Data strategies and modernization initiatives misaligned with the overall business strategy—or too narrowly focused on AI—leave substantial business value on the table.

Modernizing data with strategic purpose

Preface

"Modernizing data with strategic purpose" is an MIT Technology Review Insights report sponsored by Thoughtworks. To produce this report, MIT Technology Review Insights conducted a global survey of senior executives across countries and industries. The report also draws on in-depth interviews conducted with experts on data strategy and data modernization.

Denis McCauley was the author of the report, Teresa Elsey was the editor, and Nicola Crepaldi was the producer. The research is editorially independent, and the views expressed are those of MIT Technology Review Insights.

We would like to thank the following executives for their time and insights:

Heath Bland, Manager, Upstream Data Office, ExxonMobil
James Morgan, Chief Data Officer, The Crown Estate
Conrad Pozsgai, Managing Director, Chief Information Officer, and Chief Operating Officer, Payback
**Danilo Sato,** Global Head of Technology, Data & AI, Thoughtworks
John Spens, Managing Director, Data & AI, North America, Thoughtworks

About the survey
The survey forming the basis of this report was conducted by MIT Technology Review Insights in January and February 2024. The survey sample consists of 350 senior data and technology executives.

The respondents work in organizations based in the US, the UK, Germany, Singapore, and Australia. Eight industries are represented in the sample: financial services, energy, health care and life sciences, manufacturing, public sector, publishing and media, retail, and travel and transport. All respondents work in organizations earning $500 million or more in annual revenue, and nearly half (47%) in organizations with revenue of $10 billion or more.
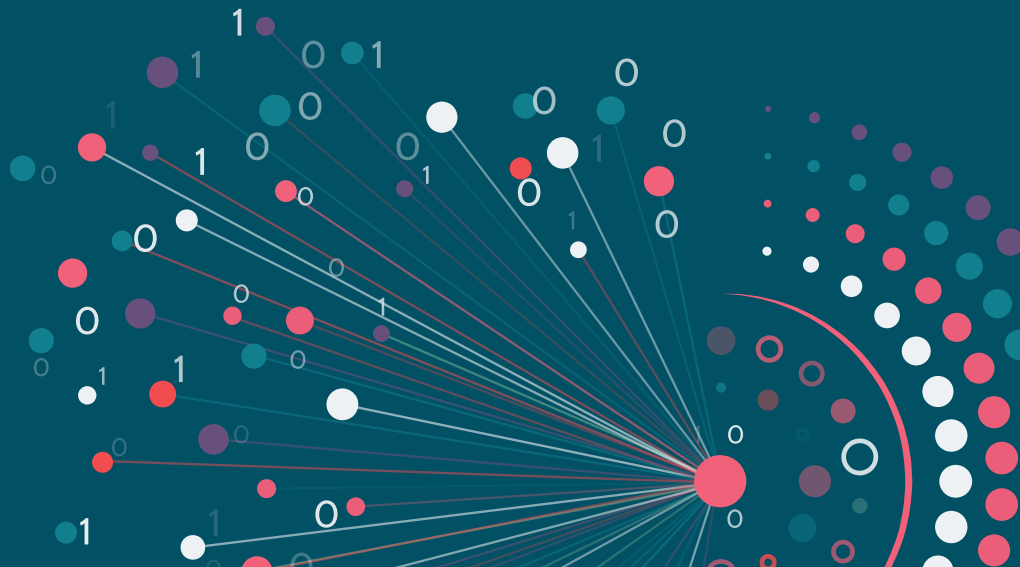
Foreword

Data has become an increasingly critical component to business success. The ability to make timely and informed decisions, create tailored customer experiences, and identify new opportunities has enabled the rise of today's leading digital companies. And with the recent explosion of interest in AI, the demand for data has only increased.

No matter what an organization hopes to achieve, success is impossible without ready access to high-quality data. Despite advances in technologies, extracting and transforming enterprise data into a usable asset remains a tremendous challenge for most organizations.

While challenges like improving data quality and enabling effective governance have existed for decades, this report demonstrates that today's technology leaders are exploring new ways to address these challenges. While new technologies alone may not solve the problem, modern, cloud-based data and analytics solutions enable technology organizations to embrace modern data engineering practices, product-oriented teams, and a more holistic approach to managing data. The report findings highlight why it's crucial for data leaders to look beyond their systems and embed processes, practices, and ways of thinking that can help their organizations solve persistent data challenges in a repeatable and scalable way.

Across Thoughtworks' extensive data modernization experience, we've seen how leveraging modern engineering practices that once revolutionized software engineering and applying them to the data discipline — using data as a product, building evolutionary architectures, embracing practices like CI/CD and DataOps, and evolving team and governance structures — can help organizations achieve a high level of data readiness and maintain it in the long term.

With all those elements in place, organizations can ensure they're always able to drive value from data at speed and scale — no matter how their governance requirements and data demands evolve.

As AI pushes data modernization to the top of more organizations' agendas, it will also become increasingly important to ensure that the data strategy is closely aligned to the broader business strategy and data leaders are able to clearly articulate how data and analytics can help business units achieve their objectives. This will not only ensure critical buy-in, but also put the organization in the best position to achieve its goals and lay a solid foundation for continuous improvement, evolution, and value creation.

Kalyanasis Banerjee
Global Head of Data and AI, Thoughtworks

CONTENTS

01 Executive summary

Data modernization is squarely on the corporate agenda. In our survey of 350 senior data and technology executives, just over half say their organization has either undertaken a modernization project in the past two years or is implementing one today. An additional one-quarter plan to do so in the next two years. Other studies also consistently point to businesses' increased investment in modernizing their data estates.[1]

It is no coincidence that this heightened attention to improving data capabilities coincides with interest in AI, especially generative AI, reaching a fever pitch. Indeed, supporting the development of AI models is among the top reasons the organizations in our research seek to modernize their data capabilities. But AI is not the only reason, or even the main one.

This report seeks to understand organizations' objectives for their data modernization projects and how they are implementing such initiatives. To do so, it surveyed senior data and technology executives across industries. The research finds that many have made substantial progress and investment in data modernization. Alignment on data strategy and the goals of modernization appear to be far from complete in many organizations, however, leaving a disconnect between data and technology teams and the rest of the business. Data and technology executives and their teams can still do more to understand their colleagues' data needs and actively seek their input on how to meet them.

Following are the study's key findings:

• **AI isn't the only reason companies are modernizing the data estate.** Better decision-making is the primary aim of data modernization, with nearly half (46%) of executives citing this among their three top drivers. Support for AI models (40%) and for decarbonization (38%) are also major drivers of modernization, as are improving regulatory compliance (33%) and boosting operational efficiency (32%).

• **Data strategy is too often siloed from business strategy.** Nearly all surveyed organizations recognize the importance of taking a strategic approach to data. Only 22% say they lack a fully developed data strategy. When asked if their data strategy is completely aligned with key business objectives, however, only 39% agree. Data teams can also do more to bring other business units and functions into strategy discussions: 42% of respondents say their data strategy was developed exclusively by the data or technology team.

• **Data strategy paves the road to modernization.** It is probably no coincidence that most organizations (71%) that have embarked on data modernization in the past two years have had a data strategy in place for longer than that. Modernization goals require buy-in from the business, and implementation decisions need strategic guidance, lest they lead to added complexity or duplication.

- **Top data pain points are data quality and timeliness.** Executives point to substandard data (cited by 41%) and untimely delivery (33%) as the facets of their data operations most in need of improvement. Incomplete or inaccurate data leads enterprise users to question data trustworthiness. This helps explain why the most common modernization measure taken by our respondents' organizations in the past two years has been to review and upgrade data governance (cited by 45%).

- **Cross-functional teams and DataOps are key levers to improve data quality.** Modern data engineering practices are taking root in many businesses. Nearly half of organizations (48%) are empowering cross-functional data teams to enforce data quality standards, and 47% are prioritizing implementing DataOps (cited by 47%). These sorts of practices, which echo the agile methodologies and product thinking that have become standard in software engineering, are only starting to make their way into the data realm.

- **Compliance and security considerations often hinder modernization.** Compliance and security concerns are major impediments to modernization, each cited by 44% of the respondents. Regulatory compliance is mentioned particularly frequently by those working in energy, public sector, transport, and financial services organizations. High costs are another oft-cited hurdle (40%), especially among the survey's smaller organizations.

