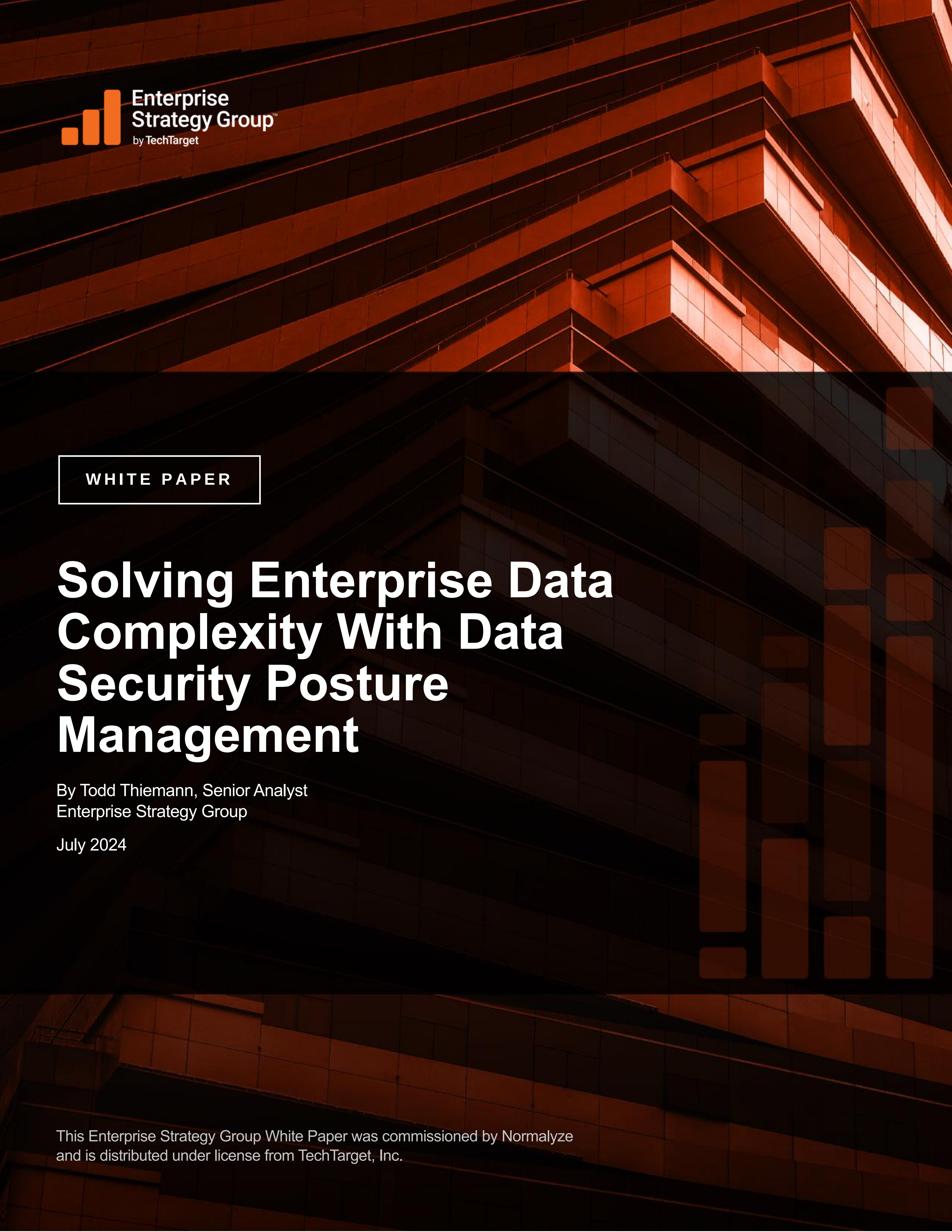


WHITE PAPER

Solving Enterprise Data Complexity With Data Security Posture Management

By Todd Thiemann, Senior Analyst
Enterprise Strategy Group

July 2024

This Enterprise Strategy Group White Paper was commissioned by Normalyze and is distributed under license from TechTarget, Inc.

