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.

Hi, I’m Barb Wixom, a principal research scientist with MIT CISR. Today I’m excited to share with you the June 2022 research briefing that I co-authored with Ida Someh and Cynthia Beath—

Building Advanced Data Monetization Capabilities for the AI-Powered Organization

Large organizations increasingly aspire to become AI-powered, but such an ambitious goal requires extraordinary capabilities. MIT CISR research into data monetization capabilities identified five capabilities—data science, data management, data platform, customer understanding, acceptable data use—that at advanced levels enable this goal. In this briefing we describe how organizations build their data monetization capabilities to be strong enough to support AI initiatives. In MIT CISR’s July 2022 research briefing we will explore an AI-specific capability we call AI Explanation, or AIX—an emerging enterprise capability required for building trust in AI.

Organizations rely on the five data monetization capabilities to effectively monetize their data—in other words, to generate economic returns from it. A data monetization initiative requires the five capabilities regardless of whether it involves performance dashboards, enterprise reporting, business analytics, or artificial intelligence. An organization’s data monetization capabilities mature over time as the organization adopts more and more advanced practices associated with each capability, building on foundational practices with intermediate and then advanced practices.

The five capabilities are interdependent and need to be of similar maturity to collectively enable data monetization initiatives. Pursuing AI requires an advanced level of the data science capability, in which machine learning, specialized computational and statistical techniques (for example, time series analysis), data scientist hiring and retention, and other advanced data science practices are in play. In addition, AI projects demand that the other four data monetization capabilities are across the board more advanced.

In 2015, data scientists at Microsoft began actively exploring novel internal applications for AI. Over time the data scientists built models and accumulated expertise and techniques for reuse across the organization. Drawing on a case study about AI at Microsoft, we illustrate how an organization adopts practices that develop the maturity of its data monetization capabilities. In describing each capability at Microsoft, we indicate what practices the company utilized and whether each practice was foundational, intermediate, or advanced.

AI at MSFT: Building Interdependent Capabilities

In 2021, Microsoft was a $168 billion technology company headquartered in Redmond, Washington. The company employed 190,000 people who operated out of more than six hundred office buildings worldwide. Microsoft’s Real Estate and Facilities group, known as RE&F, was responsible for managing this extensive network of facilities. In 2015, RE&F approached Enterprise Data Science, Microsoft’s internal-facing data science unit, to inquire whether AI could be used to reduce the cost of managing these facilities. Microsoft had previously established shared services groups including Enterprise Data Science, Enterprise Data, and Enterprise Business Intelligence to help employees find ways to use data for decision making.

Data Science

The data science unit included about twenty data scientists, who developed data science solutions for Microsoft business units, and about ten software engineers. The data scientists helped RE&F identify some problems that AI could help solve. They proposed that RE&F use machine learning techniques to better understand the company’s space utilization and identify opportunities to reduce operating costs. The scientists assembled a multidisciplinary team, initially for a pilot project to investigate space utilization at a single headquarters building. Over time, the space utilization AI model the team developed was recontextualized for a variety of use cases, including building optimization, dynamic HVAC management, and parking garage optimization both for Microsoft and, eventually, its customers.

In this initiative, Microsoft used the following Advanced Data Science practices:

Develop or hire data science talent, and

Employ machine learning techniques

Data Management

As the project team identified data needed to train models, the team acquired, cleansed, and validated data to meet the project’s needs. Over time, the team integrated RE&F data about facilities with data obtained from Microsoft's Security, IT, and HR units as well as from external sources. The data scientists pulled in weather data because they believed that weather conditions, especially temperature, would be an important factor in understanding facilities-related behaviors. The team also needed to curate new data; for example, because an employee might work in a location different from their assigned office, the team had to create a mapping layer atop employee data to record an employee's actual physical location. In some cases, the team used machine learning to fill gaps in incomplete data sets.

Microsoft used the following practices associated with Data Management:

A Foundational practice: Acquire, clean, and validate data

An Intermediate practice: Integrate data, and

An Advanced practice: Curate data

Data Platform

The team established an Azure-based Occupancy data lake to host accumulating data and to support analysis efforts. The team incentivized internal data sharing beyond RE&F by offering occupancy analysis in return for contributions of new data sources. This sharing required that the team make the Occupancy data lake accessible to an array of organizational users. When models were commercialized into customer services, the project team worked with Microsoft’s product group to productize the solution and accommodate access by external users.

