobile insurance companies are collaborating through a claim-history information exchange called LexisNexis CLUE Auto. For these insurers, sharing proprietary claims data has significantly sped up the underwriting process and reduced liability risk; as a result, 99% of US underwriters now participate. The European aerospace company Airbus has also taken a collaborative approach with its suppliers and customers. In 2017, it launched the digital platform Skywise to address industrywide operational challenges, such as predictive maintenance and fleet performance, through data sharing. The platform is estimated to generate hundreds of millions of dollars in revenue and cost savings annually among all participants.

Such examples, however, are outliers. Even in instances when sharing data would solve some intractable problems and generate value for all involved, executives' fears hold them back. They are concerned about the engineering and regulatory challenges and, crucially, they worry that the data they share might be used against them by other firms. But our research shows that such perceptions are largely outdated.

Now is the right moment for savvy executives to revisit strategic data sharing.

What's changed? The technology. Compared with even five years ago, today's software and tools, as well as new forms of data, can mitigate or resolve many of the engineering and regulatory challenges that companies (rightly) cite, while also reducing the need for trust between companies that would benefit from collaboration. Now is the right moment for savvy executives to revisit strategic data sharing.

New Tech Tackles Engineering and Regulatory Challenges

Today, companies can manage the traditional engineering and regulatory challenges associated with sharing data in wholly new ways. Two commonly cited obstacles offer cases in point.

Obstacle #1: "We lack the infrastructure, common standards, and talent." Over the years, data engineering challenges have taken different forms—for example, companies might not have the internal know-how or the digital infrastructure to share data efficiently and securely. These deficits have been made more pronounced by the lack of universal data formatting standards and common sharing protocols.

While these obstacles have stalled data sharing in the past, new technology provides solutions. Data sharing services offered by companies such as Databricks are now widely available. These services provide secure platforms for organizations to store, share, and analyze data, reducing the need for in-house tech expertise and infrastructure. Moreover, the digital world is far more standardized than it once was, even without an

explicit universal standard for sharing data. Connectors between systems, such as application programming interfaces, are more homogenous and their documentation more easily accessible than even five years ago.

Emerging technologies may also soon provide a workaround for companies dealing with talent shortages. For example, before data can be exchanged, it must be cleaned and formatted. This has traditionally been a time-consuming task. Today, generative AI can be programmed to identify inconsistencies and errors within the data and generate scripts to fix these issues, which enable companies to automate parts of the process.

The rise of strict regulatory frameworks, ranging from data privacy to antitrust to cross-border data flows, has made data sharing feel like more trouble than it's worth.

Obstacle #2: "It's too difficult to comply with regulations." The rise of strict regulatory frameworks, ranging from data privacy to antitrust to cross-border data flows, has made data sharing feel like more trouble than it's worth for many companies. They have struggled to find a straightforward path to compliance.

Companies have long been wary of sharing data because of the possibility that sensitive, personally identifiable information (PII) could slip through the cracks. Along with reputational damage, this kind of leak has the potential to violate data privacy regulations like the EU's General Data Protection Regulation. Today, however, numerous vendors offer tools that can help companies handle sensitive data. For example, companies can use discovery tools to scan and analyze their data repositories to identify hidden sensitive or confidential information before they share their data. Companies can also use modern data anonymization tools to remove or encrypt PII in an irreversible way, ensuring it can't be linked back to an individual.

Antitrust regulation can also be a concern, but it shouldn't deter efforts to share data. As already mentioned, virtually all US insurers collaborate on the LexisNexis CLUE Auto platform. Although this partnership could have attracted the scrutiny of regulators, it hasn't—because the objective is to address legitimate industrywide issues. Likewise, the largest automotive manufacturers in Germany cocreated the Catena-X initiative, a collaborative data exchange network that—far from being censured on antitrust grounds—received the active support of the German government.

