

TDWI Snapshot Series

# 2023 State of Data Management

By Fern Halper, Ph.D.

TDWI's research reveals critical areas for data management success, as well as challenges organizations face in data management and best practices for overcoming them

# 2023 State of Data Management

By Fern Halper, Ph.D.

© 2023 by TDWI, a division of 1105 Media, Inc. All rights reserved. Reproductions in whole or part are prohibited except by written permission. Email requests or feedback to [info@tdwi.org](mailto:info@tdwi.org).

Product and company names mentioned herein may be trademarks and/or registered trademarks of their respective companies. Inclusion of a vendor, product, or service in TDWI research does not constitute an endorsement by TDWI or its management. Sponsorship of a publication should not be construed as an endorsement of the sponsor organization or validation of its claims.

This report is based on independent research and represents TDWI's findings; reader experience may differ. The information contained in this report was obtained from sources believed to be reliable at the time of publication. Features and specifications can and do change frequently; readers are encouraged to visit vendor websites for updated information. TDWI shall not be liable for any omissions or errors in the information in this report.

## Table of Contents

Research Methodology . . . . .	2
What Is Data Management, and Why Is It Important? . . . . .	2
The Overall State of Data Management . . . . .	3
The State of Organizational Commitment and Resources for Data Management . . . . .	4
The State of Platforms, Tools, and Architecture for Data Management . . . . .	7
Architectural Approaches . . . . .	8
Tools and Technologies for Data Management . . . . .	9
The State of Data Governance and Security for Data Management. . . . .	11
Considerations and Best Practices for Data Management Maturity. . . . .	12
References . . . . .	15
About Our Sponsors. . . . .	15
About the Author . . . . .	17
About TDWI Research. . . . .	17

## Research Methodology

This TDWI State of Data Management Report examines where organizations are in their data management journeys. It highlights critical areas for data management success: organizational commitment, resources, architecture and infrastructure, and data governance. It examines challenges organizations face in data management and provides best practices for overcoming them.

For this report, TDWI utilized several surveys and assessments that we run throughout the year. Data in this report comes primarily from the 2022 TDWI Data Management Maturity Model Assessment, which asks about 75 questions across the five categories that form the dimensions of the TDWI Maturity Model. One hundred and thirty respondents from various industries and company sizes participated in the assessment. Additional data from the 2022/2023 TDWI Data and Analytics Survey and recent Best Practices Report surveys was also used. TDWI conducted interviews with subject matter experts in this research.

## What Is Data Management and Why Is It Important?

Organizations are collecting increasing amounts of diverse data for many reasons, such as analyzing it to derive insights and take action, collaborating more effectively, monetizing the data for resale, using it in applications, and other use cases. According to TDWI research, many organizations are collecting tens and hundreds of terabytes of data. Some are collecting petabytes. This data consists of structured data as well as other data types—text data, machine data, log data, demographic data, other third-party data, real-time streaming data, geospatial data, and much more.

To derive value from this data, organizations must be able to ensure that they can effectively utilize their data and make informed decisions, comply with regulations, protect sensitive information, and use resources efficiently. Data management is the process of collecting, storing, organizing, maintaining, and using data effectively and efficiently. Data management encompasses several data-oriented technical disciplines, including data integration, data quality, master data management, data architecture, and metadata management. The goal of data management is to ensure that data is accurate, consistent, accessible, and secure and that it can be used to support the organization's needs. It is the key foundational element of a successful data and analytics program.

As organizations try to keep up with new and complex data types and to become more advanced analytically, they must evolve their data management practices. At TDWI, we see that many organizations have made the move to cloud platforms, often resulting in hybrid on-premises and cloud environments. Some enterprises are using cloud data stacks to store and manage their data; others are looking to unify their data using a data fabric approach to stitch data together across multiple environments—on premises, in the cloud, and even across multiple cloud environments.

Organizations are moving to collect and process data at the edge for certain use cases. They are deploying data catalogs to better understand and manage their data for consumption and are looking at new ways of orchestrating data services and governing all data and new organizational models for success, including DataOps. Enterprises are improving access to data with self-service tools and automated and augmented approaches. In fact, the majority of respondents to a recent TDWI survey were either modernizing their data management strategy or planning to do so.<sup>1</sup>

This State of Data Management Report examines the current state of data management, the top challenges organizations face, and best practices for success.

The majority of respondents to a recent TDWI survey were either modernizing their data management strategy or planning to do so.

## The Overall State of Data Management

Numerous interrelated factors form the current state of data management. Data management maturity is not simply a matter of the platforms used. Data management involves people, processes, and technologies. In the TDWI Data Management Maturity Model, we measure the status of five categories of data management: organizational commitment, resources, architecture, data life cycle, and data governance. The overall average score for the 2022 model is 54.6 out of 100, which puts respondents at stage 3—in the *early* established stage of maturity (Figure 1).

During this stage, executive leadership begins to recognize the importance of data management and starts to establish a data strategy, emphasizing the need for data literacy. Collaboration across functional silos starts with sharing and reusing data assets. Organizations allocate funds to prioritize data management and invest in data management talent and education, including data literacy programs.

