REPORT EXCERPT

The following is an excerpt from an independently published 451 Research report, "DataOps and the Evolution of Data Governance," released in October 2020.

To purchase the full report or to learn about additional 451 Research services, please visit *https://451research.com/products* or email *451sales451@spglobal.com*.

451 Research®
Now a Part of
S&P Global Market Intelligence

THOUGHT LEADERSHIP

OCT 2020

DataOps and the Evolution of Data Governance

Paige Bartley, Senior Research Analyst, Data Management

Organizations are under pressure to maximize value from data, while also facing new regulatory and privacy restrictions. There is a need to align business objectives in order to accelerate data-driven outcomes, and attitudes toward data governance are evolving – ultimately viewing it as an enabler of business value.

©Copyright 2021 S&P Global Market Intelligence.
All Rights Reserved.

REPORT EXCERPT

Table of Contents

1. DataOps Methodology: Data Governance as an Enabler 1

Defining DataOps in the Context of Data Governance 1

Modern Perceptions of Data Governance 1

Figure 1: Degree That Data-Driven Decision-Making Influences Data Governance Perceptions 3

2. Forces Driving Data Governance Efforts 4

The Democratization of Data Access and Use 4

Proliferation of Data Sources Leveraged for Insight 5

Figure 2: Data-Driven Organizations Have More Data Silos 6

Evolving Regulatory Landscape 7

Uncertain Times Call for Certainty in Data 8

Figure 3: Certain Data-Driven IT Initiatives See Acceleration During Coronavirus Pandemic 8

3. All Hands on Deck: Organizational Dynamics 10

The Growing Number of Data Stakeholders 10

Complex Processes and Interdepartmental Efforts 11

The Role of the Chief Data Officer: A Matter of Influence 12

Figure 4: CDO Reporting Structure and Generalized Characteristics 12

4. The Need for Alignment of Business Objectives 14

False Dichotomy: Proactive vs Reactive Data Initiatives 14

Figure 5: Data Privacy Concerns Rank High as a Perceived Barrier to Being Data-Driven 15

Two Sides of the Same Data Governance Coin 16

Leveraging Common Requirements for Business Acceleration 16

©COPYRIGHT 2021 S&P GLOBAL MARKET INTELLIGENCE.
ALL RIGHTS RESERVED.

5. Automation as Data Governance Table Stakes **18**

Speed, Scale and Consistency: The Case for Automation 18

Figure 6: Key Challenges in Manual Data Governance Efforts 19

Automation in Action: Current Usage 19

Is Metadata Management the Answer? 21

6. Conclusions **22**

7. Further Reading **24**

8. Index of Companies **25**

©COPYRIGHT 2021 S&P GLOBAL MARKET INTELLIGENCE.
ALL RIGHTS RESERVED.

2. Forces Driving Data Governance Efforts

The Democratization of Data Access and Use

To realize the potential benefits of becoming more data-driven as an organization (as guided by DataOps methodology), any type of employee with decision-making capacity in their role (no matter how minor) ought to be able to guide those decisions with the help of quality data and relevant informational resources. More employees are accessing and leveraging data than ever before.

Technology refinements have helped. Enterprise software and applications have, over the past 10 years or so, undergone dramatic consumerization, with the increased focus on ease of use and introduction of low-code and no-code interfaces for functions that used to be the sole realm of technically skilled end users. The causative reasons for this are up for debate – employees' software preferences influenced by the simplicity of mobile apps certainly may have been a factor – but what is clear is that organizations have largely embraced the trend, as it allows more individuals to ultimately participate in digital workflows with less training.

Software providers have been eager to advance this trend, with visualization tooling as an early example of this democratization. The hallmark of many of these products has been their organic adoption and spread within organizations: first implemented by one department or for a limited number of end users, word-of-mouth popularity often pressures greater adoption as knowledge employees reap the benefits of more accessible data-derived insights and enhanced productivity.

