Information Governance – the Foundation for Information Security

Deborah Juhnke[1]

Abstract:

Reducing the amount of data under management is appropriate and necessary to help improve an organization's security posture. The identification, classification, and segregation of information, coupled with routine disposition of detritus, will yield less information requiring protection and a better ability to apply tiered safeguards. The solution is not simply technical, however, and requires legally-defensible guidance, executive mandate, and changes to culture. We explore a data management model based on Information Governance principles and propose a triage process that focuses on the elimination of ROT (redundant, obsolete, and trivial data) using legally-validated retention schedules and policy guidance. We review various information security standards that support the inventory and management of information assets, with an eye toward practical applications.

DARK DATA IS BECOMING an information governance nightmare (Shetty 2017). Unstructured and uncontrolled for decades, "[e]mail, instant messages, documents, ZIP files, log files, archived web content, partially developed and then abandoned applications, [and] code snippets" (Shetty) not only impact costs, but also the ability to apply effective security controls. A recent survey suggests that fifty-four percent of data in organizations is stale, and that seventy-four percent of organizations have over one thousand stale sensitive files (Varonis 2018, 11).

This paper offers a roadmap on how to solve this problem. It explores an information management model based on the Information Governance principles of Structure, Direction, Resources, and Accountability, and proposes a triage process that focuses on the elimination of unnecessary information using legally-validated retention schedules and policy guidance.

[1] Senior Consultant, Information Governance Group, LLC. djuhnke@infogovgroup.com. 4324 Belleview, Suite 201, Kansas City, MO 64111

Better information governance yields better information security.

By reducing the volume of unstructured data under management to a fraction of its current information inventory, an organization will free up storage, reduce licensing costs, shorten backup cycles, and drastically cut e-discovery preservation costs. More importantly, a reduction will diminish the footprint for potential compromises. Less, better-categorized data offers a smaller attack surface and limits vulnerabilities arising from redundant, orphaned, obsolete, forgotten, transitory, and hidden data stores (ROT, or **r**edundant, **o**bsolete, and **t**rivial data). The availability of ROT in systems opens the door for external penetration, exploitation, and internal compromise.

Insider threats—both intentional and inadvertent—are responsible for a significant number of data breaches. This has justifiably led to more training regarding password management and recognition of phishing attacks. Overlooked, however, is the fact that the ROT that lies dormant in unstructured systems, and that is *created* by insiders, offers up a cornucopia of treats for hackers: files containing business confidential information, credentials in plain text files, Intellectual Property (IP), sensitive Protected Health Information (PHI) and Personally Identifiable Information (PII), and more. A focus on eliminating ROT through retention and rule enforcement will mitigate many of the vulnerabilities that come from excess and unmanaged data. Insiders are the soft underbelly of information security, especially given the vast amount of unprotected, unstructured data that exists in most organizations.

Reducing the amount of ROT under management is appropriate and necessary for businesses generally, but particularly for all critical infrastructure sectors, and begins with a simple proposition:

> ***Identification, classification, and segregation of information + routine disposal of detritus = less to protect + better ability to apply tiered protection***

The solution, however, is not simply a technical one. It requires engagement of senior management and end users and will benefit greatly from the support of an organization's legal, risk, compliance, privacy, and audit functions. Such groups may be engaged to identify common goals and to leverage budgets and bandwidth. These siloed groups have similar concerns, yet often struggle to make an isolated business case for change. Like puzzle pieces, aggregating these concerns creates a complete picture, most often with enough clarity and unified purpose to get an executive commitment and budget for change.

Current State of Information Governance

According to the Compliance, Governance and Oversight Council's Information Governance Benchmark Survey of 2018, even though there is evidence showing that information governance (IG) programs have increased support, there continues to be a lack of measurable progress (CGOC 2018, 6). Although roughly seventy-five percent of respondents report progress in their IG programs and have an appropriate level of executive sponsorship and leadership, only a third have an automated defensible disposition program in place (even though in 2010 ninety-eight percent of respondents identified defensible disposal of information as a desired benefit).

The CGOC report suggests that problems and barriers include a lack of data classification, data silos that make it difficult to link retention schedules to data, and the fact that retention, preservation, and disposal are often not considered prior to provisioning new systems. External pressures also play a role. Vendors of IT storage and cloud solutions promote sales to their clients of unlimited space for email and documents because it increases their revenue, with little consideration for the risks their clients will face from over-retention. More storage is the short term, easy answer to rampant data growth, but not a good one. More storage and unlimited email repositories only exacerbate the problem.