Modern data engineering practices are taking root in many businesses. These sorts of practices, which echo the agile methodologies and product thinking that have become standard in software engineering, are now proving their value in the data realm.
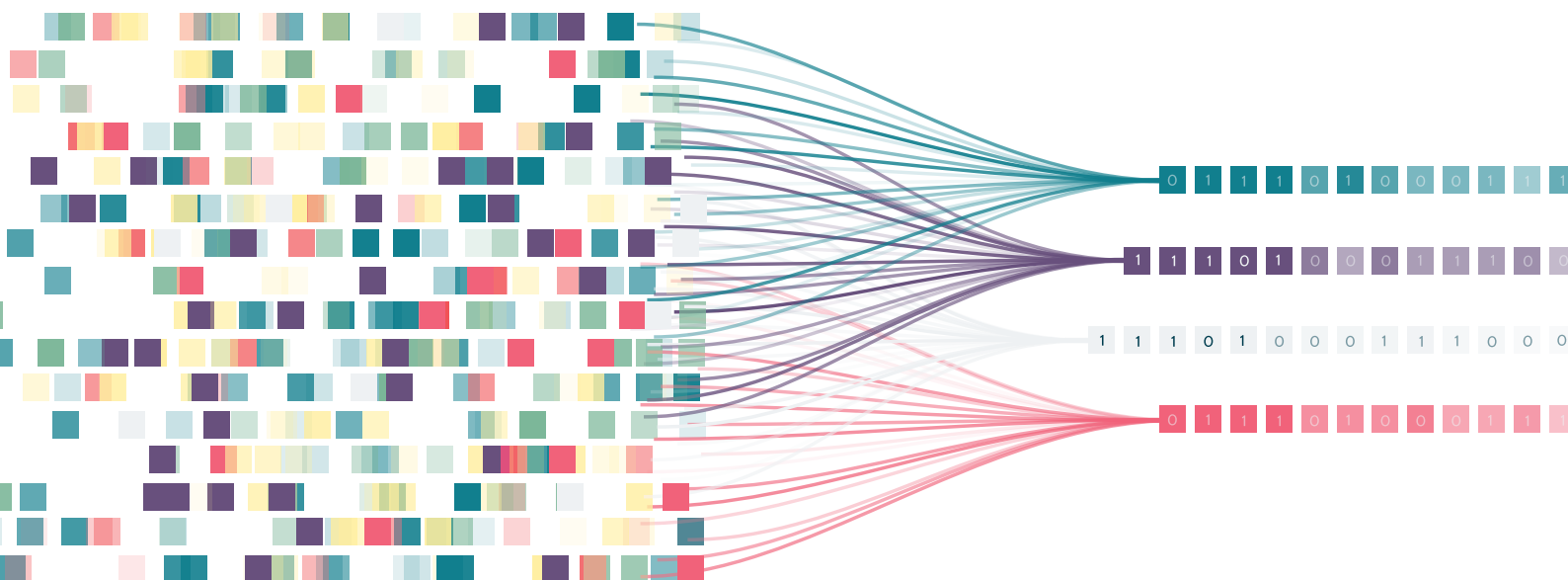
02

A vision for data

Organizations' growing adoption of AI in the past decade has heightened attention on the state of their data. AI models, after all, produce little of value without sufficient data and may prove harmful with data of poor quality.

The recent emergence of generative AI models and applications has further expanded organizations' understanding of the tremendous potential of data, and particularly the unique value they may find in their own internal, unstructured, or otherwise neglected data. It has also uncovered many of the difficulties of moving forward with these opportunities when their underlying data foundations are immature.

But it shouldn't have required enterprise-ready AI to convince management that their data is a strategic asset. Many critical enterprise operations, from risk management to demand forecasting, relied on a strong data foundation long before AI became a useful business tool.

Developing a data strategy

The importance of data is widely understood at this point, and accordingly, the majority of organizations have a fully developed data strategy. Only around one-fifth of the executives in our survey (22%) say their enterprise does not.

This, however, is a recent development. Only 27% of the executives surveyed say their organization's data strategy has been in existence for more than two years (see Figure 1).

The smaller organizations in the survey (those with annual revenue between $500 million and $9.9 billion) are more likely than the larger ones (with revenue of $10 billion or more) to lack a data strategy and much less likely to have had one for more than two years. These are the first of several divergences in response between organizations of different sizes. Taken together, they suggest that organization size and resources make a difference in certain facets of strategy and modernization.

"Data strategy should naturally address technology and skill needs, but these should be tied to the organization's overall business objectives. Otherwise the strategy won't deliver the value the organization really needs."

**Danilo Sato**, Global Head of Technology, Data & AI, Thoughtworks

Figure 1: Most organizations have a data strategy, but this is a recent development

Which of the following statements apply to your organization's data strategy?

We have yet to fully develop a data strategy

22% Total

15% Larger organizations

28% Smaller organizations
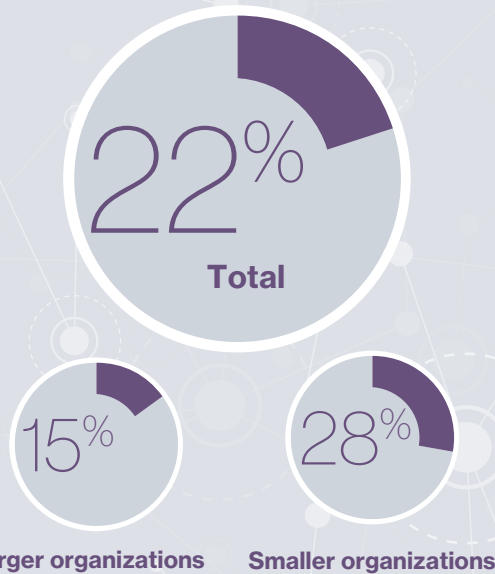
It has been in existence as a fully developed strategy for more than two years

27% Total

44% Larger organizations

12% Smaller organizations
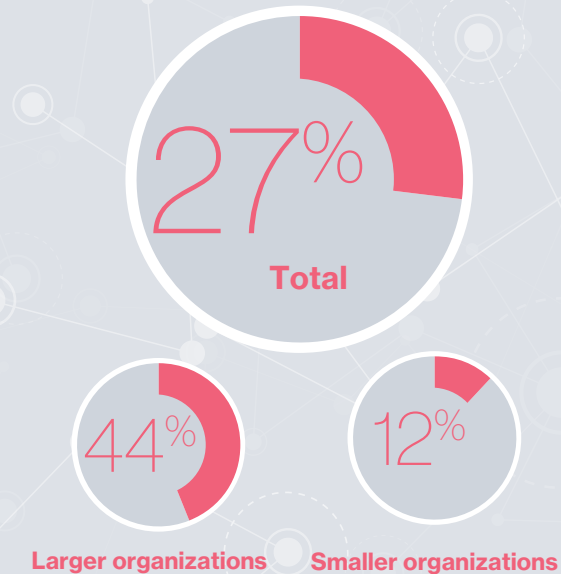
Source: MIT Technology Review Insights survey, 2024

"Every organization needs a data strategy," says James Morgan, chief data officer with The Crown Estate in the UK. "You absolutely must have a vision. And there's no point having a vision if you can't work out how you're going to get there." Morgan's team has developed an action plan for its data strategy: "We're unlikely to do all of it, as business priorities and needs will change, but you've got to have a plan to put the strategy into action."