Furthermore, decreased reliance on IT is an enormous motivator for many line-of-business (LOB) units and end users to pursue consumerized applications and software – removing the perceived bottleneck of IT control over data. Particularly in the early days of the self-service era, this circumnavigation of IT could undermine existing data governance efforts. Technology providers sensed this increased involvement and motivation of business users, and strategic acquisitions such as Informatica's pickup of governance specialist (and business user-focused) Diaku are designed to harness and address this energy in a productive way.

©COPYRIGHT 2021 S&P GLOBAL MARKET INTELLIGENCE.
ALL RIGHTS RESERVED.

The result is an expanding self-service ecosystem of users engaging in everything from dashboarding and visualization of data sources, all the way upstream the data management pipeline to functions such as data prep and ETL. Increasingly, technical data management functions are becoming accessible by a broader number of less-technical employees. The lowered barrier to entry for participation in data-driven workflows – via ease of use and low-code or no-code functionality – has prompted an explosion in number of employees accessing and using data. Under these conditions, data governance becomes essential – not only to provide the right data at the right time to all these individuals, but to provide proper guardrails so that less-technical personas are guided to manipulate data in appropriate ways.

Proliferation of Data Sources Leveraged for Insight

As technology evolves, more data sources are created, and data can be generated and collected more frequently. An example of this has been the effect of smartphone market penetration on data-driven customer experience initiatives. With smartphones generating a much broader and more frequent stream of activity-based data than feature phones, organizations have been able to increase the number of data sources analyzed and leveraged to provide customized offers and other customer engagement touchpoints. Based on our *VotE: Customer Experience & Commerce, Digital Transformation 2019* survey, roughly half of organizations have seven or more distinct 'islands' (sources of customer data) alone.

It is difficult to estimate the total number of data sources any given organization collects and processes across all business units, as the number is typically high and prone to frequent change. However, business software applications – which typically generate data – help provide some perspective. According to our *VotE: Customer Experience & Commerce, Digital Maturity 2019* survey, respondents estimate their organizations have a mean of 185 separate software apps in use.

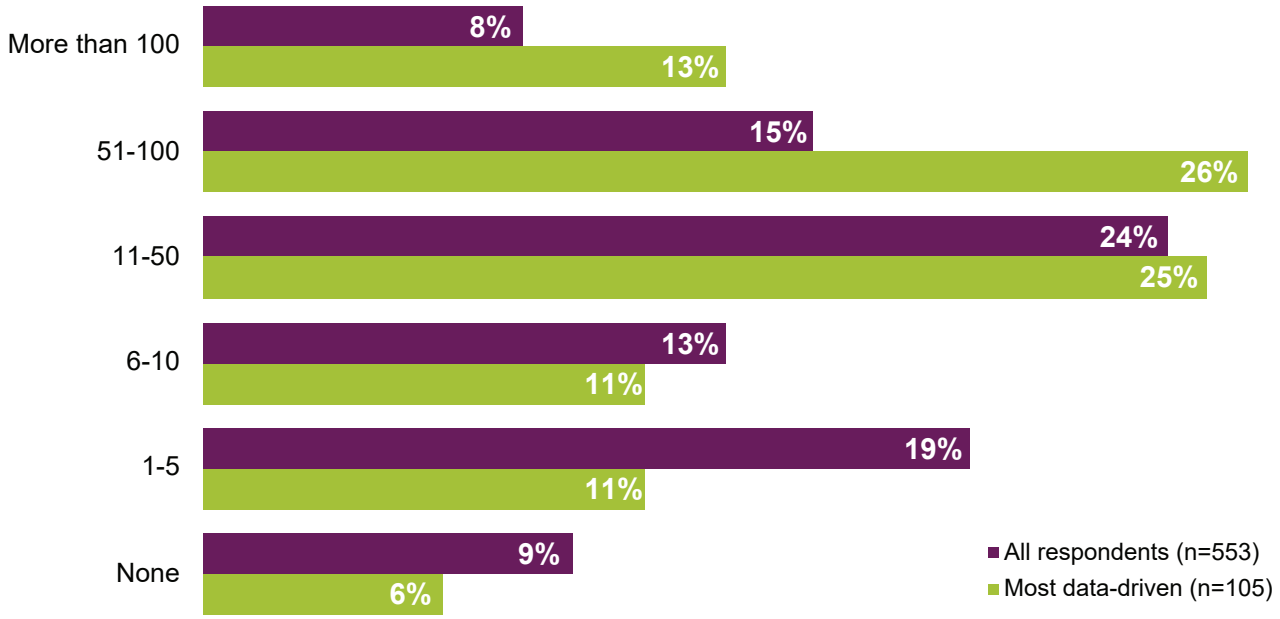
The proliferation of data silos can also serve as a useful gauge of data source number and complexity. In our VotE: Data & Analytics, 2H 2019 survey, data silos are defined isolated repositories of data that are controlled by one department and are not shared with the rest of the organization (in this sense, sharing some conceptual overlap with software applications). Some individual silos may have multiple sources of data, while other silos may have duplicate data overlapping with other silos. More silos can be generally assumed to mean more overall data types being collected and used, though this is not an absolute measure.

