



# DATA LIQUIDITY LEVERS AT CATERPILLAR

**Barbara H. Wixom, Principal Research Scientist**

**MIT Center for Information Systems Research (CISR)**

**Joaquin Rodriguez, Assistant Professor of Information Systems**

**Grenoble Ecole de Management, and Academic Research Fellow, MIT CISR**

**Gabriele Piccoli, Professor, Louisiana State University,**

**and Academic Research Fellow, MIT CISR**

**Cynthia M. Beath, Professor Emerita, University of Texas at Austin**

**and Academic Research Fellow, MIT CISR**

Data liquidity—the ease of data asset reuse and recombination—matters to business performance. In a recent MIT CISR survey,<sup>1</sup> organizations with high data liquidity<sup>2</sup> reported 14 percent higher scores for customer experience and time to market than organizations with low data liquidity. High-liquidity organizations also outperformed low-liquidity organizations in other ways: data sharing within the enterprise and beyond was 28 percent easier, data monetization impact was 24 percent greater, and employees spent nearly twice as much time gleaming insights from data.

How do organizations build data liquidity?<sup>3</sup> Our previous research identified important foundational tactics: Organizations invest in data monetization capabilities, using practices such as data cataloging and data lineage management as well as adopting cloud-based database technology and implementing security controls.<sup>4</sup> And they design and deploy modular data assets.<sup>5</sup>

Caterpillar®, a heavy equipment manufacturer, built highly liquid data assets as a cornerstone of its digital vision to double its 2016 \$14 billion services revenues by 2026.<sup>6</sup> By 2025, Caterpillar had established a digital platform that enabled capabilities including e-commerce, fleet management, and predictive and preventative maintenance. Also by 2025, Caterpillar had grown services from \$14 billion to \$24 billion. In this briefing, we describe how Caterpillar successfully executed foundational data liquidity tactics by managing three data domain practice areas we call data liquidity levers.

## DATA LIQUIDITY LEVERS

In our research, we identified three domain practice areas that influence the ease of data asset reuse and recombination for an organization: data architecture, data preparation, and data permissioning (see figure 1). These areas function as levers to increase or limit data liquidity.

- 1. Data architecture: establishing a modular platform comprised of data objects.** This enables an organization to cost efficiently deliver reusable and combinable data assets inside and outside the enterprise.
- 2. Data preparation: curating accurate data assets that are broadly relevant and strategically important for the enterprise.** This requires investing in practices that generate high-priority, reliable data assets that will appeal to people across the organization.
- 3. Data permissioning: establishing guardrails for acceptable data use.** This requires data ownership, as well as security and oversight practices that enable safe, widespread democratization of data assets.

An organization must manage all three of these levers to achieve high liquidity. For example, an organization with an

1 2024 MIT CISR Data Monetization: Generating Financial Returns from Data and Analytics Survey (N=349).

2 The sample was split into quartiles by composite data liquidity scores. High quartile (mean score of 3.7/5, N=89) and low quartile (mean score of 1.9/5, N=86) data liquidity organizations were compared using ANOVA.

3 This briefing draws on four research studies: (1) in 2020, 73 interviews with data and analytics leaders MIT CISR researchers and collaborators conducted to learn about data investments as a part of strategic digital initiatives; (2) in 2022, 33 interviews with members of the MIT CISR Data Research Advisory Board and conceptual development described in Gabriele Piccoli, Joaquin Rodriguez, Barbara Wixom, and Ida Asadi Someh, “Data Liquidity: Conceptualization, Measurement, and Determinants,” Proceedings of the International Conference on Information Systems (ICIS), December 2022, <https://aisel.aisnet.org/icis2022/entren/entren/16/>; (3) the 2024 MIT CISR Data Monetization survey; and (4) from July 2023 to December 2024, the case study on Caterpillar. For the Caterpillar study, the authors conducted 56 interviews with 42 stakeholders. The authors supplemented those interviews with information from publicly available sources, informative decks provided by Caterpillar, and iterative content review by Caterpillar case participants during development of the case narrative.

4 B. H. Wixom and G. Piccoli, “Build Data Liquidity to Accelerate Data Monetization,” MIT CISR Research Briefing, Vol. XXI, No. 5, May 2021, [https://c isr.mit.edu/publication/2021\\_0501\\_DataLiquidity\\_WixomPiccoli](https://c isr.mit.edu/publication/2021_0501_DataLiquidity_WixomPiccoli).

