451 Research PATHFINDER REPORT

S&P Global Market Intelligence

Now a Part of

# Accelerating Outcomes with Data Governance

COMMISSIONED BY

#### **NOVEMBER 2020**

©COPYRIGHT 2020 451 RESEARCH. ALL RIGHTS RESERVED.

## About this paper

A Pathfinder paper navigates decision-makers through the issues surrounding a specific technology or business case, explores the business value of adoption, and recommends the range of considerations and concrete next steps in the decision-making process.

#### ABOUT THE AUTHOR



#### PAIGE BARTLEY

SENIOR RESEARCH ANALYST, DATA MANAGEMENT

Paige is a Senior Research Analyst for the Data, AI & Analytics Channel at 451 Research, a part of S&P Global Market Intelligence, covering core data management topics. Key themes include data privacy, data governance, data integration, metadata management, data quality and master data management. She has experience covering a broad range of technologies spanning the Data, AI & Analytics stack, from databases to self-service analytics.



# **Executive Summary**

Times and technology change, and organizations must constantly adapt to new challenges. Unparalleled shifts in customer demand and consumption patterns have become the new norm, and organizations are struggling to make sense of their existing data to either simply stay afloat or to maintain their competitive advantage.

While data governance has long been valued for its foundational control in helping organizations achieve 'reactive' needs such as regulatory compliance, it is increasingly being viewed through a clearer lens as a true multiplier of business value. Organizations are becoming more datadriven, and all data-driven outcomes – whether 'proactive' or 'reactive' – ultimately depend on the integrity of data that consistent, enterprise-wide governance can deliver.

Data governance is not without its challenges, as both organizational and technical barriers persist today. Data silos remain a common culprit; however, a broad array of cultural factors play a part as well. Fragmented efforts toward governance, via well-managed but isolated 'islands' of data, can provide the illusion of control, though these efforts are usually fragile in nature. What is needed is more of an enterprise-level vision of governance and visibility into informational assets – both for reactive and proactive use cases.

Existing enterprise initiatives, particularly in regulatory compliance, suggest a current lack of control of data that is mismatched with today's positive perceptions of data governance. There is opportunity in this mismatch between reality and perception. Common-denominator needs across organizations underscore these opportunities, particularly for increased automation. By applying AI and machine learning technology, organizations can accelerate and scale their governance efforts, supporting downstream dependencies such as worker productivity with information.

#### **Key Findings**

- Organizations today largely view data governance as an enabler of business value rather than a cost center, challenging historical perceptions.
- Functional barriers to achieving cohesive enterprise-wide data governance remain. These barriers can be both organizational and technical.
- The proliferation and persistence of data silos remains a common challenge in achieving a single high-level view and unified control of enterprise data.
- Benefits of being more data-driven can directly add business value, help respond to external forces or help improve efficiency. All benefits of being more data-driven are ultimately dependent on data governance.
- Organizations' current compliance capabilities or lack thereof suggest a lack of insight into existing enterprise data and, in all likelihood, a shaky data governance foundation.
- Automation is a necessary accelerator, both for data management tasks such as classification and for broader worker productivity with information-centric tasks.



# The Data Governance Renaissance

Data governance is experiencing a renaissance, despite some of the perceptions historically associated with it. The practice of data governance has certainly seen periods of urgency, but traditionally, it has been viewed as a 'reactive' function within organizations that was initiated or sustained primarily due to external forces or requirements. In many cases, data governance directly supported use cases such as regulatory compliance, risk management and legal defensibility. Data consumers within organizations – those who wanted to use data to derive insight – did not see direct downstream benefit from these activities. Data governance's strong association with these traditionally conservative business functions gave it a reputation for red tape and restriction.

Today, as organizations look to maximize the value of their informational resources, they are realizing that highly 'proactive' data-driven business functions also depend on a foundation of well-managed data. The aggregate population of data consumers has grown larger in nearly all organizations and includes a higher number of data consumers that have less-technical skill sets. Consistent data access policies, data curation and quality become paramount for successful results in such an ecosystem.