Without a coherent approach to using data to solve business problems, data teams will forever be in reactive mode, according to John Spens, managing director for data and AI (North America), at Thoughtworks. The lack of a strategy, especially in large organizations where individual business units have sizable budgets, can lead to duplication and needless complexity. "This is how so many companies have found themselves operating in multiple clouds," he says. "Not through a thoughtful strategy, but through someone buying a solution residing on one vendor's cloud while others buy solutions from other vendors."
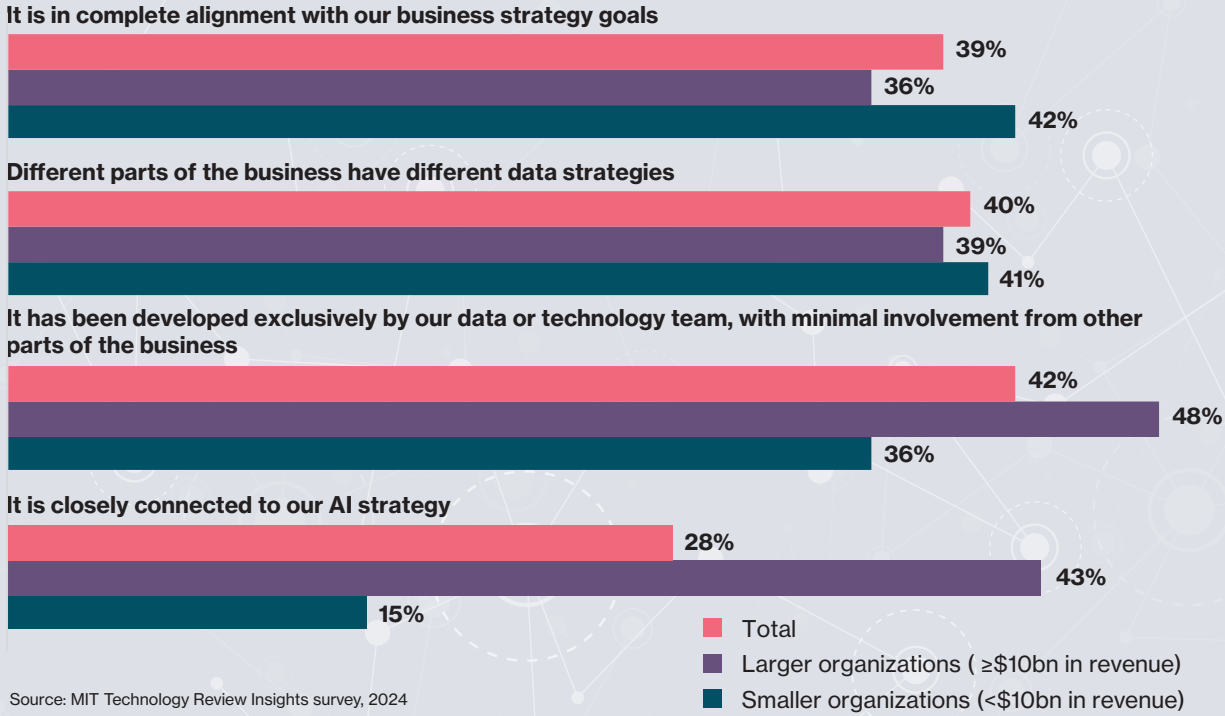
Incomplete alignment

Most of the surveyed organizations have a data strategy, but a sizable share have not fully aligned one across the enterprise (see Figure 2). At 40% of the respondents' organizations, different business units have their own data strategies, though that does not necessarily mean a lack of alignment. At energy provider ExxonMobil, for example, business-line data strategies are developed at corporate, business capability, and business line/ asset levels, according to Heath Bland, manager of the company's Upstream Data office. They're all coordinated, however, by the Central Data Office. "This approach ensures that the different strategies complement and build on each other, which we think is crucial for maximizing our data's value," he says.

A sizable share (39%) of surveyed executives say their data strategy is in complete alignment with their business strategy goals. For the majority who do not, however, that misalignment could be costly. Incomplete alignment leads to organizations failing to unlock value from their data strategy, according to Danilo Sato, global head of

Figure 2: Data strategy alignment is incomplete

Which of the following statements apply to your organization's data strategy?

It is in complete alignment with our business strategy goals

- 39%
- 36%
- 42%

Different parts of the business have different data strategies

- 40%
- 39%
- 41%

It has been developed exclusively by our data or technology team, with minimal involvement from other parts of the business

- 42%
- 48%
- 36%

It is closely connected to our AI strategy

- 28%
- 43%
- 15%

Legend:
- Total
- Larger organizations (≥$10bn in revenue)
- Smaller organizations (<$10bn in revenue)

Source: MIT Technology Review Insights survey, 2024

technology, data and AI, at Thoughtworks. "Data strategy should naturally address technology and skill needs, but these should be tied to the organization's overall business objectives," he says. "Otherwise the strategy won't deliver the value the organization really needs."

One reason for this apparent lack of alignment is that data strategy is often developed exclusively by the data team or IT, with minimal involvement from other parts of the business. Forty-two percent of survey respondents say this is the case in their organization.

The chief data officer (CDO) or their team, or in some cases the chief information officer (CIO), is very likely to lead the development of an organization's data strategy, but this should not happen in a vacuum. "As CDO, you may sit at the same table with all of the other CXOs, but you are also a service provider to them," says Spens. "Your responsibility is to understand their needs and be able to ensure that the strategy you're developing is aligned to those needs."

Such coordination is not a one-time effort, says Conrad Pozsgai, who is managing director, CIO, and chief

operating officer at Payback, a European multi-partner loyalty program and marketing platform. "It is a constant challenge to coordinate and align data strategy with other business functions," he says. "There's a lot of communication needed and a lot of reinforcement."

Organizations that are developing AI capabilities will also find it worthwhile to closely coordinate their data and AI strategies. Just over one-quarter of respondents (28%) say the two are closely connected in their business, though larger organizations (43%) are considerably ahead of smaller ones (15%) in this regard.

Experts see such integration becoming more commonplace. "It makes a lot of sense to coordinate closely," says Sato. He offers the example of teams using generative AI based on open-source large language models (LLMs). "If those have been trained on external data, teams may want to refine them by integrating some of their own data," he says. "Everyone's trying to figure out how to overlay their expert knowledge to leverage the LLM for their own purposes. To do that, you do need to think about data and AI together."

03

Purpose in modernization

A coherent, aligned data strategy is particularly vital as organizations set about modernizing their data capabilities. The majority of those in our survey have either modernized multiple elements of their data estate within the past two years (23%) or are in the process of doing so (31%). Another 23% plan to modernize within the next two years. Most of the organizations that have already modernized (71%) have also had a data strategy in place for more than two years (see Figure 3).
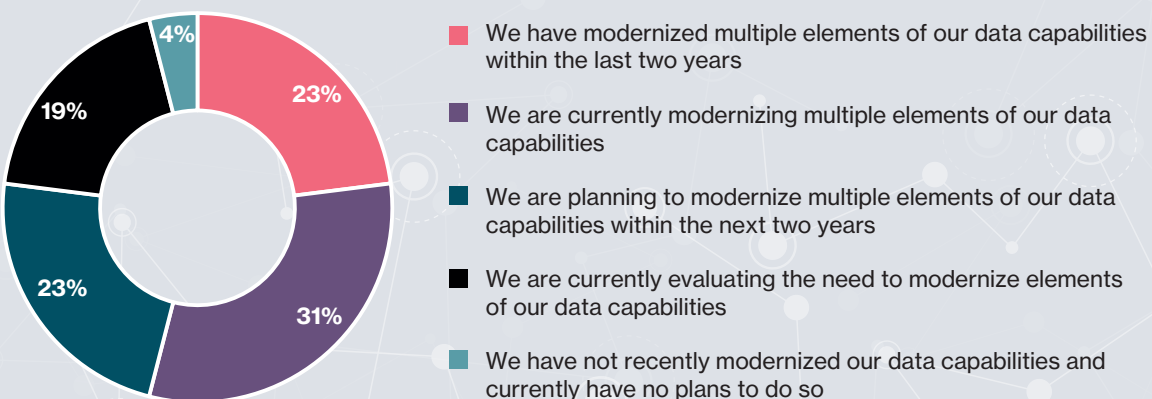
Data modernization can require considerable investment. This helps explain why the largest organizations in the survey, those with $10 billion or

more in annual revenue, are considerably more likely to have begun modernizing than smaller ones (see Figure 4). These investments also require some vision and patience to recoup: Sato says that, despite large data-related investments over the years, many companies have only recently begun reporting higher success rates on those investments.

The ability to demonstrate early returns, whether in financial terms or other key metrics, however, is vital to sustaining modernization buy-in. Modernization plans designed to show early successes often create momentum that aids with the plans' full execution. "While investments in modernizing data usually have a two- to five-year horizon to pay off, the approach

Figure 3: Status of data modernization: all respondents

Which of the following best describes your organization's current approach to its data capabilities?