To meet the increasing demand for analytics, organizational architecture moves towards self-service, and traditional data warehouses are no longer central to the architecture. Emerging platforms, including the cloud, play a vital role in this transformation. Data integration becomes critical, requiring data pipelines to support the growing volumes and velocity of data. Data project delivery becomes more agile, taking into account data assets used for analytics. The growing reliance on shareable data underscores the need for metadata and data curation. Organizations establish a data governance team to operate on new platforms, such as the cloud, and analytics is integrated into the plan.

# The phases of TDWI's Data Management Maturity Model and average scores per dimension

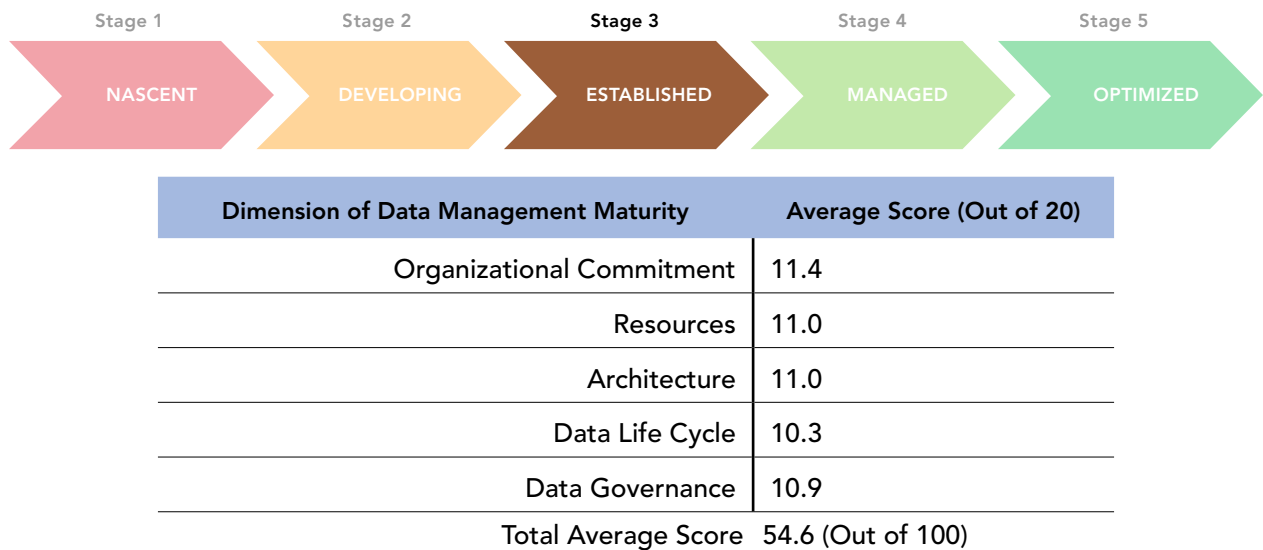


Figure 1. The stages of data management maturity along with participants' average scores for each category.

Again, respondents were early in this stage; we will see that the current state of maturity reflects this.

## The State of Organizational Commitment and Resources for Data Management

Organizational commitment is critical for data management success. At TDWI, we see that although executives and others may appreciate the value of data, that doesn't necessarily mean they will provide the help needed to build a data strategy, work on a culture to support it, fund it, or communicate the effort (Figure 2). For instance, in the Data Management Maturity Assessment (DMMA), 73% of respondents agreed that executives value data and 71% agreed that the organization values data, but only 19% had a strong data strategy in place.

Furthermore, 45% said their data strategy wasn't communicated, and only 37% believed a data-driven culture was defined in their organization. In fact, some of the lowest scores in organizational commitment had to do with culture. A median score of 1 (on a scale of 1-5, where 1 indicates least agreement) was given to the statement, "Executive leadership in my company has defined and enabled a data-driven culture."