There is one caveat: in many parts of the world, regulations governing cross-border data flows are highly fragmented and restrictive. This often limits data sharing to within regulatory blocs, such as the EU or the US—making data exchanges across these blocs challenging. We see a path forward, however, with new types of data that have emerged with the rise of AI, such as features, embeddings, or AI model parameters. These new forms of data, used in lieu of raw data, could allow for safe sharing that respects regulators' objectives, such as protecting individual data privacy—although this would require regulators to update their policies to take these new forms of data into account.

Trust Is Still Paramount

Advances in technology make handling sensitive data more secure, but companies are still squeamish because of perceived strategic risk. This perception creates a collective action problem; companies see both the value and the risk, but rarely an incentive to be first movers. That sense of risk can take different forms depending on the data sharing partner, whether that's a direct competitor, a customer or supplier in the supply chain, or an aggregator.

> **When data sharing counterparties are also competitors, companies fear revealing their IP or "secret sauce" hidden in the data.**

When data sharing counterparties are also competitors, companies fear revealing their intellectual property (IP) or "secret sauce" hidden in the data. For example, the US auto insurers contributing to the LexisNexis CLUE Auto database worry that they could indirectly reveal confidential information. Claim data contains information such as the amount paid or vehicle type which, when put together, might reveal valuable information about a company's customer base. However, LexisNexis, in its role as data sharing intermediary, also provides trust-creating features. Crucially, only members that report their data to the platform are allowed to withdraw information, and the service strictly controls which information can be withdrawn.

Data sharing therefore works within the industry, despite the risk, because (1) the auto insurers have sufficient trust in one another and the data sharing system; and (2) the substantial benefits of the data sharing system outweigh any lingering strategic risk.

> **Despite the existence of strategic risk, sharing is made possible by the trust that Airbus, its suppliers, and customers each place in the data sharing system and its ability to improve the supply chain's functioning in a neutral way.**

In other instances, such as when companies share data within their supply chains, executives worry that suppliers and customers could use shared data to gain an edge in commercial negotiations. In the case of the airplane industry's Skywise database, participating airlines are sharing sensitive details about their operations, such as in-flight and engineering data. These details could be used by Airbus or other companies in the supply chain to evaluate things like a company's headroom ahead of pricing negotiations. Yet despite the existence of strategic risk, sharing is made possible by the trust that Airbus, its suppliers, and customers each place in the data sharing system and its ability to improve the supply chain's functioning in a neutral way.

This trust can be explained by multiple mechanisms. For example, Airbus built transparency into the platform through data governance, which makes participants' behavior observable to other participants. Airbus has also created an ecosystem of trustworthy participants through the training and certification that Skywise delivers to partners it has vetted through an intensive verification system.

Conversely, when companies don't trust each other enough, the perceived risk of data sharing outweighs the expected benefits. Consider the aggregator service Order with Google. This service collects data, including pricing and most popular menu items, from multiple food delivery services. But Google also has internal teams that use this data to understand the marketplace and offer the food delivery services insights to improve their clients' business. This may be a primary reason why Uber Eats, for example, does not participate in Google's aggregator service. Uber would likely benefit from the insights and incremental orders, but arguably would not want its data used to help competitors.

How Technology Makes Trust Easier

The good news is, when trust doesn't come naturally, technology can help to lower the "trust threshold." In other words, technology can act as a partial substitute for trust, especially at the start, to ignite the relationship between prospective data sharing partners in the following ways:

- **Technology provides transparency into partners' data governance and usage.** Technology can help increase trust among firms by creating data sharing systems that are trustworthy. For example, when companies decide to collaborate, they establish data sharing agreements. These agreements specify minimum data governance standards—the processes and policies that each partner must implement to manage and protect any data that falls into their hands. Historically, it's been difficult to monitor how these standards get implemented in practice, but today, modern software can help provide transparency for all involved parties. For instance, specialized tools can create irrevocable records of data transaction history to automate various aspects of data governance monitoring, including the review and enforcement of each partner's data governance policies.