- We have modernized multiple elements of our data capabilities within the last two years
- We are currently modernizing multiple elements of our data capabilities
- We are planning to modernize multiple elements of our data capabilities within the next two years
- We are currently evaluating the need to modernize elements of our data capabilities
- We have not recently modernized our data capabilities and currently have no plans to do so

Source: MIT Technology Review Insights survey, 2024

"We believe in an incremental delivery approach — in thin slices — that demonstrates early returns. While smaller in scale, early wins demonstrate confidence in the modernization strategy and get internal buy-in that helps propel the longer-term roadmap."

Danilo Sato, Global Head of Technology, Data & AI, Thoughtworks

an organization takes to get there can make a big difference," says Sato. "Instead of front-loading infrastructure, platform, and data ingestion efforts, and pushing use case delivery after data is available, we believe in an incremental delivery approach — in thin slices — that demonstrates early returns. While smaller in scale, early wins demonstrate confidence in the modernization strategy and get internal buy-in that helps propel the longer-term roadmap."
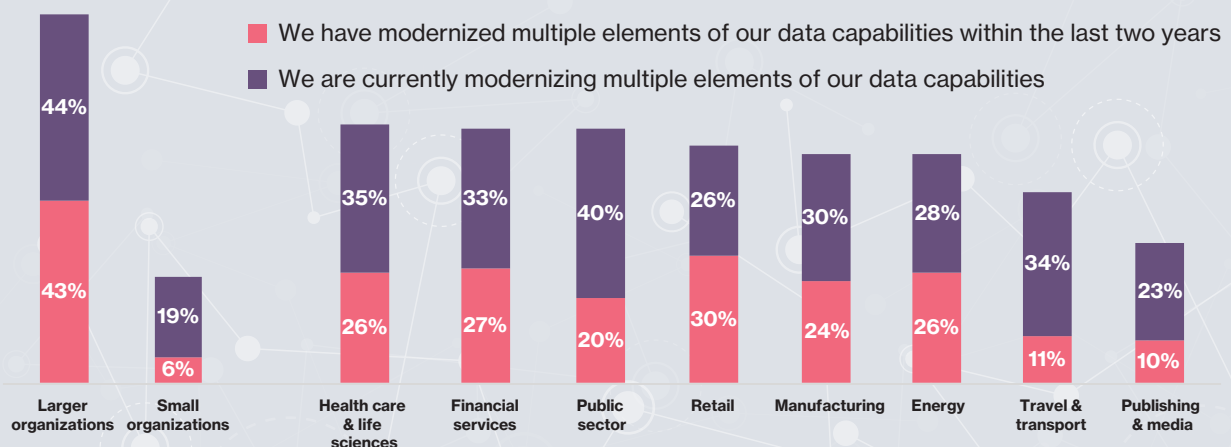
Impediments to modernization

High cost figures among the chief impediments to data modernization: 40% of respondents cite this as a barrier. Not surprisingly, it appears to be a greater burden for the smaller organizations in the survey (cited by 46% of this group). But one-third of the larger organizations report that cost is a factor as well (see Figure 5).

The most frequently mentioned impediments to modernization, however, are regulatory compliance considerations and security concerns (both cited by 44% of respondents). In heavily regulated industries such as financial services and health care, compliance considerations may, for example, proscribe the unification of certain data repositories or limit the data that can be stored in public clouds. Regulation looms particularly large as a modernization impediment for respondents in energy, public sector, transport, and financial services organizations.

Security concerns also hinder some modernization efforts. In the survey, these are emphasized especially strongly by public-sector respondents as well as by those working in the financial and retail industries — all common targets for cyber criminals seeking critical personal and banking information.

Figure 4: Status of data modernization: by organization size and industry
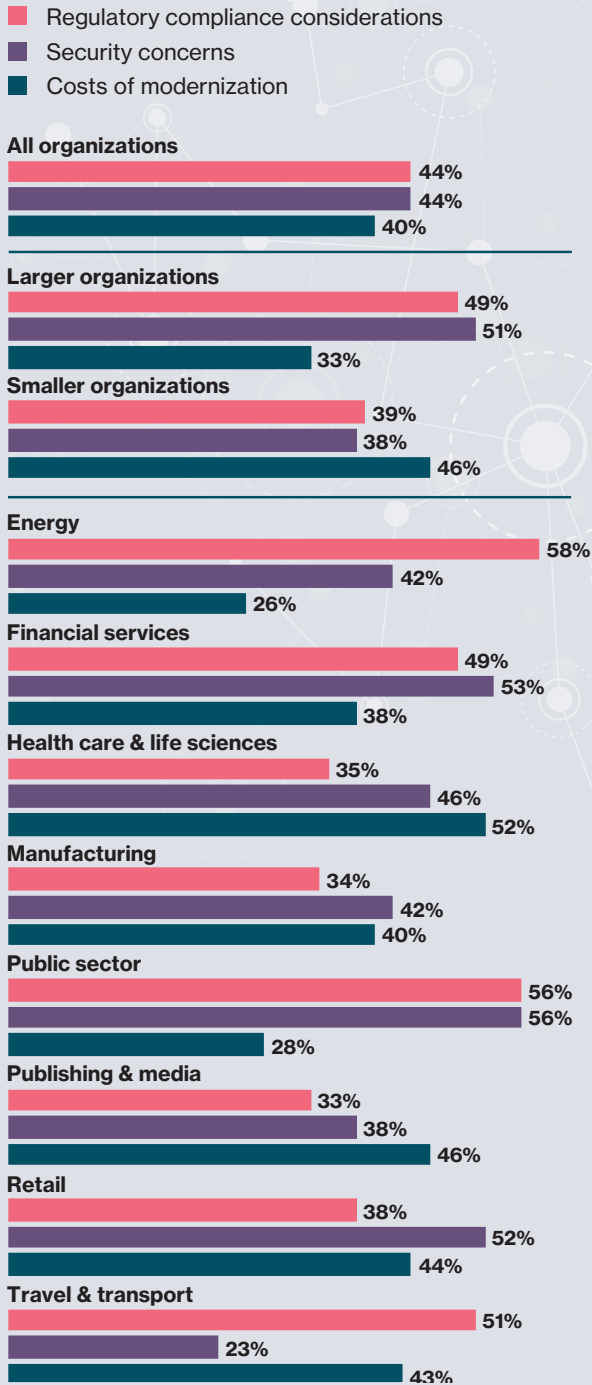
Which of the following best describes your organization's current approach to its data capabilities?

Legend:
- We have modernized multiple elements of our data capabilities within the last two years
- We are currently modernizing multiple elements of our data capabilities

	Larger organizations	Small organizations	Health care & life sciences	Financial services	Public sector	Retail	Manufacturing	Energy	Travel & transport	Publishing & media
Currently modernizing	44%	19%	35%	33%	40%	26%	30%	28%	34%	23%
Modernized	43%	6%	26%	27%	20%	30%	24%	26%	11%	10%

Source: MIT Technology Review Insights survey, 2024

Figure 5: What's holding data modernization back?

Which of the following are the chief impediments to modernization of your organization's data capabilities? (Top 3 responses.)

- ■ Regulatory compliance considerations
- ■ Security concerns
- ■ Costs of modernization

All organizations
- 44%
- 44%
- 40%

Larger organizations
- 49%
- 51%
- 33%

Smaller organizations
- 39%
- 38%
- 46%

Energy
- 58%
- 42%
- 26%

Financial services
- 49%
- 53%
- 38%

Health care & life sciences
- 35%
- 46%
- 52%

Manufacturing
- 34%
- 42%
- 40%

Public sector
- 56%
- 56%
- 28%

Publishing & media
- 33%
- 38%
- 46%

Retail
- 38%
- 52%
- 44%

Travel & transport
- 51%
- 23%
- 43%

Source: MIT Technology Review Insights survey, 2024

Understanding cloud costs

Cloud is a frequent focus when discussing the costs of data modernization. Business leaders may fear that ever-expanding volumes of data needed for AI, coupled with the ease of pay-as-you-go scalability in the cloud, will lead to costs spiraling out of control. The transparency of cloud costs, as compared to those of the infrastructure cloud replaces, can also make cloud's price tag particularly apparent.

