GIGAOM

Data Governance v1.0

The Key to Integration, Balance, & Transformation

ANDREW J. BRUST

TOPIC: **DATA & ANALYTICS**

CREDIT: GERD ALTMANN

GIGAOM

Data Governance

The Key to Integration, Balance, & Transformation

TABLE OF CONTENTS

1. Summary

The current era of analytics is one of extreme contrasts. More than ever, organizations are being encouraged to, or already have bought into the idea that they must leverage the full value of their data for strategic and competitive advantage. They must build data literacy, encourage a data-driven mindset, and instill in their team the discipline of analyzing data to the point that doing so is almost second nature. But organizations are conditioned also to be vigilant in protecting their data, to comply with new privacy regulations such as GDPR and CCPA, and some have learned the hard lessons of high-profile data breaches.

This combination of maintaining a proactive data-driven mindset and a defensive data protection regime can, at times, seem almost schizophrenic. So how can these seemingly conflicting goals be harmonized and brought into equilibrium? Data governance is the key to establishing a blended, balanced approach to address the needs of data culture with the requirements of data protection. By combining curation of data to make it discoverable, and control over that data to keep it safe, organizations can, with hard work and careful planning, rationalize these needs and make it all work. Enlightened organizations can enable data democratization with governance across on-premises data lakes, cloud-based databases and analytics platforms to make the data within those repositories available and drive new business insights.

To be sure, this combination of curation, control, and data democratization is easier said than done. But getting there is well within reach, attainable by most any organization that is serious about attaining it. And those organizations that do will find themselves in a virtuous cycle of continuous growth in data availability, accompanied by a maximizing of that data's value. The ethos of data governance and data-driven thinking yields a perpetual data dividend.

2. Facets of Governance

Data governance is not a monolith. It is a fabric of several technologies and methodologies. Essential components of a governance regimen include security and access control, data classification, and data discovery. Beyond these components themselves, a dedication to cultural change in the organization is also needed, in order for data-driven thinking to be evangelized and adopted through patient, dedicated mentoring.

Security and access control is perhaps the very nucleus of data governance, although many concentric rings will surround it. Making sure users in specific roles see all of the data they are authorized to see, and none of the data for which they are not authorized, is the very foundation of successful data governance. Open source software such as Apache Ranger provides core functionality for such role-based access control, and it can work hand-in-hand with Apache Atlas, focused on governance more broadly, including data discovery at scale.

Open-source frameworks like these provide a standardized foundation upon which vendors can base their platforms and to which they can add value. For example, a vendor platform may use Ranger and Atlas as its "engine," then provide its own user experience and integration with other elements in the governance platform.

Data Classification

One such additional element is data classification, sometimes called data fingerprinting. Data classification is a fundamental data governance technology that can scan databases and data lakes, then determine what *kind* of data various columns contain. Such identification goes beyond the raw data type (text, dates, integer or decimal numbers, etc.), and extends to business types, including part numbers and customer IDs; geographic entities like cities, regions, and countries; and generic entities like company names.

Data classification can be useful also in identifying personally identifiable information (PII), and thus it is a critical component in sensitive data protection and data privacy. Some platforms may use data classification *only* for sensitive data detection/protection, though it applies to data of all types.

Data Discovery

Data discovery functionality in a governance system can detect relationships in data, be they codified in the data's schema or completely undocumented. These relationships may exist between tables in a single database, data sets in a data lake, across tables in different cloud databases, or even among multiple on-premises and cloud data services.

Data discovery platforms work by looking not just at the metadata/schema of the database, but at the actual data stored in the tables and datasets. Part of the process involves looking for statistically significant matches between columns in different tables and data sets.

Distinctions and Nuance

To some, governance is all about security. But, as we have seen, there is much more to it than that. Federal governments perform more than law enforcement; they build roads, create rules for commerce, develop and maintain parkland, and facilitate recreational activities within those properties.

Similarly, data governance is not just about security and access control. It also implements data curation, discovery, and lineage to smooth the path for business users to find the data they need, and enables them to trust that it has been vetted. Similarly, while a data catalog might provide a simple metadata repository, a full-blown governance platform provides data classification, data discovery, and more, all while providing base access control and metadata management, to boot.

3. Culture Shock?

All of these governance platform elements can be combined to catalyze, facilitate, and promote data-driven operations within an organization. But they cannot do that by themselves. Instead, a symbiosis between the technology, on the one hand, and work culture evolution on the other, is necessary. Deputizing business workers to do analytics in their organization must be accompanied by training and mentoring as well to build their skills, their confidence, and their success.

