

MIT CDOIQ Symposium Focuses on AI Ethics, Data Sharing, Federated Architectures, and CDO Success Strategies

By Wayne Eckerson with Kevin Petrie and Joe Hilleary

This trip report was created courtesy of [Eckerson Group](#), a Boston-based consulting, advisory, and research firm that transforms organizations with data.

This year's MIT CDOIQ Conference, held virtually July 20-22, focused on several new themes, including AI ethics and data sharing, while retreading older ones of a practical nature, such as how to devise a data strategy, empower data analysts, design a modern architecture, and succeed in the ever-evolving role of chief data officer (CDO).

This year's Symposium attracted more than 2,000 registered attendees who listened to presentations from more than 150 speakers and panelists. More than 30 software vendors and service providers sponsored the event.

Focus on AI Ethics

The event's keynote speaker, Archbishop Vincenzo Paglia, appealed to the audience to endorse and contribute to the "Rome Call for AI Ethics", issued in February 2020, which seeks to create a movement to promote the responsible use of artificial intelligence. "We have a responsibility to avoid disastrous consequences that directly touch the lives of individuals and all society. We must construct a society where humans are the masters of technology, not its slaves," said the Archbishop.

The presentation spurred many questions from the audience seeking basic information about the intersection of data, AI, and ethics. "Do ethical standards for data exist, like something similar to GAAP in accounting?" asked one attendee. Another posited, "Is there—or will there ever be—an international "AI Ethics Regulator" who can help determine "good" versus "bad" versus "illegal" implementations of AI-driven implementations?"

These questions and others were addressed in subsequent panel discussions with CDOs and AI champions, as well as several sessions dedicated to AI ethics, including a panel session titled "Trusted AI: Data Executives Are Building an Ethical Foundation for AI" and "Women in Data Ethics". (All sessions were recorded and available to registered attendees for replay and view.)

Data Sharing

More than eight sessions focused on emerging strategies and technologies for sharing and monetizing data with external constituents, including suppliers, customers, partners, and consumers. The sessions ranged from best practices for integrating open data sources to strategies for creating a data marketplace to the ethics of data sharing. The underlying message was not whether organizations should share data—that was a given—but how to do it effectively, profitably, and responsibly.

Speakers framed three questions to help shape an external data-sharing strategy: 1) Why do I need this data? 2) Where will I get it? and 3) How may I use it? Data remains expensive to acquire—not just in terms of license cost, but also in the time and resources required to integrate it with internal systems. The key is to let business needs drive data acquisition, and only buy data if you know what questions it will answer.

Technology. The number of data providers multiplies every day as new data exchanges and marketplaces make it easier for companies to monetize and distribute data. Sorting through these providers is challenging. Look for platforms that pre-vet or even directly manage data assets to increase the likelihood of finding data that will meet your needs. Several sponsors of MIT CDOIQ Symposium offer data sharing solutions, including Dawex and Explorium.

Privacy. Finally, in a GDPR world, there are real consequences for companies that take liberties with consumers' personal data. Several speakers mentioned the importance of getting legal teams involved from day one in major data acquisition projects. Just because data is for sale doesn't mean it's okay to use for any purpose.

CDO Success Strategies

Operating Model. Many CDOs at the event spoke about their success factors and lessons learned. Dr. Matthew Graviss, Chief Data Officer at the U.S. State Department said his organization combined data management and analytics under one roof, to improve value for its internal users. "You need analytics to pay the bills since most people don't understand data management and are less willing to fund it. This organizational alignment helps us deliver more value."

Drilling in further, Kjersten Moody, CDO of Prudential Financial, said their data team uses a hub-and-spoke operating model where the corporate data team (i.e., the hub) manages enterprise data and the departments (i.e., the spokes) use the data to drive business value. "We play defense through hub teams and offense through spoke teams, and that has worked out very well."