'Proactive' data initiatives generally include situations where the goal of data use is to drive immediate business value; however, such initiatives also encompass Skunk Works scenarios where workers are allowed to use data in exploratory ways without strictly defined targets, with the overarching objective of identifying valuable new use cases for information. This contrasts with more 'reactive' initiatives, which tend to be repetitive actions involving data that are in response to a predictable (and typically externally mandated) requirement. Proactive data initiatives, broadly speaking, can be broken down into tiers based on how predetermined the objectives are.

- Free data experimentation. In the least predetermined scenario, data consumers (such as a data science team) are allowed to freely play around with any and all business data in the hope of stumbling across fundamentally new uses and insight. Discovering 'unknown unknowns' is the ultimate prize in this scenario.
- **Insight and value prospecting.** Slightly more specific, an insight-and-value-prospecting approach narrows down data to a specific domain. It is still exploratory in nature in the sense that the aim is to uncover deeper insight and value, without a predefined business need.
- Focused value discovery or creation. When proactive analysis of data is directly tied to a very specific, and predefined, business or organizational need, it can be thought of as focused value discovery or creation.

While free data experimentation initiatives get a disproportionate amount of attention, the reality for most organizations is likely much more benign. Free data experimentation for most organizations – particularly those that do not have strong existing data governance – is akin to a lottery system. When these initiatives are successful, they can be enormously so. However, limitations such as poor data quality and data that isn't representative of the situation being modeled mean many of these exploratory projects remain relatively small-scale and isolated within the business – 'science fair' projects. Without certain ROI, some organizations are reluctant to invest much in these projects to begin with.



The bulk of proactive data initiatives still tend to be relatively focused in nature, domain-specific, and tied to existing objectives or needs. This doesn't make them any less valuable; use cases with data where needs are predetermined can be thought of as the workhorses of business insight. However, there remains room to grow in more exploratory and experimental situations, and rate of success in this realm depends heavily on the availability of high-integrity data.

Generally speaking, it is not productive to get caught up in the differences between 'proactive' and 'reactive' use cases. Despite perceived cosmetic differences, proactive and reactive initiatives have more in common than what may initially meet the eye.

## Proactive and Reactive Efforts: Two Sides of the Same Coin

The concepts of proactive and reactive data-driven use cases are not opposites but, rather, two complementary sides of the same data governance coin. Whether retrieving and presenting data for a regulatory request or allowing self-service users to iteratively explore relationships across data assets, all use cases are dependent on the underlying integrity and quality of data.

So just as a skilled sports team must leverage a mix of offensive and defensive tactics to win a game, organizations can benefit from conceptualizing proactive and reactive functions as complementary components of a broader competitive strategy, rather than as superficial opposites. In the realm of data, both proactive and reactive functions ultimately depend on full control and understanding of informational assets. Modern perceptions of data governance are beginning to reflect this realization, with sentiment toward data governance generally being quite positive.



#### Figure 1: Data governance is viewed as an enabler of value

Source: 451 Research's Voice of the Enterprise: Data & Analytics 2H 2019

Q: To what extent do you agree or disagree with the following statement: In my organization, data governance is seen as an enabler of business value rather than a cost center?

% of respondents (n=361)



Despite a certain degree of historical 'baggage' associated with data governance terminology, organizations today understand the value and potential of systematic governance practices – acting as the foundation for consistent data integrity for the enterprise data consumer masses. In a recent 451 study, 72% of enterprise practitioners said they either 'completely' or 'mostly' agree that data governance is an enabler of business value, rather than a cost center, within their organization. For organizations that self-identify as 'highly data-driven,' that number jumps to 82%.

However, there may still be a mismatch between organizations' perception of data governance – which is generally positive – and the reality of execution. In short, despite having positive perceptions of data governance, many organizations still struggle with the 'nuts and bolts' of successful data governance orchestration.



## Barriers Remain in Achieving Data Governance

Despite generally positive sentiment toward data governance, functional barriers remain in achieving data governance as a true business accelerator. For some organizations, this positive view may be aspirational, reflecting the realization that governance is a valuable concept; however, the realities of existing enterprise environments may not reflect this true value.