The Association for Information and Image Management, an information governance industry association, recently published The State of Intelligent Information Management: Getting Ahead of the Digital Transformation Curve (AIIM 2018). The survey found that on average forty percent of respondents reported that organization of their Office documents, email, scanned documents, design files, and intellectual property assets was “chaotic” or “somewhat unmanaged” (AIIM 2018, 10). The same was true for over fifty percent of web content, social media, photos, and instant messaging. Nearly half of respondents also rated the effectiveness of their organization in managing, controlling, and utilizing electronic information as toward the “terrible” end of the scale, as opposed to “excellent.” Most telling is that the needle has barely budged toward “excellent” for the same survey question in the last ten years. Recognition that something needs to change to modernize information management strategies is strong, however, at ninety-two percent.

The CGOC survey shows that there is a fundamental disconnect between desired outcomes and true progress, as respondents still report after eight years that *sixty percent of all stored data has no business, legal, or regulatory value* (CGOC 2018, 8). Shifting the focus to improved information security is a way to bridge this gap.

Data Lakes Are Not the Answer

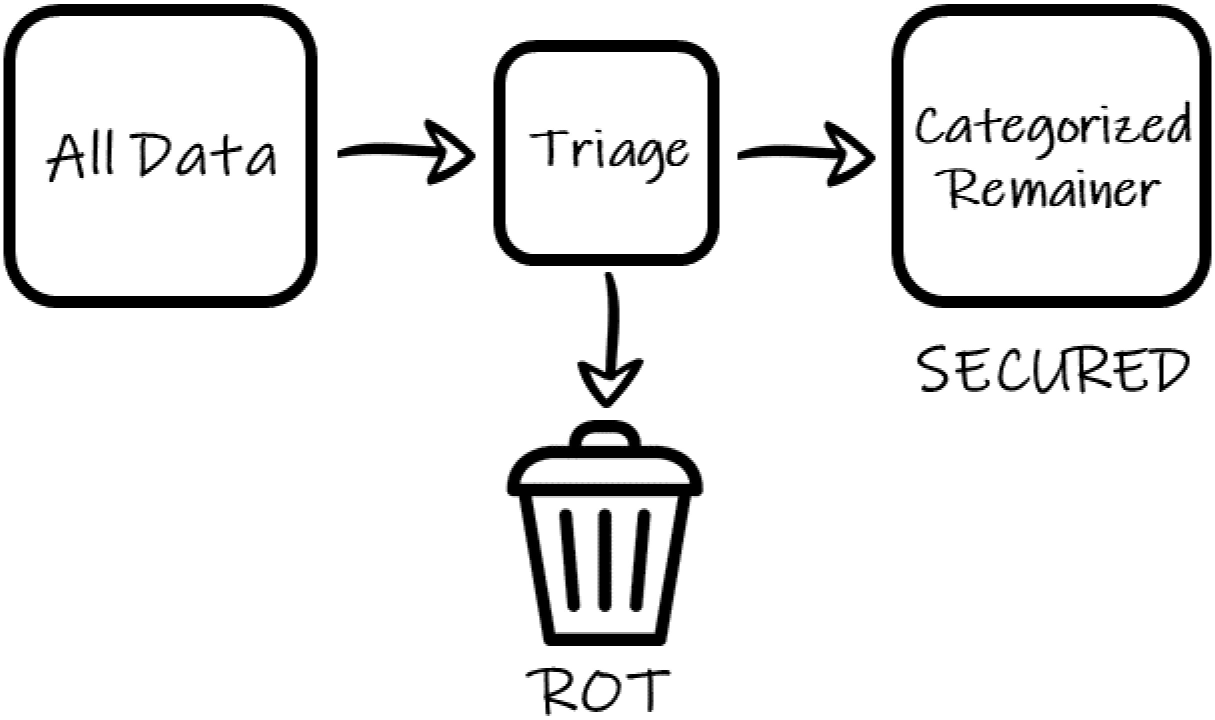
Although it may be tempting to create larger pools of unrelated data to simplify application of security controls, data lakes do not solve the problem of too much data. Simply aggregating poor quality, irrelevant, or obsolete data diminishes the value that may be gained from targeted data mining of curated information. Routinely cleaning out the chaff makes the remaining data more valuable—just one of many benefits of good governance.

Unstructured data also commonly holds highly sensitive information that, if exploited, can yield everything from unencrypted passwords to sensitive business and industrial system information. In addition, unnecessary storage of seemingly inconsequential information can enable inference attacks by allowing access to files from which more robust information about sensitive databases may be inferred, leading to unauthorized access and exfiltration.