Contents

Executive Summary

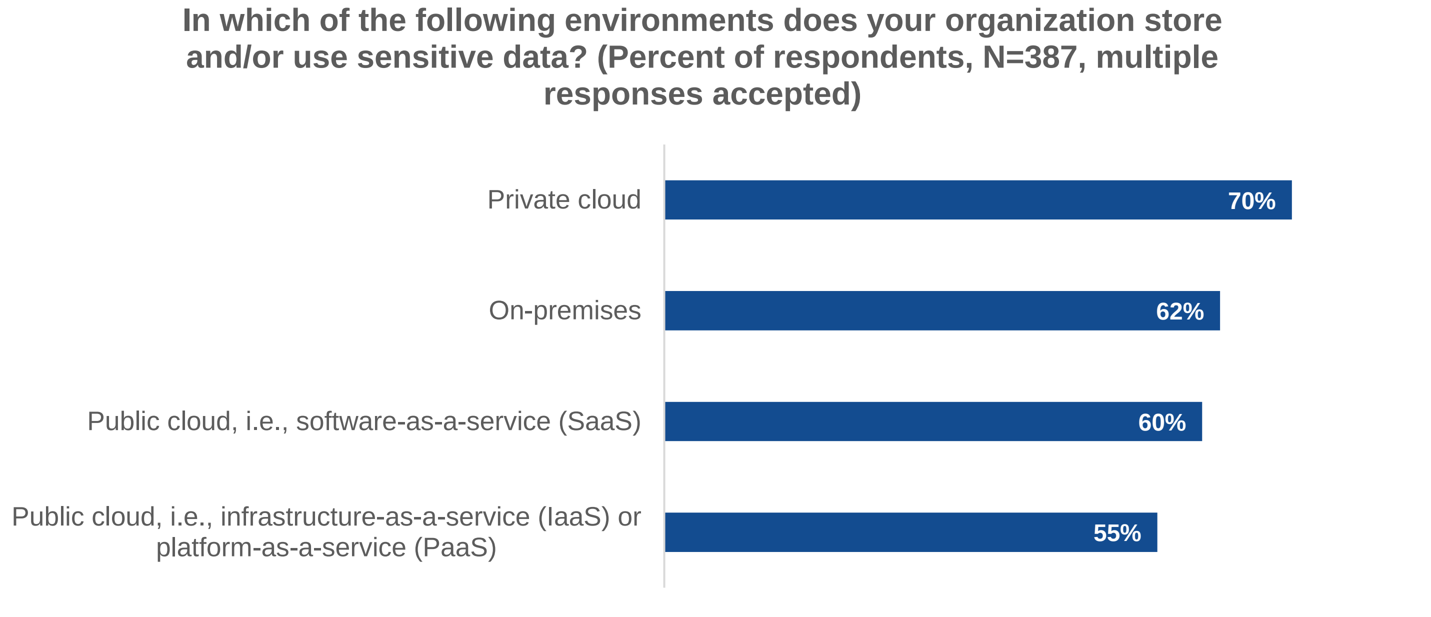
Enterprise data stores are rapidly growing, with valuable and sensitive data spread across on-premises, SaaS, PaaS, multi-cloud, and hybrid environments. Enterprise security data teams strive to control risk but frequently lack answers to these questions: Where is my data? Which data stores contain valuable or sensitive data? Who/what has access to those? How is this data accessed/accessible? How valuable is this data to hackers? Data security posture management (DSPM) has emerged as a modern solution for getting those questions answered—continuously and with the power of AI and automation.