ORGANIZATIONAL BARRIERS TO DATA GOVERNANCE	TECHNICAL BARRIERS TO DATA GOVERNANCE
Lack of buy-in or support from leadership	Proprietary technology and lock-in
Organizational communications barriers	Disparate tools and user experiences
Competing or misaligned business objectives	Difficulty in integrating legacy systems/data
Departmental or 'territorial' power dynamics	Data silos and divergent control mechanisms
Skills limitations and shortages	Maintenance; keeping components up to date

Data silos, in particular, continue to be a thorn in the side of organizations. They are a problem as old as IT architecture itself, and they remain a constant challenge for any business trying to gain a more consistent view of informational assets.

#### Figure 2: Data silos affect data-driven organizations

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



These data silos remain a predominant challenge in enterprise-wide data governance efforts, and they disproportionately affect both large organizations and highly data-driven organizations. The second group is notable and perhaps paradoxical, given such organizations' perceived prowess in deriving value from data.



One-third of organizations with 1,000+ employees have 50+ distinct departmental data silos. This is perhaps unsurprising since larger organizations tend to have more expansive and complex IT ecosystems. For highly data-driven organizations, that number is 39% regardless of organization size.

Highly data-driven organizations are often early movers in terms of IT, collecting new data sources and adopting new technology as soon as it becomes available. This philosophy can lead to more data silos over time. While data silos are generally thought of as a complication and barrier to consistent governance, they can inversely be thought of as an opportunity. More silos generally translate to more untapped value of data. And the more silos the organization currently has, the more the organization stands to gain from comprehensive data governance efforts.

## The 'Island Mirage' of Fragmented Governance

Just because an organization has data silos doesn't mean that it hasn't attempted data governance. In fact, those efforts likely exist, and they are likely duplicative. Individual silos of data may, in fact, be managed reasonably well in isolation, giving the illusion that the house is in order when the organization is examined with a magnifying glass rather than with binoculars. But fragmented efforts at data governance have a major impact on the business as a whole.

A data 'island' analogy is useful here. Individual islands of data are formed by departments/teams or systems/applications. They may or may not involve more than one data silo. These individual islands may be well-governed in isolation, giving the impression, when viewed at close range, that broader governance is less of a concern.

But little islands of perfection, while idyllic onshore, can look like a fragmented archipelago when viewed from the enterprise perspective. Some islands may be completely isolated, while others are precariously bridged to others via differing structures built through the team's own efforts – or with the expensive help of a systems integrator (SI). These structures typically are prone to collapse if anything changes with the governance of either respective island.

The questions for the resident data 'islanders' would then be:

- What are the typical areas within the island archipelago where they waste time, effort or insight?
- What value is being lost from not being able to cooperate consistently as a cohesive island nation?

So, in enterprise reality, data governance itself may be fragmented. Governance is not necessarily synonymous with unified control, at least in practice. Lack of this unified control can result in data inefficiencies and limit higher-level applications and leverage of data. Island life, as it were, is not always paradise.



# The Benefits of Being Data-Driven Depend on Governance

Broadly speaking, the benefits of being more data-driven as an organization are widely recognized. Although identifying inefficiencies and reducing costs may be low-hanging fruit, the majority of identified benefits to being more data-driven as an organization are oriented toward net value creation.

#### Figure 3: Being more data-driven has numerous benefits

Source: 451 Research's Voice of the Enterprise: Data & Analytics, Data Platforms Q: What are the most significant benefits your organization would expect from being more data-driven? Please select up to three. Base: All respondents (n=500)



Contrasting with the traditional reactive perception of data governance, modern benefits of being more data-driven tend to be directly additive in value. The objectives of lowering costs and improving compliance indeed rank as identified benefits of being more data-driven, but they pale in comparison to benefits associated with optimizing the customer experience and improving business agility – additive or proactive objectives.

