Speaker 1: Welcome to the MIT CISR Research Briefing series. The center for information systems research is based at the Sloan School of Management at MIT. We study digital transformation.

Stephanie Woerner: Hi, I’m Stephanie Woerner, principal research scientist and director of MIT CISR. Today I’m pleased to share with you a research briefing from the archives—the October 2020 research briefing authored by Barb Wixom, Ina Sebastian, and Robert Gregory:

Data Sharing 2.0: New Data Sharing, New Value Creation

A 2017 survey conducted by MIT Sloan Management Review found that higher levels of data sharing with customers, suppliers, and competitors are associated with higher levels of innovation. MIT CISR research finds that simply sharing data in old ways does not pay off. Instead, companies must share new data in new ways to realize the full value potential of digital opportunities.

Historically, companies have shared minimal data—and only when it was required—to execute a transaction, solve a well-defined problem, or comply with regulation. They engaged in data sharing 1.0: sharing transactional data assets and capabilities to enable or preserve an existing value proposition.

Today, companies are beginning to appreciate data sharing as an opportunity rather than an obligation, opening doors for new value creation. Companies that sell products through platform marketplaces see opportunity in data about the characteristics and behaviors of all platform consumers, including those who looked at their products but did not buy them. Companies that manufacture sensored products see opportunity in data about how and why their customers are using the products—and how customers are making or saving money from that use. Companies that amass rich data assets that reflect customer journeys, supply chain activities, and manufacturing operations see opportunity in start-ups, crowds, and partners who can contribute analytics, skills, leading-edge tech, and outside-the-box thinking to help solve hard problems and search for blue ocean ideas. In order to exploit these opportunities, companies need to engage in data sharing 2.0: sharing complementary data assets and capabilities to create new value propositions.

For the past three years, MIT CISR has investigated interorganizational data sharing at twenty-three distinct organizations. We explored why and how companies share data, and how data sharing activities contribute to value and innovation from digital initiatives. This briefing illustrates several data sharing 2.0 use cases from the research and describes three sets of data sharing practices that position companies to fully exploit digital opportunities.

Complementary Data Assets and Capabilities

The magic of data sharing 2.0 is rooted in the cross-company sharing of complementary data assets and/or capabilities, which fills data gaps and allows companies, often collaboratively, to develop innovative or transformative solutions.

PepsiCo and its retail partners share data about consumer behavior and product categories to grow retailer sales. In one case, a retailer provided PepsiCo with data about its purchases of fountain drink syrup from other providers, which PepsiCo combined with data on its own sales to the retailer. Then, drawing on advanced analytics capabilities and insights from the company’s rich trove of anonymized shopper data, PepsiCo identified attributes about sales, stores, and shoppers that influenced syrup usage. Analyzing the complementary data using machine learning produced specific action items intended to generate a win-win outcome for consumers, the retailer, and PepsiCo by optimizing sales of both the company’s own and other providers’ beverages.

Schneider Electric and its business customers share data via Schneider’s IoT platform EcoStruxure. The platform hosts data and algorithms that, when combined with customers’ sensor and operations data, optimize energy management for customers. As a greater number of customers leverage the platform’s data science capabilities, the underlying models learn and improve. For example, energy is the second largest operational cost (after labor) for EcoStruxure customer Hilton Hotels & Resorts. Hilton has achieved three percent savings per year in energy procurement, and more in energy efficiency, using the Schneider’s Ecostruxure Resource Adviser digital offering that shows real-time consumption and pricing data and helps optimize resource efficiency.

Health insurance provider Anthem and a variety of health partners share data to co-create AI-based solutions for complex healthcare problems. Anthem assembled a data set of twelve years of de-identified medical claims, prescription claims, and lab data associated with 45 million individuals. Then, the company made the data available using its Data Sandbox, a platform certified to legally and acceptably provision the data to vetted users. The health partners—academic and commercial organizations and NGOs—contributed algorithms, analysis techniques, and novel perspectives. The Data Sandbox currently hosts fifteen projects completed or in process.

In all of these examples, companies identified and created new value by unleashing the synergistic combination of complementary data assets and capabilities.

