THOUGHT LEADERSHIP SERIES

MEETING THE **GROWING CHALLENGES** OF **DATA SECURITY** AND **GOVERNANCE**

database
TRENDS AND APPLICATIONS

2022 | APR

HOW DATA SECURITY AND GOVERNANCE EXPAND DATA'S VALUE

With cyberattacks, ransomware, and the potential for high costs in terms of lost customer trust and regulatory penalties, data security and governance rank high on the priority list for today's data-driven organizations.

However, despite greater awareness of threats, protecting data has not become easier in recent years. At the average enterprise, data is distributed across an array of on-prem data centers, cloud databases and applications, and smart devices. The result is a risk environment that is growing steadily in both size and complexity. So, while the goals of data security and governance remain clear, the challenge is that maintaining effective processes to optimize how information is accessed and used—while also improving regulatory compliance and minimizing risk—is anything but simple.

DATA RISKS

In 2021, the overall number of data compromises was up more than 68% compared to 2020, according to the "2021 Annual Data Breach Report" issued by the Identity Theft Resource Center (ITRC). And, on top of that, research conducted annually on the cost of a data breach found that data breaches now cost surveyed companies $4.24 million per incident on average—the highest cost in the 17-year history of the survey.

In particular, the number of fines issued for violations of the EU's GDPR, one of the most widely known data-handling regulations, was up substantially in 2021. According to law firm DLA Piper's annual "General Data Protection Regulation (GDPR) Fines and Data Breach Survey," EU data protection authorities issued a total of almost €1.1 billion ($1.2/£0.9 billion) in fines in 2021 representing nearly a seven-fold increase over the previous year's total.

EMERGING THREATS

Companies whose data is not protected adequately run the risk of having their businesses harmed irreparably. According to the NCC Group's November 2021 "Threat Pulse" report, there has been a marked increase in the use of "double extortion," where bad actors not only steal a company's data, but then increase the pressure by threatening to publish it. There has also been a notable increase in companies being targeted by PYSA (protect your system amigo) ransomware, a trend that is also affecting government sector victims.

Healthcare organizations, insurance companies, and financial services vendors are among the companies whose treasure troves of PII data make them particularly high-profile targets for cyberattacks and fraud, notes a 2022 report on the insurance industry cyberthreat landscape.

MORE DATA, MORE CLOUDS, MORE COMPLEXITY

Amidst this trend toward greater risk, data is being created at a rapid clip and there are more ways than ever before to store it. According to Statista, the total amount of data created, captured, copied, and consumed globally is forecast is projected to grow to more than 180 zettabytes in 2025, up from 64.2 zettabytes in 2020.

A recent study by Unisphere Research, a division of Information Today, found that many organizations see cloud as the way forward. The survey report authored by lead analyst Joe McKendrick explored the strategies or tactics being employed to reduce the cost of database maintenance, thus freeing up funds for new types of initiatives. Moving to virtualization or cloud-based solutions tops the list of approaches, cited by 66%.

With rapid data growth and the use of more cloud platforms by global organizations, the data governance and security challenges are also getting greater.

SECURITY AND GOVERNANCE PRIORITIES

Despite the complexity, experts suggest that strong data governance and security practices can actually support organizations' data analytics goals. Security and governance have become one the critical requirements of every client's request for data strategy and architecture in recent years, observed John O'Brien, principal advisor and CEO of the research and advisory firm Radiant Advisors. "Companies want to have a data architecture that empowers their people to work with data in a secure and proper way."

A common mistake companies make is to say they'll take care of data security and governance after they get the data and analytics delivered, pointed out O'Brien. "In reality, there's always another data project that is more urgent, and retrofitting security and governance into the solution is difficult because it can break or reverse aspects of projects already delivered. How I handle this is to separate technical security and governance from business-oriented aspects." For example, he said, it is possible to inherit security controls from the data source applications that have already been established and tag all data for governance and data catalogs as accounted for, but not reviewed or defined.

Data is rushing through today's enterprises as if "in a firehose," and identifying and capturing the information that is of most material importance to the business can be challenging, added Unisphere's McKendrick. "Solutions such as data warehouses and lakes have attempted to archive or tame some of this data, but it's difficult to manage the torrent. At the same time, while data managers may be doing their due diligence to secure data, it is often duplicated and sent to backup sites, clouds, and development teams, where security is not always assured."

In work with customers, O'Brien said he incorporates steps to "architect for SLAs" that include security aspects in the physical implementation of modern data architectures as modern data infrastructure. One of his most popular workshops focuses on architecting for secure data, in which he explores patterns for securely transporting secure data, processing it into data center secure zones, utilizing public key encryption technologies, storing data encrypted at rest, and delivering data via a secure framework. "An overarching best practice that I incorporate is how to interweave this into an overall data architecture that still allows for self-service data exploration and development," he noted.