Data silos disproportionately affect two types of organizations: those that are large (1,000+ employees) and those that identify as highly data driven. In organizations with 1,000+ employees, 33% of survey respondents report that their organization has 50+ data silos, relative to 23% overall (see Figure 2). For the most data-driven organizations, 39% of respondents report their organization has 50+ data silos.

©COPYRIGHT 2021 S&P GLOBAL MARKET INTELLIGENCE.
ALL RIGHTS RESERVED.

Figure 2: Data-Driven Organizations Have More Data Silos

Source: 451 Research's Voice of the Enterprise: Data & Analytics, 2H 2019
Q. How many data silos would you estimate exist across your organization?
Base: All respondents



Larger organizations, by nature, tend to have more complex and diversified IT ecosystems. Highly data-driven organizations are often early movers in terms of IT: collecting new data sources and adopting new technologies as soon as they become available. It's also worth speculating that highly data-driven organizations – being more ambitious with data – may be more comfortable living with the byproduct of numerous data silos rather than attempting to conduct drawn-out consolidation exercises – ventures that may be unproven to yield better insight or may be prone to failure and high cost. Early adopter organizations may also be more likely to be using technologies that enable them to analyze data across silos without fully resolving the underlying architectural disparities. However, this strategy can ultimately lead to silo propagation over time, especially if accompanying data governance practices are not rigorous or come as an afterthought.

©COPYRIGHT 2021 S&P GLOBAL MARKET INTELLIGENCE.
ALL RIGHTS RESERVED.

Evolving Regulatory Landscape

It is difficult to argue with the conclusion that we have entered a new era of data-driven regulations. The EU's General Data Protection Regulation (GDPR), which went into effect in May 2018, is widely considered the vanguard in this new wave of requirements, focused on data privacy and data protection for consumers. Around the world, regions have followed in kind with their own regulations, frequently mirroring key principles of GDPR. Despite similarities, all regulations have their own nuances that can be difficult for organizations to track and comply with. Many data privacy and protection regulations, in particular, are horizontal in nature (affecting all industries) and have extraterritorial reach, impacting businesses that may be based halfway across the world. More regulations are proposed and emerge in regions worldwide at an accelerating pace.

Organizations' biggest difficulty with data-driven regulations isn't necessarily their technical requirements; it is their inherent level of ambiguity, and the need to interpret that ambiguity in order to craft suitable compliance strategies. By nature, most regulations cannot be highly prescriptive in their requirements around technology. Because the legislative process is deliberate and relatively slow-moving, any technology regulation must have language and definitions that can endure over time as the technology landscape rapidly evolves in new and unexpected directions. Regulations need to have a stable shelf life in order to be useful. Specific technologies are rarely mentioned in regulatory language, or if they are, they are very general and well-established categories of technology, such as encryption or high availability. Because of this, organizations must often use ample human judgment to interpret regulatory guidelines as they apply to their specific IT environments.