Data Sharing 2.0 Accelerators

Although opportunities enabled by data sharing 2.0 can be incredibly exciting, companies hesitate to engage in interorganizational data sharing beyond requisite transactions. Among the reasons why: more and new data sharing comes with added cost and risk, particularly when sharing requires new systems and governance. And the nature of data is sensitive—for example, company intellectual property, or data related to personal health or financial performance. Also, the nature of desired data usage is novel—for instance, for use in data wrapping. It takes time for companies to figure out what people, processes, and technologies are required for safe, effective execution—and to put them in place. MIT CISR research identified three sets of practices that help companies more easily and confidently accelerate data sharing 2.0: curated content, designated channels, and repeatable controls.

Curated Content

Creating, purposely assembling, and curating data for interorganizational sharing accelerates data sharing 2.0 by establishing a prepared set of material with which to innovate. Curation can include removing personally identifiable information, converting field values to conform with industry standard codes, compiling records into higher levels of aggregation, and extracting insights for sharing that exclude underlying data. Curation can also include pre-vetting the company’s ownership of the data and its right to share the data for specified use cases.

In 2011, BBVA created a data set comprising four million anonymized credit card payment transactions for its innovation team to use in developing new information-based services. The company aggregated and masked the data and performed legal and compliance inspections on it so that the team could feel assured of its appropriateness for market-facing use. Initially, the static data set informed pilot solutions for urban planners and disaster recovery strategists. Today, the data set is one of many that supports informational solutions available via the company’s API market.

Designated Channels

Establishing channels that enable secure data exchange accelerates data sharing 2.0 by simultaneously protecting data and easing data access and transfer. One way channels can manifest is as a data platform, either internal—such as an API-enabled data lake—or external—such as a third-party hosted platform. Another way is as secure data movement, whereby data is exchanged using blockchain technology or some form of masking and encryption. Yet another is as data-at-rest sharing techniques, which manage access rights across tenants of a cloud offering or common platform (like Snowflake), or federate analysis such that only the results of queries and algorithms and not the underlying data are available to participants.

In December 2018, Maersk and IBM launched the TradeLens platform that was purpose-built to share data about global shipping events and documents, with the goal to facilitate the end-to-end journey of a container. The underlying blockchain architecture provides high levels of data security. Ocean carriers can host and manage a blockchain node and become “trust anchors.” The permissioning structure (essentially, who can access what data) is based on a United Nations standard, and only the hash value of commercially sensitive information (typically documents) is stored on the blockchain; authorized companies can see if information has changed without seeing the underlying data. The TradeLens platform currently plays a key role in convincing diverse stakeholders, including ocean carriers that compete with Maersk, to engage in the collaborative ecosystem arrangement. As a result, approximately 175 ecosystem members (including ocean carriers, ports and terminals, inland transportation, and customs authorities) and customers (importers, exporters, and freight forwarders) currently participate in the interorganizational data sharing.

Repeatable Controls

Moving from time-consuming, bespoke governance measures to standard operating procedures for oversight accelerates data sharing by enabling alignment in data and outcomes across the company. Many executives in our research described a heavy reliance on their company’s security and legal teams to help create processes, procedures, and activities that could be generalized across sharing relationships. In some of these cases, project teams captured lessons learned, exceptions, and repeatable steps to build an onboarding playbook used to guide (and expedite) subsequent data sharing arrangements.

In 2016, a collection of financial institutions, including ANZ, National Australia Bank, and Westpac invested in a company called Data Republic that has evolved to serve the market in a data sharing escrow capacity. Data Republic offers a wide range of data sharing oversight services, such as a legal framework and licensing workflows, governed data downloads and transfers, and full audit trails and reporting of sharing activities. ANZ chose to use Data Republic for some of its own data cross-company sharing needs in part because of the latter’s stringent security and governance control. After six months of deployment, ANZ reported a reduction in time to negotiate data sharing agreements and methodologies from eighteen weeks to one week and rapid creation of shared data assets to help ANZ uncover potential opportunities.

Primed for Opportunity

Companies will increasingly encounter exciting digital opportunities, but not all companies will be poised to create value from them. MIT CISR research indicates that data sharing 2.0 accelerators can help companies break through cost, risk, and timing obstacles to focus on new value creation. Get set for opportunity—perform accelerator activities and prepare to create value.

Speaker 1: Thanks for listening to this reading of MIT CISR research, and thanks to the sponsors and patrons who support our work. Get free access to more research on our website at cisr.mit.edu.