Modernization efforts that lead to more use of data can magnify existing inefficiencies. "Cloud costs are growing for companies of all sizes," says Danilo Sato, global head of technology, data and AI, at Thoughtworks. "For example, efforts to expand employees' access to data—a key enabler of data democratization—incentivize more cloud usage, which makes the cost of long-running queries or low-performance code visible quickly. Taking a modern approach to good data design and writing high-performance code, therefore, can lead to cost efficiencies that were previously invisible."

Beyond good data engineering practices, taking financial advantage of cloud may require a more dedicated way of looking at its costs. FinOps, a portmanteau of "finance" and "DevOps," refers to an emerging set of practices that bring finance and IT experts together to manage data, AI, and cloud costs. A FinOps approach can help an organization feel confident that its cloud costs are monitored, predictable, and in line with the business's overall strategy, with a goal of maximizing the organization's return on its cloud investments, not just cutting costs.

A variety of benefits for the business

What are organizations looking to achieve with data modernization? Ranked first is improved decision-making across the business. Nearly half (46%) of the surveyed executives cite this (see Figure 6).

This is a core aim of The Crown Estate's modernization effort. "We need to make the right decisions about the optimum usage of all of our assets for the benefit of the nation, whether parcels of land or properties in London," says Morgan. "And it's not just an optimum financial output. It needs to be a balance of societal, environmental, and value-creation benefits."

The second most-important modernization aim cited by respondents is supporting the development of AI use cases (40%). Cloud can be an important strategic option here, providing access to hardware and infrastructure that enables AI model training on a pay-as-you-go basis. "Supporting AI models is definitely a driver of modernization for us," says Pozsgai. "That's part of why we're moving into the cloud, because it's too expensive to run models on-premises. We also need the technology and other capabilities of our cloud provider."

This varies significantly with organization size: the larger organizations surveyed are two-and-a-half times more likely to say support for AI is a goal of data modernization than their smaller brethren (at 58% and 23%, respectively).
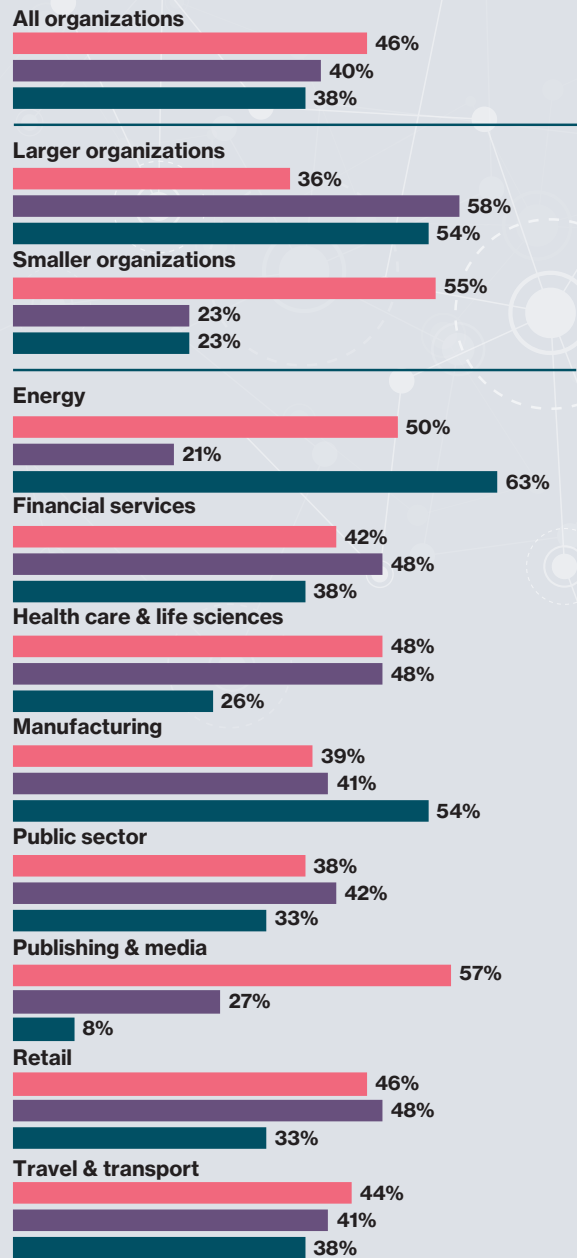
"We need to make the right decisions about the optimum usage of all of our assets for the benefit of the nation. And it's not just an optimum financial output. It needs to be a balance of societal, environmental, and value-creation benefits."

James Morgan, Chief Data Officer, The Crown Estate

Figure 6: The chief aims of data modernization

What are your organization's main motivations for modernizing its data capabilities? (Top 3 responses.)

- Improve decision-making across the business
- Support AI use cases
- Support decarbonization and reduction of environmental footprint

All organizations
- 46%
- 40%
- 38%

Larger organizations
- 36%
- 58%
- 54%

Smaller organizations
- 55%
- 23%
- 23%

Energy
- 50%
- 21%
- 63%

Financial services
- 42%
- 48%
- 38%

Health care & life sciences
- 48%
- 48%
- 26%

Manufacturing
- 39%
- 41%
- 54%

Public sector
- 38%
- 42%
- 33%

Publishing & media
- 57%
- 27%
- 8%

Retail
- 46%
- 48%
- 33%

Travel & transport
- 44%
- 41%
- 38%

Source: MIT Technology Review Insights survey, 2024

"We're changing a lot with our data infrastructure and processes. It's part of a broader effort to become what I call a product delivery organization."

Conrad Pozsgai, Managing Director, Chief Information Officer, and Chief Operating Officer, Payback

Ranked third as a modernization driver (by 38%) is supporting decarbonization and other efforts to reduce the business's environmental footprint. For those organizations that have already undertaken data modernization in the past two years, supporting AI and supporting decarbonization have been by far their key motivations (cited by 70% and 60% of this subgroup, respectively). Improving regulatory compliance (33%) and operational efficiency (32%) are other important modernization drivers.

Immature data capabilities

As data modernization efforts get under way, executives should have a clear view of the weaknesses in their existing data capabilities, particularly those that could jeopardize achievement of their stated modernization objectives. Among many possible weaknesses listed in the survey, executives narrow in on several causing particular concern. These include talent and skills gaps in their data teams, the data strategy (which may relate to gaps in existing

Payback: Moving faster with data

Business success at Payback is predicated on designing effective loyalty programs and multi-channel marketing campaigns for its retail clients in Germany and 10 other countries. Considerable responsibility rests with the company's technology and data teams, headed by Conrad Pozsgai, to deliver new program features and new campaigns at an ever-increasing pace. "Our main goal as a technology organization is to be quicker to market with new features," says Pozsgai. "And of course if we're quicker, we also lower the cost of developing them."

To help gain the desired speed, Payback has launched a major modernization of its data estate, including a shift to modern data engineering practices. "We're changing a lot with our data infrastructure and processes," says Pozsgai. "It's part of a broader effort to become what I call a product-centered delivery organization."