Furthermore, individual regulations are subject to adjustment as the legislative process is carried out, even after the regulation technically goes into effect. The California Consumer Privacy Act (CCPA) provides a good example of this, as we cover in our *CCPA report* from 2019. While CCPA went into effect January 1, 2020 and enforcement began July 1, 2020, amendments were still being proposed, evaluated and finalized through September 30, 2020. California will additionally have a ballot measure – the *California Privacy Rights Act* – that citizens will vote on November 3, potentially further changing the requirements associated with Californian data privacy and protection requirements. When we asked organizations what their compliance status is with CCPA as part of our *Voice of the Customer: Macroeconomic Outlook, Corporate IT Spending Q4 2019* survey, 35% of respondents simply answer that they don't know: a possible testament to the regulation's complexity.

CCPA is an example of the shifts associated with *one* regulation in *one* US state; for multinational organizations, keeping up to date with the evolution of all relevant data-driven regulations around the world is an endeavor with exponential complexity. Each regulation is potentially a moving target, yet nearly all of them implicitly underscore the enterprise need for full control and understanding of data being collected and processed. For many organizations, compliance complexity has served as an impetus to revive stagnant data governance efforts with new urgency and budget.

©COPYRIGHT 2021 S&P GLOBAL MARKET INTELLIGENCE.
ALL RIGHTS RESERVED.

Uncertain Times Call for Certainty in Data

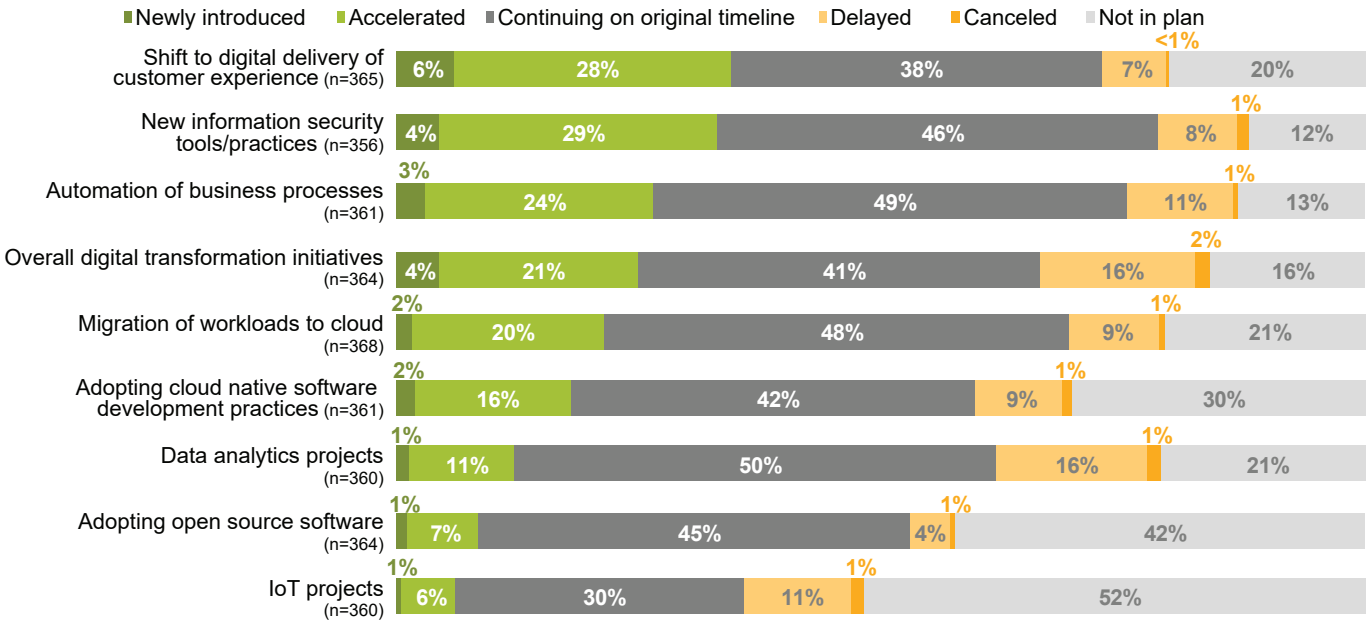
The global coronavirus pandemic has had a measurable impact in the way many businesses operate, with certain sectors (manufacturing, retail, travel/hospitality, etc.) seeing disproportionate shifts in critical business variables such as supply chain stability and customer demand. Many organizations are struggling to simply ride out a period of economic uncertainty, while others are looking to adapt business models to meet new customer demands and purchase patterns.

