database
TRENDS AND APPLICATIONS
ONE COMPLETE MARKETING PROGRAM

DATA GOVERNANCE and SECURITY for the MODERN DATA ECOSYSTEM

database
TRENDS AND APPLICATIONS

Best Practices Series

DATA GOVERNANCE and SECURITY for the MODERN DATA ECOSYSTEM

Best Practices Series

Data governance and security were once something few, beyond the data or IT department, worried about. However, in today's data-intensive business landscape, it needs to be everyone's job across the enterprise. Still, data managers and professionals need to provide the guidance and expertise required to assure that their organizations are getting the most out of their data resources.

Achieving successful data governance and security for modern data ecosystems requires looking at the problem from several different directions, encompassing established data management, cloud platforms, partner, and customer data-sharing. Data governance and security have been transforming in recent years in an effort to meet regulations and mandates along with the increase in security attack surfaces, the sophistication of analytics, and the drive to democratize data and analytics.

Efforts to assure the quality and viability of data are on the rise, a recent survey by Unisphere Research, a division of Information Today, Inc., has found. A majority of respondents, 58%, report having a data quality strategy, up 5 percentage points from the previous survey. In addition, there has been a significant rise in organizations recognizing the importance of data quality to their data strategies—56% consider this to be a top priority, up from 50% in the previous year's survey.

The digital era now upon us calls for new approaches to data governance and security.

The survey has also found some measure of greater confidence in data integrity—33% now express "complete" confidence, up from 30% in the previous survey. While there's greater confidence in the quality of enterprise data, managers admit it's more of an issue than it was in the previous survey. More than one-third, 35%, see it as a constant issue that needs to be dealt with, up 9 percentage points from a year ago (Information Today, Inc./Unisphere Research, "Data Quality Challenges and Strategies for the 2020s," November 2022).

The digital era now upon us calls for new approaches to data governance and security. The following are best practices to design and develop data governance and security for a modern data ecosystem:

- **Don't outsource governance and security**. For many data managers,

There are many layers to information, with some being of critical material importance to businesses, especially with analysis that puts it into relevant context.

cloud provides opportunities to offload many of the rote or onerous tasks associated with data management, provisioning, and even security. However, data quality and integrity must, by necessity, remain within the purview of enterprise customers. If anything, moving to the cloud has increased the issues associated with data quality and integrity. The complexities introduced by cloud to manage data quality have reportedly increased, the Unisphere Research survey has found. Forty-six percent report increased challenges to their data quality efforts, up from 42% in the previous survey.

- **Develop an active metadata layer.** As high-end data analytics comes into play, this necessitates an active metadata layer, which provides a comprehensive view of the data assets that exist across the enterprise and beyond. Metadata, or data about data, is a resource that often gets overlooked. It provides an active accounting of data lineage and provides a picture of what data resources the enterprise may leverage or lack. It enables the monitoring of data resources, including changes that are taking place, as well as enhances data observability. An active metadata layer spotlights what datasets may be hidden away in parts of the enterprise and data that is underused or unused. It proactively alerts or notifies data teams or business users about changes taking place within enterprise data.
- **Establish and understand data lineage.** An active metadata resource, powered through high-end analytics, sets the scene for fully understanding and visualizing data lineage. With an understanding of data lineage, data teams and business users will know the sources of their data assets, the consumers of the data, and what happens to the data as it moves from source to consumer—the entire data lifecycle. It establishes the trustworthiness of the data—an important factor in gaining the trust of its ultimate consumers.
- **Establish a framework for flexible and responsive governance and security.** Previous data governance approaches were intended for information within corporate walls and based on limited data sources. Such governance structures strain as data sources explode, reliance on cloud platforms expands, and the complexity of data environments increases. A viable data governance and security strategy should include policies and procedures for acquiring, managing, analyzing, and storing data assets. Such a framework needs to be responsive to all relevant compliance requirements.
- **Prioritize the importance of data to the business.** There are many layers to information, with some being of critical material importance to businesses, especially with analysis that puts it into relevant context. Business leaders and analysts need to be fully engaged at every stage of the planning process for data governance and security efforts.
- **Keep employees and partners up-to-date with governance and security developments, and provide training and education.** A viable data governance and security strategy will only work if end users find data resources easy to access and employ in their day-to-day work. To facilitate ease of use, employees and other end users should receive training to understand data lineage, management, and security best practices.
- **Audit early and often—and put data observability in place.** Most organizations employ audits to track data usage and anomalies, but often only catch breaches or events well after the fact. In an era where real-time information is shaping business decisions and automated actions, it's critical to be aware and be able to address incidents, preferably before they cause major problems. Data observability, an emerging best practice, provides this capability as well as assures the trustworthiness of data.
- **Keep up with data security developments and offerings.** There are a range of strategies and technologies that can be employed to provide guardrails for the data moving in and outside of the organization, from data encryption to the use of synthetic data tied to well-managed backup and recovery strategies.