Data Culture. Like many CDOs, Vipin Gobal of Eli Lilly and Company believes that fostering a data-driven culture is the number one goal of a CDO. That culture manifests itself when leaders ask for data in business meetings to answer questions. To make that happen, CDOs are investing heavily in education at all levels of the organization, from the board down to front-line workers. "We have spent the past year investing in top-down and bottom-up education, starting with the board, to help them understand the value of data, ML, and AI, and their potential pitfalls," said Tracy Daniels, CDO of Truist Financial.

Data Strategy. Maria Villar, head of enterprise data strategy and transformation at SAP, presented "10 Tips for Creating an Effective Data Strategy" where she iterated the importance of aligning data strategy with business strategy. This is something that CDOs understand, but find hard to implement. Villar encouraged CDO's first to "do their homework" by researching the business strategy in annual reports, SEC filings, and board meeting minutes. Then they need to orient the data strategy to business outcomes and define value using business metrics, not data

metrics. Finally, they need to tell a compelling data story that generates excitement and engagement among executives and staff.

Data Consistency. To address data consistency problems, Cathy Doss said Capital One created a “metrics lexicon” to store key metrics and their definition, including how they are calculated and what data is required to populate them. Going further, the company attached a seal to any report, dashboard, or PowerPoint presentation that properly embedded the metrics. “The seals gave executives confidence when making decisions with the data,” she added. “And if a report didn’t contain a seal, they would ask why. It worked out very well.”

Data Efficiency. Doss, who is now CDO of the Federal Reserve Bank of Richmond, also gave three suggestions for avoiding churn caused by “data wrangling”—where data analysts spend an inordinate amount of time fixing data issues and reworking solutions instead of analyzing data. To improving operational data efficiency, Doss said CDOs need to “coach, change the culture, and build it right.”

Rather than talk in technical terms to businesspeople, Doss recommends explaining how business processes can wreak havoc on data professionals in terms that businesspeople can understand. “Senior executives don’t get ‘data management’ but they do understand the need to ‘manage data’, especially if you can give them examples that hit home, like using a person’s first name or nickname when sending correspondence rather than first, middle, and last name.” She also encouraged businesspeople to ask how hard a request will be to fulfill and for data professionals to push back when they see an opportunity to consolidate or refactor requests to make them easier to deliver.

Communications Skills. [Wayne Eckerson](#), veteran data & analytics consultant and thought leader, underscored a similar theme in his MIT CDOIQ Symposium presentation, *Making Self-Service Work: How to Align, Empower, and Retain Data Analysts*. He encouraged CDOs to continually educate data analysts about how to ask the right questions when working with businesspeople:

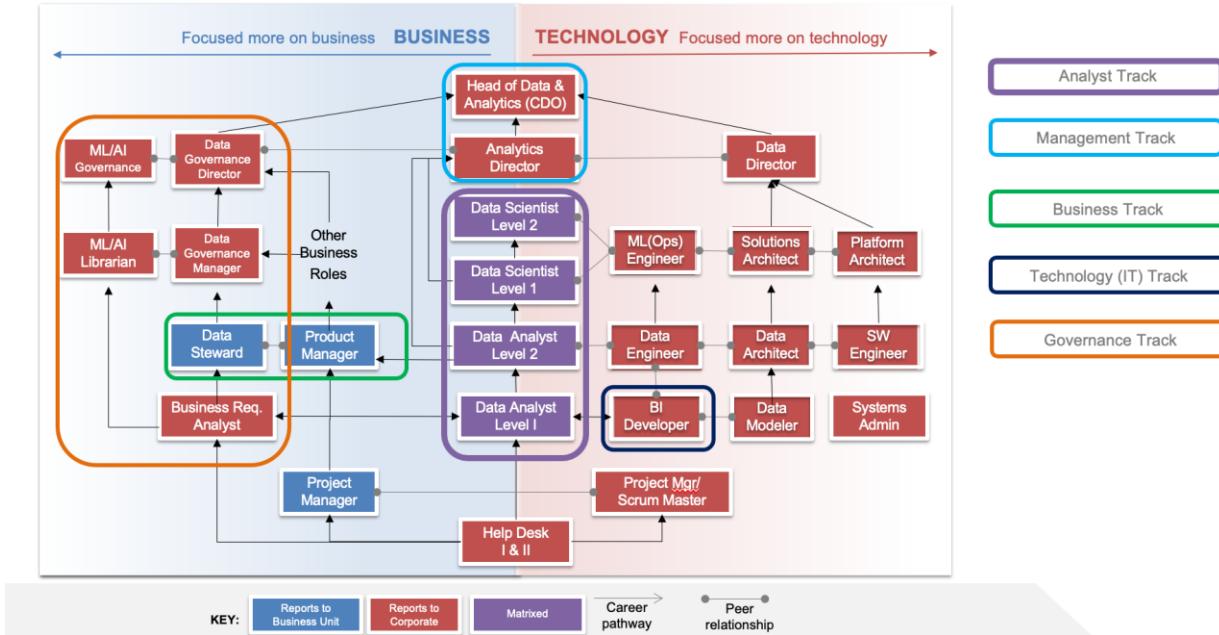
- What are you trying to accomplish?
- What’s it worth to you?
- What actions will you take as a result?
- Are you ready to implement the actions?