Data generally and sensitive data in particular originates from across the enterprise, from on-premises repositories to the cloud, and that data ebbs, flows, and changes over time. This paper delves into the transformative potential of DSPM solutions, offering insights into how they can empower organizations to safeguard their most valuable asset—data. By exploring key trends, challenges, and innovative strategies, this report underscores the essential role of DSPM in the modern cybersecurity landscape, providing a comprehensive guide for businesses striving to enhance their data security posture.

Enterprises Don't Know Enough About Their Valuable and Sensitive Data

Valuable and sensitive data is continuously being created and stored in hundreds of environments across the enterprise, and both data and security teams need to understand where their valuable and sensitive data live in order to secure it and maintain compliance with external regulatory internal data governance requirements. Enterprises store sensitive data in a variety of locations, as shown in Figure 1, with private clouds and on-premises data stores typically holding the most sensitive data.[1]

Figure 1. Sensitive Data Lives in All IT Environments

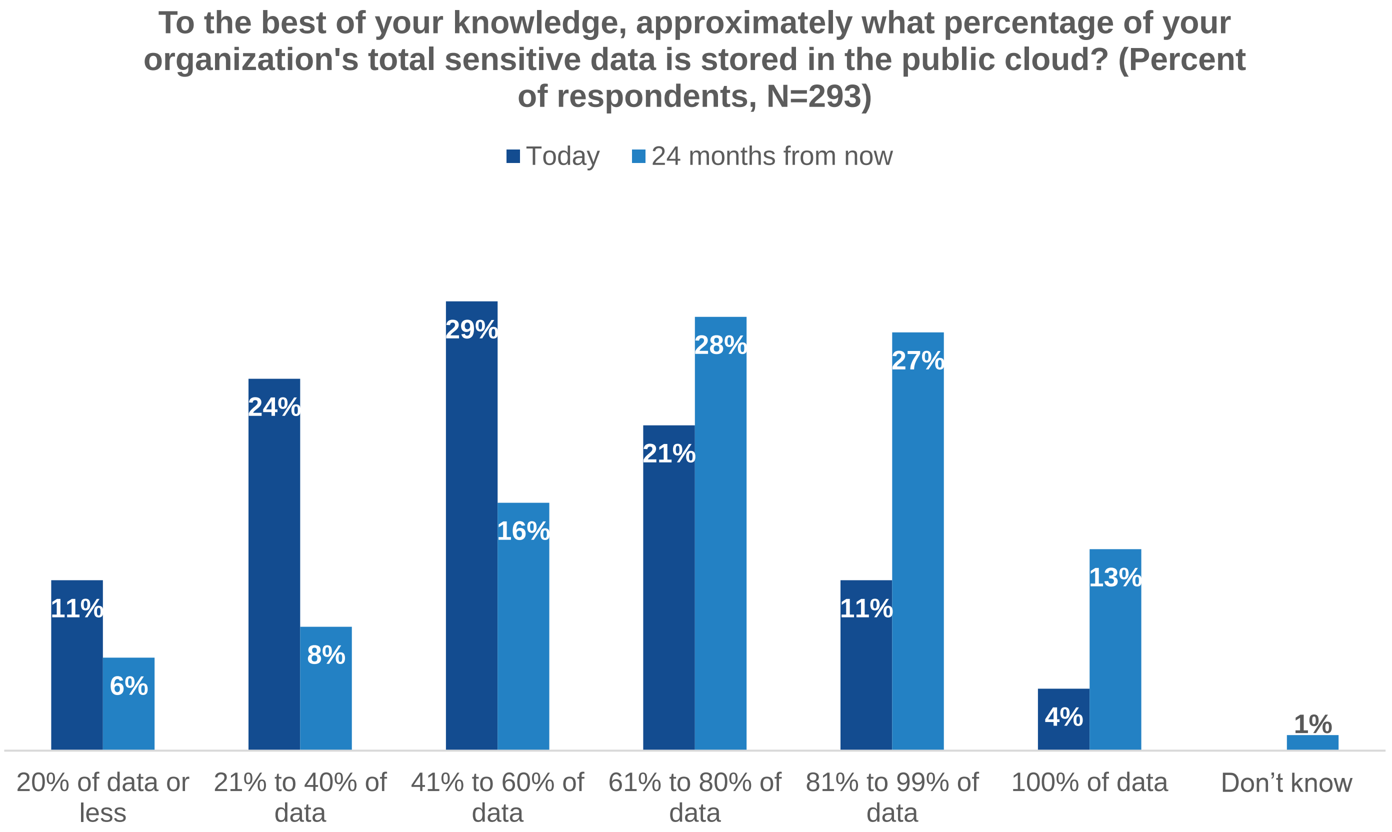


Source: Enterprise Strategy Group, a division of TechTarget, Inc.

[1] Source: Enterprise Strategy Group Complete Survey Results, *Operationalizing Encryption and Key Management*, February 2024.

While all data stores typically grow, the growth of cloud and SaaS data stores has been particularly fast. Digital transformation initiatives, business intelligence, data analytics, and the rapid explosion in the use of generative AI have accelerated the migration of sensitive data assets to the cloud. Specifically, Figure 2 reveals that 36% of respondents to a recent survey by TechTarget's Enterprise Strategy Group said that more than 60% of their organization's sensitive data resides on public cloud services today. This is expected to increase to 68% of organizations within 24 months. More surprisingly, 4% of organizations store all their sensitive data in the cloud, which is expected to more than triple to 13% of organizations within 24 months.[2]

Figure 2. The Amount of Cloud-resident Sensitive Data Is Growing



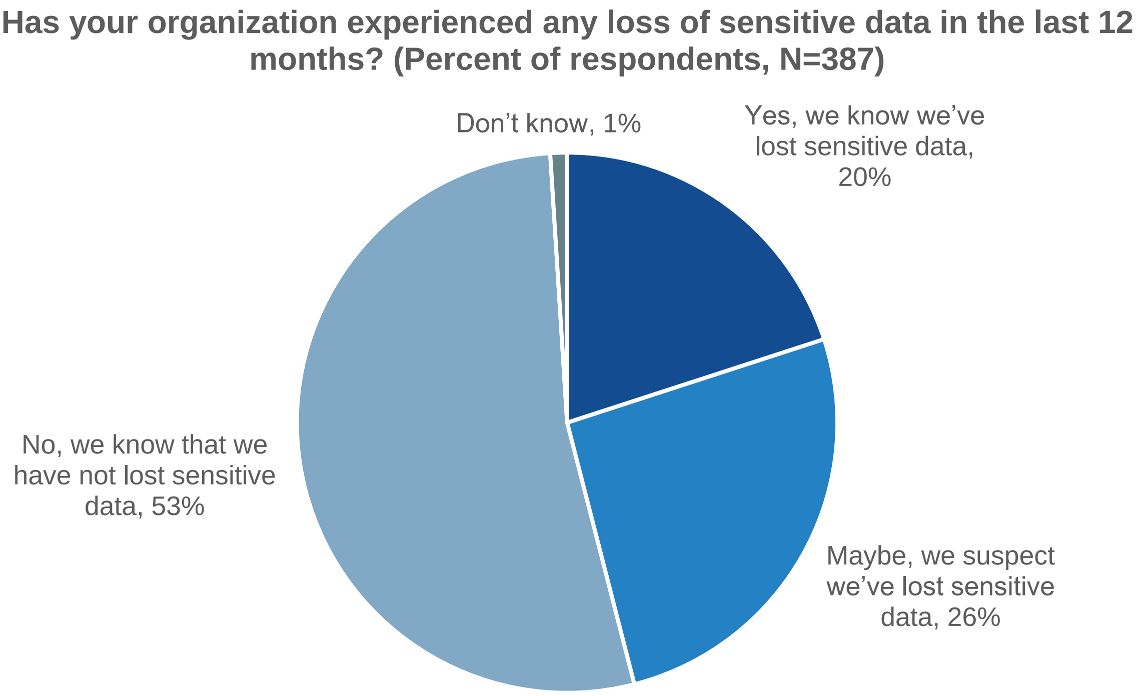
Source: Enterprise Strategy Group, a division of TechTarget, Inc.