In particular, as organizations look to maintain competitive viability amid business disruption, they increasingly turn toward the bedrock of their profitability: customers. The global coronavirus pandemic has mandated digital delivery of the customer experience, and customer preferences across verticals have – for the most part – shifted. Optimizing the customer experience, via data, is no longer a simple competitive advantage; it is a baseline expectation.



# The Reality, Part I: Evidence of Lacking Data Governance

Despite the overwhelming evidence that data governance and supported data-driven initiatives have a net positive impact on business value, the reality for most organizations is not as bright. Most still struggle to gain a cohesive, unified understanding of their data landscape. While not definitive, current reported compliance stances are telling in this regard. It is impossible for individuals or organizations to understand the true extent of their compliance capabilities if they first do not understand the nature of the data under management. One needn't look further than reported compliance stances to understand the depth of 'data misunderstanding.'

#### Figure 4: Uncertainty is common with regard to compliance

Source: 451 Research's Macroeconomic Outlook, Corporate IT Spending, Q4 2019 Q: What is your organization's current status with regard to the California Consumer Privacy Act? Base: All respondents (n=571)



Reported compliance status with regard to the California Consumer Privacy Act (CCPA) can be used as a rough barometer for current data governance efforts. 451 Research's macroeconomic survey, conducted in late 2019, asked IT-aware enterprise practitioners about their organization's CCPA compliance status. Since CCPA went into effect in January 2020, one would expect that organizations subject to the regulation would be well on their way to implementing sufficient controls for data. Instead, a sizable minority of over one-third – 35% – reported they simply didn't know their organization's compliance status.

While an individual not knowing compliance status could stem from multiple nuanced organizational and technical factors, the most obvious barrier is complexity of the IT ecosystem and lack of governance. An organization simply cannot understand compliance capacity if it does not understand what data it has under management, or how that data is managed.



Many organizations struggle to simply understand what data they have. Not only does this lack of understanding slow down proactive initiatives to leverage data in new innovative ways, it hamstrings basic attempts to gain control for reactive needs. In this sense, everyone loses.

# The Reality, Part II: Evidence for Common Needs in Governance

While the need for data governance is generally accepted, methods – and technology – for achieving it can vary across organizations. What is generally emerging as a pattern, however, is the need for automation in information-centric tasks.

Current organizational use cases for machine learning (ML), for example, are telling. While customer-facing AI and ML use cases such as chatbots tend to steal the spotlight, current enterprise use of ML suggests that the technology is more frequently being applied to internal data management challenges. When organizations today are asked their top business reasons for using ML, 41.7% report 'data management and classification.' This use case far exceeds more glamorous customer-facing use cases such as product recommendation.

The need for more automation is also evident in workers' preferences. When asked, workers report that the things that could be most improved with automation are, indeed, processes and tasks that deal heavily with information.

#### Figure 5: Automation is desired for information-centric tasks

Source: 451 Research's Voice of the Enterprise: Workforce Productivity & Collaboration, Work Execution Goals & Challenges 2020 Q: Which of the following, if any, could be significantly improved with more automation? Please select all that apply. Base: Workforce respondents (n=915)

The integration of data from different applications into a workflow 27% Surfacing information when its needed 23% Assigning people to projects based on their skills and availability Reporting the progress of work against goals 20% The creation of documents, spreadsheets and slides 18% Prioritization of your different work commitments 17% The creation of your own daily workflows 13% The creation of emails and messages 13% The scheduling of your meetings 13% None of the above 21%

Hampered worker productivity is just a symptom of a broader data management syndrome. If the organization can gain better control and insight into data, it stands to accelerate the performance of the business from the ground up. As individual workers and contributors increasingly become dependent on data to navigate their daily responsibilities, any marginal improvement in that data's governance will help accelerate productivity.



Furthermore, insight into data usage (typically via metadata), derived through machine learned or intelligent automation, turns the status quo on its head. Rather than a user trying to find the right data, the right data will come and find the user.

