INFORMATION PAPER

Governance, Privacy, and Exposure

Understanding Information Governance, Data Privacy, and Data Breach Exposure Mitigation

By Matthew Miller, Senior Vice President of Information Governance and Data Privacy, HaystackID

HAYSTACK®

Objective

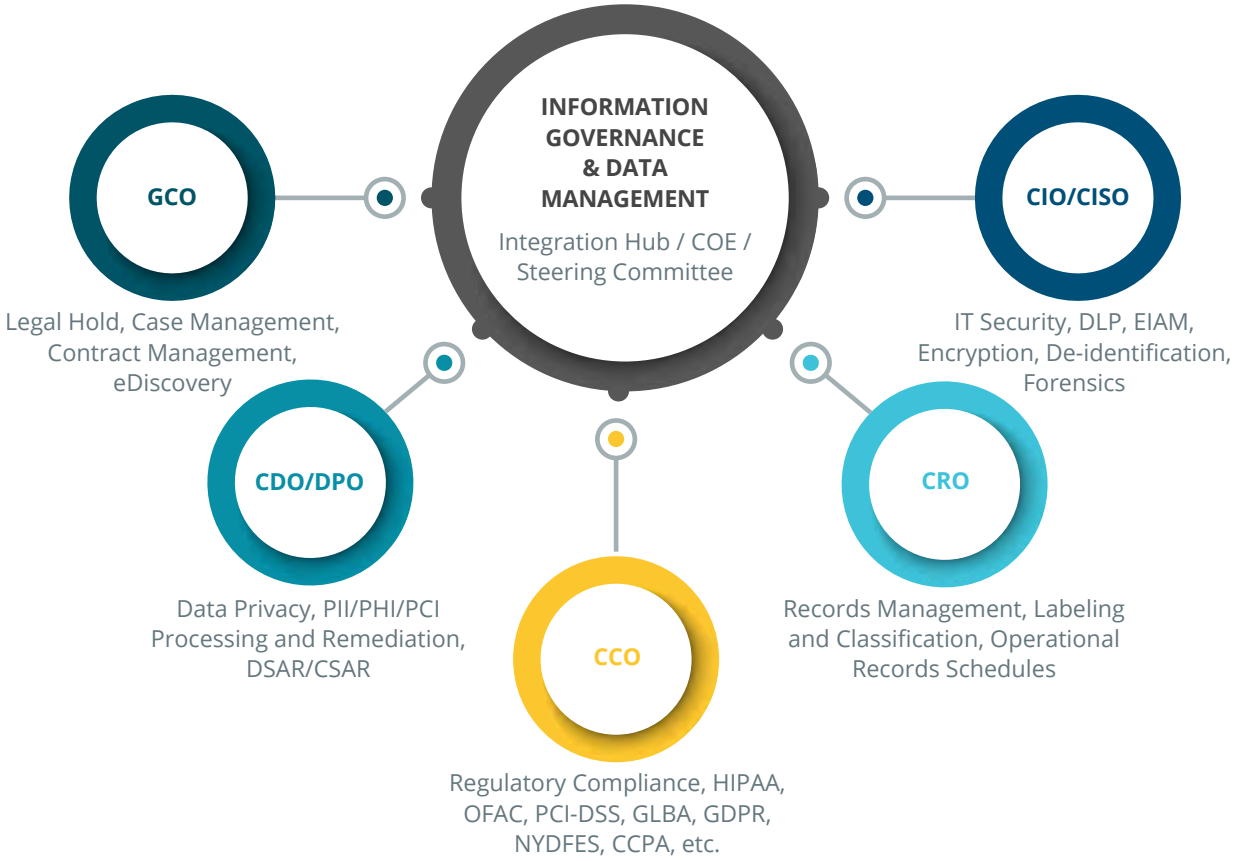
This paper, prepared by information governance and eDiscovery expert Matthew Miller of HaystackID, presents a framework and solutions for deploying or enhancing information governance programs. When adequately implemented, these programs should ensure cyber-incident preparedness and demonstrate, at minimum, reasonable security measures for sensitive and critical information assets.

Introduction: Balancing Data Value and Risk

Organizations today, both private and public, should be excited about what is happening with their data and their networks, as advancements in technology have finally caught up with the requirements necessary to handle and manage via controls and classification, any volume of data, regardless of magnitude, in any repository, in any geography, behind the firewall, or in the cloud or a hybrid environment. Due to technology and data mining, organizations can get more value from information assets. However, the rewards of data value must be considered with the risk of a network compromise and the related data privacy and cybersecurity obligations imposed by legal, regulatory, and business retention requirements.