Microsoft used these practices in connection with Data Platform:

A Foundational practice: Establish a data lake

An Intermediate practice: Enable internal data lake access, and

An Advanced practice: Enable external data lake access

Customer Understanding

During model development and training processes, the project team regularly discussed results with key stakeholders to get their perspective on what the model was doing and how the results should be interpreted. When solutions were deemed to have potential commercial application, the project team worked directly with Microsoft Consulting and external customers to develop a proof of concept. Then, Microsoft’s product group productized the solution or created a service that Microsoft Consulting Services could deliver and manage over time at scale.

Practices Microsoft used for Customer Understanding included:

A Foundational practice: Involve stakeholders in analysis

An Intermediate practice: Prototype products with customers, and

An Advanced practice: Manage and evolve products at scale

Acceptable Data Use

As the project team amassed more and more data, several notable data governance challenges arose that required consultation with Microsoft’s privacy and legal teams. For example, combining data in new ways surfaced the potential to reverse-engineer individual identities. The teams spent significant time establishing privacy and security practices and policies for the data. Later, the teams had to harden various practices and policies to ensure they were sufficient for commercialized services. Eventually, the project team used hashing to conceal employee identity while allowing for an individual to be identified when it was permissible.

Microsoft used these practices in ensuring Acceptable Data Use:

A Foundational practice: Establish data protection practices

An Intermediate practice: Protect data for commercial use, and

An Advanced practice: Develop scalable data protection practices

An Enterprise Capability Perspective

Organizations have been focused on deploying select AI initiatives to demonstrate that achieving the “impossible” with AI is possible. AI projects teams have learned via such initiatives how to build and deploy models that fruitfully power work tasks. These organizations now aspire to move beyond the learning mode and build data monetization capabilities that allow the organization to harness the power of AI in sustained and pervasive ways.

They can do this by taking an enterprise capability—rather than local capability—perspective. Enterprise capabilities result from the accumulation of expertise and practices into resources—people, tools, applications, and routines—that are available across the organization. Enterprise capabilities, once established, can be reused or refined by both local and centralized AI project teams as the teams build and manage their AI models.

Taking an enterprise capability perspective produces complementary benefits. When capabilities are established at the enterprise level, their underlying practices don’t have to be repeatedly “discovered” by new teams; this speeds up new AI projects. At the same time, enterprise capabilities are more discoverable within the organization, which increases the odds that AI models might be recontextualized for new use cases; this reuse reduces AI project costs.

At Microsoft, for example, once the data scientists had an AI model that could identify building occupancy, they were able to quickly launch subsequent smart occupancy projects. Completion of these AI projects was accelerated because needed capabilities were already well established: The company’s data science capability supplied expertise and techniques. Its data management capability maintained data that had been cleansed and curated by earlier efforts. Its data platform capability provided a cost-effective way to process subsequent initiatives. Its acceptable data use capability had surfaced regulatory and ethical concerns that it monitored. And its customer understanding capability ensured that the customer voice was heard.

In addition, as Microsoft employees became aware of the Occupancy data lake, how much data it contained, and how easily that data could be integrated and used together, they suggested ideas for related applications of the AI model. For example, being able to identify employee location led to an idea for a building safety application. The project team developed a model that could detect danger in an area of a particular building and know whether that area was crowded or empty. The goal was to detect employees who needed to be evacuated and to generate tailored alerts to them.

Next Steps for Advanced Enterprise Data Monetization Capability Building

Although there are clear benefits to having advanced data monetization capabilities that are discoverable and reusable across the organization, it is not a foregone conclusion that the capabilities you build will be enterprise capabilities. To ensure that they are, you need to (1) mature your capabilities by adopting more and more advances practices; (2) develop enterprise capabilities, not just local ones, by concentrating expertise and practices into resources; and (3) evangelize capabilities across the organization to ensure discovery and reuse.

To assess whether the capabilities you are establishing to fuel an AI-powered organization are advanced enterprise capabilities, consider your top AI initiative: What data monetization capability gaps exist within this initiative? If other initiatives share a gap, implement a solution at the enterprise level for use across the organization that both fills the gap and circumvents silos. What challenges do you face with this initiative in sustaining capability gains? Develop policies or incentives to encourage reuse. What AI project results from the initiative point to practices that worked well? Communicate these widely to promote involvement in practice development and reuse.

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@cisr.mit.edu.