While enterprise security teams frequently have visibility into data loss, Enterprise Strategy Group research indicated that 46% of them suspect or are certain they have experienced data loss, while 53% of respondents believe they have not lost data (see Figure 3).[3] In the event of a breach, enterprises need to comply with SEC regulations for material events and could spend considerable energy understanding the magnitude of the breach so they can determine whether it is material or not. Reputational and business risk can increase without adequate visibility into where data stores are located and what they contain. For example, Marriott International in 2018

[2] Ibid.
[3] Ibid.

announced a data breach affecting 500 million guests at its Starwood hotels but subsequently adjusted to that number to 383 million guests in January 2019.[4]

Figure 3. Loss of Sensitive Data Is Common and Suspected



Source: Enterprise Strategy Group, a division of TechTarget, Inc.

Data Security Complexity

Enterprise data is hard to manage given its volume, variety, and velocity. Enterprises have massive data stores from different sources (e.g., customer data, financial records, operational data, product data), and that data is constantly changing. Sensitive data is valuable to an organization as well as to adversaries, which means that it can also be vulnerable. Managing sensitive data is a challenge with multiple use cases. What follows are some of the most typical use cases for data security posture management.

[4] Source: David Sanger, "Marriott Concedes 5 Million Passport Numbers Lost to Hackers Were Not Encrypted," NYTimes.com, January 2019.

Shadow Data

Related to the uncertainty around whether or not valuable or sensitive data has been lost is the topic of "shadow data." IT and information security teams have visibility into most data stores but frequently lack visibility into unknown, hidden, or overlooked copies of sensitive information that exist outside the purview of an organization's IT security measures and data governance policies. Such data is created, stored, or shared without being formally managed or governed by the relevant IT and security teams. This might be from a line of business application development or from business users setting up Dropbox or Box repositories, Google Drive, or Teams folders.

Shadow data can reside across the enterprise in unstructured files, structured databases, cloud storage, or personal devices, often without the IT department's knowledge or control. For example, it can include data copies in test environments, unmanaged backups, abandoned databases, data extracted by insiders, and data leakage through third-party apps.

Data Security Posture Management Typical Use Cases

- **Shadow Data** – Locating data which may contain sensitive information that exists outside the control and oversight of IT and information security.
- **Abandoned Data** – Identifying data no longer in use for which there is no lifecycle plan.
- **Generative AI and LLMs** – Managing the data informing GenAI and LLMs that lacks relevant categorization and may result in a data breach.
- **Controlling Risk and Ensuring Compliance** – Understanding data posture as it changes over time to minimize risk and address regulatory requirements.

Abandoned Data

IT teams usually do backups of data. When it comes time to delete the data, they might delete the source/primary data but can often forget to remove the backups. If there is no lifecycle plan associated with the data, it becomes "abandoned" and an area of risk. There are other ways critical data can be "left behind," including during merger and acquisition events or following layoffs. The creators of that data might have departed an organization, and nobody within the organization knows that the data exists.

Generative AI and Large Language Models

An emerging risk that enterprises are starting to grapple with comes from the adoption of generative AI (GenAI) and large language models (LLMs). Across today's enterprise, different teams are figuring out how to improve efficiency and gain competitive advantage using generative AI. Recent Enterprise Strategy Group research showed that the top current or planned use cases include marketing, software development, research, customer service, and product development.[5] GenAI applications need data, and that data will include sensitive data, regardless of whether using such data is planned or inadvertent.

While GenAI is improving productivity and unleashing innovation, it also poses security challenges if sensitive data is inappropriately used to inform LLMs. If valuable or sensitive data is inadvertently included in model input and subsequent output, organizations face the risk of that critical data landing in the wrong hands (data leakage) or violating governance or compliance obligations by inappropriately using sensitive data.

