PKWARE

pkware.com

Metadata Management

What It Is and Why Your Business Needs It

2022

Contents

Introduction

According to the World Economic Forum (WEF), it is expected that by 2025, over 463 exabytes of data will be created globally each day. The WEF also predicts that there is currently over 44 zettabytes of data on the internet, which is more than 40 times the number of stars in the entire observable universe. (Desjardins, 2019) In the business sector, the amount of data that a single organization generates can often be large enough to create an entirely different category of data—metadata.

Metadata is data in the context of who, what, when, where, why, and how. This information helps users identify the content of the larger data, including what it might mean and how it can be used. One easy (though not all-encompassing) way to think about metadata is that it's the data about the data. Metadata curates the crucial information needed for analytics and accurate business insights. Additionally, using metadata to classify, manage, and organize massive amounts of enterprise data makes it easier to understand what the underlying data is to deploy the right resources. It is most often created via cataloging and classification tools.

Businesses find indescribable value from finding practical applications for data that they have. For example, the curation of metadata for analysis can support insight into the efficiencies or effectiveness of a given process, or the profitability and sustainability of a particular product. Aside from performing analysis to define metrics such as risk or performance indicators, metadata is also valuable for its capability to categorize and classify similar data.

Data by itself—regardless of the quantity—is just data. But once that data is put into context, it adds immense value, and as with all valuable things, risk. A string of nine unrelated numbers on its own most likely can't do a lot of damage. However, once those numbers are identified as a Social Security number or a phone number, it may become sensitive and be classified as Personal Data or Personal Information and is therefore within the scope of state or country data protection legislation.

Any organization that collects, handles, processes, and/or stores sensitive information may be obligated to protect it by law or regulations. Across the United States, individual states such as California, Virginia, and Colorado have begun crafting and passing data protection and compliance laws. Some countries or unions of countries—Canada, Brazil, South Africa, China, the European Union, to name a few—have nationwide laws that protect citizen data wherever it resides around the globe. Challenges associated with governing and protecting both data and its metadata will continue to be a focus

of organizations as they are continually modernizing data governance programs to keep pace with emerging technologies.

As the amount of data present within the organization grows, it is important to understand how to meaningfully define and align to the lifecycle the organization has defined for it. Metadata also plays an important role in how an organization manages, monitors, creates, protects, and destroys data—the data lifecycle. The amount of metadata generated by an organization also has to have a lifecycle to govern its use.

The Importance of The Data Governance Foundation

Data governance is how an organization manages specific aspects of data such as availability, usability, integrity, and security. Done well, data governance can produce a comprehensive and centralized approach to enabling teams to define and document data and its requirements. Organizations use data governance to ensure that data is consistent, trustworthy, and used appropriately. But data governance is not just a technology play: It must be a shared responsibility across the organization.

One major consideration when establishing any data governance program is removing the perception that data governance will inhibit any ability of the business. This perception may result in a lack of stakeholder buy in or funding of the data governance program, and eventually even breed resentment. It is important that security teams implement thoughtful controls to manage data in a way that both enables the business and creates awareness of the importance of data governance. Each type of data set should be protected according to its individual needs. Data should exist under a well-balanced umbrella of protection and usability. For example,

data can remain usable for analysis while also being protected.

Data governance and security governance must be able to converge and work together to form data security governance (DSG), a more balanced approach to implementing security through data protection and privacy. Many of the activities performed in both data security and privacy have synergies, which actually promote efficient and effective program management. As cross-functional teams further collaborate, it improves governance across the organization, furthering the business' capabilities to manage the data lifecycle. When teams can work together, the business will also realize financial benefits as efficiencies drive a greater level of maturity within data governance—without an uplift in resources.

Understanding Data with Metadata Governance

A core part of data governance is understanding the data that you are protecting. This understanding comes from metadata that is collected by the organization and used to develop data sets that can be analyzed by the business. Metadata can be used to help organizations categorize or catalogue data into classifications—which may also determine what specific controls must be used to protect it. Metadata may also contain information that contains Personal Information or sensitive data or may be conjoined to create sensitive data when aggregated. Which means then that metadata must also be included in data governance plans—and protected.

