Redefining data governance

More and more enterprises are adopting data governance as a way to create order from the chaos of skyrocketing data growth. Unfortunately, garden-variety data governance often isn't enough to help all organizations become truly data-driven. Business users need a data governance platform that serves as a true system of record to engage all users—the Data Citizens—by enabling them to access, understand and trust enterprise data.

The undeniable explosion in data volume, combined with its associated expansion of velocity and variety, has put business users between a rock and a hard place. Their excitement over the possibility of uncovering new insights in order to make better decisions has often been thwarted by their inability to get the data, or even to make sense of it.

That dilemma—along with data-centric issues such as compliance, legal requirements and the advent of new data sources—has been a big contributor to the surge of enterprise adoption of data governance solutions. Research indicates that the global data governance market will grow from $863 million in 2016 to a whopping $2.2 billion by 2021—a 21% compound annual growth rate.[1]

Definition: a Data Citizen is any person who relies on access to trusted data to perform job responsibilities, make business decisions and ultimately drive business transformation.

While traditional data governance is often a good first step for enterprises finding themselves overrun by data and for business users looking for a roadmap to smarter decisions, it's really just the beginning of the process. True data governance delivers the context, relationships, lineage and access that Data Citizens need to make better decisions. Think of traditional data governance as table stakes, similar to eating your vegetables in order to improve your physical health. True data governance gives you a well-balanced diet of essential nutrients that maintain your data health and give you a competitive edge.

1. "Data Governance Market by Component (Solution and Service), Application (Incident Adjustment Management, Risk Management, Sales & Marketing Optimization), Deployment, Vertical, Business Function & Region - Global Forecast to 2021," MarketsandMarkets, February 2017

But to succeed, enterprises need to embrace the concept of a comprehensive, enterprise-wide data governance platform. And the platform should include an integrated data catalog to make it easier for business users to locate data, assess it in its proper context, learn from colleagues' similar searches and feel fully confident in its accuracy and relevance. By embracing this comprehensive approach to data governance, business users are transformed into Data Citizens, empowered through automated, yet straightforward, systems to make smarter and better decisions.

So what do organizations need to do to manage the business implications of these changes and become truly data-driven? How they answer this question will determine their overall competitive advantage.

Challenges and implications

The runaway growth in data volume, along with the increasing velocity and variety of data, represent a double edged sword for Data Citizens. On one hand, there's more data to help all data users make better, smarter and faster decisions. On the other hand, the process of locating, sorting and analyzing the data is more challenging.

For instance, consider the simple fact that up to 60% of enterprise storage capacity is dedicated to accommodating copy data.[2] The emergence of a slew of new data sources—applications, repositories, devices and others—ensures that data doesn't reside in just one place. Data is siloed across the organization. And, because that data lives in many places, it becomes very difficult to find.

Now, add in the challenges of the "connected enterprise," with users creating, using, and sharing data across social media, cloud services, mobile devices and the Internet of Things (IoT). Artificial Intelligence is another major trend that has dramatically spiked data volume because of its increasingly widespread use in analytics, Machine Learning and cognitive computing.

This makes it difficult to actually use all that data being created. And since data is only valuable if and when it is used, organizations need to shorten the path to finding the data for Data Citizens.

The runaway growth in data volume, along with the increasing velocity and variety of data, represent a double edged sword for Data Citizens.

A true system of record—an enterprise-wide data governance platform—breaks down silos by connecting all relevant data, regardless of the data sources.

Moreover, as these changes accelerate—will most likely be exacerbated by the emergence of new, still unknown, developments—organizations need flexibility. That's because this is just the beginning of a dizzying spiral of complexity and confusion.

As a result, the implications of not getting a handle on this data complexity are profound and becoming increasingly so. According to the 2016 Data Science Report, three out of every five data scientists surveyed spent the largest part of their time cleaning and organizing data, often referred to as "data wrangling."[3] Not surprisingly, there are big costs associated with this—especially in opportunity costs borne when users aren't doing "productive" work with the data.

The lack of clear context and definition of organizational data means organizations need a better way to normalize data that exists in different formats and contexts. Without it, organizational agility is compromised, especially considering enterprises' reticence to throw manpower at these problems or to invest heavily in third-party technical consultants.