[5] Source: Enterprise Strategy Group Research Report, *Beyond the GenAI Hype: Real-world Investments, Use Cases, and Concerns*, August 2023.

Whether data stores are managed by IT or not, enterprise data needs to be identified and categorized so appropriate steps can be taken to control the security and governance risk posed by sensitive data. That data also needs to be governed on an ongoing basis.

Controlling Risk and Ensuring Compliance

To understand their information risk and adhere to regulatory mandates, data security teams need to understand their data. They need to know where data stores are located and what is inside those data stores so they can ensure that it is appropriately secured with encryption, masking, or some other control. They also need to understand how those data stores change over time as information gets added or modified during the data lifecycle. Understanding where sensitive data is located and how it changes lays the groundwork for securing that data as well as ensuring governance of the data. In the event of a data breach, security teams can quickly and accurately determine the magnitude of the incident so that appropriate notification can take place. Accurately understanding a breach helps the business to report the right information once and avoid the reputational damage that can come with having to update an initial report that lacked a comprehensive understanding of the data involved in the incident.

What's Needed

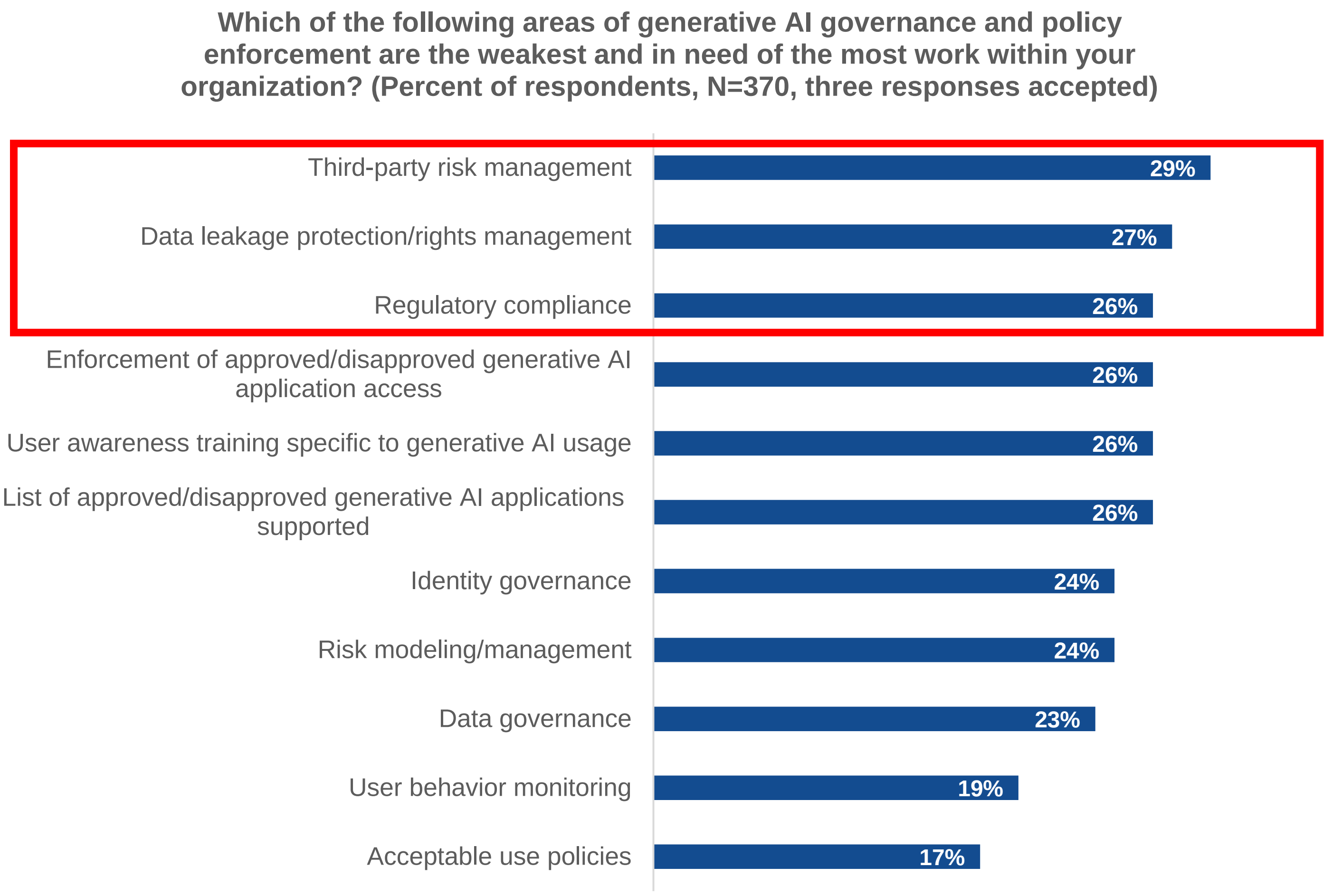
DSPM has begun to demonstrate value to security leaders among some key use cases. As IT and security teams undertake various projects, DSPM enables them to locate and categorize sensitive data to facilitate those projects. Projects can be done without DSPM, but they can be done more efficiently and effectively with DSPM. It is important to note that DSPM is an ongoing process that facilitates operations as new data is created, saved, moved, or accessed and as existing data stores evolve. For example, a data store might not contain sensitive data at its inception, but data changes over time, meaning that sensitive data might be added to the data repository.