In addition to adopting a new, modular data architecture, the modernization also involves considerable structural change. Pozsgai explains
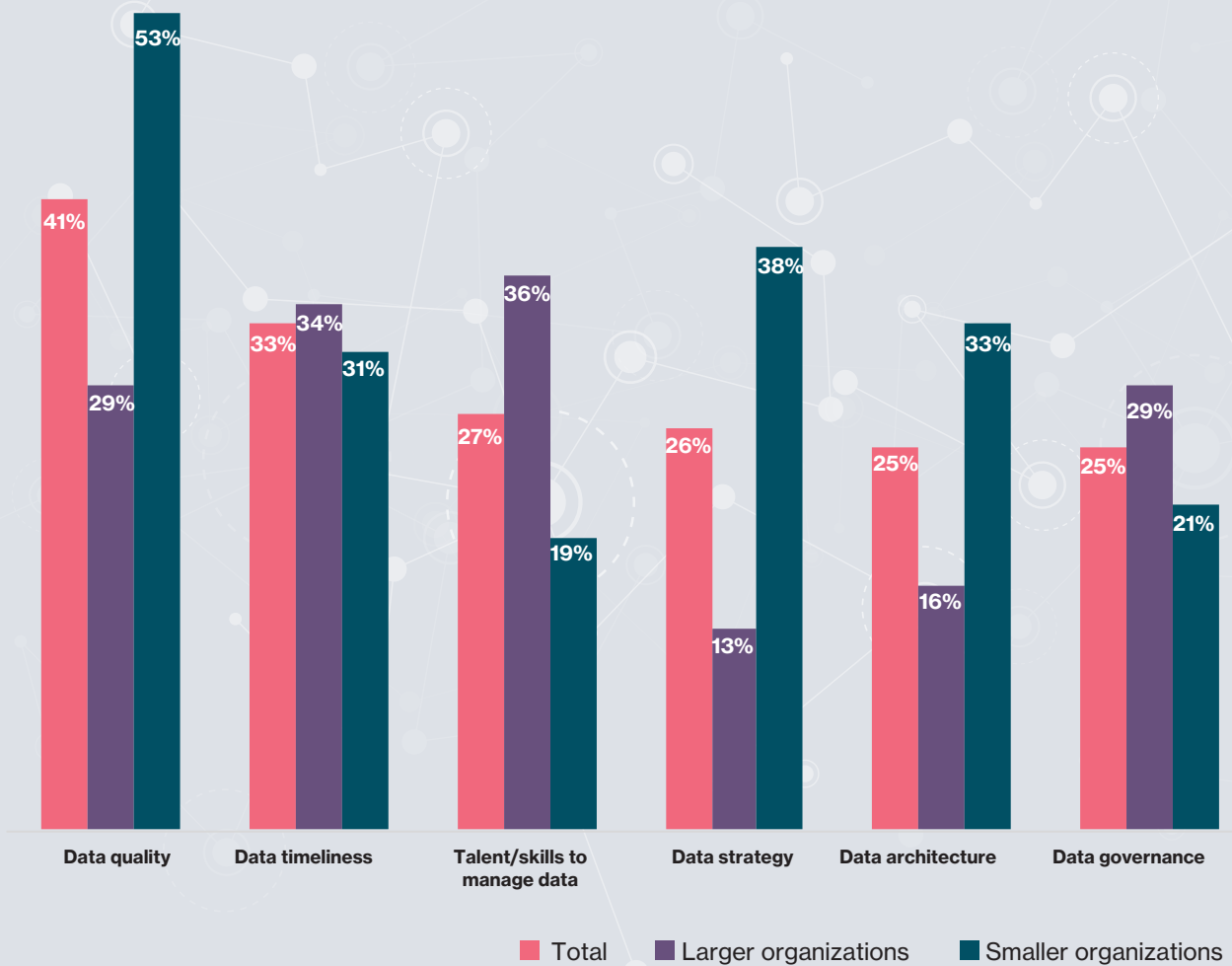
that the data organization is transitioning from a centralized structure to one that is domain-driven: "This mirrors our business areas, so we're gradually shifting our data teams into the same domains. They'll work cross-functionally for a product or product cluster; that's a big change for all our data people."

A domain-driven structure accords the data teams autonomy to work semi-independently, with Pozsgai stressing that they need to be able to build data products on their own. It's not all happening at once, he says: "For now, all our teams are fully occupied with the change and will need central support. But over time they will become increasingly independent."

Data analysis and other data work can no longer be performed as it was in the past. Pozsgai says, "Transactions and communication are happening in real time, and data analysis must increasingly be done on the fly. So we need to integrate data, like any other technology area, into the overall delivery process."

**Figure 7: Weak links in organizations' data capabilities**

Which elements of your organization's data capabilities are least ready today to support your data modernization objectives? (Respondents who ranked in their top 3.)

	Total	Larger organizations	Smaller organizations
Data quality	41%	29%	53%
Data timeliness	33%	34%	31%
Talent/skills to manage data	27%	36%	19%
Data strategy	26%	13%	38%
Data architecture	25%	16%	33%
Data governance	25%	29%	21%

Source: MIT Technology Review Insights survey, 2024

strategy or even the absence of one, as discussed earlier), and the state of their organization's data architecture (see Figure 7).

What concerns survey respondents most, however, is the quality of their data (41%) and its timeliness (33%). Fifty-three percent of survey respondents from smaller organizations identify data quality as a top weakness. As we will discuss below, executives see their efforts to improve governance as key to addressing these data quality and timeliness issues.
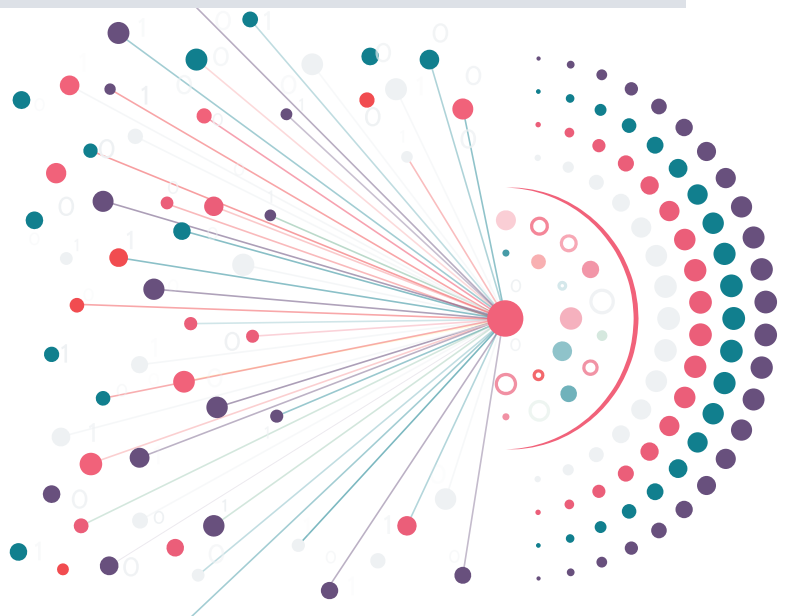
ExxonMobil: Profit and pain in data modernization

The eighth-largest company in the world in terms of global revenue, ExxonMobil generates, processes, and analyzes enormous flows of data from its upstream and downstream oil, natural gas, and petrochemical operations and assets. Attempting to modernize a data estate of such size and complexity is not for the faint-hearted.

According to Heath Bland, manager of ExxonMobil's Upstream Data Office, the organization has faced several challenges as it implements modernization, but two have presented particular difficulties. One is a limited scope of vendor solutions. "Despite the availability of cloud offerings, the scope of available solutions for a business in our industry remains narrow," he says. "Our market doesn't attract new players to develop solutions, due to the limited customer base and the uniqueness of the industry."

The other major challenge, says Bland, is the current state of the company's data culture: "This, combined with limited awareness and alignment across organizations of the technical challenges and the lack of direct, visible business value impact [of modernization], has led to insufficient funding, and constrained resources to enable data analytics."

But, says Bland, the rewards ExxonMobil is starting to see from modernization are tangible. He lists five key ones:

- **Deploying self-service analytics**: "We've empowered users to access and analyze data independently, which has enabled the training of AI models."
- **Making the enterprise's data FAIR**: "This means enhancing the findability, accessibility, interoperability, and reusability of our data."
- **Reducing the influence of siloes**: "We've been able to blend data across siloes to gain a global view, leading to new insights."
- **Improving industry data-sharing**: "We've adopted the OSDU data standards[2] and data platform for industry-level data integration."
- **Implementing industry data standards**: "This is enabling us to transform and modernize our business digital processes and to make data cheaper through consistency."

"We are starting to see some of these benefits already, but some will take more time," says Bland. The progress that's already been achieved, he says, is due to the company's pivot toward data-as-a-product thinking and the maturity of its data governance.

"Limited awareness and alignment across organizations of the technical challenges and the lack of direct, visible business value impact [of modernization], has led to insufficient funding, and constrained resources to enable data analytics."

Heath Bland, Manager, Upstream Data Office, ExxonMobil

04 Governing for quality

"Ensuring data quality is our biggest pain point," says Morgan. With the diverse types of data that The Crown Estate ingests, and with the range of formats it comes in, making sure the data is in one place and can be accurately measured and utilized is a constant challenge.

Substandard data quality – caused for example, by missing values, duplicate data (a common byproduct of siloes), errors in manual data entry, or other factors – leads to user questions about data trustworthiness. Timeliness is a related factor, adds Sato. "Failure to access data when enterprise users need it detracts from data quality," he says.