Understanding today's data security and governance problems is the first step to solving them. Here are 10 key tenets for improving data security and governance in today's complex data environments:

1. **Identify and catalog data:** Knowing where data resides and what data you have is a critical step to protecting the data, understanding the risk, and employing appropriate restrictions on access.
2. **Put the right tools to work:** Passwords, firewalls, encryption, masking, anonymization, role-based permissions are critical, but information—including documents and contracts—can also be altered deliberately or unintentionally by authorized users. Blockchain can help to improve the security profile of an organization.
3. **Don't pass the buck on security in the cloud:** "Organizations have to recognize that data security is part of their domain, and cannot be outsourced to cloud or third-party providers," McKendrick noted. "In addition, enterprises need to get more aggressive about educating and training users to be security-aware. Employing the latest tools, such as encryption and automation, is also important. But security is now left too much to the IT department alone."
4. **Leverage DataOps to enhance data engineering:** The DataOps methodology helps to address data quality and governance issues by routinizing and operationalizing data processes and access. Data can be limited by type, user, geography, department, and other criteria to ensure only data that is needed for a job/role/level/partnership is accessible. This helps organizations adhere to individual company and partner requirements; industry mandates such as PCI-DSS, SOC 2, and HIPAA; and domestic and regional data-handling requirements such as CCPA and GDPR. Having a comprehensive view of data pipelines and processes to manage data lineage and access is an essential element of data security and governance.
5. **Take advantage of automation to enable data sharing in a cloud environment:** If data is locked away securely, it's true that it is protected but, in today's data-driven environment, data must be analyzed, shared, and explored. According to a recent survey, 75% of respondents said that sensitive data is important to analysis. In addition, nearly half said they are actively using sensitive data in analytics. With the importance of sensitive data for analytics, and the increasing use of multiple platforms—not to mention a variety of databases being used on those platforms—data access and governance can become more complex. Automation can help to support cross-platform data sharing and effective data analytics.
6. **Bake in security from the beginning:** Data security and governance are key components of a data strategy and architecture, noted O'Brien. "I strongly emphasize an approach that accounts for both from the very beginning of data ingestion to be the foundation upon which data and analytics are to be built."
7. **Consider a data hub platform that supports governance:** A data hub platform can help an organization bring together data with the critical element of governance, so that the data can be trusted. Key governance elements include a business glossary, lineage, policy definitions and enforcement, and rules and processes.
8. **Uncover shadow IT:** Unauthorized applications can add unknown risk. Organizations must have the necessary visibility to monitor the network and all the apps on it so there are no unexpected surprises, and to support staff members by understanding the needs that are not being met already by organization-sanctioned applications.
9. **Take steps to improve cloud and hybrid architecture incident response:** A 2021 IDC incident response survey found that there are four types of telemetry that are very important to forensic analysis for incident response: cloud applications and workloads; operational technology (OT); netflow, network traffic analysis (NTA); and endpoint detection and response. On average, nearly 60% of respondents believe that cloud workload and application telemetry are very important to cloud incident response.
10. **Evaluate an XDR (extended detection and response) approach:** According to Gartner, the trend continues for security and risk management (SRM) leaders to seek security vendor and product consolidation to help decrease risk and improve security operations productivity and XDR if evolving to provide these benefits. Gartner predicts that by year-end 2027, XDR and secure access service edge (SASE) will be used by up to 50% of end-user organizations to reduce the number of security vendors they have in place, up from less than 5% today.

IMPROVING YOUR DATA GOVERNANCE AND SECURITY POSTURE

A recent survey found that 52% of data teams plan to adopt two or more cloud platforms in the next 12-24 months. With the use of more platforms, more databases on those platforms, and more sophisticated users seeking access to data, data security and governance protocols must be enforced across all data pipelines.

Ultimately, while data governance and security require a combination of expertise, discipline, and technology, it is well worth the effort. Without well-orchestrated procedures, data cannot be trusted and relied upon for decision making. Moreover, a lack of control and oversight may put the whole enterprise at risk. 🟩

—Joyce Wells

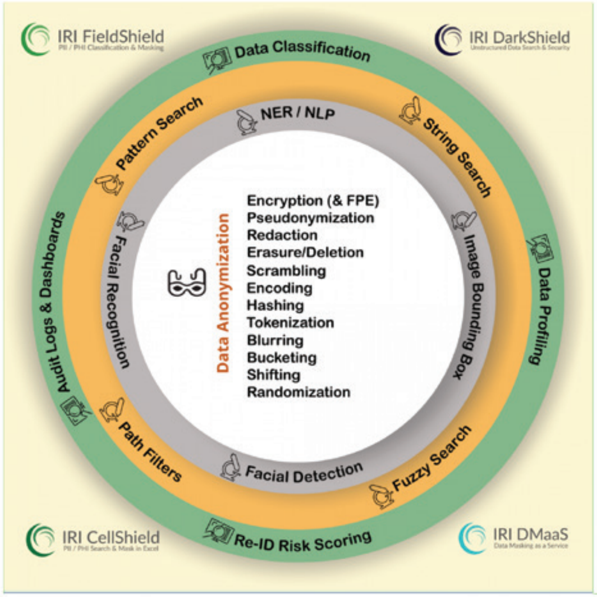
Defining Data Security Governance

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