5 J. Rodriguez, G. Piccoli, and B. H., Wixom, “Increase Data Liquidity by Building Digital Data Assets,” MIT CISR Research Briefing, Vol. XXI, No. 11, November 2021, [https://c isr.mit.edu/publication/2021\\_1101\\_DigitalDataAssets\\_RodriguezPiccoliWixom](https://c isr.mit.edu/publication/2021_1101_DigitalDataAssets_RodriguezPiccoliWixom).

6 The case study on Caterpillar in this briefing draws from B. H. Wixom, J. Rodriguez, G. Piccoli, and C. M. Beath, “Caterpillar’s Digital Data Journey,” MIT CISR Working Paper No. 467, August 2025, [https://c isr.mit.edu/publication/MIT\\_CISRwp467\\_CaterpillarDataJourney\\_WixomRodriguezPiccoliBeath](https://c isr.mit.edu/publication/MIT_CISRwp467_CaterpillarDataJourney_WixomRodriguezPiccoliBeath).

effective data platform and a suitable data quality program may have low data liquidity because employees struggle to gain permission to access data assets.

The following case study of Caterpillar describes how the organization has employed these levers in its data liquidity journey.

## CATERPILLAR'S DATA LIQUIDITY JOURNEY

In 2024, Caterpillar Inc. (Caterpillar) was a 112,900-employee, \$64.8 billion global manufacturer of construction and mining equipment, off-highway diesel and natural gas engines, industrial gas turbines, and diesel-electric locomotives.<sup>7</sup> Over its 100-year history, the company had accumulated an extensive network of more than 150 independent dealers and had over 4 million Cat® products at work around the world.<sup>8</sup>

In 2019, Caterpillar announced the goal of doubling its 2016 \$14 billion service revenues by 2026. Then-CEO Jim

7 Caterpillar Inc., 2024 Annual Report, May 13, 2025, 2, from the Caterpillar website, <https://s7d2.scene7.com/is/content/Caterpillar/CM20250506-c118a-5d3cb>.

8 Caterpillar Inc., 2024 Annual Report, 2.

Umpleby<sup>9</sup> committed to establishing a robust digital ecosystem as a core enabler of this growth and hired Ogi Redzic as chief digital officer and head of the new “Cat Digital.”

The leaders of Cat Digital envisioned a platform that would manage every digital interaction with a customer, any data point generated by a piece of equipment, and any service event performed by a dealer. The cornerstone of this digital vision was what they called Helios: a single, reliable data source and scalable software development platform. In 2019, Redzic established a Helios team and made a seasoned internal data leader, Brandon Hootman, accountable for the platform's success.

By 2025, Helios underpinned strategic initiatives, such as AI-enabled Prioritized Service Events that delivered timely service leads to dealers, more than \$15 million per business day in dealer parts sales to users through e-commerce channels, and digital capabilities serving previously underserved small operators.

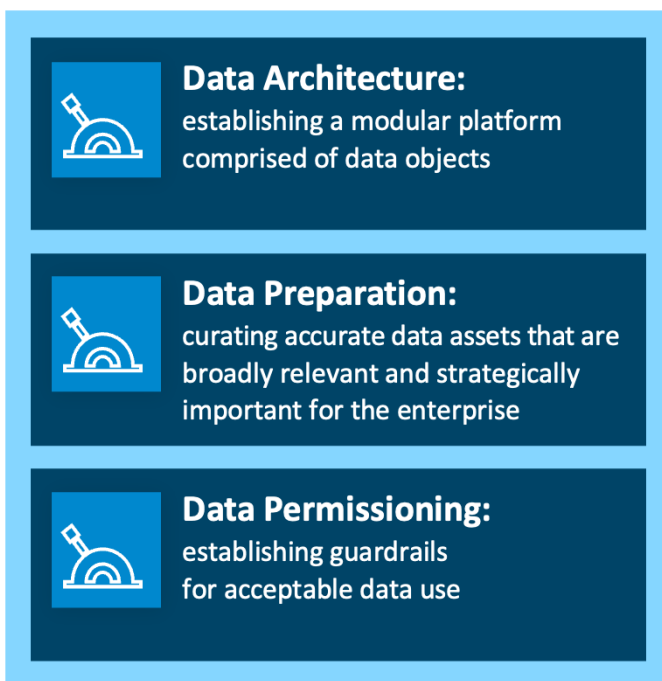
## Data Liquidity Lever One: Data Architecture

In the mid-2010s, Caterpillar was managing a complex data environment. Caterpillar's divisions used siloed applications, all with different data cleansing and transformation rules, to run their businesses. The company relied on about 200 different interfaces with dealers to exchange a range of business data, such as orders, warranties, and product pricing. Cat equipment, whose telematics systems collectively generated several million messages daily, had varied “asset IQs”: Some equipment reported fuel level and fault codes, while more sophisticated equipment reported exactly when the machine started, stopped, or was idle.