Data privacy, legal, IT security, compliance, and records management teams must work together to minimize data breach exposure by classifying and managing unstructured data with real-time reporting and continuous monitoring to manage both legacy and new or modified data on an ongoing basis with tighter budget scrutiny than in the past. Working together is crucial given the challenge of global network visibility and growing shortages of internal resources for securing and protecting critical and sensitive information assets.

HaystackID's IG Integration Hub Approach to Satisfy NIST. SP.800-53r5, 3.13 Program Management, PM-23 Control



Reducing exposure to cybersecurity-related privacy events makes the myriad intertwined data challenges a primary focus for organizations as we head toward 2022, continuing to combat a global pandemic that has altered how organizations, employees, and consumers interact with data. Enabling a coordinated and proactive approach toward complying with courts, regulators, and auditors, in addition to the ever-changing landscape of data privacy laws, is no longer a "nice-to-have" but a "must-have." This paper aims to present a framework and solutions for deploying or enhancing an information governance program, which, when adequately implemented, will ensure cyber-incident preparedness and demonstrate, at a minimum, reasonable security measures for sensitive and critical information assets.

The NIST Cybersecurity Framework 2.0: Integrating Cybersecurity and Privacy Risks (Aug. 8, 2023)

Associated with cybersecurity incidents arising from loss of confidentiality, integrity, or availability.

Associated with privacy events arising from data processing.

Other ERM Risks: Financial, Legal, Operational, Physical Security, Reputational, Safety

Anywhere and Anytime: All Networks Are Vulnerable

The commonality of data breaches, business email compromise, and ransomware attacks since the beginning of 2020 has dramatically increased. Data points supporting this increasing commonality include:

- The FBI recently reported that the number of complaints about cyberattacks to The FBI Internet Crime Complaint Center (IC3) is up from 10 to as many as 4,000 a day. A 400% increase post-coronavirus. (MonsterCloud)
- Phishing or social engineering attacks have jumped to 20,000 to 30,000 daily in the U.S. alone. (Microsoft)
- Ransomware attacks are up 800% during the pandemic. (Entrepreneur.com)
- In December 2020, multiple U.S. agencies were successfully targeted, including the Departments of State, Treasury, Commerce, Energy, and Homeland Security, as well as the National Institutes of Health.
- In October 2019, hackers attached malware to a software update from Austin, Texas-based company SolarWinds, whose tools monitor computer networks, and approximately 18,000 of their 300,000 clients received the update. More than 100 major private organizations were impacted.
- FireEye, a leading incident response and data breach remediation company, identified a compromise as a supply chain attack where attackers gained remote network access via malware inserted into previously unknown software vulnerabilities. Unfortunately, FireEye was also compromised by the attack.
- In December 2020, a data security incident involving Accellion, a third-party provider of hosted file-transfer services, affected numerous organizations, including law firms and state governments, whose impacts are currently under investigation.

These data points are just a sample of the extent to which cyber incidents have impacted businesses, governments, and the global economy.

According to the IBM-sponsored *2021 Ponemon Institute Cost of a Data Breach Report*, which analyzed data breaches that took place between May 2020 and March 2021 (the COVID-19 outbreak was declared in March 2020), the average total cost of a data breach increased by nearly 10% year over year. This was the most significant single-year cost increase in the last seven years. The report lists some extremely pertinent data breach facts based on countries and industry sectors. These facts demonstrate the dramatic increase in advanced persistent threats (APTs) that organizations face daily.

Cost of a Data Breach by Country/Region

The top five countries and regions with the highest average total cost of a data breach in both 2020 and 2021 were:

01	02	03	04	05
UNITED STATES	MIDDLE EAST	CANADA	GERMANY	JAPAN

Cost of a Data Breach by Industry Sector

Healthcare was the top industry with the highest average total cost for the 11th year. The top five sectors with the most increased average total price were:

01	02	03	04	05
HEALTHCARE	FINANCIAL	PHARMA	TECHNOLOGY	ENERGY

The average total cost for healthcare data breaches increased from $7.13 million in 2020 to $9.23 million in 2021, a 29.5% increase. Energy dropped from the second most costly industry to fifth place, decreasing in cost from $6.39 million in 2020 to $4.65 million in 2021, a 27.2% decrease. Other sectors that saw significant cost increases included services (7.8%), communications (20.3%), consumer (42.9%), retail (62.7%), media (92.1%), hospitality (76.2%), and public sector (78.7%).