# Organizational factors for data management

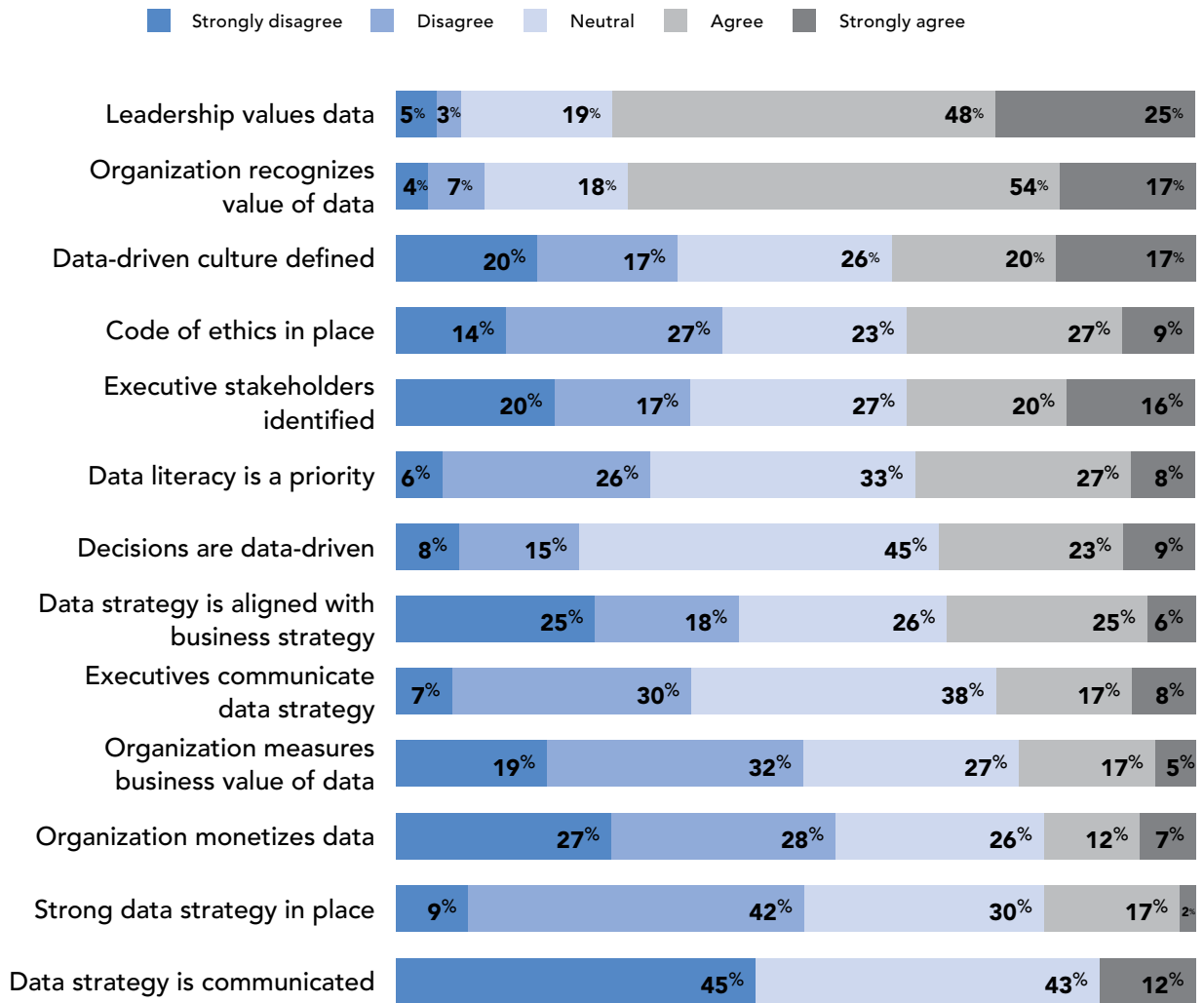


Figure 2. Based on over 100 responses from the 2022 TDWI Data Management Maturity Model Assessment, ordered by highest combined "Agree" and "Strongly agree" answers.

Likewise, many respondents to the DMMA suffer gaps in data management funding. Additionally, many organizations have employees with skills for traditional data management but not for new platforms or data types. On top of that, many lack adequate data management education to build these skills. Education was one of the top data management challenges and one of the lowest-scoring questions.

In other words, some organizations are doing well in putting the organizational underpinnings in place for a solid data management strategy, but many organizations are not. Leadership is key, and leaders must do more than just say data is important. This conclusion is in line with other TDWI research. For instance, in a recent unpublished TDWI survey, we asked, "What is the most important area for organizational improvement for data management?" The top answer was improving data management leadership (28%) followed by communication with the business (21%).

In other TDWI research, we've seen that organizations are starting to fill the role of chief data officer (CDO), whose job it is to derive *value* from data and set the strategy to do so. The CDO may lead data governance and data literacy programs and may be responsible for data management architecture and tools. In a 2022 Best Practices Report, 20% of respondents had a CDO or a chief analytics officer (CAO) in charge of data management and 25% had the VP or director of IT filling that role. However, those with CDOs tended to be more satisfied than those with other roles leading data and analytics efforts. Additionally, enterprises with a CDO as part of the C-suite enjoy a greater potential for success.<sup>2</sup>

In addition to putting a CDO in place for data and analytics, some organizations are implementing DataOps—a structured process in which data is continuously integrated, transformed, and prepared for deployment into business intelligence, reporting, and ad hoc analytics applications. DataOps encompasses a wide range of functional components for handling and managing data throughout its life cycle. These include a host of pipeline services for discovering, acquiring, extracting, transforming, profiling, cleansing, augmenting, staging, loading, replicating, delivering, indexing, searching, versioning, and protecting data. These functions are sometimes known collectively as data engineering processes.

DataOps encompasses a wide range of functional components for handling and managing data throughout its life cycle. These functions are sometimes known collectively as data engineering processes.

In several recent TDWI surveys, more than half of respondents have data engineers in their organization.<sup>3</sup>

Another step some organizations are taking is to move to a “data-as-a-product” and “data product” model. This is still a relatively new concept. *Data-as-a-product* is the principle that data has value and should be treated as a product. That means that customers of that product should be satisfied, and data must be trustworthy, discoverable, etc. *Data products* are derivative assets created from data. These products can run the gamut from enriched data sets provided to a customer to a dashboard that provides the output of machine learning models to external partners, or apps that use derived data for a specific industry.