Cat Digital's leaders recognized that Helios needed to support different data consumption patterns, use cases, and requirements across a mix of technologies and approaches. Over time, the Helios team landed on a platform that comprised a thin application layer,<sup>10</sup> a service layer that handled the bulk of application functionality, and a data layer that included multiple kinds of reusable data components in a variety of data formats, such as relational and time series data.

In the data layer, the data components moved through stages. In the first stage, data coming in from a data source was stored in a raw data staging area, with very little if any

Figure 1: Three Levers Influence Data Liquidity



Source: 33 interviews in 2022 with members of the MIT CISR Data Research Advisory Board and conceptual development described in Gabriele Piccoli, Joaquin Rodriguez, Barbara Wixom, and Ida Someh, “Data Liquidity: Conceptualization, Measurement, and Determinants,” Proceedings of the International Conference on Information Systems (ICIS), December 2022.

9 In May 2025, Jim Umpleby transitioned from chief executive officer to executive chairman of the board of Caterpillar.

10 The application layer included “single-page” applications, meaning that applications did not have a persistent back-end layer but instead were entirely dependent on the Helios platform for data and integration.

transformation. In the second stage, the platform applied transformation rules to validate, remediate, and normalize data and create data objects. In the third stage, the platform converted data objects into master datasets or combined them into derived datasets. The master datasets were stable, and each one had a data owner and data stewards; a derived dataset was an amalgamation of one or more of master data, other data objects, and non-Helios data within Caterpillar.

The resulting modularized, scalable platform provisioned reusable data and related services. In one example, the Helios team created a fleet list-derived dataset that enumerated the Cat equipment customers owned, operated, or rented. That derived dataset and its related services were used across many applications and analytic products, saving development time and improving the consistency of customer experiences.

### Data Liquidity Lever Two: Data Preparation

To select which data the platform would offer for reuse, the Helios team prioritized data that were strategic for realizing the company's service revenue target and frequently were linked together: customer, contact, and asset (i.e., equipment) master datasets. Customer master data was company-level descriptive information about the entity that legally owned a piece of Cat equipment. Contact master data described company contacts with whom Caterpillar needed to communicate, such as for marketing purposes. Asset master data reflected the full customer fleet list that Caterpillar might support, service, or eventually replace. When combined, the three master datasets together could generate answers to questions like, "Which contact at which customer is responsible for replacing which machines?"

Hootman established a data science data quality group to help drive out poor-quality data and fill data gaps. The data quality team established four data quality levels and validated Helios data against those levels using algorithmic, statistical, and machine learning-based processes that were built as reusable services for Cat Digital's data pipelines. Data quality was measured at the time of ingestion, as a part of an overall monthly review of Helios data objects, and during ad hoc "smart data quality" checks of key attributes and relationships. When the data quality checks uncovered "data grief"—records that did not meet data quality criteria—those records were flagged for Caterpillar experts or data stewards to resolve.

After data quality foundations had been established, the data quality team engaged in developing specialized services that solved especially difficult data quality problems. Take, for example, a service that validated assets' serial numbers. When

Caterpillar collected asset information, the serial number field often contained errors; it was not unusual for people to mistype the number, perhaps because it was hard to see or had worn off. Because transactions like repair services and equipment sales required a serial number, the data science team created an API-enabled asset validation service fueled by five machine learning algorithms for which Caterpillar eventually received a patent.

### Data Liquidity Lever Three: Data Permissioning

About a dozen vice presidents across Caterpillar's businesses owned the company's data domains. In addition to receiving information monthly about the quality of their data, they were also kept informed about the applications that consumed their data.

