

DELIVERING ON THE PROMISE OF BIG DATA THROUGH DATA GOVERNANCE



Today's Enterprise: Struggling to Extract Value from Data

Organizations today are faced with a new challenge: how to leverage the power of big data to drive business advantage. In addition to all the data about customers, products, sales, employees, inventory, and vendors with which organizations have been wrestling for years, business users are now faced with even more and bigger data from sources including social media, Internet of Things, and other types of behavioral data like website clickstreams.

Organizations know that somewhere within reams of raw data lie answers to important questions about current and potential customers and what they really want and need. Answers to questions about competitors, vendors, regulators, markets, and territories. Answers to questions like: How are we doing? What are we doing right? What are we doing wrong? How much better could we be doing? Where should we be going? What should we do next? What should we do next year?

This volume of data in the work world will only continue to expand as more business becomes digital. Further complicating matters, new data sources are often unstructured or semi-structured formats. External data sets also pose a threat to privacy and regulatory compliance. When these different characteristics of big data vs. traditional data are ignored, organizations end up with data that is either used inappropriately due to its inaccuracy and lack of precision, or is missing context and cannot be interpreted for analysis. And critically: if anything about the data or its analysis is unclear to business users – the data citizens who need this information in their daily work life – they'll fail to trust that data, especially if the data suggests letting go of long-held beliefs or work practices. And since the point of big data analysis is to change the way organizations work and behave, this lack of trust means the analysis fails to impact the operation of the organization and its people, and delivers little value.

That's the paradoxical peril and promise of big data: It's the swollen stream that can drown your organization, but it's also the gold mine of useful information your organization can leverage to gain competitive advantage.

Data governance is the missing link between big data and its value.

Enter Data Governance: Mining the Gold

Before any data gold can be mined, and before all the answers it contains can be understood, analyzed, and used for the good of the organization, the data must be accessed. And before it can be accessed, it has to be input, stored, organized, and shared in logical, efficient, streamlined processes that make sense for each individual organization.

This is the role of data governance: to create a sophisticated and systematic approach to managing the deluge of big data while ensuring its availability, usability, integrity and security. Data governance gives data citizens the information they need to find, know, and trust the data. And it gives organizations the capabilities to take care of the data and improve its usability, quality, and value.

Companies often accrue and store their data in an ad-hoc fashion. In some cases, different departments know (or think they know) which data they need to gather, which data to keep, and how to keep it. These "little data" processes are typically adequate for day-to-day procedures: if accurate bills are sent out on time and payments are recorded correctly, then the billing department may think its data is managed just fine.

But across the hall, the marketing department may be in need of some of this same behavioral data to enable them to predict which customers are likely not to renew after their initial contract period is up. Unfortunately, this group may not have access to this vital organization, even though it's in daily use by colleagues in other departments. This

silos, which happens all too often in today's enterprise, can prevent a company from realizing the full potential of its data. For instance, client data stored in the billing department could offer the marketing department insight on how to design new product offers that attract and retain lucrative customers.

Big data offers the perfect use case for data governance. The tremendous volume and variety of data and the enormous potential value of analytics is combined with the frequent need to combine data across organizational silos.

Going Beyond the Duct-Tape: The Value of An Automated Approach to Data Governance

Most organizations may already have some basic data management in place for storing, cleaning, consolidating, and reporting data. But the control processes around the data are often duct taped together through spreadsheets, e-mails, and meetings, with no larger system or team in place to ensure enterprise-wide accessibility, accountability, flexibility, and consistency.

Because big data is typically so vast and hidden within independent business departments, functions, and/or processes, gathering and organizing it for analysis is difficult. Consequently, the process of analyzing the data may not only be logistically challenging, the results might be biased and lack context, the analysis too often based on convenience samples or subsets defined by disparate groups across the organization. In other words: garbage in, garbage out.

Data governance is the way an organization gathers and enriches information about its data, identifies relationships between data and its uses, and corrects problems with that data. Data governance assures that the data is the right data, used the right way, by the right people, and means what business users think it means.

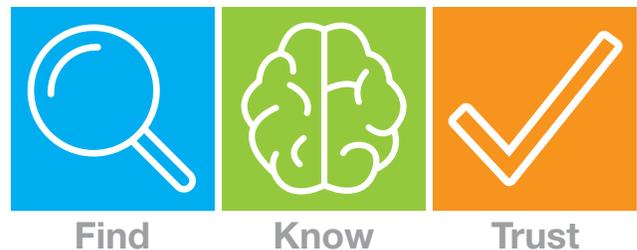
Effective data governance creates policies and rules that enable data citizens throughout the enterprise to easily, logically, consistently, expeditiously, and securely store their data. And most important, data governance enables data citizens to access that data and analyze it to achieve organizational goals. Implementing data governance is the crucial step to unlocking the true value of a company's data – to mining the data gold and using it to shape business strategy and achieve organizational goals.

