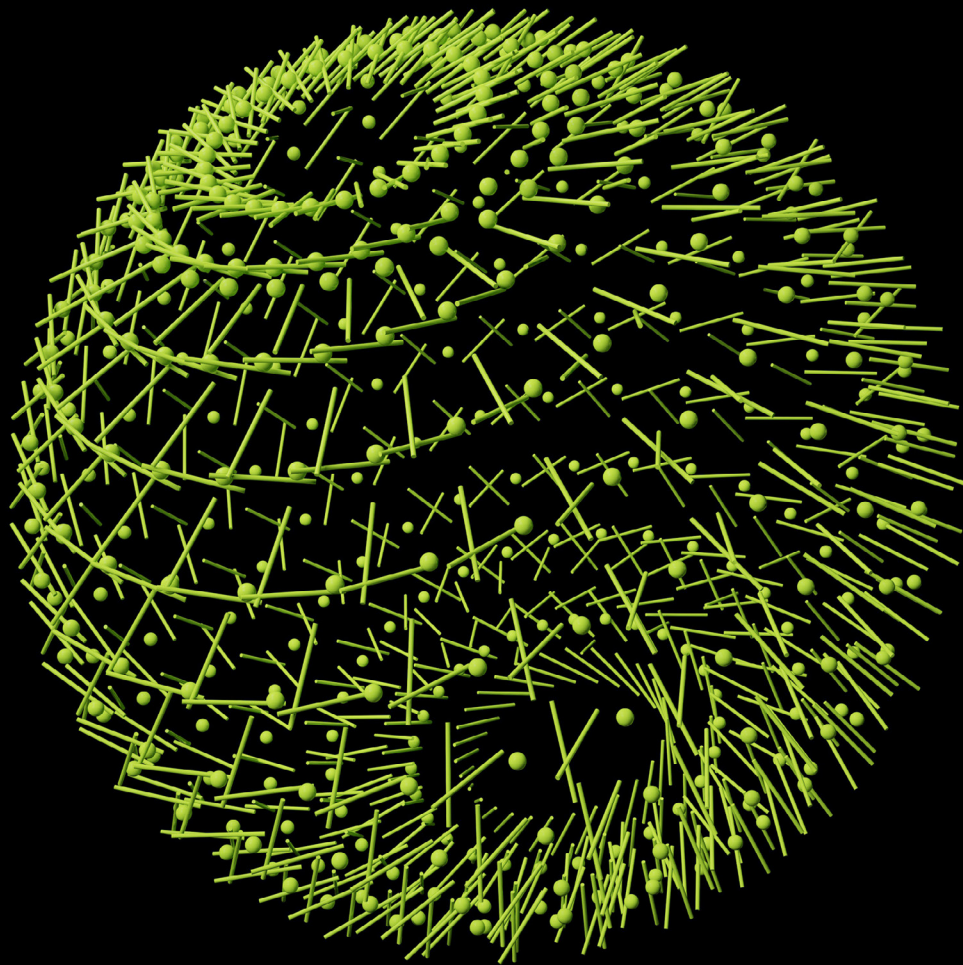


Deloitte.



**Data governance for
next-generation platforms**

Today's explosion of data—and the insights revealed—is not only highly valuable to organizations from a strategic standpoint, but also it presents challenges in storage, management, and adherence to regulatory and legal requirements. Developing an effective data governance program for the next-generation platforms staged to manage this landscape is essential to harness the data's potential and to help minimize risks.

With great power comes great responsibility

Data is significantly changing the game in today's world. Pulling as much of that data into a single place and organizing it so advanced digital technologies can serve up insights puts innovative companies at the significant advantage of being able to see what others can't.

New cost-saving and revenue-promoting innovations become possible because next-generation technologies, like sensors, IoT, AI, and robotics have access to the big data systems they need to operate. Without this, it substantially lowers the potential for new digital technologies to be successful. The sheer volume of data empowers them in ways that just aren't recognized with smaller sets of data.

