COSMOS
Navigate the Compliance Universe

Compliance Today - December 2021
How to successfully manage ungoverned data for risk mitigation

By Dean Gonsowski, Esq.

Dean Gonsowski (dean.gonsowski@activenav.com) is the Chief Revenue Officer at ActiveNav, located in the Washington, DC, metro area.

- linkedin.com/in/dean-gonsowski-2a469/
- @dean_gonsowski

Are we currently experiencing a data *explosion*, as many pundits claim? If the general volume of data is truly exploding, it is the longest and continuous explosion that we have ever seen. While organizations are struggling to control this data proliferation, they're also facing increased regulatory compliance issues, ranging from state-based privacy initiatives like the California Privacy Rights Act to New York's cybersecurity regulation, to even more novel regimes like Illinois's new biometric rules. In concert, headlong data and regulatory growth means an even larger burden for regulated entities in the healthcare, financial services, and energy sectors.

"Unstructured" data can comprise as much as 90% of a company's data environment.[1] Unstructured data is data that doesn't fit into predefined data models and is therefore inherently full of risk because it's hard to search, analyze, and control. No one knows what's hiding in this dark data, which is often created by users and usually includes sensitive information that hackers look for, such as personal identifiable information (PII), Social Security numbers, and credit card information.

Simultaneously, the looming threat of data breaches continues to grow. Alarmingly, in 2020, despite huge investments in cybersecurity technologies, the number of compromised records was the highest on record.[2] On average, the cost per record breached was $146. That cost went up to $150 if the record contained PII, which is still the most compromised type of record, with PII being breached in 80% of cases.[3] Healthcare has been hit especially hard with compliance concerns during the pandemic. Today, healthcare data records are almost 50 times more valuable[4] than the next highest-value record, and the average cost of a data breach for healthcare institutions stands at $7.13 million—the most expensive of any sector.[5]

Cybersecurity alone is insufficient

Companies have been and continue to invest heavily in cybersecurity technologies like endpoint protection, identity and access management, and third-party risk management to improve their security posture—as they should. Despite these prophylactic measures, data breaches continue to occur at an alarming rate, even at some of the most sophisticated companies in the world. Certain sectors are being targeted at an even higher rate than average entities. For example, more than one in three healthcare organizations globally reported being hit by ransomware in 2020, with this sector experiencing a 45% uptick just since November 2020, according to *Health IT Security*.[6]

Information governance can complement cybersecurity efforts to help better manage information prior to and when the inevitable occurs. Information governance is the suite of activities and technologies that organizations

can employ to maximize the value of their information, while minimizing associated risks and costs. When done effectively, information governance reduces the amount of data under management and provides visibility into what data was breached if it occurs—or, rather, *when* it occurs.

Information governance vs. information management

Traditional approaches to governing information are no longer working. Different functional areas within companies have unique objectives and value information differently, which is particularly true in regulated industries like healthcare that have mature compliance regimes like HIPAA. These functional areas tend to establish policies and practices that satisfy their line of sight. The problem is that this often leads to conflicting silos between sales, marketing, finance, operations, etc.

Where information management focuses on *how* information flows through an enterprise, information governance asks *why* we have the information in the first place. We need to govern information in a way that works for everyone, no matter their role or department.

In our experience, the best and most efficient way to do this is to link information creation, use, and disposition to business objectives. Organizations must ask these questions:

- What are our organizational objectives (business, legal, and regulatory)?
- What information is needed to achieve our objectives?
- How long is that information useful?
- While it is useful, how does it need to be organized (access, security, privacy)?
- What do we do when information is no longer useful?

The key here is to link the potential value of retained information with the realized value, which is the value actually being derived from information. Otherwise, the attendant risk of breaches and ransomware will likely outweigh the potential value and put the organization in a tenuous governance position.

You can't get value from data you can't access

Too much data makes everything less efficient, yet organizations still tend to hoard data "just in case." Other problems created by collecting and storing too much data include increased regulatory burdens, storage costs, and security risks. Known risks include the exfiltration of data after a breach and the costs of legal review due to electronic discovery issues.

Data remediation addresses these problems. Data remediation is a process directed at bringing order to information, and that doesn't just mean deletion. The goal of remediation is to help ensure companies retain only valuable information (i.e., information that is necessary to meet the organization's business, legal, or regulatory objectives and obligations). Remediation helps to ensure that information that is no longer useful to the organization is deleted in a defensible manner.

A framework for remediation

Data discovery: The first step is to gain visibility into your data estate, particularly for sensitive data. You can't protect and manage what you don't know you have. Enterprise software solutions such as file analysis can help you gain insight into all your sources of data, what's being collected, and where it's being stored. By discovering and mapping your data, you'll be able to find out what is actually there. After all, most business processes and

collaboration activities are not captured with a top-down approach.