Data Breach Cost Categories

In 2021, lost business continued representing the seventh year's most significant share of data breach costs. Of the four cost categories, at an average total cost of $1.59 million, lost business accounted for 38% of the average total data breach cost. Lost business costs include business disruption and revenue losses from system downtime, the cost of lost customers and acquiring new customers, reputation losses, and diminished goodwill. The second most costly was detection and escalation costs, which had an average total cost of $1.24 million, or 29% of the total cost. The other cost categories are notification (6%) and data breach response (27%).

Working Toward the Zero Trust Architecture

One primary, privacy-based information governance strategy to combat the exponential increase in malicious activity by nation-state bad actors and criminal enterprises is implementing a Zero Trust Architecture (ZTA) approach to cybersecurity. The goal of ZTA is to prevent unauthorized access to data and services coupled with making access control enforcement as granular as possible. How is ZTA related to information governance? The connection between information governance and ZTA is based on minimal or least privileged access. Least privileged access means restricting resources to those needing access and granting only the minimum privileges to those needing to perform required tasks.

To make this ZTA approach a reality, organizations need to know what data they process and maintain at the file, email, and database levels. Implementing this approach requires global network visibility and resources (i.e., people, process, technology) to handle data identification, classification, inventory creation, and remediation to mitigate risk.

The National Institute of Standards and Technology (NIST) published *NIST.SP.800-207, Zero Trust Architecture (August 2020)*, which describes the following seven tenets of zero trust:

1. All data sources and computing services are considered resources.
2. All communication is secured regardless of network location.
3. Access to individual enterprise resources is granted on a per-session basis.
4. Resource access is determined by dynamic policy—including the observable state of client identity, application/service, and the requesting asset—and may include other behavioral and environmental attributes.
5. Enterprise monitoring and measuring of the integrity and security posture of all owned and associated assets.
6. Resource authentication and authorization are dynamic and strictly enforced before access is allowed.
7. Enterprise collection of as much information as possible about the current state of assets, network infrastructure, and communications and used to improve its security posture.

These seven elements are essential for achieving a ZTA approach to security and information governance.

Continuous Monitoring and Data Supervision

Information governance technology and techniques combined with IT security reporting enable the ZTA requirement of robust monitoring and reporting systems that provide actionable data about the current state of enterprise resources. Implementing a ZTA should establish a continuous diagnostics and mitigation (CDM) system. For ZTA to operate, there is a need for a constant cycle of obtaining access, scanning and assessing threats, adapting, and continually reevaluating trust in ongoing communication. Organizations must know their data and who can and should have access to critical and sensitive information at any given time. An enterprise implementing a ZTA would be expected to have an identity, credential, and access management (ICAM) and asset management systems in place. When looking at the data lifecycle, organizations can limit the amount of data subject to these restrictions by eliminating redundant, outdated, or trivial data from the network. This practice of data minimization via a defensible data disposition program can genuinely narrow down the attack surface requiring data protection measures.

An enterprise should collect data about asset security posture, network traffic, and access requests, process that data, and use any insight gained to improve policy creation and enforcement. This data can also be used to provide context for access requests from subjects.

Similarly, *NIST.SP.800-53r5 Security and Privacy Controls for Information Systems and Organizations, September 2020, Updated 12/10/20*, states in Control CA-7 Continuous Monitoring that organizations should develop a system-level continuous monitoring strategy and implement continuous monitoring under the organization-level continuous monitoring strategy. Continuous monitoring at the system level facilitates ongoing awareness of the system security and privacy posture to support organizational risk management decisions. The terms “continuous” and “ongoing” imply that organizations assess and monitor their controls and risks at a frequency sufficient to support risk-based decisions. Having access to security and privacy information continuously through reports and dashboards allows organizational officials to make effective and timely risk management decisions, including ongoing authorization decisions.