Yet, it's going to be a delicate balance between control and access. For decades, many organizations spent their time, money, and resources on defenses and procedures designed to not only keep external cyber adversaries out of their networks, but also manage access internally. These were often elaborate systems and procedures geared toward maintaining more control across legacy systems. However, today, the advent of new technologies, which are paramount to business and have the potential to make overall operations more intelligent and automated, require more access to data and at higher velocities. It's causing organizations to adapt and change their modus operandi from controlling data to managing it.

The new way of managing data is more cost-effective and efficient, scales, and most importantly, can integrate with next-generation technologies and applications it couldn't before, but with it comes great responsibility. No matter what the best mix may be for an organization's data management system, the data itself will always remain the responsibility of the organization, and detailed data governance and security plans are of utmost importance. The consequences of a compromise are just too large not to focus on them.

Now, couple that governance and security challenge with the regulatory and legal restrictions that companies are going to have to adhere to on a global scale. From the 2018 deadlines for the European Union's General Data Protection Regulation to recent Australian privacy laws, these country-specific responses are complex, and companies can face penalties if they fail to abide by the rules.¹ Many organizations are realizing they don't know—much less have control over—what data exists within the enterprise, where it sits, and how it is being used across business units, geographies, or with third parties.² Data governance and security, and its driving change, should be top of mind for every C-level executive.

Next-generation platforms, which are being built in response to exponentially expanding data, need to support business operations, regulatory requirements, process automation, and digital transformation. However, with the use of next-generation platforms comes the challenge of the increasing scale and complexity of enterprise and data architecture—and subsequently, the governance around it. **When processing large data sets, unstructured data, or low latency data, it becomes more difficult to understand the context, importance, and associations of data, and thus to ensure the data's reliability and security. Without a shift in data governance mind-sets, companies will likely find it challenging to leverage all the business benefits of next-generation platforms and successfully manage risks.**

Why it can't stay the same

Here are the drawbacks of traditional data governance. Organizations developed traditional data management techniques when data volumes were still relatively small.

In this simpler time, structured business data typically lived in tables or basic systems.³ It went one of two routes. Either organizations built one-off systems to address specific needs or they created strict processes that ruled the day for all company data. The latter worked well until data became unstructured and originated in a variety of external and internal sources. Both have a high possibility of failure as we enter a true data tsunami.⁴ Traditional data governance can't handle either the variety or scale of data that is headed our way.

International Data Corporation (IDC) offers a macro view, predicting that by 2025, the world will create and replicate 163 zettabytes (ZB) of data annually (a ZB is 1 trillion gigabytes), representing a 10-fold increase over the annual amount of data generated just nine years earlier.⁵ Data governance becomes increasingly pressing for companies as data grows. Currently many companies are struggling to do it, and the problem is about to compound.

What we got right

It's not all lost. Data governance capabilities for traditional platforms are well-defined, and the processes are mature in organizations that are data driven. Foundational capability activities include the identification of an enterprise's data governance organization and who is responsible for identifying data-related policies, standards, and processes. These activities are subsequently followed by the identification and definition of key data domains that support a business purpose, like HR or sales.

From there, critical data elements are identified in each domain (i.e., payroll or CRM data), followed by the definition, collection, and monitoring of key data elements' metadata. Metadata helps categorize data elements, so humans and machines can easily recognize its original source. Metadata will play a significant role in tomorrow's managed data environment.

From a technical standpoint, traditional platforms typically manage structured data and use relational databases and ETL tools to move and transform data from source to target systems and ultimately feed analytical platforms. The variety of technical and operational metadata assets that need to be captured is relatively small, which makes the metadata collection process structured and suitable for automation. However, large decentralized organizations can have a complex and ambiguous data architecture, which brings additional challenges even in traditional data governance implementations. Either way, typically the company's foundation for data governance will help an organization move forward as they transition to new modernized platforms and capabilities. It just needs to shift and expand to adjust to new technology requirements and the volume, regulations, and velocity tied to big data.



What makes a stellar data governance program in the new world of data?

