The Role of Information Governance in Data Security

By John C. Montaña – ISSA member, Denver Chapter

This article discusses the need for inclusion of information governance in assessing the risk associated with implementing a data security framework such as the NIST Framework for Improving Critical Infrastructure Cybersecurity.

Abstract

This article discusses the need for inclusion of information governance in assessing the risk associated with implementing a data security framework such as the NIST Framework for Improving Critical Infrastructure Cybersecurity, and its role in the program resulting from implementation of the framework. Strong information governance is a necessary component of a comprehensive data security program, and omitting it from the program incurs significant risk and reduces the effectiveness of the program.

Data security incidents are becoming increasingly common—and costly—and organizations are spending increasing amounts of money and resources to combat them. The scope of the risk is likewise increasing. In the European Union very substantial penalties for data breaches are now being levied under the data privacy rubric (a data breach involving PII being by definition a privacy violation),[1] and data security is increasingly a regulatory compliance matter as administrative agencies promulgate regulations addressing it, with all the usual fines and other penalties for non-compliance.[2] In response, the security armory becomes ever more well-stocked with increasingly sophisticated tools. But often overlooked are the basic tools of information governance.

This can be a costly oversight. Data breaches and other data loss can often be traced back to poor information governance practices, and even when poor governance is not the primary cause, it often makes the consequences of a data security incident worse than they might otherwise have been. A good data security program, therefore, looks to sound information governance as one of the tools that should be included in that arsenal, and used regularly. But, this aspect of the landscape is often either overlooked or assumed.

Consider the US National Institute of Standards and Technology (NIST) Framework for Improving Critical Infrastructure Cybersecurity.[3] It expressly contemplates, in the "Identify" function, that an organization identify and inventory the devices, systems, data flows, and other data management characteristics that "enable the organization to achieve business purposes." [4] This identification step is a critical first step in the development of a security program because any subsequent activities are necessarily predicated on the ability to identify relevant information assets. Implicit is the idea that systems themselves can be identified and the information in them is sufficiently well-identified and organized to effectively protect.

In practice, however, this identification and inventory process is extraordinarily difficult for most organizations, and few, if any, organizations have full knowledge of their information assets. For many a substantial portion of their information assets are poorly identified, poorly organized, and as a result not well-positioned for implementation of data security controls. In an era where our information assets are measured in petabytes and exabytes, that can be a vast amount of poorly controlled data.

The NIST framework attempts to address this issue by reference to COBIT 5, ISA 62443-2-1:2009, and similar sources. But, like the NIST framework these are high-level structures, intended to provide a framework within which an effective program can be built. They do not attempt to articulate or enforce sub-processes such as information governance on a

[1] BBC, "British Airways Faces Record £183M Fine for Data Breach," BBC – https://www.bbc.com/news/business-48905907.

[2] "Cybersecurity Requirements for Financial Services Companies," New York State Department of Financial Services – https://www.dfs.ny.gov/docs/legal/regulations/adoptions/dfsrf500txt.pdf.

[3] "Framework for Improving Critical Infrastructure Cybersecurity," NIST – https://www.nist.gov/sites/default/files/documents/cyberframework/cybersecurity-framework-021214.pdf.

[4] Ibid, p. 19, table 2.

detailed level. Information governance and data management can therefore be thought of as specific tools and artifacts within these frameworks that facilitate the accomplishment of specific tasks that support strategic objectives.

What are information governance and data management?

At its most fundamental level, information governance includes basic organization and management. Absent the usual professional gloss, this means that information is identified, organized, and managed so that you can effectively and efficiently use it for other processes, including day-to-day business processes, audits and investigations, litigation and discovery, and more recently data security and data privacy. Much of these are very basic organizational processes and rules for day-to-day activities surrounding the data. Without these basic management processes in place and enforced the functioning of higher-level control frameworks is significantly impaired, sometimes disastrously so, for the simple reason that all of these high-level frameworks presuppose a substantial degree of low-level control of the information assets in question. Absent that control, the assumptions behind the framework fail. Common points of information governance failure include the following basic issues.

There's simply too much data in the first place

The average organization maintains far more information than it needs for business and compliance purposes. This is the result of several factors.