Implement Defensible Disposition and Remediation

A defensible disposition program aims to organize or classify data and eliminate data as appropriate to reduce corporate risk, control legal and business costs, and gain insight into and appropriately manage business documentation. The defensible disposition also helps an organization consider data efficiently as a business asset and manage compliance with identified and current preservation obligations, whether imposed by national laws and regulations, corporate standards, or business terms defined within commercial contracts. Implementing an effective defensible disposition program enables organizations to:

- Reduce both enterprise risk and litigation risk.
- Reduce eDiscovery, investigation, cyber incident response, and storage IT costs.
- Improve data protection.
- Improve employee productivity.
- Improve response times and accuracy for regulatory and compliance obligations.

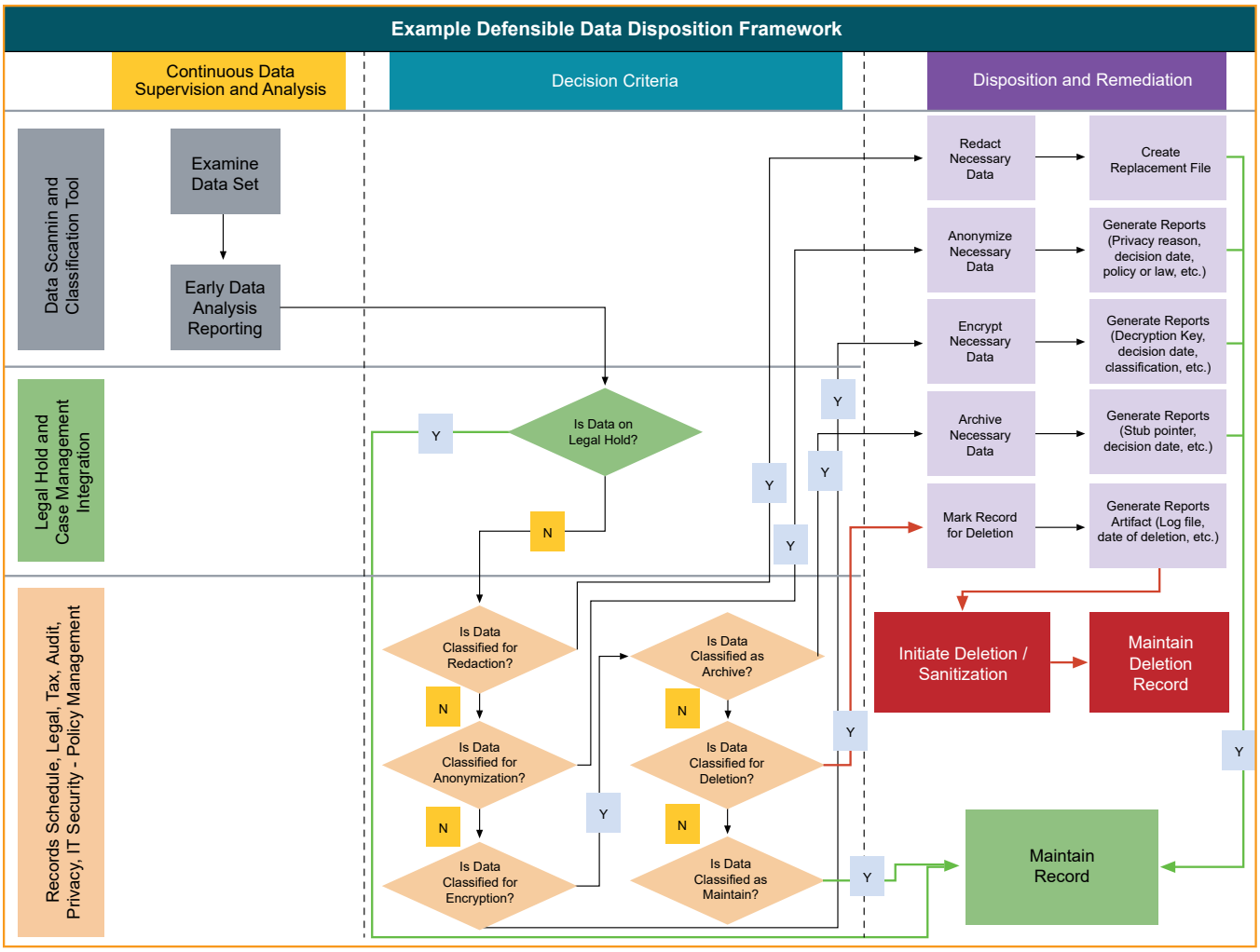
Developing and Implementing a Defensible Data Disposition Framework

The first step in establishing defensible data disposition is for the organization to define a framework. To do this, an organization's information governance steering committee should agree to the framework, taking into account the similar and competing obligations from different lines of business, including but not limited to:

Frameworks are organized into process groups, processes, decision points, criteria, and considerations related to methods and decision points.