Enabling Secure Generative AI

Whether experimenting with LLMs or deploying a generative AI application such as Microsoft Copilot or ChatGPT, organizations are relying on training data. Enterprise Strategy Group research showed that enterprise IT and security teams rank third-party risk management, data leakage protection, and regulatory compliance as the top three weakest areas of GenAI governance (see Figure 4).[6]

[6] Source: Enterprise Strategy Group Complete Survey Results, *Generative AI for Cybersecurity: An Optimistic but Uncertain Future*, April 2024.

Figure 4. Weakest Areas of GenAI Governance



Source: Enterprise Strategy Group, a division of TechTarget, Inc.

Data security posture management can improve the governance and security of data around GenAI by identifying and categorizing the relevant training data so that sensitive data does not inadvertently make its way into an LLM model.

Democratizing Data in PaaS Environments

PaaS environments enable the rapid development and deployment of applications, but locating and classifying unstructured data including JSON blobs is a difficult task, particularly when it involves large volumes of streaming data.

Reducing Storage and Data Consumption Costs

Data teams need to optimize spending by locating duplicate or abandoned data stores that can be deleted or backed up to more cost-effective cold storage. This requires identifying data stores, categorizing the data inside, and understanding the data owner so that the data can be appropriately dispositioned.

Improving Data Access Models

Data teams have to make continuous decisions about which data to allow users to access, and such decisions might need to be made in real time without knowing what a data store contains. DSPM can inform those models with appropriately categorized models.

Minimizing the Risk Footprint

Data and security teams need to collaborate to understand their data posture and minimize risk, whether the focus is on avoiding a potential breach or regulatory requirements. Minimizing the risk footprint is particularly important in sectors such as highly regulated industries or companies engaged in merger and acquisition activity.

Normalyze for DSPM

The Normalyze platform offers agentless One-Pass Scanner for rapid data discovery and accurate scanning and classification across any data store, including those in SaaS, PaaS, public or multi-cloud, on-premises, or hybrid environments. This enables enterprises to get comprehensive context around their data security posture, wherever the data is located.