As data is often viewed as an asset *and* a product, some organizations are starting to fill the role of data product manager. This person is responsible for making sure that data consumers are happy with the product as well as developing new products with data. In a 2022 TDWI Best Practices Report survey, about 30% of respondents had a data product manager in place.<sup>4</sup> Less than 20% of the respondents to the DMMA were actually monetizing data last year, but organizations believe it is very important. For instance, in a 2023 TDWI survey, more than 70% agreed with the statement, “Data productization/monetization is critical for companies to compete in the future.”<sup>5</sup>

The current state of organizational support for data management is in transition. Organizations are beginning to support data management, including hiring CDOs or implementing DataOps to deal with data pipelines.

## The State of Platforms, Tools, and Architecture for Data Management

Organizations are currently making use of a range of data management architectures, platforms, and tools, both on premises and in the cloud. When we asked respondents in 2022 about platforms in use for data management and analytics, on-premises relational databases led the way (Figure 3). However, cloud platforms are also mainstream, and many organizations use multiple providers in a multicloud environment. In that survey, 61% of respondents had an on-premises data warehouse, and 54% were already using a

What data management and analytics tools does your organization currently use? Please select all that apply.

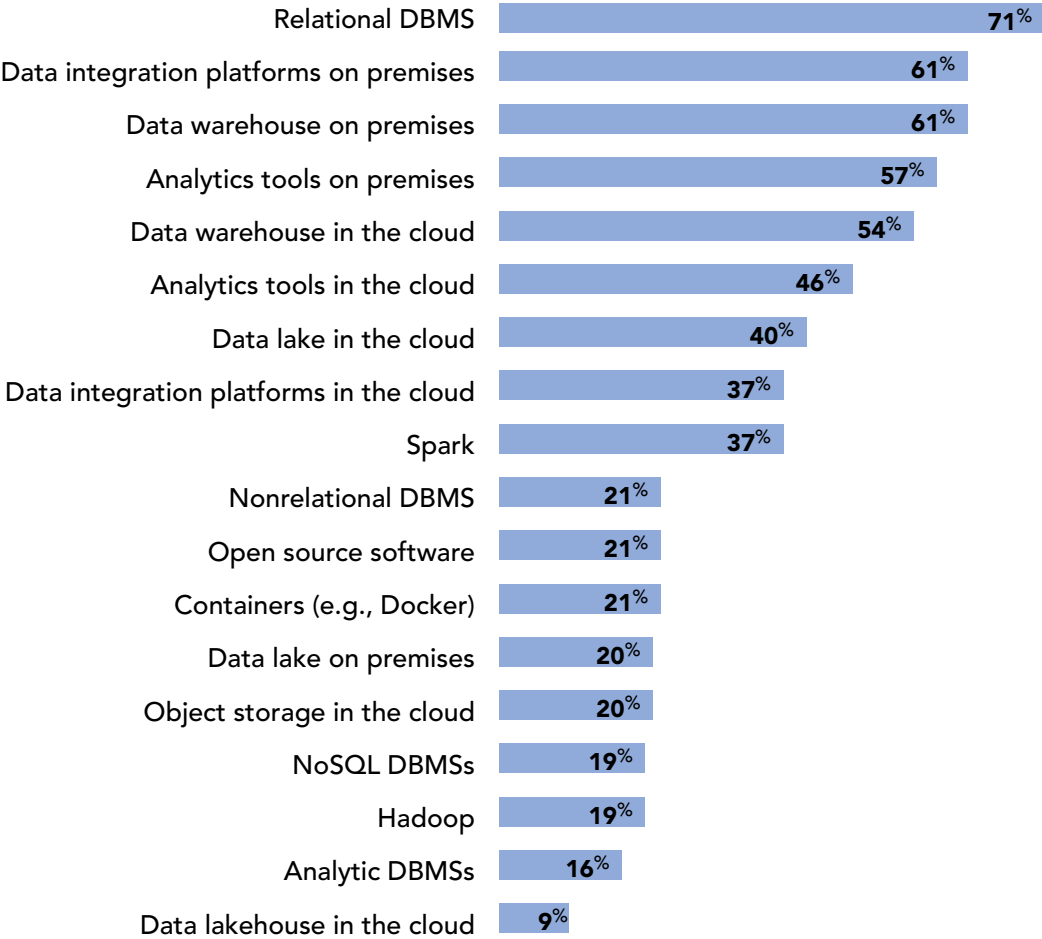


Figure 3. Based on approximately 100 respondents from the 2022 TDWI Data and Analytics Survey.



cloud data warehouse. Additionally, about 20% of respondents had an on-premises data lake, but over 40% were using a cloud data lake.

The percentage of respondents using cloud data platforms is even higher in 2023 surveys. This is most likely to support high volumes of diverse data, such as text data or machine data, for analytics. In one 2022 survey, a large percentage of respondents (about 50%) were utilizing both cloud data warehouses and on-premises data warehouses in a hybrid fashion (not shown). The prevalence of hybrid environments is not surprising. Although the cloud is mainstream, many organizations have not moved all their data to the cloud, nor do they intend to do so. Yet, organizations are looking to unify their data silos in some way for analytics and development.