Whether an organization is looking to improve operating efficiency and reduce costs, to gain competitive advantage amid sector shakeup or accelerate the delivery of products and services to customers that have shifted their engagement to digital channels, there is one common denominator: dependence on data. In many cases, organizations are sitting on vast quantities of underutilized data, and turning to this data for additional insight can be a relatively low-cost way to sustain business viability in uncertain times. Strong data governance and all that it entails – data quality, data relevancy and appropriate data delivery and access – increases the likelihood of success for these efforts.

Figure 3 suggests that organizations are, in some cases, initiating and accelerating investment in certain IT initiatives due to the effects of coronavirus, rather than halting or delaying them.

Figure 3: Certain Data-Driven IT Initiatives See Acceleration During Coronavirus Pandemic

Source: 451 Research's Voice of the Enterprise: Digital Pulse, Coronavirus Flash Survey June 2020
Q. For each of the following types of technology initiatives, please indicate how they were affected at your organization, if at all.
Base: All respondents



©COPYRIGHT 2021 S&P GLOBAL MARKET INTELLIGENCE.
ALL RIGHTS RESERVED.

Based on our *VotE: Digital Pulse, Coronavirus Flash Survey June 2020*, 34% of respondents report their organization has newly-introduced or accelerated their shift to digital delivery of the customer experience, relative to only 7% that report their organization has delayed or canceled these efforts. Overall digital transformation initiatives see 25% of respondents reporting newly introduced or accelerated efforts. Digital delivery of the customer experience, and of course overall digital transformation, are both highly data-dependent and contingent on underlying governance efforts.

When organizations are facing uncertain market conditions, a business decision based on faulty insight derived from poorly managed data can potentially be fatal rather than inconvenient. Stakes are high for data governance.

©COPYRIGHT 2021 S&P GLOBAL MARKET INTELLIGENCE.
ALL RIGHTS RESERVED.

5. Automation as Data Governance Table Stakes

Speed, Scale and Consistency: The Case for Automation

As mentioned in the case of compliance, one of the most challenging aspects of implementing proper governance controls can be the human judgment required to establish functional and appropriate policies for data. In this sense, the human element of data governance is not going anywhere; laws will continue to shift and require interpretation, and organizations will continue to rely on subject matter expertise to understand how business context and processes affect the leverage of informational assets.