The past few years have seen the start of an explosion in data volumes and types, accentuated by data lakes, low-cost storage available through cloud services, and high-density disks, which provide immense opportunities, as well as potential risks, to organizations. In an era in which organizations rely on having the right data at the right time, a comprehensive governance and security framework and initiative will deliver competitive advantage. This not only allows for data quality and viability to parts of the business that need it, but also opens its availability for next-generation applications. ■

—Joe McKendrick

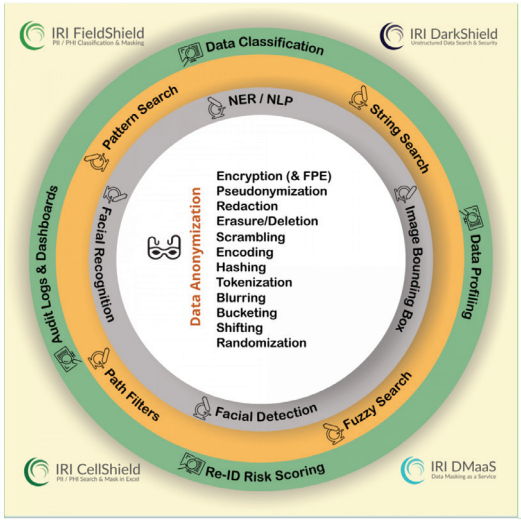
Implementing Modern Data Security Governance

IRI Voracity
An Insatiable Appetite for Data

Gartner defines data security governance (DSG) as the "*subset of information governance that deals specifically with protecting corporate data (in both structured database and unstructured file-based forms) through defined data policies and processes.*"

You define the policies and the processes; there is no one-size-fits-all solution to DSG. There is also no one product that meets all of the needs of DSG. You must look at your data and weigh which areas have the greatest importance to your organization. You take data governance into your own hands to avert disaster: information is your responsibility.

While there are multiple pathways to safeguarding data—logical, physical, and human—three primary software methods that IRI customers successfully employ are the classification, discovery, and de-identification (masking) of personally identifiable information (PII) and other data considered sensitive.



DATA CLASSIFICATION

To find and protect specific data at risk, it should first be defined in named categories or groups. Data so classified can be cataloged not only by its name and attributes (e.g., US SSN, 9 numbers), but through computational validation (to distinguish it from other 9-digit strings), and sensitivity attribution .

In addition to those assignments, data classes or groups can be characterized by where they are located and/or how they should be found (search method/s) if their locations are unknown. Also possible is the global assignment of a remediation, or masking function, so that de-identification can be carried out consistently for all members of the class, regardless of location, preserving its referential integrity.

DATA DISCOVERY

To find sensitive data, search operations that may or may not be associated with data classes can be performed. Examples of IRI data discovery techniques include RegEx or Perl Compatible Regular Expression (PCRE) searches, exact and fuzzy matching algorithms to values in a lookup file, source-specific column or path filtering logic, named entity recognition (NER), and facial detection.

It is also possible to leverage machine learning in the recognition process. IRI supports semi-supervised machine learning in NER model building, for example, in its DarkShield product.

DATA DE-IDENTIFICATION

One of the ways to reduce, and even nullify, the risk of data breaches is by masking data at rest or in motion, with field-level functions that render it protected but still usable to some extent.

According to Gartner analyst Marc Meunier in, "How Data Masking Is Evolving to Protect Data from Insiders and Outsiders:"

Adopting data masking helps organizations raise the level of security and privacy assurance for their sensitive data—be it protected health information (PHI), personally identifiable information (PII) or intellectual property (IP). At the same time, data masking helps meet compliance requirements with security and privacy standards and regulations.

Most enterprises—either by virtue of internal rules or data privacy laws—have been, are now, or will soon be, making data masking a core element of their overall security strategy.

PROVEN SOFTWARE SOLUTIONS

IRI provides static and dynamic data masking solutions for structured databases, flat files, proprietary mainframe and legacy application sources, and big data platforms (Hadoop, cloud stores, etc.) in its FieldShield product or Voracity platform, as well as data at risk in Excel via CellShield.

For data in semi-and unstructured sources like NoSQL DBs, free-form text files and application logs, MS Office and PDF documents, plus image files (and faces), you can use IRI DarkShield to classify, discover, and de-identify it.

These 'shield' products use data masking functions like blurring, deletion, encryption, redaction, pseudonymization, hashing, and tokenization—where certain functions can be reversed to restore original data. Voracity (which includes all three products), can also fold data masking into ETL, data cleansing, migration, replication, federation (virtualization), reporting, and wrangling for analytics.

Built-in data discovery, classification, metadata management, risk scoring, and auditing features facilitate automatic and external assessments of the re-identifiability of affected records.

See: www.iri.com/solutions/data-masking and www.iri.com/solutions/data-governance for more information, and contact info@iri.com for help creating or enforcing your DSG framework through sensitive data classification, discovery, and de-identification. ■

IRI, THE COSORT COMPANY
www.iri.com/voracity