He also provided guidelines for how analysts should answer business questions.

- State the question or hypothesis
- Provide the answer with supporting data
- Never describe the analysis steps
- And the kicker: put everything on one slide!

Career Paths and Rotations. Finally, he said CDOs should rotate analysts through departments to gain domain experience and implement career pathways to keep young ambitious analysts engaged, focused, and motivated. (See figure 1.)

Figure 1. Data & Analytics Career Pathways in a Large Organization



Logical Versus Consolidated Data Architecture

Several sessions underscored a difficult architectural question that many organizations struggle to answer: is it best to consolidate data in a single place or provide access to it wherever it resides? The answer varies by use case, data volume, and enterprise environment.

Aaron Carreras and Nate Weisz of the Financial Industry Regulatory Authority (FINRA) described how their organization consolidates massive volumes of data into an Amazon S3 data lake to address analytics use cases. There they transform the data with Apache Spark, query it with Presto, and run a variety of data science algorithms on it. The data volumes are so high that FINRA needs to have all the data in one place to reduce processing overhead. The FINRA team also contains cloud compute costs by leveraging the spot market for compute cycles that will lower costs, while adjusting resources to meet its service level agreements.

The other option is to analyze data wherever it resides, whether different databases, systems, or locations. The emerging paradigm of the data mesh supports this approach by creating a distributed environment of datasets and data owners. Starburst CEO and co-founder Justin Borgman hosted an educational panel discussion on this topic with Jonathan Reardon at Liberty Mutual, Karthik Kirubakaran at CVS Health, and Srini Sankar with Hanover Insurance.

Eckerson Group believes that if the processing overhead is not too high, it's ok to use distributed query engines, such as Presto, to query federated environments. This helps avoid the costs of ingesting and consolidating data in a single store. But Srinivasan Sankar, enterprise data leader at Hanover Insurance, in a metaphorical observation, said chipmunks gather acorns and hide them

in many places as they prepare for winter, but to survive, they need to remember where they put them. The point is that enterprises need to track and document where all their data resides to make a federated approach successful.

Ultimately, each organization needs to evaluate the architectural tradeoffs and develop a strategy that makes sense for its environment and culture.

Summary

Planning for next year's MIT CDOIQ Symposium is already underway with track themes and managers already assigned. The hybrid event will take place online and MIT's campus in Cambridge, Massachusetts. For more information, see <https://cdoiq.mit.edu>.

[Eckerson Group](#) helps large organizations around the world develop enterprise data strategies, modernize their data infrastructures, develop and enhance data governance programs, gain traction for data science initiatives, and build self-service strategies that work. [Schedule](#) a 30-minute call.