Sensitive data location is an increasingly important task for most organizations. Results from a new Osterman Research report corroborate this new sense of urgency. This recent study found that "90% of respondents say discovering sensitive data is a high or medium priority in their organization. Organizations in three industries—healthcare/pharma, technology, and financial services—were more likely to say discovering sensitive data was a high priority."[7]

Data classification: The goal of data classification is to make content easier to govern and to access. Data classification automates the ability to audit policies, request responses, and protect data. Leverage taxonomies, similarity clusters, and predictive learning classifiers to make it easier to understand the sensitivity and value of the data you have. Data can be classified in several different ways depending on business needs, but common categories include:

- Content
- Security
- Confidentiality
- Privacy
- Regulatory
- Legal hold

Metadata management: Metadata is simply "data about data." Creating key metadata helps with the ongoing governance of content. Metadata helps determine how content will be discovered. My advice is to make use of existing metadata as much as possible. From there, develop rules for identifying and extracting additional metadata that can help with further data classification.

Data remediation: Once you have discovered and classified your content, you then need to decide what you want to do with that data. You might decide to keep, tag, migrate, quarantine, or delete given pieces of information depending on factors such as age or redundancy. No matter what you decide, you must ensure that there is a defensible audit trail upon disposition.

Ongoing governance: Information is constantly flowing in and out of an organization, meaning that governance is not a one-and-done type of project. There needs to be a process, one that's embedded into the culture of an organization. Establish a program to continuously identify and delete content that no longer provides value to the business and find ways to get the end users excited about and on board with your governance program, increasing chances of long-term success.

Where to start?

Start with the proverbial low-hanging fruit. Take bite-sized chunks of data—maybe that's a single data repository or business unit to begin with—and build from there. Decide your success metrics ahead of time, whether that's efficiency gains, risk mitigation, or cost savings. If you're able to demonstrate your success, it will be easier to gain stakeholder support for larger projects.

Information governance provides insight into workflows and business processes. These insights can then be used to optimize workflows and processes, leading to even more efficiency and risk mitigation. By beginning small, measuring success, and communicating often, effective information governance can be achieved.

When to start? Be proactive, not reactive

As the saying goes, "The best time to plant a tree was 20 years ago. The second-best time is now."

The data breach is inevitable. Additional privacy regulations will be signed into law. Security threats will evolve. It takes time to plan out an information governance program that will be effective for your specific organization, gain stakeholder approvals, find and deploy the tools needed to execute the initiative, and train the business on proper governance practices. Don't wait to start the process as a reaction to an event. Be proactive about establishing and implementing a robust data governance program so that when an event occurs, your organization is prepared.

Conclusion

Effective data and information governance *is* possible, but it requires a coordinated approach among interested stakeholders.

When implementing an information governance program, start by asking yourself:

- Do we understand what data we have, where it is, and how many copies we have of it?
- Are we safeguarding our most valuable and sensitive data and prioritizing data protection and stewardship?
- Are we positioned to confidently respond to our organization's business, legal, or regulatory objectives and obligations?
- Can we ensure that information that is no longer useful to the organization is deleted in a defensible manner?

Satisfactory resolution of these questions will help mitigate the risks associated with a security incident. Information governance and cybersecurity need to work together to minimize the impacts of data breaches.

Takeaways

- Unstructured data is the hidden threat in digital businesses and one that cannot be ignored.
- Information governance and cybersecurity need to work together to minimize the impacts of security incidents and data breaches.
- Data remediation, which doesn't just mean deletion, is key to address numerous problems caused by storing and collecting too much data.
- Create a community of information governance advocates and stakeholders to support information governance programs and efforts.
- Information governance needs to be embedded into a company culture, which means you need to ensure the conversation is happening at an executive level, too.

1 Dwight Davis, "AI Unleashes the Power of Unstructured Data," *CIO Magazine*, July 9, 2019, https://bit.ly/3pc1xUI.
2 Dan Lohrmann, "2020 Data Breaches Point to Cybersecurity Trends for 2021," *Lohrmann on Cybersecurity*

(blog), *Government Technology*, January 22, 2021, https://bit.ly/3iPHmKq.
3 IBM Security, *"Cost of a Data Breach Report 2020,"* accessed October 8, 2021, https://www.ibm.com/downloads/cas/RZAX14GX
4 Ellen Neveux, "Hackers, breaches, and the value of healthcare data," SecureLink, last updated October 1, 2021, https://www.securelink.com/blog/healthcare-data-new-prize-hackers/.
5 IBM, "IBM Report: Compromised Employee Accounts Led to Most Expensive Data Breaches Over Past Year," news release, July 29, 2020, https://ibm.co/3tIHyiB.
6 Jessica Davis, "Healthcare Accounts for 79% of All Reported Breaches, Attacks Rise 45%," *Health IT Security*, January 5, 2021, https://bit.ly/3DKXDbR.
7 Osterman Research, "Sensitive Data Discovery Rises as a Top Concern for Organizations – White Paper," September 22, 2021, https://bit.ly/3oPgiib.