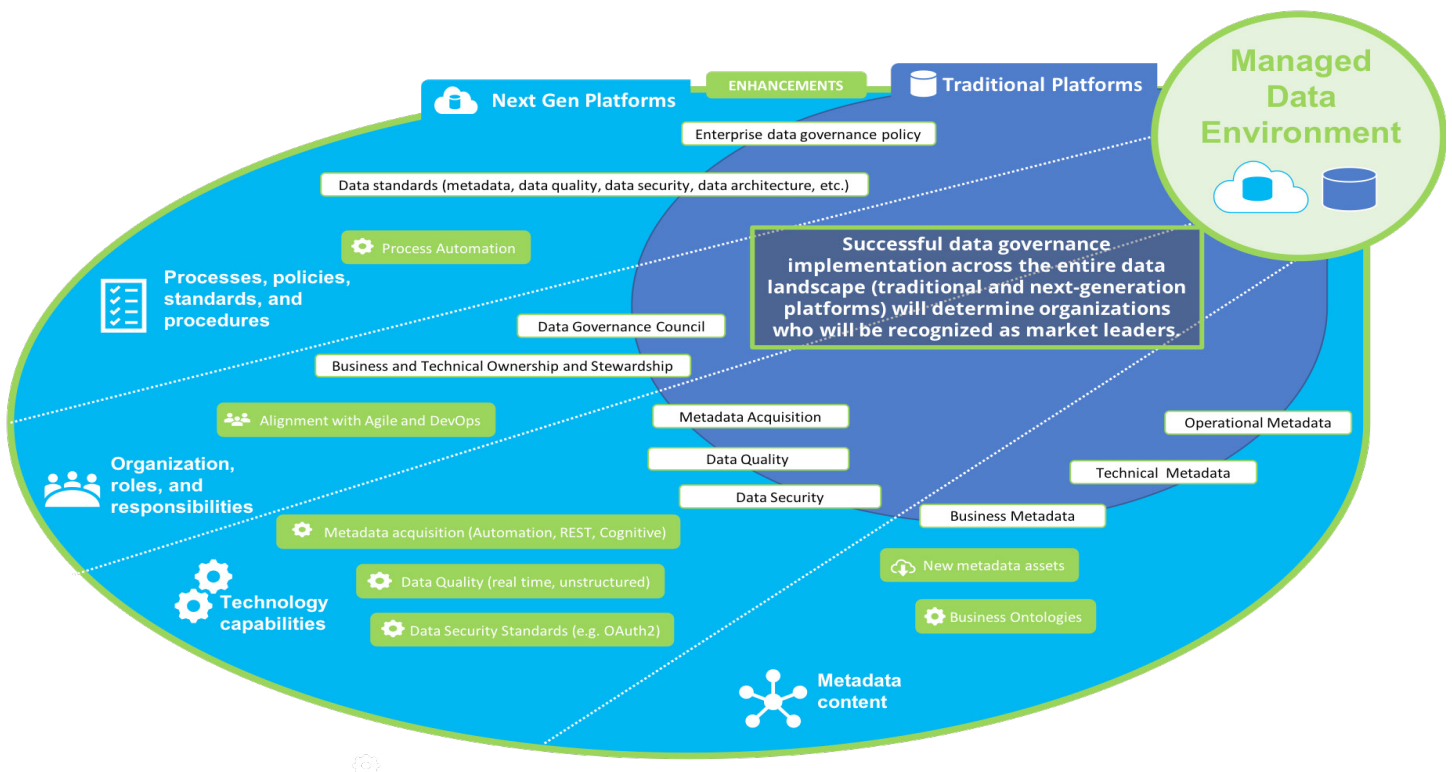
There are four pillars.

Data governance of tomorrow is not only about maximizing the value of data for operational effectiveness, decision making, and regulatory requirements, but also about minimizing the risks associated with poor data management. As a fundamental capability of data management, data governance relies upon four pillars:

- Processes, policies, standards, and procedures
- Organization, roles, and responsibilities
- Technology and tool capabilities
- Metadata content (or data catalog)

Well-established data governance capabilities for traditional platforms is a solid starting point for bringing next-generation platforms under the enterprise data governance umbrella. However, the increased scale and complexity of the enterprise data architecture requires enhancements across the four key data governance pillars (see figure 1).