Maintaining less data can limit entry points, limit the scope of a breach, limit the exposure of sensitive information such as IP, PHI and PII, and significantly minimize e-discovery costs in the event of litigation or regulatory investigation. Perimeter security is useful, but once breached, a far greater amount of aggregated data becomes exposed than would be if culled and sequestered in appropriately protected tiered systems.

Proposed Model

Below is a simplified diagram illustrating an improved information governance approach.



The **All Data** box represents the entire body of an organization’s information, including structured and unstructured data, archives, and local and cloud-based storage. More than half of this data may commonly be classified as ROT—redundant, orphaned, obsolete, forgotten, transitory, and hidden data. But “All Data” also includes business critical information, IP, PHI, PII, and information required to be retained pursuant to regulation or statute. The goal is to eliminate the ROT permanently while ensuring that the remainder are identified, segregated, protected, and retained appropriately.

Triage occurs through a series of processes designed to identify, classify, and apply rules. Identification of information assets may sound like an obvious task and one that many would assume has been done. In fact, most organizations do not have a firm grasp of what information they hold and where it is. Creating a basic inventory of both systems and data is the first step. The inventory should include not only active data, but also data held in archives and off-line storage.

The rules to apply take the form of policies (such as for data classification, records retention, and legal holds), a retention schedule, and guidance documents regarding segregation and storage of sensitive information.

Applying these rules will enable organizations to cull **ROT**, and the methods used can vary. For example, large scale culling may sometimes be applied to known data sets such as unstructured files of terminated employees, ad hoc backups of data, redundant “just in case” archives, and transitory data. In some cases, a more refined approach must be taken (particularly in regulated industries that have rigorous retention requirements), but in no instance should anyone, including users, be required to sift through files one-by-one. There are numerous software products that support the inventory, categorization, review, and triage of unstructured data stores, most of which also provide for migration or disposition of data.

During the triage process, **Categories** of data will emerge, some of which will require one or more levels of security controls, and others of which will not. Key to efficient categorization is limiting the options to no more than three or four. For example, based on the sensitivity of the information, the categories to use might include Public, Business Confidential, and Restricted.

Once categorized and culled, information may be segregated to **Secure** and **Other** locations, where appropriate security controls are applied. The "crown jewels" will warrant having the most layered and stringent controls, while Public data may have fewer and less-sophisticated controls. Segregation is the operative word, ensuring that any eventual compromise is contained.

Beyond the expected compliance and security benefits, following the above process gives great visibility into an organization's information assets, and can uncover additional opportunities for streamlining workflows and eliminating unnecessary creation or duplication of information. Information governance, however, is not a one-time project. It is evergreen and demands periodic, (e.g., at least bi-annually), refreshing of regulatory, statutory, and business-need retention requirements, as well as internal audits to ensure adherence to policy.

Legal & Compliance Support for Information Governance

Because the Information Security function cannot decide in a vacuum what to manage and what to dispose of, the Legal department can be a great ally and facilitator of change. Most unstructured information that exists in file shares, SharePoint sites, dormant databases, archives, and email systems is at best redundant, and at worst obsolete. "Last accessed" dates offer a simple measure of the volume of ROT, though they may not always be available or reliable enough alone to trigger disposition. A "last accessed" date may, however, be a useful metadata element as part of a more nuanced set of review criteria such as "last modified" and file extension. Lawyers understand that *some* information must be retained according to various statutes and regulations and that *some* information has business value beyond retention requirements. They also understand that the remainder falls under the categories of convenience copies or duplicates, non-business data, and obsolete copies of what were once *bona fide* records. The reality is that as much as eighty percent or more of most organizations' information falls into these latter categories. One caveat: In the case of an impending or existing lawsuit or investigation, data that is not otherwise required to be kept, but which is pertinent to the matter, must be *preserved* until the matter is finally resolved.

Appropriate use of terminology here is an important and critical distinction. *Retention* is applied to data in the normal course of business. *Preservation* is applied to data pertinent to a lawsuit or investigation, regardless of its value or retention requirements, and supersedes any disposition mandate. This preservation duty is commonly effected through a "legal hold" issued by an organization's internal or outside legal counsel. Legal holds remain in effect until formally lifted by legal counsel. Consequently, inventory and triage efforts *must* consider any existing legal holds when designating data for disposition.

