



BUILDING ADVANCED DATA MONETIZATION CAPABILITIES FOR THE AI-POWERED ORGANIZATION

Barbara H. Wixom, *Principal Research Scientist*

MIT Sloan Center for Information Systems Research (CISR)

Ida A. Someh, *Lecturer in Business Information Systems*

UQ Business School at The University of Queensland, Australia

Cynthia M. Beath, *Professor Emerita*

University of Texas, Austin

Large organizations increasingly aspire to become AI¹-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, and acceptable data use—that at advanced levels enable this goal.² 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—i.e., generate economic returns from—their data.³ 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 (e.g., 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 (see table 1).

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, Microsoft⁴ 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 (RE&F) group 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.⁵

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.

1 We define artificial intelligence (AI) as applied analytics models that have some level of autonomy.

2 This research draws on a Q1 to Q2 2019 asynchronous discussion about AI-related challenges with fifty-three data executives from the MIT CISR Data Research Advisory Board; more than one hundred structured interviews with AI professionals regarding fifty-two AI projects from Q3 2019 to Q2 2020; and five AI project narratives published by MIT CISR.

3 B. H. Wixom and L. Owens, "Digital Data Monetization Capabilities," MIT Sloan CISR Research Briefing, Vol. XIX, No. 4, April 2019.

4 B. H. Wixom, I. A. Someh, R. M. Ionescu, and D. E. Leidner, "Microsoft's AI Scaling Journey for Occupancy Prediction," MIT Sloan CISR Working Paper No. 453, February 2022.

5 I. A. Someh and B. H. Wixom, "Microsoft Turns to Data to Drive Business Success," MIT Sloan CISR Working Paper No. 419, July 2017.

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.

Associated practices

- Advanced: Develop or hire data science talent; 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.

Associated practices

- Foundational: Acquire, clean, and validate data
- Intermediate: Integrate data
- Advanced: 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.

Associated practices

- Foundational: Establish a data lake
- Intermediate: Enable internal data lake access
- Advanced: 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.

Associated practices

- Foundational: Involve stakeholders in analysis
- Intermediate: Prototype products with customers
- Advanced: Manage and evolve products at scale

Table 1: Requirements for Advanced Data Monetization Capabilities for AI

	Data Science	Data Management	Data Platform	Customer Understanding	Acceptable Data Use
AI projects need:	People, processes, and technology that can build, train, deploy, and manage machine learning models over time	The ability to curate data sets that are accurate, well understood, comprehensive, voluminous, and dynamic	A scalable on-demand platform that can regularly ingest, integrate, update, and process data and support external access	Access to tacit and digitized subject matter expertise that can inform market needs; experience in co-creation and experimentation with users	Frictionless oversight that can minimize decision risk, bias, and unintended consequences

Source: B. H. Wixom, I. A. Someh, A. Zutavern, and C. M. Beath, "Explanation: A New Enterprise Data Monetization Capability for AI," MIT Sloan CISR Working Paper No. 443, July 2020.

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.

Associated practices

- Foundational: Establish data protection practices
- Intermediate: Protect data for commercial use
- Advanced: 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.⁶ 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 advanced 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 in sustaining capability gains? Develop policies or incentives to encourage reuse.
- What AI project results point to practices that worked well? Communicate these widely to promote involvement in practice development and reuse.

6 B. H. Wixom, I. A. Someh, and R. W. Gregory, “Scaling AI to Generate Better and Different Outcomes,” MIT Sloan CISR Research Briefing, Vol. XXI, No. 12, December 2021.

MIT SLOAN CENTER FOR INFORMATION SYSTEMS RESEARCH

Founded in 1974 and grounded in the MIT tradition of rigorous field-based research, MIT CISR helps executives meet the challenge of leading dynamic, global, and information-intensive organizations. We provide the CIO and other digital leaders with insights on topics such as business complexity, data monetization, and the digital workplace. Through research, teaching, and events, the center stimulates interaction among scholars, students, and practitioners. More than seventy five firms sponsor our work and participate in our consortium.

CISR RESEARCH PATRONS

AlixPartners
Avanade
Axway, Inc.
Collibra
Pegasystems Inc.
PricewaterhouseCoopers
Standard Bank Group
(South Africa)
The Ogilvy Group

CISR SPONSORS

Allstate Insurance Company
Amcor
ANZ Banking Group
(Australia)
Australian Taxation Office
AustralianSuper
Banco Bradesco S.A. (Brazil)
Banco do Brasil S.A.
Bank of Queensland
(Australia)
BlueScope Steel (Australia)
BNP Paribas (France)
Bristol-Myers Squibb
Cabot Corporation
CarMax

Caterpillar, Inc.
CEMEX (Mexico)
Charles River Laboratories,
Inc.
CIBC (Canada)
Cochlear Limited (Australia)
Commonwealth
Superannuation Corp.
(Australia)
Credit Suisse (Switzerland)
Cuscal Limited (Australia)
CVS Health
DBS Bank Ltd. (Singapore)
Doosan Corporation (Korea)
Fidelity Investments
Fomento Economico
Mexicano, S.A.B., de C.V.
General Mills, Inc.
General Motors Corporation
Henkel AG & Co. (Germany)
Hitachi, Ltd. (Japan)
HSBC Technology & Services
(USA) Inc.
Johnson & Johnson (J&J)
Kaiser Permanente
King & Wood Mallesons
(Australia)
Koç Holding (Turkey)

Mercer
National Australia Bank Ltd.
Nomura Holdings, Inc. (Japan)
Nomura Research Institute,
Ltd. Systems Consulting
Division (Japan)
OECD
Pacific Life Insurance
Company
Pioneer Natural Resources
USA Inc.
Posten Norge AS (Norway)
Principal Financial Group
Procter & Gamble
QBE
Raytheon Technologies
Reserve Bank of Australia
Santander UK/Grupo
Santander
SC Global Tubular Solutions
Scentre Group Limited
(Australia)
Schneider Electric Industries
SAS (France)
Scotts Miracle-Gro
State Street Corp.
Stockland (Australia)
Suncorp Group (Australia)

Teck Resources Limited
(Canada)
Tetra Pak (Sweden)
Trinity Health
Truist Financial Corporation
UniSuper Management Pty
Ltd (Australia)
USAA
Webster Bank, N.A.
Westpac Banking Corporation
(Australia)
WestRock Company
Wolters Kluwer
Zoetis Services LLC

MIT CISR is funded by Research Patrons and Sponsors, and we gratefully acknowledge their financial support and their many contributions to our work.

Sponsorship and benefits: cistr.mit.edu/community/sponsor-and-patron-benefits

MIT CISR research publications: cistr.mit.edu/research-library



MIT Sloan School of Management
Center for Information Systems Research

245 First Street, E94-15th Floor
Cambridge, MA 02142

t 617-253-2348 | e cistr@mit.edu

cistr.mit.edu |  

Team | Christine G. Foglia *Associate Director*, Nils O. Fonstad, Dorothea Gray-Papastathis, Cheryl A. Miller, Ina M. Sebastian, Nick van der Meulen, Peter Weill *Chairman*, Barbara H. Wixom, Stephanie L. Woerner *Director*