Unnecessary collection

First, most organizations collect information that's unneeded in the first place. There are an assortment of reasons for this, ranging from poorly thought through (and increasingly automated) information collection processes, to the simple human desire to capture and maintain information for just-in-case or CYA purposes, or the increasingly popular notion of mining the data for further value. But whatever the reason, now it's there, often poorly managed and frequently in vast repositories and potentially exposed to a data breach should one occur. Everything else being equal, twice as much data means twice the exposure in the event of a security incident. That in and of itself should give an organization pause when considering the risk assessment function of the NIST framework.[5] This is likewise a failing under many privacy laws that mandate data minimization, thereby increasing overall risk, regardless of an actual breach.[6] But, as we shall see, everything else is not equal, and other very basic failings make things worse for many organizations.

Keeping it too long

Second, organizations maintain far too much information for far too long. One of the basic tools of information governance is the retention schedule, or disposition schedule. In some cases this is a requirement written into law,[7] and this will increasingly be the case over time. It's pretty much what it sounds like: policy that sets forth the classes of information an organization maintains, along with retention and disposition rules for each class. An example might be:

- Personnel files – two years after termination of employment
- Customer orders – two years after fulfillment

The goal is that for each such class information is periodically purged and destroyed so that the information set being maintained for that class is bounded and controlled. The rule itself is derived from a combination of factors including:

- Legal requirements, often multiple legal requirements for one information class, particularly if the organization is multi-jurisdictional
- Risk management considerations such as audit requirements, statutes of limitation, and similar secondary legal authority
- Risk management considerations based upon the organization's risk history, risk tolerance, and other soft factors
- Business need and business utility
- The ability of IT systems to ingest and enforce retention rules

Inquiry often reveals that 40 percent or more of an organization's information set is past due for disposition, either because the organization does not have a retention schedule or because they do not enforce it.[8] And once again, more data necessarily increases the level of exposure in the event of a security incident.

Poor data organization

At its most basic level, data management is about developing and enforcing very basic tools such as file structures and metadata schemes, naming conventions for data objects, and other organizational tools. In the United States it is uncommon to find an organization-wide data structure scheme in place, even less common to see it vigorously and consistently enforced. Thus there is no institutional knowledge of the overall data assets of the organization, making the inventories contemplated by the NIST framework difficult or impossible to make, which in turn cascades down into all subsequent steps within the framework, which assumes that such a high-quality inventory is both possible and completed.

[5] Ibid., p. 22.

[6] "What Is the Minimum Necessary Rule in HIPAA?" HIPAA Security Suite – https://hipaasecuritysuite.com/2018/09/13/what-is-the-minimum-necessary-rule-in-hipaa/.

[7] Colorado Revised Statutes Title 6 Consumer and Commercial Affairs, § 6-1-713 Disposal of Personal Identifying Documents: "Each covered entity in the state that maintains paper or electronic documents during the course of business that contain personal identifying information shall develop a written policy for the destruction or proper disposal of those paper and electronic documents containing personal identifying information." Colorado General Assembly – http://leg.colorado.gov/sites/default/files/images/olls/crs2018-title-06.pdf.

[8] Ursula Talley, "Proactive eDiscovery: The Key to Reducing Litigation Risks and Costs," Information Security Today – http://ittoday.info/Articles/Proactive_eDiscovery.htm (DuPont study reveals 50 percent of documents produced for discovery in a lawsuit were past retention period); "Unused Data Is a Virtual Goldmine," Digital Signage Today – https://www.digitalsignagetoday.com/blogs/unused-data-is-a-virtual-goldmine/ (97 percent of data within organizations is unused).

Dark data and uncontrolled data

Then there's the matter of dark data and uncontrolled data. Most organizations have lots of it, in the form of badly or unmanaged share drives, uncontrolled SharePoint directories, and abandoned servers. In all of these cases, there is little or no institutional knowledge of what's in there, no meaningful ownership and control, and no real management going on. Again, the data volumes can be huge. In such situations it is understatement to say that inventory and risk assessment are difficult or impossible and that the outcome of a risk assessment is likely to be undesirable.

The net effect of these factors

As noted above, the NIST framework and other data security frameworks assume that a substantial level of identification and control over information assets is achievable or has been achieved. And the basic information governance failures noted above are present to one degree or another in most organizations. To the extent they are present, they impair or prevent the implementation of the NIST standard from initial prioritization and scope through risk assessment, gap prioritization, and plan implementation. If you can't find it or don't know what it is, you can't effectively manage it for any purpose. And if there's twice as much as there needs to be, everything you do is going to cost more.