Figure 1. Data governance capability enhancements to support next-generation platforms





Pillar 1

Processes, policies, standards, and procedures

Without well-defined policies, chaos ensues

Processes, policies, standards, and procedures

include the definition of different abstraction layers and requirements of how data must be managed. A data governance policy should contain an organization's strategic direction and high-level descriptions of both the desired behavior it is trying to achieve and the desired outcome. Standards define the rules required to ensure that the data fits the purpose. Processes and procedures provide details about how standards will be executed. Typical standards include metadata management, data quality, data security, data architecture, and data modeling.

⇒ *Implication for next-gen platforms*



Increased scale makes automation key

Some traditionally manual data governance processes cannot keep pace with the increased scale

of next-generation platforms. Therefore, these processes require automation in metadata discovery, ingestion, interpretation, and enrichment. In some cases, automation requires the use of disruptive technologies, such as cognitive technologies and artificial intelligence (AI) to help keep pace.



Distributed data intensifies data security efforts

Data security must be managed differently in a distributed environment, where copies of the

same data may be stored in different physical locations. In addition, new data architecture patterns, such as microservices, are more susceptible to

compromise due to a highly distributed environment of independent components. Therefore, the data security features across different components of the enterprise data architecture must be standardized to achieve a feasible solution across the organization. Otherwise organizations may be faced with significant operational, regulatory, and reputational risks.



Unstructured data presents new challenges for data quality

The increased scale and complexity of data has a direct impact on how organizations should consider

performing critical data quality activities. High volumes, varieties, and velocities of data significantly increase the time required to perform routine data-quality assessments and may change the organization's approach to data-quality itself. For example, the quality of unstructured data can be assessed by recognizing patterns, transforming it into structured data sets, and validating it with trusted sources. That way, organizations can increase trust in unstructured data, take advantage of the information, and leverage enhanced decision making.



Evolutionary analytics make new data-quality innovations necessary

Additionally, a large volume of data processed by next-generation

platforms enables interactive, experimental, and evolutionary analytics, unlike traditional platforms where usage is primarily for a defined functional goal like regulatory reporting. Hence the scrutiny of data quality and its associated time and effort should be aligned accordingly. Analytical applications should have inherent capabilities to work around the acceptable quality issues in data.

Pillar 2 Organization, roles, and responsibilities

Who's held accountable?

The buck stops here. Companies should carefully define formal bodies, roles, and responsibilities within an organization to facilitate, oversee, and perform business processes in the context of data management and governance. Consider forming a data governance council, which manages the development and implementation of processes, policies, standards, and procedures. Choose both business and technical owners and stewards who will be accountable and responsible for the data governance content.

⇒ *Implication for next-gen platforms*



Don't get stuck holding the bag

Next-generation platforms still require well-defined data governance bodies, roles, and responsibilities with the organization, but anticipate those roles to evolve significantly with new technologies. The changes in how organizations perform their data operations will require a greater injection of data governance requirements, not just in everyday processes but into the software development life cycle itself. Furthermore, the introduction of process automation and disruptive technologies used to address increased scale and complexity of big data will impact how certain roles and responsibilities will be defined. With more data and more access comes more risk. Don't get stuck holding the bag because organizational roles weren't well-defined in the first place.





Pillar 3

Technology and tool capabilities

New tools can help manage data governance

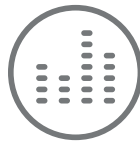
New technology capabilities include platforms, tools, and the tool subject-matter experts required to support a sound data governance process. Typical tools associated with data governance are metadata management, data quality, workflow management, and data security tools, and they have changed significantly. Today's tools take on a proactive rather than reactive approach to data governance. In addition, traditional data governance tools are limited in that they can't support the variety of metadata sources and devices next-gen platforms require.