Depending on the data source, the defensible data disposition plan will detail the appropriate quality control procedures, approvals, and remediation execution steps with strategic remediation recommendations based on prioritized risk listings, including:

- Integrating legal hold systems and policy with data lifecycle management and developing plans for suspending the lifecycle for specific information relevant to a legal or tax matter, etc.
- Developing a defensible deletion framework to delete redundant, obsolete, or trivial data.
- Developing a system for anonymizing, quarantining, or redacting sensitive or critical data.
- Performing tokenization, masking, de-identifying, or removing personally identifiable information from systems of record.
- Migrating or archiving select information to the cloud or retaining information in an archiving system based on business and legal requirements.
- Deciding and tracking the destruction of information at the end of its lifecycle.



Once it has been determined that information should be in the Deletion/Sanitization workflow, *NIST SP 800-88 Rev. 1 Guidelines for Media Sanitization* must be considered. Here, based on the security categorization of the confidentiality of information contained on the media, not the type of media, organizations can make a decision on what type of sanitization is best for that use case and then, based on the media type, determine the technique used for the sanitization. Information disposition and sanitization decisions occur throughout the information system lifecycle.

Types of Sanitization

Adapted from *NIST Draft Special Publication 800-88 Rev 1: Guidelines for Media Sanitization*; Section 2.5

The categories of sanitization are defined as follows:

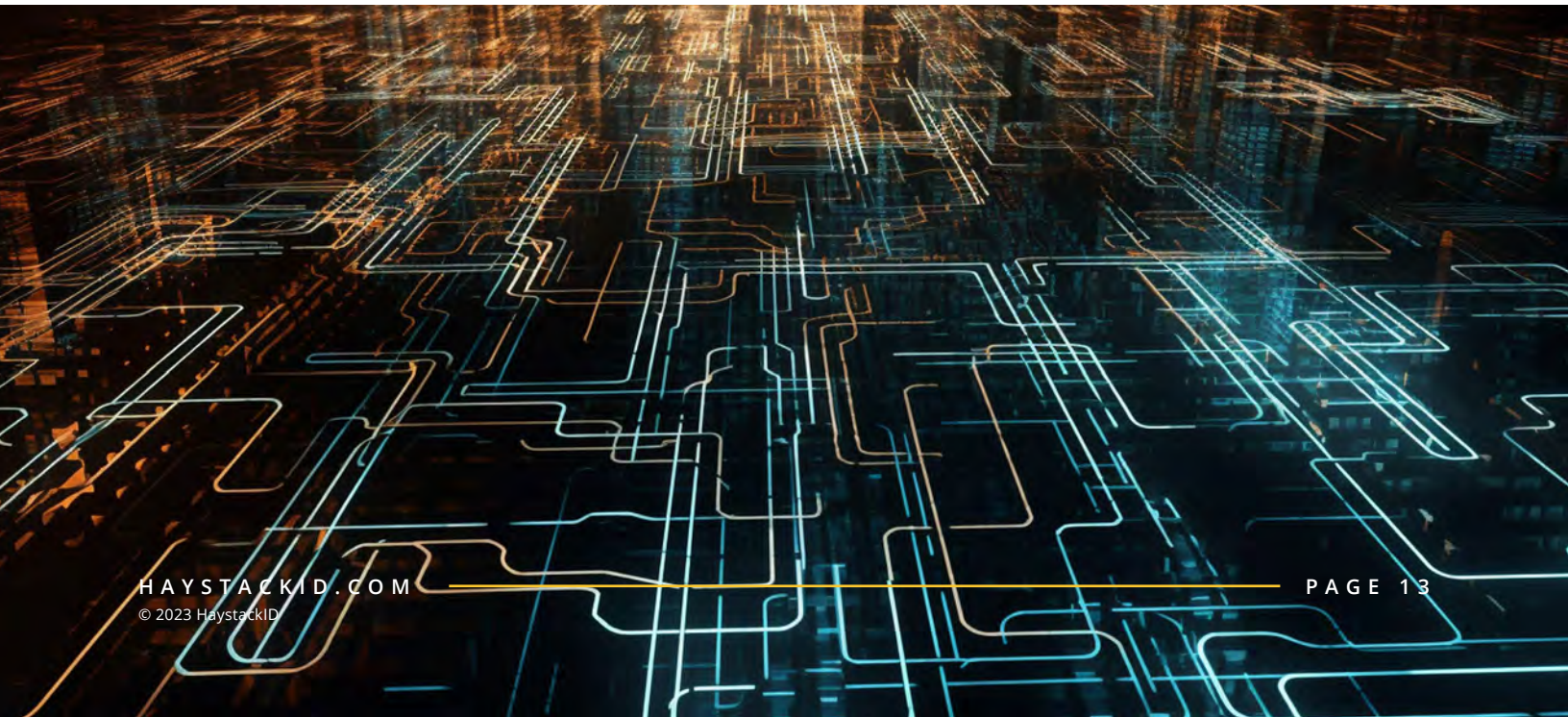
- **Clear:** Applies logical techniques to sanitize data in all user-addressable storage locations for protection against simple, non-invasive data recovery techniques; typically applied through the standard read and write commands to the storage device, such as by rewriting with a new value or using a menu option to reset the device to the factory state (where rewriting is not supported).
- **Purge:** Applies physical or logical techniques that render target data recovery infeasible using state-of-the-art laboratory techniques.
- **Destroy:** Renders target data recovery as infeasible using state-of-the-art laboratory techniques and results in the subsequent inability to use the media for storage of data.