There are also regulatory authorities in virtually every critical infrastructure industry for recordkeeping and other compliance requirements. A thorough legal review and summary of these authorities will yield a records retention schedule—the roadmap to compliant disposal of data. It is the primary basis for decision-making regarding what to keep and what to toss and, if well crafted, will be an authoritative source of guidance for defensible disposition.

Lawyers know the value of enforcing disposition of ROT: improved compliance, reduced risk, improved security, and cost savings. They also know how to draft policies and gather executive support for information governance initiatives.

Building Blocks to Improved Security

Before controls may be applied, good information security requires: (1) knowledge of what information exists, (2) where it is, and (3) the legal and compliance requirements for its retention, all to enable compliant disposition. This information will help dictate what policies to put in place, what tools to acquire, what training to provide, and what other technical, administrative, physical, and operational controls to apply.

Foundational elements of information governance include:

- **Structure**
- **Direction**
- **Resources**
- **Accountability**

Structure supports the understanding of what information exists, where it is and in what form, how long to retain it, and when and how to dispose of it. A **Record** is defined as, "[r]ecorded information, regardless of medium or characteristics, made or received by an organization that is evidence of its operations, and has value requiring its retention for a specific period of time." It is common to consider only record-worthy information when performing a data inventory, but to be effective, *all information* must be identified and classified. Record retention is as much about segregating and managing the lifecycle of non-record information as it is about retaining information required by law.

An actionable, current, and legally-validated records retention schedule codifies not only legal requirements, but also business needs for retention and disposition. Note that a retention schedule is not simply a policy. It is a detailed, legally annotated framework that identifies bundles of information and record types and how long to retain them. File plans capture further detail about the specific types and locations of business records, typically on a departmental basis, and can enhance the framework for classification and segregation of sensitive information.

Direction comes from policies and processes that enable employees to comply with information governance requirements. Email and computer use policies, records management policies, and privacy and security policies all inform employees of what they should do. Processes, such as document creation guidance, storage guidance, and periodic clean-up days, tell employees how they should do it.

Identifying the right **Resources** is an indispensable aspect of good information governance. The right people, training, and technology all play a role. Because security is not simply an IT issue, it is important to engage personnel at many levels, including executive mandate and oversight, departmental liaisons, end users, internal subject matter experts, and those in the legal, compliance, privacy, risk, and audit functions. The necessary cultural change required to accept and effectuate policy, discussed further below, is achieved in part through training both in new or improved processes, and training to support behavioral change generally. The range of technology tools available to support information governance and security is vast, but certain classes of tools are particularly useful. These include content management systems, auto-classification tools, and data identification and culling tools.

Without **Accountability**, efforts to improve information governance usually fall short. There must be a clear executive mandate and a strong audit function. Individual accountability must be driven by a combination of policy and cultural change. Workers may only be held accountable, though, if they are informed, trained, and supported. Organizations must work to instill and support the self-discipline required to rein in the indiscriminate creation and retention of information.

Making the Cultural Leap

Controls for the creation, management, retention, and disposition of data have not kept pace with the ability to create and store it, opening the door for compromise of critical infrastructure systems through unmanaged unstructured data. Because employees have for decades been left to create and store data indiscriminately, a culture and practice of data hording proliferates. Yet most data that is not otherwise required to be kept loses its value in a relatively short time—as soon as one to two years (CGOC, 2013). Further, storing excess data can compromise the ability to find the most current and accurate version. Still, users often keep data "just in case" by default.

To be successful beyond the scope of a one-and-done information clean-up project, it is imperative that executive management lead the way to cultural change and lead by example. This means that not only must they set out expectations and guidance, they must themselves subscribe to the change, particularly since compromise of executive information poses the greatest risk.

Security Standards Support for Information Governance

Various security standards support the concept of information asset management as a key component of information security. Among these are the National Institute for Standards and Technology (NIST) Framework for Improving Critical Infrastructure Cybersecurity, Ver. 1.1 (NIST 2018); International Organization for Standardization 27001:2013, Information technology — Security techniques — Information security management systems — Requirements (ISO 2013); NIST Special Publication 800-53, Rev. 4, Security and Privacy Controls for Federal Information Systems and Organization (NIST 2015); and NIST Special Publication 800-37, Rev. 2, Risk Management Framework for Information Systems and Organizations: A System Life Cycle Approach for Security and Privacy (NIST 2018).