Pozsgai provides an example from Payback's operations: "We run a lot of real-time marketing campaigns. Customer actions in a shop or online will often trigger a push message from us in real time. It's critical that we have the data about the customer's actions available so we can act on it immediately."

Improving timeliness is critical in the oil and gas industry, says Bland. "The value of operations and subsurface data increases with real-time or near real-time access," he explains. "As we make progress in advanced analytics and AI, the importance of data timeliness will continue to grow and create new opportunities to transform business processes. We've already seen advancements in the form of increased network capacity and speed, digital twins, data platforms, data acquisition, and process automation, to name a few."

Improving data quality and enforcing the maintenance of data practices that ensure integrity are largely a function of good data governance. This may explain why organizations undergoing data modernization efforts have prioritized improving governance. Asked about the measures they've taken in the past two years to modernize their data capabilities (see Figure 8), respondents most commonly cite updating their governance models (45%).

"As we make progress in advanced analytics and AI, the importance of data timeliness will continue to grow and create new opportunities to transform business processes."

Heath Bland, Manager, Upstream Data Office, ExxonMobil

Figure 8: Time frames for data modernization measures

When will your organization take the following measures to modernize its data capabilities?

Review and update data governance

| 45% | 21% | 24% | 8% | 2% |

Adopt a new data architecture

| 36% | 32% | 20% | 8% | 4% |

Widen the use of cloud data services

| 31% | 29% | 25% | 11% | 4% |

Decentralize data storage and management

| 23% | 23% | 21% | 19% | 14% |
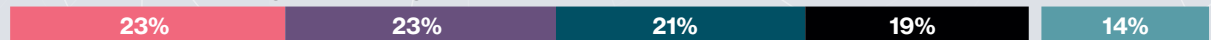
Revise the data organization structure

| 15% | 28% | 31% | 22% | 4% |
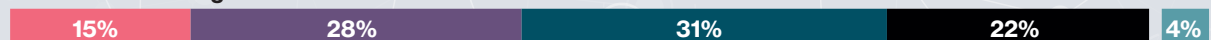
Re-engineer data processes

| 13% | 31% | 31% | 22% | 3% |
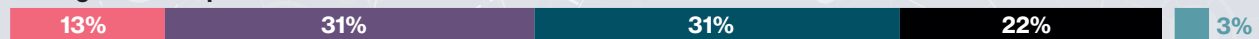
Consolidate the number of data repositories

| 12% | 33% | 33% | 16% | 6% |
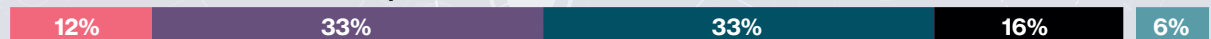
Legend:
- Have done in past 2 years
- Will do in next 6 months
- Will do in 6–12 months
- Will do in 1–2 years
- Not planning to do, not sure, or not applicable

Source: MIT Technology Review Insights survey, 2024

Computational governance—using software to automate the application of governance policies—is an emerging technique. "A modern approach to implement data governance is to support it computationally within the platform," says Sato. "With automation, organizations can either enforce or monitor their agreed policies—such as data security, data quality, access control, and data privacy—and shift data governance from a process-heavy to an enabling approach."

The next most common data modernization measure, completed by 36% of respondents, has been to adopt a new data architecture. There are a variety of views among organizations about the ideal data architecture model. Some favor a domain-driven approach, with a large degree of autonomy granted to cross-functional data teams. A key feature of Payback's data modernization initiative is a shift toward such a model. Other organizations are pursuing a more unified architecture with centralized platforms and data repositories. Still others seek to apply a hybrid approach combining centralized and decentralized elements.

There is a large degree of consensus, however, when it comes to governance models. To ensure the highest levels of data quality, there needs to be an overarching governance framework for the entire organization. As Payback moves toward a domain-specific structure, Pozsgai believes it is all the more important to maintain a centralized data governance model. "Of course, domain owners should monitor their data quality for their teams, but there are a lot of dependencies between different domains and teams," he says. "There needs to be an overarching governance framework."

Improving data quality and timeliness

The surveyed organizations are also taking more specific measures to improve data quality and timeliness (see Figure 9). Nearly half (47%) are prioritizing implementing DataOps for this purpose. Among the core principles of DataOps methodology is the automation of data quality management, including testing, monitoring, and anomaly detection. DataOps also emphasizes cross-functional collaboration by data teams, and, separately, 48% of respondents say their organizations are empowering such teams to enforce data quality standards.

Figure 9: Improving data quality and timeliness

What are your organization's top priorities for improving the quality and timeliness of the data used by different parts of the business?

48%
Empower cross-functional data quality teams to enforce quality practices

47%
Implement DataOps

35%
Improve data validation processes

31%
Implement enterprise-wide training on data quality and standards

29%
Redefine data quality standards

23%
Bring automation into data quality management

Source: MIT Technology Review Insights survey, 2024

For Spens, stewardship is the other vital ingredient to improving data quality: "The business must, of course, invest in technology measures designed to fix data quality problems. At a higher level, however, active stewardship from the top—the idea that I, as a business leader, care about the quality of our data—is just as important."

Organizational structure and the data team

All the foregoing highlights the importance of organizational and structural change to data modernization. "So much of modernization is rethinking where data teams sit within the enterprise and who they consist of," says Sato. The cross-functional teams to which the respondents attach so much importance should be part of any organizational set-up, he says. "Cross-functional teams are needed to drive projects. If you've got the domain experts, the data scientists and engineers, and the analysts in one team, you'll reduce handover delays and optimize for flow. The teams will get to value quicker."

Diversity of all kinds, including skills, is critical in data teams, says Morgan. This applies even just to the subset of data specialists within them. "There are people that I need in the team who are very good at coding, but who may not truly understand the business context," he says. "I then need other people to work with them who are the storytellers. They can understand enough about the technical, but they can speak truly about business outcomes and value."

What is DataOps?

DataOps is to data what DevOps is to software. Both are sets of philosophies and principles (rather than specific technologies), based on agile methodology and emphasizing automation and continuous improvement. The main goal of DevOps is to deliver better software at a faster pace. The main goals of DataOps are to improve data quality and deliver better insights to business end-users, while simultaneously reducing the cost of data management.

DataOps methodologies are intended to optimize how data teams design, build, and deliver, helping them more quickly create scalable, flexible, and reliable data products and architectures. A team using DataOps methodologies is likely to focus on rapid cycles of feedback and improvement via continuous delivery, enabled by automated testing and deployment.

Implementing DataOps also assumes the need for cultural change, which breaks down data siloes and fosters collaboration across teams and functions to ensure that data delivers value to the business.

The surveyed organizations appear to be leaving structural change until later in the modernization process. While many have updated data governance and adopted a new architecture in the past two years, just 15% of respondents say their companies have also addressed issues of organizational structure. The majority, 59%, say this will become a focus within the next year, while 22% expect at least a year will pass before they address such changes.

Modernizing the data estate is a big undertaking, and executives may feel it unwise to try to introduce too much change too fast. To the extent that changing people's roles and reporting structures are the toughest changes of all, overcoming the resistance that will inevitably emerge may require substantial attention from the organization's data and technology leaders.

"The business must, of course, invest in technology measures designed to fix data quality problems. At a higher level, however, active stewardship from the top is just as important."

John Spens, Managing Director, Data & AI, North America, Thoughtworks

The Crown Estate: Conquering data complexity

James Morgan and his team of data professionals have a unique remit: to deliver value, create efficiencies, and optimize decisions from the data generated by and about The Crown Estate's natural and built assets. Those holdings are highly diverse, ranging from agricultural lands to coastal environments and the seabed to commercial and residential buildings in London and other cities. Morgan has teams managing data relating to seabed mapping and modeling, carbon emissions from its farmlands, materials used in its residential buildings, footfall and turnover at its commercial properties, and numerous other categories. The organization's data systems, says Morgan, must support analytics, reporting, data science, and decision-making about all those assets.