There also are potentially crippling compliance and legal problems that arise, due to inaccurate or incomplete data. All of these issues can cascade into a scenario where enterprises lose competitive advantage because they either can't make decisions quickly and confidently, or end up making the wrong decisions. Additionally, organizations can lose another invaluable asset: customer trust. All of these challenges can—and often do—result in a hit to an organization's business performance.

Defining the solution

Helping business users sift through a huge maze of data, without any real insight into the context, accuracy, timeliness, or relevance of the information, requires new kinds of tools. Often, traditional data governance solutions only look at data across individual applications and data sources don't act a system of record for all data across the enterprise.

A true system of record—an enterprise-wide data governance platform— breaks down silos by connecting all relevant data, regardless of the data sources. In doing so, the platform approach makes the data meaningful for all users, as well as improves its accessibility. Equally important, the data governance platform must be able to scale seamlessly and efficiently in

3. "2016 Data Science Report," CrowdFlower, April 2016

order to accommodate ever-increasing data volumes. Scalability is essential because the varieties of data are rapidly changing; spreadsheets and Word documents are no longer the primary sources for the most essential data.

The platform must take an extensible approach that can encompass evolving data capabilities, such as those stemming from IoT, Artificial Intelligence, Machine Learning, cognitive computing and blockchain.

And then, business users need even more functionality and capability if they are going to fully leverage their data. They need the ability to understand the data's definitions, context and lineage. The data must also be accessible in a straightforward way. An integrated data catalog is a vital and essential element within this system of record; after all, without a data catalog, data governance is nothing more than a collection of lists.

In addition to providing the location, context, lineage and meaning of the data, a data catalog should take advantage of "crowd-sourced" metadata to provide a more comprehensive view of data using input and experience from relevant colleagues within the organization.

How Collibra addresses these issues

Delivering both a comprehensive system of record and a context-aware data catalog can seem like a tall order for many executives to fulfill. Too often, Data Citizens have struggled to achieve their goals using legacy data governance systems.

Fortunately, Collibra offers both an enterprise-wide data governance platform and an integrated data catalog. Collibra puts people first by building its approach on the core principles of helping all Data Citizens access, understand and trust their data.

The Collibra Platform acts as the system of record for Chief Data Officers and all Data Citizens who must find the right data at the right time. The platform breaks down legacy data silos to locate and surface all relevant data and it enables Data Citizens to understand the context and relevance of the data. This approach spans data sources across the entire enterprise, helping to squelch potentially debilitating data brawls that often erupt when users get conflicting information in their searches. The end result of this process is the ability to have a sense of trust that the data provides the right insights for the right business problem in the right context.

It's intuitive, easy to use and built with the business user in mind, in order to help bridge the gap between IT and the business.

Collibra also offers the Collibra Catalog, a governed, context-aware application that gives business users the ability to understand the lineage, ownership, usage, and context of data. Collibra Catalog is built on top of the Collibra Platform with governance and privacy embedded, ensuring users always have access to the most accurate and trusted data across the enterprise. The catalog also enables self-service data access so users can quickly and easily find the trustworthy, pre-defined and pre-approved data they need to do their jobs without waiting in the IT queue.

This integration of the Collibra Catalog into the enterprise-wide data governance platform is unique in that it gives experienced data scientists, IT and all Data Citizens the ability to capture all relevant, up-to-date information; it's intuitive, easy to use and built with the business user in mind, in order to help bridge the gap between IT and the business. It also helps to ensure the data's accuracy and relevance in order to instill a feeling of trust and confidence. This integration is crucial because enterprise-grade capabilities are vital to support organizations with complex operating models, structures and data architectures.

Conclusion

Too many Data Citizens find that, despite their best intentions and their best efforts, they come up short when trying to use the right data, at the right time, for the right purpose. All too often, that disconnect is driven by the lack of a comprehensive, enterprise-wide data governance platform powered by an integrated data catalog.

The sheer volume and size of data sets, made even more challenging and complex by the inability to locate exactly the right data among scores of data sources, demands a data catalog that incorporates machine learning to automate data discovery and personalize search results.

The Collibra Platform, with Collibra built on top of it, helps organizations use their data to achieve meaningful business outcomes. As a scalable, cloud-based platform, Collibra is positioned to securely enable departments across the whole organization, accelerate digital transformation and empower Data Citizens.

Collibra | TechTarget Custom Media White Paper

For additional questions, contact us at:

United States
+1 646 893 3042

United Kingdom
+44 203 695 6965

All other locations
+32 2 894 79 60

By email
info@collibra.com