Metadata sets the foundation for a business' ability to utilize the data it has in a meaningful way. All business initiatives can be supported using the results of data analysis to fuel intelligent decisions for both technical and business projects. It also makes larger data sets more manageable by reducing the amount of information needed to process the data. Another use could involve extracting key components of a data set to reduce the amount of information that has to be sifted through for analysis, thus creating a lower level of effort or technical debt to perform the task.

Due to the rise in cyber-attacks, data breaches, legislation, and increased risk awareness at the board level, data governance has become top of mind for many organizations. This has resulted in the need for many organizations to confirm they have adequate data protection mechanisms in place to support the privacy and security of their data. Typically, these

controls are based—fully or in part—on regulations or compliance standards the business is required to adhere to, such as GDPR, CCPA, PCI DSS, and HIPAA.

One common strategy organizations use to help reduce the risk present with sensitive and private data is to extract only the non-sensitive data and metadata from their environment when performing analysis. In the event there are compliance regulations surrounding a specific data set, remediation options such as masking, redaction, and encryption may also sufficiently remove data from the audit scope while still rendering data usable for analysis.

Acceptable Use and Data Governance policies help an organization ensure data stays protected while the company is responsible for protecting it. This is typically defined by contractual obligations, compliance regulations, and internal data governance policies. These protections traditionally must confirm the data is protected throughout the environment, both at rest and while in transit.

As with all governance programs, it is important that the key consideration is enabling the business to meet its goals and accomplish its objectives. Data governance is not only looking at the security of the data, but also the operational efficiency that can be gained through any insights the data provides.

Successful metadata governance begins with a comprehensive data security governance framework that incorporates both people and processes, encourages open communication, and fosters an organization-wide data-centric culture.

Identifying Necessary Security Controls for Data

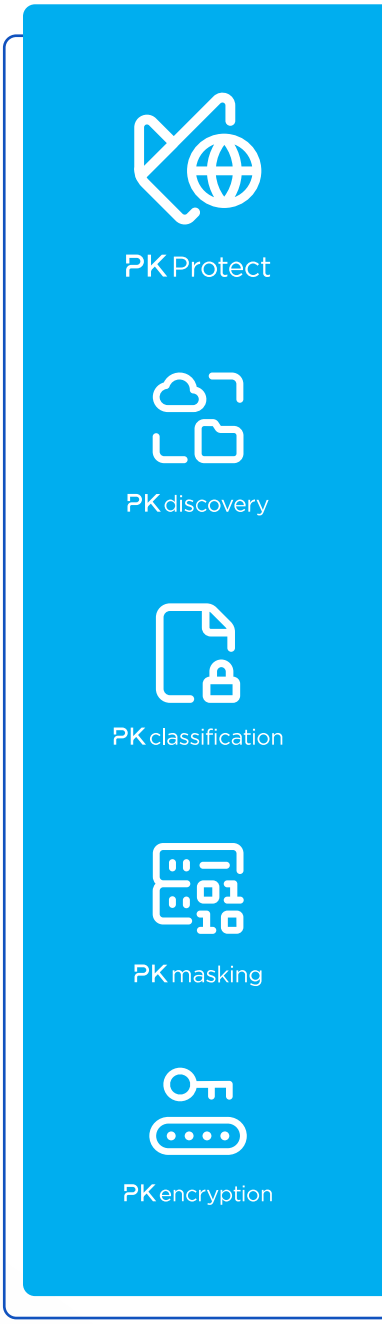
1. **Prevent:** Control begins with knowing what data and metadata the organization has where and who can access it in what state. Uncover where all sensitive information resides and apply the appropriate protection—masking, redaction, encryption, access controls, etc.—to avoid data access by audiences that should not have it.
2. **Detect:** Once data access privileges are mapped, they should be continuously monitored to ensure the right people have access to the right data—and no one has any access they shouldn't. Keep an updated account of how business application users and admins are granted access to datasets.

3. **Respond:** Any inconsistencies or changes to data access or activity must be monitored and addressed. It's also important to understand how data flows across geographic jurisdictions—for instance, to the EU or across a US state border into California—that may dictate compliance needs and result in necessary policy changes.
4. **Predict:** Gaps or inconsistencies in governance policies can leave information vulnerable. Even though they may be a colleague, someone that should not have access to particular data sets running across them due to policy gaps still presents great risk to the company and the sensitive or private information. Continually monitor data access and keep audit logs.
5. **Extend:** Data and metadata governance does not just pertain to the four walls of the organization. All data and metadata governance policies must also extend to any third parties that access and/or process data from your business.