Ultimately, the creation of a data culture hinges on the acquisition of habits, instinct, and standard procedures. Technology can accommodate and accelerate the cultural change, but the desire and effort to establish it must come from the organization. Data-driven operations and digital transformation, in general, are as much about change management as they are about tech.

4. Governance as Integration Layer

Heterogeneous technology is a fact of business life. Even companies with strict standards around platforms used within the organization will encounter exceptions as different teams pursue the best solutions for their needs. This is particularly true for organizations beginning to adopt the public cloud, especially if individual business units are taking the lead. Some organizations offer a set of technology choices, while still others embrace the variety and deputize business units to adopt their preferred platform.

Regardless of where along the spectrum of data technology variety your organization might be, data governance is key to bridging the gaps, breaking down the silos, and bringing cohesion to your data estate, whether it is generally concentrated or spread far and wide.

Your data assets may be heterogeneous, but a homogeneous data governance and access layer can bring equilibrium. A holistic governance platform allows data to remain physically near the operational systems that generate it, while providing a single point of control for discovering, documenting, securing, and connecting to it.

While the word "governance" has connotations of control, compliance and obligatory protection, a governance platform's utility extends well beyond these defensive capabilities. A good data governance implementation can function as the connective tissue among the various organs of your data ecosystem. A governance layer provides horizontal unification across what would otherwise be vertical silos of data. When governance systems are well implemented and well adopted, they weave otherwise loose threads of data into a fabric, a concept that works as both a term of art and an apt metaphor.

5. Order + Facilitation = Data-Driven Coordination

Data governance is not about authoritarian rule; it is about order, control, curation, and usability. In an era where organizations must use their data to their greatest advantage, yet protect privacy and prevent breaches, data governance is the great equalizer, bringing productive coexistence to these seemingly contradictory edicts.

The nature of operational systems is such that data is naturally scattered, siloed, and even incongruous. That is because those systems are varied, and the data is optimized for their performance, as it should be. On the analytics side, the goal is to smooth out these differences and provide a logical layer of unification over these physical and operational separations.

Data management, in general, and governance in particular, reconciles data's discrepancies, disparities, and dispersal while controlling access and ensuring protection. In doing so, it adds an element of endorsement and trust to data sets that enhances your team's confidence in the data. That, in turn, will strengthen team members' resolve to employ data-driven practices in their work and embolden them to do so consistently.

If your organization is undertaking digital transformation in an earnest, disciplined fashion, you will find data governance to be foundational to the effort. Digital transformation is not magic. Rather, it is about the rigors of coalescing your data into a rational whole and ensuring access to it by parties with the appropriate need, permissions, and passion. Ultimately, data governance uniquely marshals the balance of discipline, availability, and facilitation needed to make a digital transformation, and your organization's success, a reality.

6. About Andrew Brust

Andrew has held developer, CTO, analyst, research director and market strategist positions at organizations ranging from the City of New York and Cap Gemini to Gigaom and Datameer. He has worked with small, medium and Fortune 1000 clients in numerous industries and with software companies ranging from small ISVs to large clients like Microsoft. Andrew's resulting understanding of technology, and the way customers use it, makes his market and product analyses relevant, credible and empathetic.

Andrew has tracked the Big Data and Analytics industry since its inception, as Gigaom's Research Director and ZDNet's lead blogger for Big Data and Analytics. Andrew co-chairs Visual Studio Live!, one of the nation's longest running developer conferences. As a longtime technical author and speaker in the database field, Andrew understands today's market in the context of its longtime Enterprise underpinnings.

7. About GigaOm

GigaOm provides technical, operational, and business advice for IT's strategic digital enterprise and business initiatives. Enterprise business leaders, CIOs, and technology organizations partner with GigaOm for practical, actionable, strategic, and visionary advice for modernizing and transforming their business. GigaOm's advice empowers enterprises to successfully compete in an increasingly complicated business atmosphere that requires a solid understanding of constantly changing customer demands.

GigaOm works directly with enterprises both inside and outside of the IT organization to apply proven research and methodologies designed to avoid pitfalls and roadblocks while balancing risk and innovation. Research methodologies include but are not limited to adoption and benchmarking surveys, use cases, interviews, ROI/TCO, market landscapes, strategic trends, and technical benchmarks. Our analysts possess 20+ years of experience advising a spectrum of clients from early adopters to mainstream enterprises.

GigaOm's perspective is that of the unbiased enterprise practitioner. Through this perspective, GigaOm connects with engaged and loyal subscribers on a deep and meaningful level.

8. Copyright