## Conclusions

In a world that only seems to be growing more complex and chaotic, it can be alluringly simple to treat baseline requirements for data – such as regulatory compliance – as a net burden on the business. This view is myopic in the sense that it doesn't look at the organization as a holistic, interdependent ecosystem underpinned by and dependent on data. Any mechanism to increase the consistent, unified control of that data across the enterprise ultimately has downstream benefits to all users of data within the organization, a population that continues to grow. Both reactive and proactive use cases for data ultimately depend on this control, so viewing a traditionally reactive function such as compliance as being somehow in opposition to a more proactive function such as self-service data access is not only wrong; it is potentially detrimental to the organization.

While data governance today has a broadly positive reputation, functional barriers persist in achieving consistent control of data. Data silos, of course, remain a challenge. But the organization also must be aware of cultural/communications silos and 'abstractions' of data silos that can arise when there are different groups of people in the enterprise using different tools that connect to slightly different (but overlapping) sources of data.

Moving forward, ensuring the productivity of workers that need to access, view and leverage data in their daily roles should be a key emphasis for data governance. Any efforts to control data in a manner that creates tangible friction for these workers is bound to be met with resistance and maladaptive behavior that undermines the data governance effort as a whole. Thus, treating productivity as a top-level objective of governance, rather than an afterthought, should be a priority. Governance methodology and technology today can help facilitate this outcome, though organizations must be careful to strike a balance between 'bottom-up' organic adoption of popular tools/applications and 'top-down' or centralized control methods.

Data governance efforts need to be tightly aligned with an overarching effort to build and sustain a culture of data; data governance cannot succeed if isolated in a back room. All data stakeholders – increasingly, all workers – depend on the integrity and availability of data appropriate for their individual roles and responsibilities. Therefore, any technological intervention in the data governance effort needs to go hand in hand with ongoing cultural assessments, and if necessary, adjustments. Organizations with the greatest success fully recognize that all data-dependent initiatives – whether proactive or reactive in nature – are bound by a common need for consistently governed data across the organization.

## **CLOUDERA**

Learn more by visiting <u>the Cloudera SDX products page</u> or take a <u>video tour</u> <u>of Cloudera Data Platform</u> to see it in action.

#### **PATHFINDER** | ACCELERATING OUTCOMES WITH DATA GOVERNANCE



#### COMMISSIONED BY CLOUDERA

12



## S&P Global Market Intelligence

### About 451 Research

451 Research is a leading information technology research and advisory company focusing on technology innovation and market disruption. More than 100 analysts and consultants provide essential insight to more than 1,000 client organizations globally through a combination of syndicated research and data, advisory and go-to-market services, and live events. Founded in 2000, 451 Research is a part of S&P Global Market Intelligence.

© 2020 S&P Global Market Intelligence. All Rights Reserved. Reproduction and distribution of this publication, in whole or in part, in any form without prior written permission from S&P Global Market Intelligence is forbidden. The terms of use regarding distribution, both internally and externally, shall be governed by the terms laid out in your Service Agreement with 451 Research and/or its Affiliates. The information contained herein has been obtained from sources believed to be reliable. 451 Research and S&P Global Market Intelligence disclaim all warranties as to the accuracy, completeness or adequacy of such information. Although 451 Research may discuss legal issues related to the information technology business, 451 Research does not provide legal advice or services and their research should not be construed or used as such.

The content of this artifact is for educational purposes only. S&P Global Market Intelligence does not endorse any companies, technologies, products, services, or solutions. S&P Global Market Intelligence shall have no liability for errors, omissions or inadequacies in the information contained herein or for interpretations thereof. The reader assumes sole re-sponsibility for the selection of these materials to achieve its intended results. The opinions expressed herein are subject to change without notice.



**NEW YORK** 55 Water Street New York, NY 10041 +1 212 505 3030

SAN FRANCISCO One California Street, 31st Floor San Francisco, CA 94111 +1 212 505 3030



LONDON

20 Canada Square Canary Wharf London E14 5LH, UK +44 (0) 203 929 5700

BOSTON

75-101 Federal Street Boston, MA 02110 +1 617 598 7200