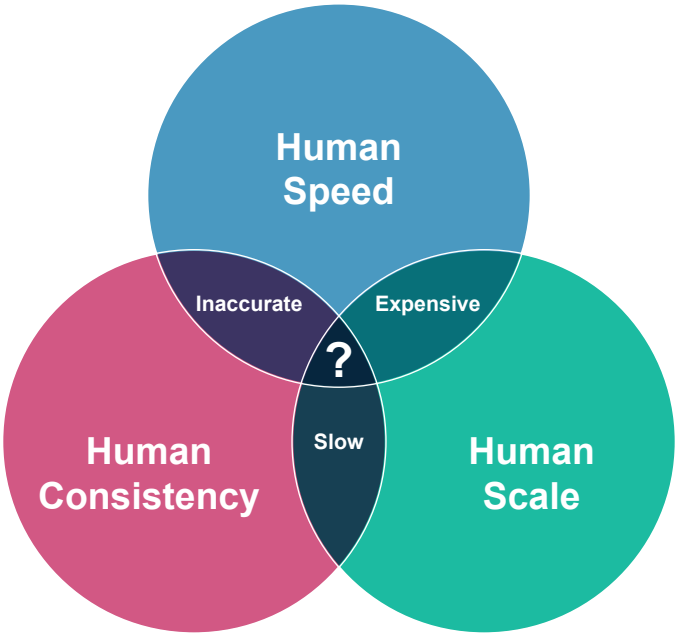
However, modern enterprise data volumes – which are continuing to grow – mean that no amount of human labor could feasibly keep pace with the number of data governance tasks required to keep all data up to date in terms of comprehensive policy, classification and other attributes. Current volumes of data under management, and their expected growth over the next two years, illustrate this reality. Participants in our VotE: Data & Analytics, 2H 2019 survey report a current median of 638TB. Expected data volume growth over the next two years is expected to be 29%.

Speed, scale and consistency are the perennial challenges of human data governance efforts. Data, particularly streaming data sources, are created and ingested by the organization faster than any human can act or make decisions. Sheer volumes of data are not feasible to assess and manage manually, regardless of how many skilled workers are dedicated to the task. And consistency is another problem; while human judgment still exceeds automation in interpreting ambiguous context, there is enormous variance in how one human may act in one situation versus another (see Figure 6).

©COPYRIGHT 2021 S&P GLOBAL MARKET INTELLIGENCE.
ALL RIGHTS RESERVED.

Figure 6: Key Challenges in Manual Data Governance Efforts

Source: 451 Research, 2020



While human speed, scale and consistency can all be improved to some degree in isolation, improving more than one at a time comes with penalty. Human speed can be marginally improved with expertise and ample training, but to achieve that in combination with human scale requires a massively expanded workforce of highly trained, experienced professionals –a prohibitively expensive proposition. Likewise, increased human speed in conjunction with increased human consistency often results in the sacrifice of higher-level human judgment in cases that are ambiguous – ultimately resulting in inaccuracies.

Improving all three simultaneously is essentially unknown territory. Here, at this convergence, is where the greatest opportunity for automation lies in data governance.

Automation in Action: Current Usage

Automation efforts in the broader data management market are a trend we have discussed previously in our *Data Management Market Map 2019*. In short, given the challenges of speed, scale, and consistency in managing data, software providers have been gradually adding in guided and 'smart' functionality – driven largely by machine learning (ML) – to help organizations augment human efforts with automation that can address the bulk of simple use cases, reserving human labor for cases of ambiguity and complexity.

©COPYRIGHT 2021 S&P GLOBAL MARKET INTELLIGENCE.
ALL RIGHTS RESERVED.

In the realm of data governance, a good example of automation exists with data discovery and data classification. Many sensitive and high-risk data types such as social security number and payment card information follow a templated pattern, and rules-based or regular expression-based automation can easily detect and classify them in most cases, allowing human efforts to be spent determining the sensitivity of more ambiguous data sources, such as unstructured and textual communications. By using automation within data management products to target the bulk of common sensitive data types, the enterprise can greatly reduce its risk profile, allocating human talent elsewhere to fine tune efforts.

Software vendors are not the only ones building in automation and providing organizations the option to buy these capabilities. Many enterprises also have ongoing efforts to build homegrown ML models via data science initiatives. While consumer and customer-facing ML and AI use cases like chatbots may get a disproportionate amount of attention, it turns out that many of the top enterprise use cases for machine learning are focused internally.