## ARCHITECTURAL APPROACHES

As illustrated in Figure 3, the evolving environment is complex and distributed. Organizations are dealing with on-premises and cloud data warehouses and lakes and other platforms. This is in addition to legacy systems and other SaaS systems from which they may want to utilize data. Additionally, as more organizations implement newer kinds of analytics models, such as predictive maintenance (using IoT data), they want to process this data at the edge, where their devices live. In the DMMA, enterprises with high overall scores do well in areas such as connectivity, data accessibility, scalability, and data pipelines and other automated approaches (e.g., automated deployment of code, scheduling, testing).

Currently, many organizations are building scalable architectures to deal with different data types. Forty-one percent of respondents to the DMMA felt their data architecture was scalable. Yet, there are issues. For instance, many organizations can only support structured data and some unstructured data with their current architecture. Less than 20% of respondents in the assessment stated they can support structured and most unstructured data types. The rest either supported structured data only (39%) or structured data and some unstructured data (43%). Likewise, 44% could not support real-time data ingestion and the rest could support it (10%) or support it with limitations (46%, all not shown).

Recently, more enterprises are unifying data platforms to provide a trusted data source. There are several architectural approaches; some are moving to a cloud stack; others are utilizing a data fabric approach. Some are doing both.

- **Cloud stack.** The cloud data stack consists of an integrated set of services. In other words, it is a stack of technologies or a set of related cloud services provisioned together for data management and analytics. The cloud provider may offer services with its partners. The idea is that you are centralizing or unifying the data in a cloud stack for analytics use. As illustrated earlier in this report, many organizations are moving to cloud platforms such as cloud data warehouses and data lakes. Some are looking to unify the two.

- **Data fabric.** Data environments are becoming more distributed. The term *data fabric* has been used to describe a way to bring together disparate data in an intelligent fashion. The data fabric maps and connects relevant application data stores with metadata to describe data assets and their relationships. One approach to data fabric design uses data virtualization, a method that integrates heterogeneous and distributed data across multiple platforms without replicating it. The approach creates a single “virtual” data layer that unifies data and supports multiple applications and users. Data virtualization can create logical views in which the data looks consolidated, although the data has not been moved or physically altered. In a recent TDWI survey, close to 50% of respondents had either implemented a data fabric approach or were planning to do so.<sup>6</sup>

Although more of a framework than an architecture, the concept of the data mesh has also gained popularity over the past few years. A data mesh supports distributed, domain-specific data consumers and views data as a product, with each domain handling its own data pipelines. The tissue connecting these domains and their associated data assets is a universal interoperability layer that applies the same syntax and data standards. The data mesh encourages distributed groups of teams to manage data as they see fit, albeit with some common governance provisions.

The data mesh encourages distributed groups of teams to manage data as they see fit, albeit with some common governance provisions.

Domain ownership, data-as-a-product, self-service, and federated governance are the four pillars of this framework. In a recent TDWI survey, fewer than 15% of respondents were currently using a data mesh, although about 20% plan to use it in the future. More are adopting some of the principles of the data mesh.

The notion of a data mesh is also starting to become part of the data management vernacular, although most organizations, so far, only utilize some of its principles.

The current state of data architecture is hybrid. Organizations are trying to unify their data for analysis, which may include a logical or physical architectural approach. It is still a work in progress.

## TOOLS AND TECHNOLOGIES FOR DATA MANAGEMENT

In addition to cloud data platforms and technologies for implementing a data fabric, organizations are also evolving their toolsets. These newer tools often use automated and augmented capabilities where machine learning or natural language processing (NLP) is built into the software. Tools and technologies currently used as part of the maturity process include:

**Data catalogs.** Data catalogs, business glossaries, and metadata repositories collect information about how data is defined and modeled, where it is located, and how models and schema may have changed. The metadata repository tends to be more technical than the data catalog, which is designed so business users can understand it. TDWI research indicates that implementing a data catalog is a top priority among organizations we survey. For instance, in a 2023 TDWI survey, 32% of respondents said deploying a data catalog was a top data management priority, second only to metadata management.<sup>7</sup>

The data catalog can provide features that help ensure data quality and privacy and that trusted data is used for analysis. Without an in-depth knowledge of data and associated metadata, organizations cannot truly safeguard and govern their data. Data catalogs also play a key role in data management by making it easier for users to search for and find data from diverse sources. Some modern data catalogs include features for automated data cleansing, classifying sensitive data, and certification of data sets by their owners. Other solutions keep track of changes to data schema or structure. Some solutions support global and local data catalogs.

In the DMMA, only about a third of the respondents had a data catalog. There are other issues, however. Although many respondents have metadata, it is not integrated into the data life cycle (median score of 1). Organizations may have a data catalog, but it isn't well maintained (median score of 2).

Maintaining the catalog calls for collaboration across technical and business teams, yet many respondents stated that there was no collaboration across these functions.

**Data pipeline tools.** Data pipeline processes and tools help move and transform data from its source to its destination. A data pipeline may include data ingestion, validation, transformation, and loading. A well-designed data pipeline can significantly improve the speed and accuracy of data processing and enable faster and more informed decision-making. Pipelines are critical for modern data architectures that involve massive amounts of data and real-time processing. A top concern TDWI sees in our research is that organizations are struggling with building one-off pipelines as well as manual pipeline processes. In fact, in the DMMA, 49% of respondents were still building manual data pipelines. These are not scalable.