In the **NIST Framework for Improving Critical Infrastructure Cybersecurity**, **Ver. 1.1**, Asset Management is the first component in the Identify section: ID.AM. "The data, personnel, devices, systems, and facilities that enable the organization to achieve business purposes are identified and managed consistent with their relative importance to organizational objectives and the organization's risk strategy" (NIST 2018, 24), with references to Control Objectives for Information and Related Technologies (COBIT) and ISO 27001:2013, among others.

ISO 27001:2013 emphasizes the importance of information asset management. Among the 114 controls in Annex A is a section dedicated to Asset Management (A.8), and another focusing on Compliance (A.18). These sections address the orderly and compliant management of information assets throughout their lifecycle.

NIST SP 800-53, Rev. 4, speaks to security categorization: "The organization… [c]ategorizes information and the information system in accordance with applicable federal laws, Executive Orders, directives, policies, regulations, standards, and guidance" (NIST 2015, F-151). This process "facilitate[s] the development of inventories of information assets, and along with CM-8 [Information System Component Inventory], mappings to specific information system components where information is processed, stored, or transmitted" (NIST 2015, F-152).

The recently released revision of **NIST SP 800-37, Rev. 2** importantly recognizes need to prepare. Among other things, it promotes the need to:

- "Maximize the use of automated tools to manage security categorization; control selection, assessment, and monitoring; and the authorization process; …
- "Decrease the level of effort and resource expenditures for low-impact systems if those systems cannot adversely affect higher-impact systems through system connections; … and
- "Reduce the complexity of the IT/OT infrastructure by eliminating unnecessary systems, system components, and services — employing the least functionality principle" (NIST 2018, vii).

It further states that,

> "Recognizing that the preparation for RMF [Risk Management Framework] execution may vary from organization to organization, achieving the above objectives can reduce the overall IT/OT footprint and attack surface of organizations, promote IT modernization objectives, conserve resources, prioritize security activities to focus protection strategies on the most critical assets and systems, and promote privacy protections for individuals" (NIST 2018, vii)

This guidance is highly consistent with good recordkeeping and information governance practices, as discussed above. Several sub-sections of NIST 800-37 speak directly to the issue:

Asset Identification

Task P-10 requires *identification of assets that require protection*. Assets are defined as "tangible and intangible items that are of value to achievement of mission or business objectives," and include "mission and business processes, functions, digital information and data, firmware, software, and services. Information assets can be tangible or intangible assets and can include the information needed to carry out missions or business functions, to deliver services, and for system management/operation; controlled unclassified information and classified information; and all forms of documentation associated with the information system" (NIST 2018, 38).

Information Types

Task P-12 requires *identification of the types of information* to be processed, stored, and transmitted by the system. "Identifying the types of information needed to support organizational missions, business functions, and mission/business processes is an important step in developing security and privacy plans for the system and a precondition for determining the security categorization. NARA [National Archives & Records Administration] CUI defines the information types that require protection as part of its Controlled Unclassified Information (CUI) program, in accordance with laws, regulations, or governmentwide policies" ((NIST 2018, 39).

Information Life Cycle

Task P-13 requires *identification and understanding* of "all stages of the information life cycle for each information type processed, stored, or transmitted by the system . . . , typically characterized as creation or collection, processing, dissemination, use, storage, and disposition, to include destruction and deletion [OMB A-130]. Identifying and understanding how each information type is processed during all stages of the life cycle helps organizations identify considerations for protecting the information, informs the organization's security and privacy risk assessments, and informs the selection and implementation of controls. Identification and understanding of the information life cycle facilitates the employment of practices to help ensure, for example, that organizations have the authority to collect or create information, develop rules related to the processing of information in accordance with its impact level, create agreements for information sharing, and follow retention schedules for the storage and disposition of information.

"Using tools such as a data map enables organizations to understand how information is being processed so that organizations can better assess where security and privacy risks could arise and where controls could be applied most effectively" (NIST 2018, 40).

Categorize

"The purpose of the ***Categorize*** step is to inform organizational risk management processes and tasks by determining the adverse impact to organizational operations and assets, individuals, other organizations, and the Nation with respect to the loss of confidentiality, integrity, and availability of organizational systems and the information processed, stored, and transmitted by those systems" (NIST 2018, 46). "The RMF *Categorize* step is a precondition for the selection of security controls" (NIST).

Task C-2 requires that systems be categorized regarding impacts to organizational operations, organizational assets, individuals, other organizations, and the Nation, as well as the security objectives of confidentiality, integrity, and availability (NIST 2018, 48). Suggested categories include high, medium, and low impact systems, and the task suggests that further refinement is possible through prioritization of systems within the same impact level.