Participants in our *VotE: AI & Machine Learning, Use Cases 2020* survey survey reporting data management/classification as a main business use case: the top reported use case, overall. While our historical data demonstrates that data management ML use cases used to be more common in larger organizations, this is no longer the case. Today, organizations of all sizes report similar usage of ML for data management/classification, at rates of 40% or more in each category of organization size, based on employee count. Meanwhile, customer-facing ML functions such as customer self-service and product recommendations or offers rank lower, with percentages reporting these use cases at 26% and 21%, respectively.

Examples of automation in data management and data governance, then, abound both in the software market as well as in the development of homegrown models. A vendor list, while by no means exhaustive, might include examples such as the Alation data catalog's use of ML for enhanced data indexing. Similarly, Hitachi Vantara's Lumada Data Catalog (formerly Waterline Data) applies a patented method of ML-driven metadata discovery. Informatica's CLAIRE AI engine, feeding off a unified metadata layer, is able to help provide automated functionality throughout the company's data management portfolio, including the Axon Data Governance offering. Octopai, a metadata management specialist, is using automation to help organizations discover data and understand lineage. Many privacy-focused providers are also leveraging degrees of automation to support organizations in their compliance and governance workflows. OneTrust's acquisition of Integris Software bolstered the company's automated sensitive data discovery and classification. SECURITI.ai looks to apply automation to all aspects of data privacy workflows, such as retrieval of data.

Examples of functions that are increasingly the target of automation include data detection, data classification, data mapping, data retrieval, data quality assessment and scoring, navigation and search via natural language processing, recommendations for next best steps in workflows such as data prep and analysis, data risk scoring, policy application and execution and movement of data between tiers of storage based on attributes such as age and usage patterns.

©COPYRIGHT 2021 S&P GLOBAL MARKET INTELLIGENCE.
ALL RIGHTS RESERVED.

Is Metadata Management the Answer?

With the typical organization struggling to govern and make sense of increasing volumes and sources of data across numerous silos, one of the common desires is to gain more of a unified view of informational resources. Harkening back to the definition of DataOps, a more cohesive understanding of the enterprise's entire diversified landscape would – in theory – help organizations align their business objectives and consistently leverage data for both proactive and reactive purposes.

There's a reemerging perception that metadata holds the potential to function as a lingua franca across heterogeneous data architecture. This is for good reason: while data formats in various repositories may not be directly compatible or feasible to fit into the same schema, the metadata derived from them provides a consistent source of technical and business context that then itself can be leveraged across sources for use cases such as assisting in data navigation or understanding data usage patterns within the organization. The reasoning goes that an enterprise-wide, managed metadata layer could provide and abstraction for the complexity of underlying silos and repositories. Furthermore, as organizations look to operationalize their data science efforts, a unified metadata layer would provide an attractive source of standardized data with which to train models. A parallel line of reasoning considers this metadata layer as an opportune architectural point to *apply* ML models for use cases related to data management and governance.

But is this concept of an enterprise-wide managed metadata layer material or a mirage? Current use and planned adoption of data catalogs can provide some perspective. Data catalogs are a topic we *have discussed before*, and while not all metadata management technology is a data catalog, all data catalogs are indeed a form of metadata management. Because of this, commercially available data catalogs can provide a useful proxy for measuring metadata management practices.

Perhaps surprisingly, this idealized concept of the enterprise-wide metadata layer has not seemed to become universal in the context of data catalog usage, despite its potential utility in the realm of data governance. Only about one quarter (24%) of respondents in our VotE: Data & Analytics, 2H 2019 survey report their organization currently uses an enterprise-wide data catalog. Nor does an enterprise-wide catalog necessarily seem to be the aspiration; when asked about adoption plans over a two-year time horizon, enterprise-wide catalogs remain stable.

What is clear is that there is intent to broaden the scope of metadata management, if not necessarily making data catalogs truly enterprise wide. While catalogs specific to data warehouses or data marts are the most commonly reported data catalog implementations today, with 39% of respondents reporting this type, they are also the type of data catalog poised to see the biggest drop in two years. Much of that difference seems to shift to data lake-specific catalogs. This raises the possibility that organizations have a pragmatic approach; while they may not consider it realistic to fully manage and catalog every data source across the entire organization, they do foresee a world in which a dedicated data lake can be fully governed and cataloged, providing workers across the organization with a single resource for well-managed data.