To manage that data more efficiently and deliver more value from it, the organization is in the process of modernizing its entire data estate. Part of that effort is focused on the operational systems that capture its data. "We've been identifying where manual workarounds have seeped in," Morgan says. "We're making sure that our systems are capable of capturing all of the information needed to run an efficient business." Those operational systems, he explains, are the lifeblood of anything his teams want to do with the data, such as analytics, reporting, or data science. "So at one end we're focusing on data creation and data change and making sure that all

of the architecture and systems around that, and the business processes that go with it, are correct and the appropriate data management and governance is in place."

Morgan continues, "Then we're modernizing our data and analytics stack, including through the deployment into a cloud environment. We're building out a data ecosystem in the cloud with a data lake and structured environments above it. This a big piece of work to bring a number of legacy siloed systems into a modern cloud data architecture."

Morgan describes the organizational model to manage the modernized data estate as a sort of hybrid. "For a business like ours, it makes sense to have a central environment where we've got the right governance, the right protection, the right controls around it, and the experts who can manage it," he says. But that doesn't that mean all of the data needs to be joined into one common model. "We want to democratize data across the organization, putting it into the hands of our colleagues to support their everyday work and decision making," says Morgan. "For example, in many instances, data about our marine environments doesn't need to be linked to our London property assets," says Morgan. "There is no benefit in linking them, so why would we?" Sometimes the costs and time of bringing data together centrally outweigh the benefit.

05 Conclusion

This research offers lessons for enterprise heads of data and technology that are embarking on, or considering, the modernization of their data estate. Foremost among those are the following:

Keep AI goals in perspective. AI offers substantial promise to add value to the business, but other modernization objectives shouldn't get lost in the excitement. Delivering higher-quality data faster and more securely serves not only the needs of AI models but of many other business-critical systems, as well as other emerging technologies. The full business case for data modernization encompasses much more than just AI.

Data leaders must demonstrate the business value of modernization. A comprehensive data strategy is one that aligns fully with the business strategy, and data strategy and modernization approaches developed in isolation are sure to lead to wasted effort and resources. It is incumbent on senior data and technology leaders to understand how data can help business units achieve their objectives and tailor their efforts accordingly. They must also communicate the benefits of data modernization in terms that are relevant to the business leaders who approve and support data investments.

Software engineering practices are coming to data. The emergence of cross-functional data teams, DataOps practices, cloud-based solutions, and a focus on data as a product, for example, echo agile practices that have become standard in software engineering

over the last two decades. As organizations seek to accelerate the delivery of value from data, they find that adopting modern data engineering practices can address challenges around data quality and usability.

Approach change prudently, but keep moving. Data modernization may call for complex changes around a company's organizational structure and data architecture. Many enterprises find it tempting to delay these until later stages of the project, when more is known. But modernization initiatives gain momentum by demonstrating early value, which means that leaders may benefit from an agile mindset, taking early bets and being prepared to pivot.

About MIT Technology Review Insights

MIT Technology Review Insights is the custom publishing division of *MIT Technology Review,* the world's longest-running technology magazine, backed by the world's foremost technology institution — producing live events and research on the leading technology and business challenges of the day. Insights conducts qualitative and quantitative research and analysis in the US and abroad and publishes a wide variety of content, including articles, reports, infographics, videos, and podcasts. And through its growing MIT Technology Review Global Insights Panel, Insights has unparalleled access to senior-level executives, innovators, and entrepreneurs worldwide for surveys and in-depth interviews.

About Thoughtworks

Thoughtworks is a global technology consultancy that integrates strategy, design, and engineering to drive digital innovation. We are over 10,500 Thoughtworkers strong across 48 offices in 19 countries. For 30 years, we've delivered extraordinary impact together with our clients by helping them solve complex business problems with technology as the differentiator.

/thoughtworks

Footnotes

1.   See, for example, "2024 AI business predictions," PwC, https://www.pwc.com/us/en/tech-effect/ai-analytics/ai-predictions.html; and "A New Architecture to Manage Data Costs and Complexity," BCG, February 7, 2023, https://www.bcg.com/publications/2023/new-data-architectures-can-help-manage-data-costs-and-complexity.
2.   The OSDU (Open Source Data Universe) Forum is a group developing an open-source, standards-based data platform for the energy industry.

Illustrations

Illustrations assembled by Tim Huxford with elements from Shutterstock.

While every effort has been taken to verify the accuracy of this information, MIT Technology Review Insights cannot accept any responsibility or liability for reliance by any person in this report or any of the information, opinions, or conclusions set out in this report.
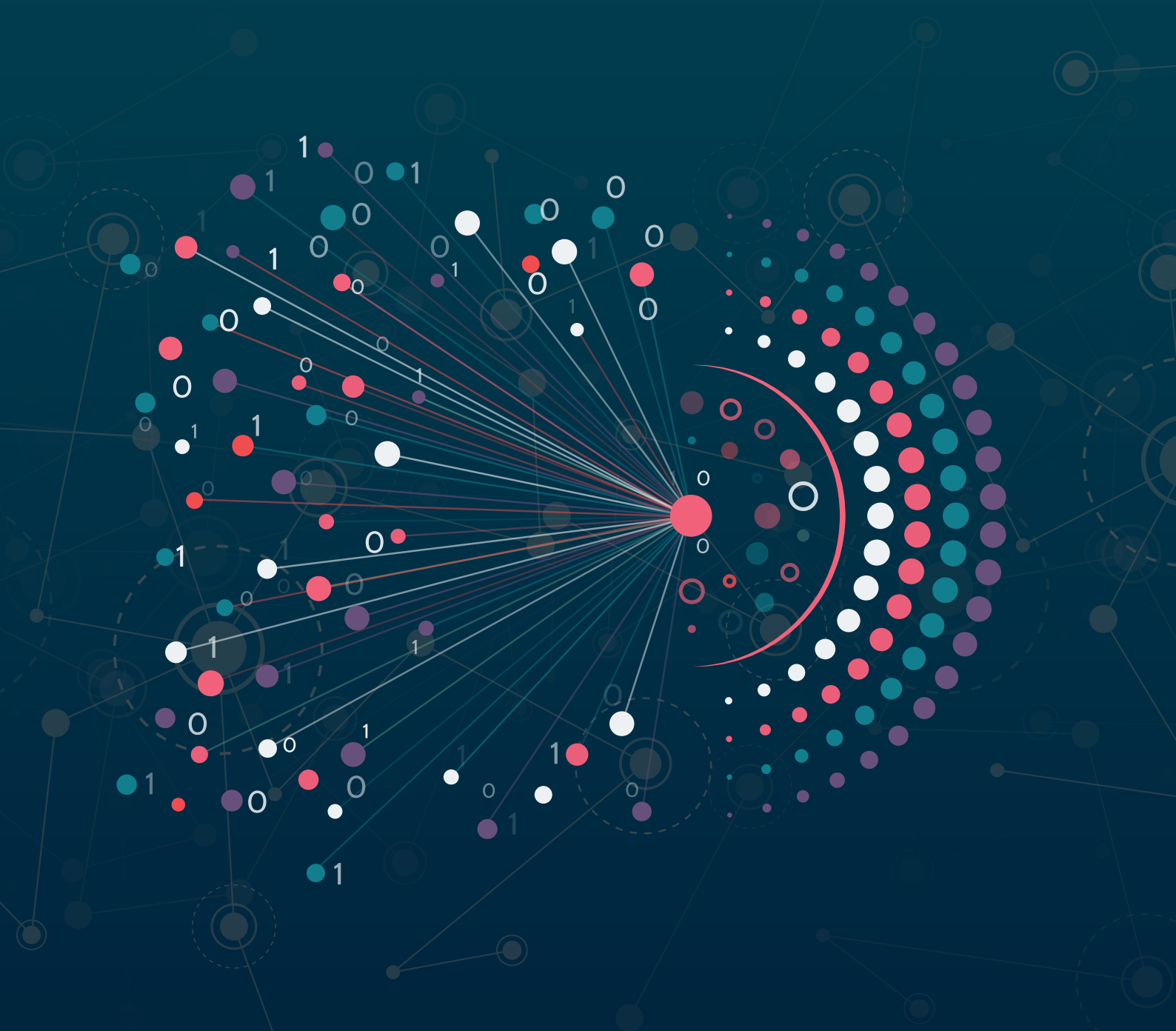
MIT Technology Review Insights

www.technologyreview.com

insights@technologyreview.com