Therefore, organizations are looking to modernize data pipeline processes. In fact, some modern pipeline tools contain the automated features described earlier, along with automated change data capture (CDC) capabilities to understand when data has been updated and the ability to schedule and reuse pipelines. Other features of modern pipelines include support for mass ingestion and hundreds of prebuilt connectors.

The data catalog can provide features that help ensure data quality and privacy and that trusted data is used for analysis. In a 2023 TDWI survey, 32% of respondents said deploying a data catalog was a top data management priority, second only to metadata management.

Some pipelines can sense and adapt to changes in schemas or other changes to sources and destinations. Some can process both batch and real-time data streams. Many are cloud-native so they can apply the computational power of the cloud. Some modern pipelines provide transparency and visibility regarding what is up, what is down, and where data is flowing. This is especially important as organizations need to monitor and shift as business changes—for instance, from one cloud provider to another.

**Self-service data preparation tools.** A major trend both in data and analytics is to provide self-service capabilities. These extend across the data life cycle. Organizations want faster access to data and they want to be able to prepare it for analytics and other use cases. This can involve transforming the data or, in the case of more advanced analytics such as machine learning, developing features used as input to the models. In TDWI research, we still see that over 50% of the time spent in analytics efforts is used for data preparation. Although self-service is a top priority in our surveys, most respondents to the data management maturity model stated that they only have self-service for some users or some parts of the organization.

**Data marketplaces.** A relatively new concept is the data marketplace, a platform where data providers can offer their data sets for sale or exchange and data consumers can purchase or access them. These marketplaces can be internal or external. Internal marketplaces allow employees to access and share data and other data products (often via a data catalog). Cloud platform providers often offer external marketplaces that provide weather data, demographic data, and industry-specific data and data products to those who use the cloud platform. These marketplaces can help enrich data sets for analytics.

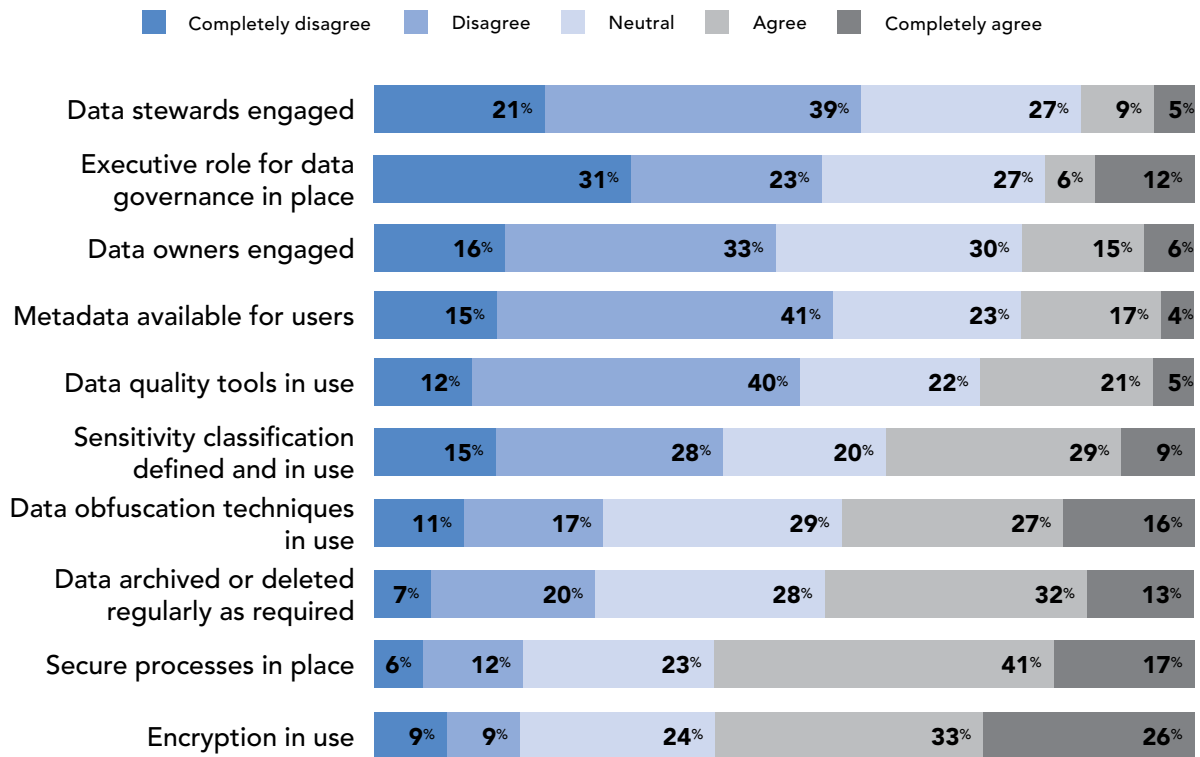
Organizations are beginning to use modern data management tools that include augmented and automated capabilities, such as data catalogs and modern data pipelines. However, it will take work to get these tools (especially data catalogs) to be widely used and maintained.

## The State of Data Governance and Security for Data Management

In TDWI surveys, we routinely see that data governance and security are top priorities for organizations—especially as they try to modernize their environments. Data governance requires people, processes, and technologies. In our assessments, we see that data governance is not advanced. In the DMMA, although organizations make use of tools such as encryption for data security and claim to have sound processes in place for data security, they don't necessarily do a good job putting people in place for oversight.