In the context of this discussion, categorization should extend beyond systems to include classes and groups of information, not only the systems on which they are stored. For example, a network share may be classified as low or medium impact for most data housed there, but there may, in fact, be highly sensitive or protected information co-mingled in that location. An important outcome of following the model above is to ferret out sensitive data that has been stored in open systems and either move it to more highly secured locations or dispose of it, leaving low impact data on low impact systems. If information is properly classified, technical controls may also be applied to enforce storage requirements.

System Disposal

Task M-7 requires that a system disposal strategy be implemented, to include execution of required actions when a system is removed from operation. Here, as above, the disposal task should be expanded to include disposal of data from systems that remain in operation. Certainly, when a system is removed from operation "[o]rganizations [should] ensure that controls addressing system disposal are implemented. Examples include media sanitization; configuration management and control; component authenticity; and *record retention* (NIST 2018, 83). *(Emphasis added.)*

It is evident that identification, categorization, and compliant disposal of information are important features of the risk management framework in NIST 800-37, as well as other standards referenced above. The most compelling arguments in favor of pursuing these tasks lie in the ability to mitigate risk associated with system ROT, to control costs, and to ensure information lifecycle management. There is an added benefit from operational efficiency gained in the ability to access the correct and most current information, as opposed to sifting through years of redundant or superseded records.

How to Get Started

1. Create rules for tools. Develop a legally-valid retention schedule to apply against information assets and understand the difference between legally-required retention and preservation for litigation. These concepts, though related, are different (see discussion above.) Be sure policies and procedures reflect the reality of data management requirements and that they are enforceable. Plan for "security by design" when considering new technology acquisitions by building in retention rules before data are created.
2. Address the human element. Training for information governance and security is critical, but its quality and impact must also be measured to be effective. Cultural "will" and the "tone from the top" will drive the success of IG initiatives. Be sure to secure executive support and consider offering periodic training.
3. Reach out to peers in other functions to find out what issues and challenges they face because of information glut. Look for synergies to gain a critical mass behind a request for change.
4. Look for opportunities to leverage triggers. It's hard to get started without a compelling argument. Look for that argument in litigation/e-discovery spend, regulatory audit findings, Board of Directors inquiries, and budget requests.
5. Resist the temptation to allocate budget for more unstructured storage. Instead, work with the legal, compliance, privacy, audit, and risk functions to establish and enforce the classification, retention, and disposition of information.

Conclusions

Organizations have for decades allowed data to proliferate unmanaged. The accumulation of ROT not only carries with it the cost of storage, but also creates tremendous security risks by increasing the footprint for compromise by both internal and external players. This paper offers a rationale and process by which ROT may be eliminated through a triage and disposition process that applies legally-validated retention rules, so that the remaining information may be categorized and stored using appropriately tiered controls.

The goal is to diminish the attack surface, while at the same time achieving improved regulatory retention compliance and reducing storage costs. The ultimate success of this approach is dependent, however, on a commitment to cultural change and on participation by all invested stakeholders, including executive management, legal, compliance, IT, privacy, risk, and audit.

Effective information governance will most certainly reduce security risk, enhance compliance, and minimize costs. Take the first step by building the right foundation.

References

AIIM, 2018. State of Intelligent Information Management: Getting Ahead of the Digital Transformation Curve. https://www.aiim.org/Resources/Research/Industry-Watches/2018/2018_May_2018-State-of-Intelligent-Information-Management.

CGOC, 2018. Information Governance Benchmark Survey 2018. https://www.cgoc.com/information-governance-benchmark-survey-2018/.

ISO, 2013. ISO 27001:2013, Information technology — Security techniques — Information security management systems — Requirements.

NIST, 2015. Special Publication 800-53, Rev. 4, Security and Privacy Controls for Federal Information Systems and Organization.

NIST, 2018. Framework for Improving Critical Infrastructure Cybersecurity, Ver. 1.1.

NIST, 2018. Special Publication 800-37, Rev. 2, Risk Management Framework for Information Systems and Organizations: A System Life Cycle Approach for Security and Privacy.

Saffady, William. 2004. Records and Information Management: Fundamentals of Professional Practice. ARMA International.

Shetty, S. 2017. How to Tackle Dark Data. https://www.gartner.com/smarterwithgartner/how-to-tackle-dark-data/.

Varonis, 2018. Data Under Attack: 2018 Global Data Risk Report From the Varonis Data Lab. https://info.varonis.com/hubfs/2018%20Varonis%20Global%20Data%20Risk%20Report.pdf.