Defensible Data Disposition Implementation Technologies and Techniques

A variety of tools may assist in the implementation of the framework. Examples include:

- **Auto Classification:** Auto classification tools attempt to classify information to determine if it is a record and automatically choose the type of record. Classified records may then be aligned to the retention schedule.
- **Duplicate Analysis:** Helps determine if the information is duplicative.
- **Data Monitoring Tools:** Helps determine the age of information, the last time it was accessed, and the owner of the information.
- **Sampling:** Supports decision-making related to a population based on a sample of the population.
- **Native Application Functionality:** The native ability of an application to manage, categorize, and move information through the framework.
- **Archiving:** Archiving may be a viable form of disposition for structured data.
- **ERMS:** Purpose-designed records management system capable of moving records through the framework.

- **Data Deletion:** Purging data and associated reference attributes that will be completely erased and unavailable for future use.
- **Data Pseudonymization:** Substituting identifiable data with a randomly generated number or token. This method offers the ability to re-identify data with additional information (such as a key). This method should be considered when storing or using personal data that may need to be accessed fully to function in required business use cases.
- **Data Anonymization:** The method for pseudonymization and anonymization is often similar—the main difference being that pseudonymization is performed to retain the ability to re-identify individuals within the data sets. Anonymization is irreversible; the original values are disposed of properly. This should be used for personal values where there is no business need to retain the data in a fully identifiable format.
- **Data Masking:** Character masking is the change of the characters of data values. Masking is typically partial and applied only to some characters in the attribute. This is to be used when the data value is a string of characters, and hiding part of it is enough to provide the extent of anonymity required.
- **Data Tokenization:** The method of replacing sensitive data with unique identification symbols or algorithmically generated numbers that retain all the essential information about the data without compromising their security.
- **Encryption:** In cryptography, encryption is the process of encoding information. This process converts the original representation of the information, known as plaintext, into an alternative form known as ciphertext. Ideally, only authorized parties can decipher a ciphertext back to plaintext and access the original information. Encryption does not itself prevent interference but denies the intelligible content to a would-be interceptor.



HaystackID Assists with Defensible Data Disposition Program Development

HaystackID will work with an organization's IT, legal, records management, IT security, compliance, and privacy teams to create or enhance a defensible, repeatable, auditable, and traceable remediation program. HaystackID experts will be responsible for developing a data disposition workflow for electronically stored information (ESI) and non-digital media, including paper documents. Our projects contain three phases:

Phase One: Initiation, Planning, and Interviews

Phase Two: Service Delivery

Phase Three: Recommendations and Strategic Roadmap

During the initiation phase, HaystackID will introduce the project team and schedule work sessions with clients and other vital resources to understand the current state of the organization's information governance or data privacy programs. HaystackID experts will review existing processes and any supporting documentation and conduct interviews with key stakeholders. By partnering with HaystackID, organizations will benefit from advice and recommendations to enhance or improve current information governance and data privacy programs.