For example, data steward is a role focusing on data quality and fitness for purpose of data assets. Only 14% had data stewards fully or partially in place. Likewise, data owners were not in place in many organizations.

# Roles or tools currently in place/in use for data governance



**Figure 4.** Based on approximately 100 responses from the 2022 TDWI Data Management Maturity Model Assessment, ordered by lowest “Agree” and “Completely agree” answers combined.

Organizations also appear to be struggling to implement tools such as data quality tools; half (52%) don’t believe data quality rules and measurements are being managed according to an agreed-upon standard. Metadata is also an issue; 56% of respondents don’t believe their organization has metadata available to users and applications to promote discovery and reuse. Additionally, as organizations move to the cloud, they must also govern this environment.

## Considerations and Best Practices for Data Management Maturity

This report has highlighted the state of data management across several areas, including organization, infrastructure, architecture, and data governance. It has detailed some top challenges. The bottom line is that many organizations are still midway on their data management journey—a journey where the goal can keep changing as organizations evolve their data and analytics requirements. The journey is one of evolution, and we’re measuring maturity at a specific point in time.

In the face of this, there are several best practices enterprises can apply to become more mature with data management now and in the future. These include:

**Get executive support.** As we saw in the results of the data management maturity model, although executives understand the value of the data, this does not necessarily translate into support or funding for data management. To build executive support, you must think about what is most important for business operations, then what data is needed to drive those business operations. If that critical data isn't managed well or managed at all, why should your executives care? Think and communicate the value to the business in business terms (not techno-speak). Getting executive support for your data management strategy is crucial now and moving forward. Lobby for a CDO or someone similar if you don't already have one.

**Think about how you manage data in the context of what is important to your business.** Building a cloud data warehouse, data lake, or even a lakehouse should be done in the context of a business need. Maintaining business context—translating the meaning of those data sets when used within business operations—is critical to success. Be proactive and talk to business stakeholders. Likewise, determine the architecture. Many organizations are starting small and maturing over time, with end-to-end data management to support multiple personas across the organization.

**Communicate the strategy.** Once you align your data strategy with the business, communicate it to the rest of the company. Those who aren't aware will become so and understand how the strategy can impact them. This can help build awareness and success.

**Become connected.** To succeed, be connected to the many data sources you have in your environment. This includes structured, semistructured, and unstructured data, on premises, in the cloud, and on the edge. If you aren't already connected to new data types, plan for this as the business demands.

**Put a priority on metadata.** As evidenced by the data presented in this report, metadata needs to be a priority. This includes developing a metadata strategy, including relevant metadata, for all kinds of data—and keep this metadata up to date in one easy-to-access place. Look into some of the metadata mapping tools currently on the market if you haven't done so already. A catalog can be a smart option but will require collaboration and buy-in.

**Make data quality omnipresent.** Poor data quality is something people understand, and it is easy to lose trust in the data. That is why data quality needs to be everywhere, both on premises and in the cloud, and in place for all data types. Make ensuring data quality (including monitoring and devising quality measures) a priority. Centralize your data quality metrics.

**Modernize your architecture.** There isn't one right architecture for every company. Determine what your best approach is given what is already in place and what your

company wants to do. Your architecture will depend on your business needs—and it will evolve. Think of a platform that is open and can work with other tools in the environment. Interoperability is key.

**Make use of new tools.** The volume and variety of new data to manage and the distributed nature of that data will necessitate using new tools for data management as well as analytics. Pay special attention to the automated and augmented tools coming onto the market. These can help to manage the complexity.

**Modernize your pipelines.** One-off pipelines cannot scale in today's complex environment. Use holistic pipeline solutions that can support governed workflows. Look for pipelines that support old and new data types, provide transparency, can deal with change, and can scale.

**Measure success.** Measuring and celebrating success is key to building a data-driven culture, which is why KPIs are important. KPIs can help you track progress and measure success. These KPIs can be used to communicate with employees to identify successes and help foster a culture of data-driven decision-making. In the assessment, the majority of organizations were not developing KPIs to measure value, nor were KPIs communicated. This needs to change if organizations want to succeed with data management. Data leaders should help ensure that the impact of data on the business is measured. This may include top-line revenue impact, bottom-line impact, or meeting regulatory risk.

**Optimize costs.** Many companies make the move to new platforms and are surprised by the cost. This has been particularly true for cloud platforms. Pay attention to some of the new cost optimization tools available. These can help you track and optimize cloud storage and use costs.

**Put data governance front and center.** Data governance is critical for any data management effort. Without it, organizations may not meet compliance mandates or build trusted data that the company will use to drive decisions. This is especially true as organizations move to cloud platforms.

TDWI sees that many CDOs like the idea of a federated data governance model. Their office is the hub, and they coordinate governance across the business units. That way, they can put standards and practices together for the organization, suggest tools, and better organize to execute for data governance.