Data Governance: What It Does

An effective data governance program will achieve the following objectives, all focused around Finding, Knowing, and Trusting valued data:

FIND:

- It will work with existing database management systems and tools to provide visibility into all your data – there is no need to reinvent the wheel.
- It will establish responsibility for big data management within the organization. It will make these responsibilities visible so data citizens can find the experts, stewards, and stakeholders.
- It will provide flexibility – it will adjust to changes in data and changes in the need for data, and determine the business impacts of those changes to ensure a company isn't working with outdated information.
- It will automate functions and processes to save time and reduce hassles.
- It will support multiple different platforms on multiple different devices, including mobile devices so data citizens can find the information they need, anywhere and anytime.
- It will be user-friendly and operate from a consumer perspective; it will, for example, support natural language searches.



KNOW:

- It will enable and encourage good behavior among employees by making protocols more logical, efficient, understandable, universal, and transparent. Data citizens across the organization will grow to value big data when it becomes easier for them to both manage and exploit its potential.
- It will encourage collaboration and data sharing across departments, divisions, and functions. Data citizens will ask: how can my data help your department and vice versa? What should we know about each other's data? How can we collaborate to leverage the promise of big data to achieve organizational goals?
- It will give the organization the tools to understand and measure its data in ways that are appropriate and meaningful for that organization and that support the organization's business goals.

TRUST:

- It will enable the organization to take care of its data, to keep it safe, secure, accurate, and "clean."
- It will limit the risks of big data management by establishing protocols that incorporate regulatory frameworks as well as privacy and confidentiality concerns – especially critical in industries such as financial services and healthcare.
- It will create explicit agreements among colleagues about how the data is used and what it is for, ensuring that the system answers needs across the organization and creating buy-in among all users.
- It will enable data citizens to present data in a meaningful way, to identify and share with each other what matters in the data, e.g., trends, directions, problems that need addressing, opportunities that should be exploited.

Five Steps to Reaping Big Data Rewards Through Data Governance

How does an organization get started on this path to reaping big data rewards through data governance? Implementing a data governance process isn't difficult, but it does require structure and automation to enable transparency and support trusted decision making.

Here are five steps to building a robust and valuable data governance process. These approaches generate the information users need to create and utilize big data analytics.

1) Put Clear Policies in Place

The first step in governing big data is to establish a set of clear policies about your data. In particular, there must be policies surrounding:

- data inventory
- data ownership
- critical data elements (CDE)
- critical big datasets (CBD)
- data quality
- information security
- data lineage, and
- data retention

This kind of policy management ensures a business consults the right stakeholders, understands the impact of changes, and formulates and enforces policies that improve efficiency and reduce errors and risk. This enables control over previously hidden or siloed enterprise data, and empowers all data citizens to go beyond just producing and consuming data to trusting and using the data to optimize value.



2) Define Data Standards to Drive Value

Once policy measures are formulated and in place, organizations then need to define data standards for those big data sets that are expected to drive the most value. Critical data elements must be identified, and the metadata and relationships they have documented. This isn't a once-and-done proposition for two reasons: the sheer volume of data makes it impossible to do in one fell swoop and also, each time a data citizen uses data in a new way, new meaning can result. Critical data often includes some structured information from the enterprise, but also things such as chat logs, Facebook data, Twitter feeds, and sensor data must be factored in to create new, value-added processes and activities.

3) Make it Flexible and Automated with Technology

The ever-increasing volume and variety of big data requires flexibility and automation. Unlike the rigorous and centralized policies of previous systems, big data activities and decisions are spread across data citizens throughout the organization. For instance, having automated ingestion functionality in place for the big data lake can ensure clean quality data, while also tracking the history and showing the impact of changes to any data standards. By their very nature, manual processes cannot scale with variety and breadth of data, or the new data that is brought into the lake almost daily. It's essential to automate the data governance with technology.

4) Leverage Analytical Models

It's important for analytical models to be part of your big data governance processes. The data governance approach, and the system that automates it, should have the flexibility to capture information about all aspects of your analytics, from map/reduce jobs to visualizations. Processes need to be simple for the data citizens and data professionals alike. If a data citizen wants to request new analytical models, there must a simple way to do this, and its status to be monitored. If not, data citizens will not trust the process.

5) Form a Dedicated Data Governance Team

Effective data governance requires buy-in and participation from across the enterprise. Organizations need to assemble a strong data governance council that includes the chief data officer and executive sponsors from key departments and functional areas. Data stewards and subject matter experts must be empowered to constantly augment, improve, and enhance the data and its information.

Collibra: Helping Organizations Reap Big Data Rewards

Organizations that want to tame the big data deluge and unlock its value turn to Collibra, the enterprise business platform for governing big data. Collibra helps Chief Data Officers, information professionals, information stewards, and data citizens across the enterprise maximize the value of their data through a cloud-based or on-premises solution expressly built to deliver big data value.

Only Collibra delivers the insight, collaboration, and ease of use that data citizens need to take advantage of big data analytics by helping all data citizens find, use, and trust their big data. The Collibra data governance platform offers a framework for setting data-usage policies and implementing controls designed to ensure that information remains accurate, consistent and accessible. It enables organizations to provide a set of information to their users, making it possible for them to leverage the power of big data.

In today's "Wild West" big data environment, Collibra enables data citizens to understand the data available to them and how to access it – turning the quagmire of big data into business insights that drive competitive advantage.



collibra.com

US +1 646 839 3042
info@collibra.com

Follow Us:
twitter.com/collibra