⇒ *Implication for next-gen platforms*



Metadata is now king

The purpose of data governance platforms and tools is to provide technology support throughout the data governance process. One of the key tool-related activities in the data governance process is the acquisition of technical and operational metadata. In traditional platforms, metadata scanners are systematic processes and sometimes vendor-specific, which can require additional effort for the acquisition. Next-generation platforms leverage standard web-based interfaces, like REpresentational State Transfer (REST) APIs, allowing the acquisition of metadata across platforms. The use of standard interfaces reduces the complexity and enables the automation of metadata provisioning. Besides metadata provisioning, metadata discovery automation should be considered as a new feature to respond to the increased volume and variety of metadata assets that need to be captured.



Data-quality tools require real-time processing

Data governance and data-quality tools must be able to integrate with next-generation platforms. Additional features required to respond to the scale and complexity of next-generation platforms include those that can perform data-quality assessments for unstructured data or real-time processing of data, as well as provide feedback to execution algorithms to remediate identified data-quality issues as they are running in production.



Data security tools? Take it up a level

Data security and privacy are increasingly important in distributed environments and often become key impediments for the migration of business-critical data from on-premise environments to cloud environments. For example, common data security threats in cloud environments are related to data breaches, insecure data interfaces, inadequate identity and access management, system vulnerabilities, and data loss. Furthermore, distributed environments that heavily rely on microservices are often vulnerable to breaches around access control and authorization handling. As a result, some industry standard protocols must be used to help ensure data security.



Pillar 4

Metadata content (or data catalog)

What does metadata look like tomorrow?

Metadata content (or data catalog) is about cataloging the business, technical, and operational characteristics of data. Metadata content is instrumental in managing, overseeing, and measuring the data governance processes of tomorrow. Existing metadata tools are more oriented to technical metadata and support a limited variety of metadata assets and their mutual associations, which should change.

⇒ *Implication for next-gen platforms*



Big data needs a variety of metadata assets

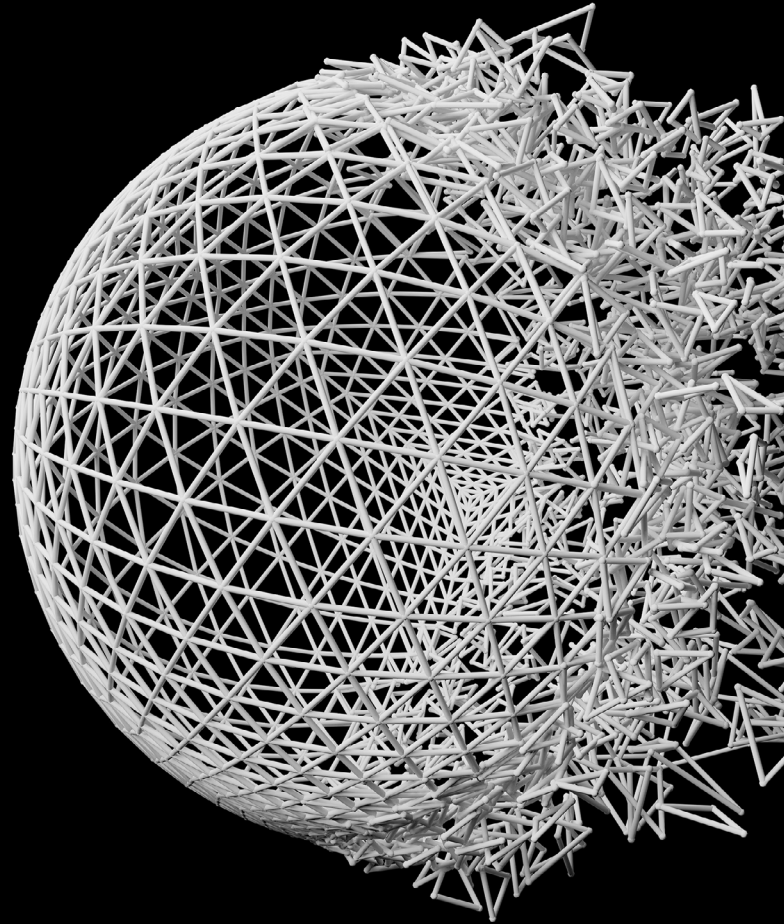
The increased variety of technical metadata assets that can be found in next-generation platforms leads to greater complexity of metadata collection, cataloging, and discovery processes and often requires advanced techniques and approaches. For instance, statistical techniques, such as the Dice coefficient or Levenshtein distance, can be used to discover the similarity between metadata artifacts in the metadata repository and still discover hidden associations.