## References

- <sup>1</sup> Unpublished 2022 TDWI Data and Analytics survey.
- <sup>2</sup> 2022 *TDWI Best Practices Report: Modernizing the Organization to Support Data and Analytics*, online at [tdwi.org/bpreports](https://tdwi.org/bpreports).
- <sup>3</sup> For instance, see the 2022 *TDWI Best Practices Report: Unifying Data Management and Analytics Pipelines*, online at [tdwi.org/bpreports](https://tdwi.org/bpreports).
- <sup>4</sup> 2022 *TDWI Best Practices Report: Modernizing the Organization to Support Data and Analytics*, online at [tdwi.org/bpreports](https://tdwi.org/bpreports).
- <sup>5</sup> Unpublished 2023 TDWI survey.
- <sup>6</sup> Unpublished 2023 TDWI Data and Analytics survey.
- <sup>7</sup> Ibid.

## About Our Sponsors



Denodo is a leader in data management. The award-winning Denodo Platform is a leading data integration, management, and delivery platform, using a logical approach to enable self-service BI, data science, hybrid/multicloud data integration, and enterprise data services. Realizing more than 400% ROI and millions of dollars in benefits, Denodo's customers across large enterprises and mid-market companies in 30+ industries have received payback in less than 6 months.

For more information, visit [denodo.com](https://denodo.com) or call +1 877 556 2531 / +44 (0) 20 7869 8053.



Hitachi Vantara, a wholly-owned subsidiary of Hitachi Ltd., delivers the intelligent data platforms, infrastructure systems, and digital expertise that supports more than 80 percent of the *Fortune* 100.

To learn how Hitachi Vantara turns businesses from data-rich to data-driven through agile digital processes, products, and experiences, visit [hitachivantara.com](https://hitachivantara.com).





As one of the world's largest providers of enterprise application software, SAP strives to help every business run as an intelligent enterprise to ultimately help the world run better and improve people's lives. The SAP Business Technology Platform accelerates innovation to unlock your business potential. SAP BTP brings together application development, data and analytics, integration, and AI capabilities into one unified environment optimized for SAP applications.

Additionally, SAP data and analytics solutions enable organizations to easily integrate, model, manage, and use data from anywhere. Using SAP's modern data stack, organizations retain the context of SAP data and have a trusted foundation that extends planning and analytics across the enterprise to make the most impactful decisions. More information is available at [sap.com](https://sap.com).



StreamSets, a Software AG company, eliminates data integration friction in complex hybrid and multicloud environments to keep pace with need-it-now business data demands. Our platform lets data teams unlock data—without ceding control—to enable a data-driven enterprise.

- Resilient pipelines adapt to constant changes in data structure, semantics, and infrastructure.
- A single design experience to create repeatable integration pipelines for all patterns—streaming, batch, and CDC.
- Reusable pipeline fragments encapsulate expert knowledge in portable elements and keep them up to date no matter where they are used.
- Python SDK lets you templatize pipelines for scale by easily creating hundreds of pipelines with just a few lines of code.
- Fifty predefined data transformation processors meet 99% of analytics requirements out of the box.
- Topologies provide transparency to see how systems are connected and data flows across the enterprise.
- Data SLAs and rules expose hidden problems in your data flows, creating guardrails throughout data pipelines for data quality, sizing, throughput performance, error rates, sensitive information leakage, and more.

StreamSets delivers analytics-ready data, improving real-time decision-making and reducing the costs and risks associated with data flow across an organization. That's why the largest companies in the world trust StreamSets to power millions of data pipelines for modern analytics, data science, and applications. Learn more at [streamsets.com](https://streamsets.com).

## About the Author



**Fern Halper, Ph.D.**, is vice president and senior director of TDWI Research for advanced analytics. She is well known in the analytics community, having been published hundreds of times on data mining and information technology over the past 20 years. Halper is also coauthor of several Dummies books on cloud computing and big data. She focuses on advanced analytics, including predictive analytics, machine learning, AI, cognitive computing, and big data analytics approaches. She has been a partner at industry analyst firm Hurwitz & Associates and a lead data analyst for Bell Labs. She has taught at both Colgate University and Bentley University. Her Ph.D. is from Texas A&M University. You can reach her by email ([fhalper@tdwi.org](mailto:fhalper@tdwi.org)), on Twitter ([twitter.com/fhalper](https://twitter.com/fhalper)), and on LinkedIn ([linkedin.com/in/fbhalper](https://linkedin.com/in/fbhalper)).

## About TDWI Research

TDWI Research provides industry-leading research and advice for data and analytics professionals worldwide. TDWI Research focuses on modern data management, analytics, and data science approaches and teams up with industry thought leaders and practitioners to deliver both broad and deep understanding of business and technical challenges surrounding the deployment and use of data and analytics. TDWI Research offers in-depth research reports, commentary, assessments, inquiry services, and topical conferences as well as strategic planning services to user and vendor organizations.

TDWI Research provides research and advice for data professionals worldwide. TDWI Research focuses exclusively on data management and analytics issues and teams up with industry thought leaders and practitioners to deliver both broad and deep understanding of the business and technical challenges surrounding the deployment and use of data management and analytics solutions. TDWI Research offers in-depth research reports, commentary, inquiry services, and topical conferences as well as strategic planning services to user and vendor organizations.



A Division of 1105 Media  
6300 Canoga Avenue, Suite 1150  
Woodland Hills, CA 91367

E [info@tdwi.org](mailto:info@tdwi.org)

[tdwi.org](http://tdwi.org)