Within Cat Digital, a data solution team owned a set of related data objects; this team deeply understood the data and recommended what data would serve new use cases. They were accountable for making data investments pay off, and they worked closely with data stewards in the business and Cat Digital teams that managed data operations and data security. Cat Digital's Data Operations team managed the ongoing upkeep of data objects once they were built by resolving "data grief" (records that did not meet data quality criteria) identified by the Data Quality team and intervening when something did not work as planned. Cat Digital's Data Security team followed a "least privilege access" tenet, giving a person the least amount of access they needed to accomplish their goal. This team also identified sensitive or confidential data and ensured that it was accessible only to people assigned certain roles. An access request portal helped people understand the data sets, entitlements, and objects available to their roles.

### ACTIONS TO LIQUEFY YOUR DATA

To realize their strategic ambitions with data, organizations must purposely liquefy their data to ensure that people across the organization can easily find, use, and trust data assets for strategic value creation. We suggest that you first assess the data asset experience for the average data asset consumer by asking them about the effort involved in the data lifecycle, from data discovery through to data use. Next, appoint a task force to recommend investments in architecture, preparation, and permissioning that will reduce the effort required from data consumers, especially by increasing data reuse. Finally, make a senior leader accountable for data reuse goals and for generating returns from data liquidity investments.

## MIT CENTER FOR INFORMATION SYSTEMS RESEARCH (CISR)

MIT CISR helps executives meet the challenge of leading increasingly digital and data-driven organizations. We provide insights on how organizations effectively realize value from approaches such as digital business transformation, data monetization, business ecosystems, and the digital workplace. Founded in 1974 and grounded in MIT's tradition of combining academic knowledge and practical purpose, we work directly with digital leaders, executives, and boards to develop our insights. Our research is funded by member organizations that support our work and participate in our consortium.

### CISR RESEARCH PATRONS

AlixPartners  
Avanade  
Cognizant  
Collibra  
IFS

Commonwealth Superannuation Corp. (Australia)  
Cuscal Limited (Australia)  
Dawn Foods  
DBS Bank Ltd. (Singapore)  
Doosan Corporation (Korea)  
Ericsson (Sweden)

Nomura Holdings, Inc. (Japan)  
Nomura Research Institute, Ltd. Systems Consulting Division (Japan)  
Novo Nordisk A/S (Denmark)  
OCP Group  
Pacific Life Insurance Company  
Pentagon Federal Credit Union

UniSuper Management Pty Ltd (Australia)  
Uniting (Australia)  
Vanguard  
WestRock Company  
Wolters Kluwer Financial & Corporate Compliance  
Xenco Medical  
Zoetis Services LLC

### CISR SPONSORS

ABN Group (Australia)  
Alcon Vision  
ANZ Banking Group (Australia)  
AustralianSuper  
Banco Bradesco S.A. (Brazil)  
Banco do Brasil S.A.  
Barclays (UK)  
BNP Paribas (France)  
Bupa (Australia)  
Caterpillar, Inc.  
Cemex (Mexico)  
Cencora  
CIBC (Canada)  
Cochlear Limited (Australia)

Fidelity Investments  
Fomento Economico Mexicano, S.A.B., de C.V.  
Genentech  
Gilbane Building Company  
Hunter Water (Australia)  
International Motors  
Jewelers Mutual  
JPMorgan Chase  
Kaiser Permanente  
Keurig Dr Pepper  
King & Wood Mallesons (Australia)  
Mater Private Hospital (Ireland)  
Nasdaq, Inc.

Posten Bring AS (Norway)  
Principal Life Insurance Company  
Ralliant  
Reserve Bank of Australia  
RTX  
Saint-Gobain  
Scentre Group Limited (Australia)  
Schneider Electric Industries SAS (France)  
TabCorp Holdings (Australia)  
Telstra Limited (Australia)  
Terumo Corporation (Japan)  
Truist Financial Corporation

### CISR ASSOCIATE MEMBERS

*MIT CISR wishes to thank all of our associate members for their support and contributions.*

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

Membership and benefits: [cistr.mit.edu/content/member-benefits](https://cistr.mit.edu/content/member-benefits)

MIT CISR research publications: [cistr.mit.edu/research-library](https://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](mailto:cistr@mit.edu)

[cistr.mit.edu](https://cistr.mit.edu) | 

**Team** | Isobela Byerly-Chapman, Margherita Di Pinto, Christine G. Foglia *Associate Director*, Dorothea Gray-Papastathis, Cheryl A. Miller, Ina M. Sebastian, Alan Thorogood, Nick van der Meulen, Peter Weill *Chairman*, Barbara H. Wixom, Stephanie L. Woerner *Director*