HaystackID will also review and digest any relevant existing policies, documentation, audits, or other artifacts (e.g., IT security policy, data privacy policy, org chart, network diagrams, etc.), in preparation for remote assessments or onsite interviews of key stakeholders.

With critical stakeholder guidance on repositories and organizational content, HaystackID can develop the program and implement the technology. Data repositories are organized into categories to facilitate the application of the defensible disposition framework. These categories include structure by:

- Type (structured, unstructured, and semi-structured)
- Content (regulated/business records vs. non-records)
- Context (legal/tax/audit hold, orphaned, duplicative, managed, unmanaged, migrating, decommissioning, etc.)

Combined with HaystackID's proprietary workflows, the IT and business stakeholders' responses yield the parameters for a defensible data disposition program. Note that several categorizations must typically be applied to process information for disposition under the invulnerable disposition framework.

Implement a Cybersecurity Framework Without Boiling the Ocean

A cybersecurity framework provides a common language and standards for security leaders across countries and industries to understand their and their vendors' security postures. It becomes much easier to define the processes and procedures that your organization must take to assess, monitor, and mitigate cybersecurity risk with a framework in place.

With NIST requirements as a backdrop, HaystackID has designed a solution to assist organizations in developing and implementing controls to identify, classify, protect, remediate, and manage data continuously. These controls provide for critical and essential operations and support an organization's assets and the privacy of individuals. HaystackID's proprietary processes reduce overall enterprise risk and minimize large-scale data management and protection costs.

HaystackID helps clients manage the growth of electronically stored information and meet regulatory compliance demands. While the NIST guidelines can lead organizations in the right direction, even more vulnerabilities may be exposed without implementation. HaystackID's information governance and data privacy solutions leverage the approaches, processes, and frameworks outlined by NIST better to secure critical and sensitive information across the enterprise regardless of volume, geography, or format.

HaystackID offers an end-to-end approach to the design, implementation, and maintenance of all aspects of an enterprise-wide information governance program, including:

- Data Classification
- Data Privacy
- Data Protection
- Data Breach Prevention and Response
- Records Management
- Technology Strategies
- Business Process and Program Development

HaystackID can augment your team with our people, bringing decades of experience in adjacent legal, privacy, and technology disciplines, enabling a measurable impact on:

- Overall information risk management profile.
- Information governance program maturity.
- Records governance program maturity.
- Data privacy program maturity.
- Legal hold and case management.
- Cyber incident response and data breach notification document review.
- IT security's ability to provide reasonable security to all critical/sensitive data.
- Legal, compliance, and audit ability to operate and protect the business.

HaystackID's information governance and data privacy advisory services enable the implementation or enhancement of required legal and operational controls, improve ongoing data protection, provide continuous data supervision, and make reasonable security for all critical and sensitive data attainable.

Learn More. Today.

Contact us today to learn more about our information governance capabilities and how we can help assess, augment, accelerate, and support your cyber, data, and legal discovery operations.

About HaystackID®

HaystackID is a specialized eDiscovery services firm that supports law firms and corporate legal departments and has increased its offerings and expanded with five acquisitions since 2018. Its core offerings now include Global Advisory, Discovery Intelligence, HaystackID Core®, and artificial intelligence-enhanced Global Managed Review services powered by ReviewRight®. The company has achieved ISO 27001 compliance and completed a SOC 2 Type 2 audit for all five trust principles for the second year in a row. Repeatedly recognized as a trusted service provider by prestigious publishers such as Chambers, Gartner, IDC, and The National Law Journal, HaystackID implements innovative cyber discovery services, enterprise solutions, and legal discovery offerings to leading companies across North America and Europe, all while providing best-in-class customer service and prioritizing security, privacy, and integrity. For more information about its suite of services, including programs and solutions for unique legal enterprise needs, please visit HaystackID.com.

About the Author

Matthew Miller is the senior vice president of information governance and data privacy for HaystackID. With a background in the legal profession before moving into the discipline of eDiscovery, Matthew previously co-developed Ernst & Young's information governance services practice and also served as the global IG advisory services leader at Consilio LLC. As an industry expert practitioner, Matthew has led highly complex incident response-related forensic investigations and multinational, petabyte-scale data governance and privacy engagements. Matthew may be contacted at MMiller@HaystackID.com.