Bottom line, uncertainty affects planning and budgeting and inevitably makes the implementation more expensive than otherwise it would have been. The net effect is at best to make things more expensive and complex than otherwise, and at worst to significantly impair planning and remediation processes. So, what should be done?

Inclusion of information governance within the inventory and risk assessment process

When conducting an assessment for the NIST framework or some other such framework, one will of course be presented with a legacy situation: *what is there is there, what is not is not*. At this stage, the goal should be to determine the presence or absence of sound governance principles and practices within the assessment process in order to make reasonable and accurate assessments of the extent to which poor information governance impacts risk and impairs achievement of data security goals. From the standpoint of documentary evidence, good governance would require the presence of at least the following:

- A standard data taxonomy covering most or all of the organization's information assets
- A retention schedule, the policy document that defines the retention periods for information assets
- Basic policies and procedures outlining and controlling basic creation, management, and disposition processes for data objects

However, the presence of these documentary artifacts alone is not enough. There is also an implementation question that must be tested and assessed. Is the taxonomy really being used? How widely? How consistently? Is the retention schedule being consistently implemented, particularly in places like large ERP systems that may not be designed or implemented in a way that facilitates sound management practices, or in unstructured repositories that are not susceptible to automated management processes? Are there large numbers of badly managed share drives, orphaned SharePoint sites, or abandoned servers? If any of these are present, they must be addressed at some point if the program contemplated by the framework is to be effectively implemented.

Goals, prioritization, and implementation

Within any large organization, the results of a thorough inventory and analysis are likely to be mixed—some data sets will be well-managed and well-controlled, some will be abandoned and uncontrolled, and much will be on a spectrum in between these two extremes.

A clear understanding of the overall situation will be necessary to address the goals, prioritization, and implementation phases of the NIST framework.[9] If, as is likely, substantial remediation must be performed, risk assessment, defined goals, budgetary and resource constraints, and the need to sequence work from a project-management standpoint will all require prioritization of the work involved.

These constraints may also require revision of goals and outcomes if possible and reasonable real-world outcomes make a desirable goal unattainable. Ideal remediation of large-scale legacy environments can quickly become cost-prohibitive, which often means falling back upon sub-optimal strategies involving some element of uncertainty and retained risk. For example, a dark data repository may be so large that detailed document-by-document analysis is cost-prohibitive. Alternative strategies could be sampling it, destroying it without analysis, or continuing to maintain it in its current state. Each strategy has significant downsides, unique to each, that must be considered in choosing the ultimate strategy for addressing it. And, the uncertainty and risk inherent in any of these approaches must necessarily be considered in overall risk analysis, goals, and outcomes to the extent that repository is relevant to them.

And because any information governance activities will be taking place within the larger framework of the data security initiative, the priorities for those activities must take their place within a larger schema of priorities that may reduce the priorities for some, or even eliminate them from the final plan. This may in turn require further adjustments to risk assessment, goals, and outcomes.

The bottom line

Information governance can't guarantee data security. It isn't a replacement for a good security program. But a well-managed data set is much easier to make secure, and if there is an incident, there's much less data at risk. Regardless of whether the security initiative is taking place within the framework of NIST, COBIT, or something else, sound, basic nuts-and-bolts information governance and data management is both assumed and demanded to drive an effective outcome for the overall security program. An information governance program should be an integral part of any data security solution. They're both facets of a larger set of issues, and both must be addressed as part of the solution.

About the Author

John Montaña, J.D. F.I.I.M, F.A.I.. is Vice President of Advisory Services at Montaña & Associates, an Access Company. His work includes analysis and advice on a wide variety of governance, compliance and management issues, including records retention scheduling, regulatory compliance, litigation and discovery, and he is recognized as one of the foremost information governance experts in the country. He may be reached at john.montana@montana-associates.com.

ISSA JOURNAL
GETTING STARTED WITH GHIDRA:
THE SOFTWARE REVERSE ENGINEERING TOOL FROM THE NSA

[9] "Framework for Improving Critical Infrastructure Cybersecurity," NIST – https://www.nist.gov/sites/default/files/documents/cyberframework/cybersecurity-framework-021214.pdf.