©COPYRIGHT 2021 S&P GLOBAL MARKET INTELLIGENCE.
ALL RIGHTS RESERVED.

S&P Global Market Intelligence

Copyright © 2021 by S&P Global Market Intelligence, a division of S&P Global Inc. All rights reserved.

These materials have been prepared solely for information purposes based upon information generally available to the public and from sources believed to be reliable. No content (including index data, ratings, credit-related analyses and data, research, model, software or other application or output therefrom) or any part thereof (Content) may be modified, reverse engineered, reproduced or distributed in any form by any means, or stored in a database or retrieval system, without the prior written permission of S&P Global Market Intelligence or its affiliates (collectively, S&P Global). The Content shall not be used for any unlawful or unauthorized purposes. S&P Global and any third-party providers, (collectively S&P Global Parties) do not guarantee the accuracy, completeness, timeliness or availability of the Content. S&P Global Parties are not responsible for any errors or omissions, regardless of the cause, for the results obtained from the use of the Content. THE CONTENT IS PROVIDED ON "AS IS" BASIS. S&P GLOBAL PARTIES DISCLAIM ANY AND ALL EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, ANY WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE OR USE, FREEDOM FROM BUGS, SOFTWARE ERRORS OR DEFECTS, THAT THE CONTENT'S FUNCTIONING WILL BE UNINTERRUPTED OR THAT THE CONTENT WILL OPERATE WITH ANY SOFTWARE OR HARDWARE CONFIGURATION. In no event shall S&P Global Parties be liable to any party for any direct, indirect, incidental, exemplary, compensatory, punitive, special or consequential damages, costs, expenses, legal fees, or losses (including, without limitation, lost income or lost profits and opportunity costs or losses caused by negligence) in connection with any use of the Content even if advised of the possibility of such damages.

S&P Global Market Intelligence's opinions, quotes and credit-related and other analyses are statements of opinion as of the date they are expressed and not statements of fact or recommendations to purchase, hold, or sell any securities or to make any investment decisions, and do not address the suitability of any security. S&P Global Market Intelligence may provide index data. Direct investment in an index is not possible. Exposure to an asset class represented by an index is available through investable instruments based on that index. S&P Global Market Intelligence assumes no obligation to update the Content following publication in any form or format. The Content should not be relied on and is not a substitute for the skill, judgment and experience of the user, its management, employees, advisors and/or clients when making investment and other business decisions. S&P Global Market Intelligence does not endorse companies, technologies, products, services, or solutions.

S&P Global keeps certain activities of its divisions separate from each other in order to preserve the independence and objectivity of their respective activities. As a result, certain divisions of S&P Global may have information that is not available to other S&P Global divisions. S&P Global has established policies and procedures to maintain the confidentiality of certain non-public information received in connection with each analytical process.

S&P Global may receive compensation for its ratings and certain analyses, normally from issuers or underwriters of securities or from obligors. S&P Global reserves the right to disseminate its opinions and analyses. S&P Global's public ratings and analyses are made available on its Web sites, www.standardandpoors.com (free of charge) and www.ratingsdirect.com (subscription), and may be distributed through other means, including via S&P Global publications and third-party redistributors. Additional information about our ratings fees is available at *www.standardandpoors.com/usratingsfees*.

CONTACTS

The Americas
+1 877 863 1306
market.intelligence@spglobal.com

Europe, Middle East & Africa
+44 20 7176 1234
market.intelligence@spglobal.com

Asia-Pacific
+852 2533 3565
market.intelligence@spglobal.com

www.spglobal.com/marketintelligence

©Copyright 2021
S&P Global Market Intelligence.
All Rights Reserved.