Get common with business semantics

Business semantics and ontologies become increasingly important in next-generation platforms where data and components are defined and described with different notations. Ontology triples are standards that can be used to develop a common language and reference model to address a variety of data architecture components and their mutual associations. For example, Financial Industry Business Ontology (FIBO), designed by the Enterprise Data Management Council, can be used as a reference ontology for the financial services industry.

The evolutionary technology shift from traditional to next-generation platforms has an impact on each of these pillars—especially the technology capabilities and metadata content pillars, which are more affected due to their technical dependency. In addition, this evolutionary technology shift does not mean that traditional platforms will completely go away and be replaced by next-generation platforms. Typically, organizations will have an enterprise data ecosystem composed of both traditional and next-generation platforms that should be managed cohesively. This new reality requires enhancements across all key data governance pillars.



Want the max out of your data?

Data governance for next-generation platforms presents challenges related to the expanded scale and complexity of enterprise architecture. Many companies have a good base for data governance, but it needs to expand significantly to handle the new data paradigm. Existing data capabilities, layered on top of traditional platforms, can be a solid foundation to start addressing this new form of data governance.

Enhancements in each of the four pillars are required to respond to the new needs. These enhancements encompass process automation; disruptive technologies such as cognitive analytics and AI; business ontologies as a reference representation model; and data catalogs and other functional improvements in metadata, data quality, and data security tools. All should address required aspects of next-generation platforms.

A solid data governance program helps companies to understand, manage, and leverage the information potential of the data stored and processed by next-generation platforms. Successful implementation of data governance across the entire data landscape—with traditional and next-generation platforms—will likely determine which organizations will emerge as market leaders. They will be the ones making data-driven decisions.

Endnotes

1. Nitin Mittal, Sandeep Kumar Sharma, Ashish Verma, and Dan Frank, "Enterprise data sovereignty: If you love your data, set it free," *Deloitte Insights*, December 5, 2017, <https://www2.deloitte.com/insights/us/en/focus/tech-trends/2018/data-sovereignty-management.html>.
2. Ibid.
3. Ibid.
4. Ibid.
5. David Reinsel, John Gantz, and John Rydning, Data age 2025: *The evolution of data to life-critical*, International Data Corporation white paper, April 2017, <https://www.seagate.com/www-content/our-story/trends/files/Seagate-WP-DataAge2025-March-2017.pdf>.

Contributors

Omer Sohail

DC Principal
Deloitte Consulting LLP
osohail@deloitte.com

Prakul Sharma

DC Senior Manager
Deloitte Consulting LLP
praksharma@deloitte.com

Bojan Ciric

DC Specialist Master
Deloitte Consulting LLP
bciric@deloitte.com



About Deloitte

Deloitte refers to one or more of Deloitte Touche Tohmatsu Limited, a UK private company limited by guarantee ("DTTL"), its network of member firms, and their related entities. DTTL and each of its member firms are legally separate and independent entities. DTTL (also referred to as "Deloitte Global") does not provide services to clients. Please see www.deloitte.com/about for a more detailed description of DTTL and its member firms.

This communication contains general information only, and none of Deloitte Touche Tohmatsu Limited, its member firms, or their related entities (collectively, the "Deloitte Network") is, by means of this communication, rendering professional advice or services. Before making any decision or taking any action that may affect your finances or your business, you should consult a qualified professional adviser. No entity in the Deloitte Network shall be responsible for any loss whatsoever sustained by any person who relies on this communication.