Analyze Data in Place

Normalyze scans data in place, eliminating the need to take snapshots and remove data from customer environments for scanning. This approach helps avoid cloud data egress costs and controls overall solution costs. Normalyze continuously discovers new data stores as they are created. Its single-pass scanner provides accurate data classification, and the unique cloud orchestration architecture allows analysis of large data volumes across multiple data stores without the need to manage individual scanners.

Prioritize Data Stores by Their Value

Data security teams face the challenge of protecting numerous data stores, knowing that not all data stores are equally important. Normalyze's DataValuator technology estimates the relative cost of a breach for each data store, helping teams prioritize their security efforts on the most critical data. Combined with insights about access and exposure risks, the risk matrix enables teams to focus on data stores with a higher likelihood of breach and greater impact if compromised.

Achieve Least Privilege Access to Data Stores

Managing and reducing overprivileged access is an ongoing concern for security teams. Normalyze analyzes identity and access management roles, permissions, database grants, and other attributes for both user and machine identities. This allows quick identification of who has access to specific data stores. Additionally, Normalyze verifies compliance with multifactor authentication and other policies, ensuring that users accessing valuable data follow secure practices. By analyzing access logs, Normalyze determines which permissions are unnecessary and can be removed, thus achieving least-privilege access and significantly reducing the potential attack surface.

Proactive 'Detect and Fix' With User Access and Attack Path Visualizations Plus Remediation Workflows

Normalyze integrates information about all compute and networking resources, as well as PaaS services used by the organization. It detects misconfigurations and vulnerabilities in these resources, organizing this information into a graph within the Normalyze platform. This allows users to visualize and detect potential attack paths to critical data. Continuous near-real-time monitoring ensures that any changes exposing data are quickly detected.

Normalyze also provides guided remediation steps for each identified risk, making it easy to address and fix vulnerabilities.

Streamline Compliance with Mapping and Reporting

Normalyze continuously identifies and highlights data privacy gaps against regulatory compliance benchmarks. Compliance violations are tagged with the relevant compliance framework and the specific control violated, providing immediate insight into the impact on compliance posture. Normalyze also offers anomaly detection to identify suspicious activities such as data exfiltration and potential account takeover by baselining user activity and detecting abnormal behavior indicative of risky actions.

Enable Secure LLMs and GenAI with DSPM for AI

As enterprises move from experimentation to implementation of generative AI applications, Normalyze ensures appropriate data and applications are used. It helps data and security teams scan for data being used in LLMs (e.g., Microsoft Copilot, ChatGPT) to prevent unauthorized data use in AI-generated content. Normalyze can also discover unsanctioned GenAI apps deployed within cloud environments, ensuring compliance and security in AI initiatives.

Conclusion

Valuable and sensitive data remaining neglected or unknown should set off alarm bells with CEOs and CISOs alike. Data security posture management is foundational technology that locates and categorizes that data, enabling IT and security teams to operate more effectively and efficiently.

You can't secure what you don't know you have, and you cannot operate effectively if you don't understand the nature of your data or who needs access to it. Normalyze provides that needed context across environments and ensures appropriate policies are followed around the data. It enables teams to streamline operations by understanding the lineage of an organization's data and who/what is accessing that data, identifying anomalies so that organizations can better protect sensitive data stores.

Information contained in this publication has been obtained by sources TechTarget considers to be reliable but is not warranted by TechTarget. This publication may contain opinions of TechTarget, which are subject to change. This publication may include forecasts, projections, and other predictive statements that represent TechTarget's assumptions and expectations in light of currently available information. These forecasts are based on industry trends and involve variables and uncertainties. Consequently, TechTarget makes no warranty as to the accuracy of specific forecasts, projections or predictive statements contained herein.

About Enterprise Strategy Group
TechTarget's Enterprise Strategy Group provides focused and actionable market intelligence, demand-side research, analyst advisory services, GTM strategy guidance, solution validations, and custom content supporting enterprise technology buying and selling.

✉ contact@esg-global.com
🖥 www.esg-global.com