  Beyond governance, data sharing agreements also typically outline how partners are allowed to use the shared data. But here too, these agreements have been hard to enforce because it's difficult to monitor what partners do with the data once they have access to it. Virtual data rooms, such as Snowflake Global Data Clean Room, are secure online spaces for data storage and distribution that can address this issue. Such spaces include tools that set permissions and restrictions on the data or track data access and usage—thus enabling the data owner to control how their data is used and analyzed even after it's been shared.

Beyond governance, data sharing agreements also typically outline how partners are allowed to use the shared data.

- **Technology reduces the risk of strategic information leaks with synthetic data.** Beyond the risk of inadvertently releasing data protected by privacy regulations, there is also the risk of mistakenly releasing raw data that contains confidential company information and IP. One emerging solution is the use of synthetic data, which is created to have the same characteristics as a real-world data set, but without including real-world data. Research has shown that synthetic data is very difficult to reverse-engineer when properly synthesized, which would enable wary companies to share data without fear of a damaging leak.

Federated learning is emerging as an alternative, decentralized approach. It uses data from multiple companies, but the data doesn't leave each individual company's premises.

- **Technology circumvents sharing company data directly by employing decentralized model training.** We are observing an ever-growing hunger for data to train AI models. Some companies are starting to realize the benefits of jointly training an AI model with other companies, in service of addressing industrywide issues. With access to a wider data set, the joint model will provide more valuable insights beyond the capacity of a single company. Joint models are typically trained by a single trusted organization—a joint venture or a third party—charged with collecting the data from each company and using the consolidated data for training. Federated learning is emerging as an alternative, decentralized approach. It uses data from multiple companies, but the data doesn't leave each individual company's premises. This allows for companies to share the insights embedded in their data and contribute to a collective effort without the risks associated with sharing the data itself with other firms.

  A recent example of federated learning is the shared platform MELLODDY, developed by a European consortium of ten pharmaceutical companies aiming to accelerate drug discovery. The companies codeveloped the platform to enhance machine learning models with data from each company, but without directly exposing their proprietary information. The interest for these pharmaceutical companies is not only to be compliant with patient data regulations, but also to enable collaboration in a low-trust environment wherein firms that collaborate gain a competitive edge.

Technology alone won't be able to fully overcome the trust gap between companies. The use of virtual data rooms, for instance, may allow companies to monitor the analyses conducted there, but they can't control how counterparties will use the insights derived from these analyses. Some risk will remain.

But the potential benefits of data sharing increasingly outweigh such risks. Companies should take advantage of today's technology to address issues across their industry by starting to collaborate with firms they may not yet fully trust.

Over time, technology itself may begin to create a “trust flywheel.” As partners benefit from the value of sharing data and gain confidence in the process, they will feel encouraged to keep sharing, and in some cases, to deepen these relationships or seek out new partners that can add more richness and depth. For some leaders, this scenario may seem far-fetched—but today’s software and tools have made it something all companies can achieve.

The authors would like to thank Gaurav Jha for his contribution to this article.

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 our website and follow us on LinkedIn and X (formerly Twitter).

Authors

François Candelon
MANAGING DIRECTOR & SENIOR PARTNER; GLOBAL DIRECTOR, BCG HENDERSON INSTITUTE
Paris

Guillaume Sajust de Bergues
LEAD DATA SCIENTIST
New York

David Zuluaga Martínez
PARTNER, BCG HENDERSON INSTITUTE AMBASSADOR
New York

Harsha Chandra Shekar
ALUMNUS

Marcos Aguiar
MANAGING DIRECTOR & SENIOR PARTNER
São Paulo

ABOUT BOSTON CONSULTING GROUP

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 work closely with clients to embrace a transformational approach aimed at benefiting all stakeholders—empowering organizations to grow, build sustainable competitive advantage, and drive positive societal impact.

Our diverse, global teams bring deep industry and functional expertise and a range of perspectives that question the status quo and spark change. BCG delivers solutions through leading-edge management consulting, technology and design, and corporate and digital ventures. We work in a uniquely collaborative model across the firm and throughout all levels of the client organization, fueled by the goal of helping our clients thrive and enabling them to make the world a better place.