Supporting the Metadata Governance Framework with PKWARE

PKWARE has worked at length to help organizations support full breadth data governance across the enterprise. Especially when it comes to the vast array of data—and data about data—that exists within an organization, protection starts with location. Businesses must know precisely where all sensitive and private data is in order to apply their policies consistently across the entire scope of the enterprise, and uniformly apply the organizations protection standards.

PK Discovery, a foundational component of the PK Protect suite, is purpose-built to detect sensitive and private data on structured, unstructured, and cloud, then perform protection on everything it finds based on assigned policies. Querying for sensitive data location and applying scores based on how much sensitive data is found within the identified systems lets organizations run risk profiling within data catalogs. PK Discovery's detailed cross-platform and device searches do exactly that.



By labeling data, classification plays a role both in providing additional access control based on the organization's definitions and in enabling enforcement. While key for data at large, this is especially important regarding metadata, as organizations can know whether something is benign, or a risk factor based on visual tags and underlying file metadata. With PK Classification, organizations can automatically apply labels to identify and protect sensitive information. PK Classification gives administrators a powerful, easy-to-use policy builder that is capable of categorizing data based on specific needs. Persistent classification provides valuable data insight that can be integrated with an organization's access control system.

Finally, PK Masking and PK Encryption empower organizations to protect according to policy by automatically applying remediation to risky data. Masking and redaction allow the protection of sensitive data while keeping business data widely accessible and usable in the organization with minimal risk. PK Encryption can enforce persistent encryption for hundreds of file types, on databases, and element level strings in files to ensure sensitive information stays encrypted no matter where it is stored, shared, or copied. Data remediation is vital to cleansing, organizing, and migrating data so that it is properly protected and best serves its intended purpose. Remediating sensitive data also reduces the organization's sensitive data footprint and decreases the risk of a potential data breach or leak.

Inconsistent data and metadata governance practices can prompt security leaders to resort to a command-and-control approach. Data may be governed chaotically between departments, with marketing data stored, cataloged, and used quite differently from sales data. While this might work for data at the department level, it falls apart at the enterprise level and could carry the additional stress of leaving the organization unnecessarily vulnerable to risk. Adding solutions such as PK Protect's PK Discovery,

PK Classification, PK Masking, and PK Encryption to the data security governance stack gives organizations the insight they need to assign the necessary level of attention to sensitive data and metadata based on the potential impact.

Data is only valuable when it is used, and data should only be governed if it has value. PK Protect can establish confidence within your governance processes that all valuable data has been identified and protected according to both an organization's own policies and applicable regulations, as well as ensure that it remains meaningful enough to support decision making and analytics for the business.

About PKWARE

PKWARE offers the only data discovery and protection solution that locates and secures sensitive data to minimize organizational risks and costs, regardless of device or environment. Our ultra-efficient, scalable software is simple to use on a broad range of data types and repositories, enabling precise, automated visibility and control of personal data, even in the fastest-moving, most complex IT environments. With more than 1,200 customers, including many of the world's largest financial institutions, retailers, healthcare organizations, and government agencies, PKWARE continues to innovate as an award-winning global leader in data discovery, security, and compliance. To learn more, visit PKWARE.com.

Enterprise-Wide Policy Management

PK Protect provides a point of control for data protection activity across the entire organization

Simple Workflow

With PKWARE, data protection is automated for end users and easy for administrators to manage

Easy Implementation

PKWARE supports a variety of deployment options, enabling organizations to implement their data protection solution without time-consuming changes to infrastructure and workflows

Protection Without Gaps

PKWARE works on every enterprise operations system and provides persistent protection that remains with data even if it's copied or shared outside organizations

Integrated Discovery, Classification and Protection

No other solution has the capability to find, classify, and protect data in a single automated workflow

Multiple Protection and Remediation Options

Organizations can take a policy-based approach to data protection and choose from action including persistent encryption, quarantine, masking, and deletion.

BOOK A DEMO

PKWARE

PKWARE.com

866-583-1795

201 E. Pittsburgh Ave.
Suite 400
Milwaukee, WI 53204