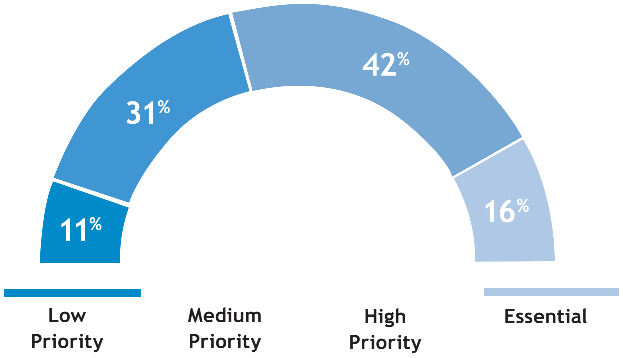
Data Classification Series Part 1

Better Data Security Through Classification: A Game Plan for Smart Cybersecurity Investments

Today, it is not uncommon for the true and core value of a state to reside in its data assets, specifically the information it collects, develops, and stores, and in the products it develops and sells that are comprised of the data, or derived from the data. We live in the information age. Information is the fuel for the engine that propels virtually every decision that is made in business today.

Data is risky business. Data is an asset. Data can truly make the difference between life and death for states. Take Indiana for example, who used data analytics to reduce the state's unacceptable infant mortality rate. It is no coincidence, then, why state chief information officers (CIOs) ranked data management and analytics—e.g. data governance; data architecture; strategy; business intelligence; predictive analytics; big data; roles and responsibilities—as a top priority for 2017. Additionally, in the National Association of State Chief Information Officers (NASCIO) 2016 State CIO Survey, 58% of state CIOs characterized data governance as essential or high on their strategic and operational plan.

Within the state CIO's strategic agenda and operational plans, how would you characterize data governance and management?



NASCIO Staff Contact:

Amy Glasscock, *Senior Policy Analyst*
Meredith Ward, *Senior Policy Analyst*

NASCIO represents state chief information officers and information technology executives and managers from state governments across the United States. For more information visit www.nascio.org.

201 East Main Street,
Suite 1405
Lexington, KY 40507
Phone: (859) 514-9153
Twitter: @NASCIO

Use and storage of this valuable data is not without risk. Data is at risk both from the inside and the outside. Employees share information without authorization, proper safeguards are not put in place, systems malfunction, phishing e-mails are sent and clicked and outside hackers infiltrate systems. A large data breach results in the compromised personal information of citizens, and costs the state time, money and the trust of citizens.

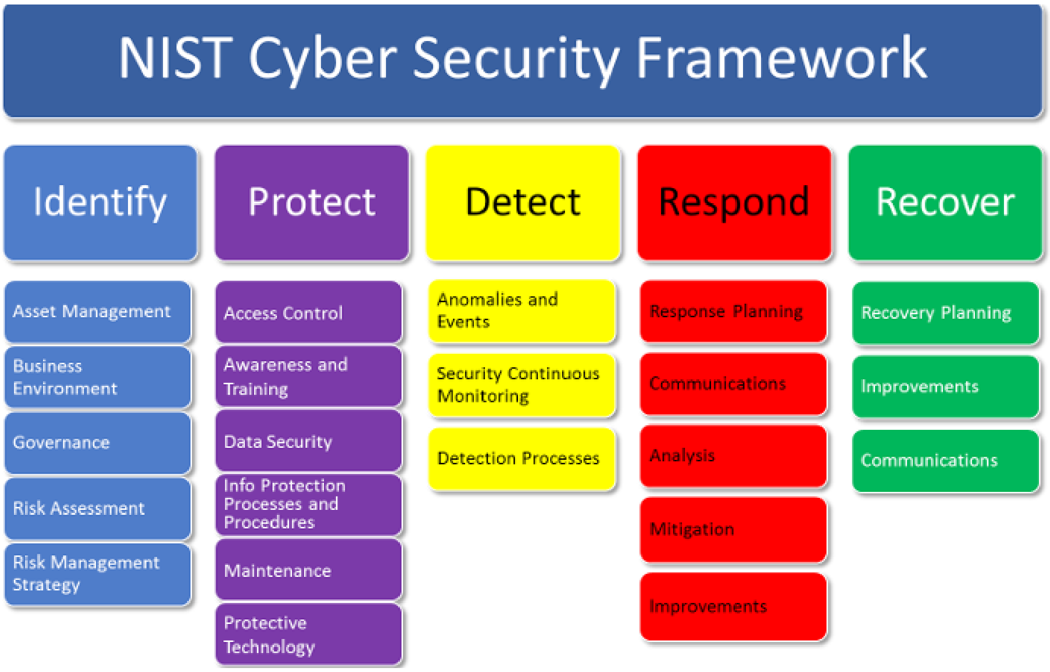
As states are collecting more and more data, citizens are feeling increasingly vulnerable. In a recent Accenture study 70% of U.S. citizens said they are concerned about the security and privacy of their personal digital data, while two-thirds would feel more confident if government agencies had better data-privacy and security policies in place.

The benefits of data classification may not be obvious to everyone, so understanding the benefits, and communicating them, will be a key success factor in getting this critical initiative underway. The purpose of data classification, and its “bottom line” attraction, is to prevent “injury” and to save money. “Injury” can be financial, reputational or psychological. A breach prevented represents a significant “intangible savings” to any state, beyond the obvious avoidance of potentially huge breach response expenses. Reputational integrity, trust, confidence and the resultant willingness to collaborate and partner with an entity, can hinge on the record it offers of responsible stewardship of its data assets. The damage to a client or constituent caused by exposure of their protected health information and personally identifiable information can be real and enduring, and the focus of legal actions. Money can also be saved with reduced storage and management costs.

State CIOs rank data management, analytics, data governance, data architecture and data strategies as a top priority, and could leverage these disciplines to advance cybersecurity’s data classification efforts. Cybersecurity covers many things from infrastructure, firewall management, system monitoring, network vulnerability and threat management, however, it is not mature in the data management space. Therefore it is important to find ways to make sure data management disciplines are incorporated in the roll-out and implementation of cybersecurity’s data classification. Data classification must also be founded at the enterprise level for cybersecurity.

For many states a proactive approach to data security is not just a good idea, it is the law. Many states already have data security laws on the books. Forty-eight (48) states, the District of Columbia, Guam, Puerto Rico and the Virgin Islands have enacted legislation requiring private or government entities to disclose and notify individuals of security breaches when the information involves personally identifiable information. At least 31 states have laws requiring entities to destroy, dispose, or otherwise make personal information undecipherable. Thirteen (13) states require private companies to maintain certain security procedures to protect personal information.

Risk assessments have not been widely implemented in states, according to the National Governor's Association (NGA) Meet the Threat campaign. However, "Identify and Document Asset Vulnerabilities" is a key step in establishing or improving a cybersecurity program, as outlined in the National Institute of Standards and Technology (NIST) Cybersecurity Framework. The NIST Framework advocates a risk based approach. It states, "it is important that organizations seek to incorporate emerging risks and threat and vulnerability data to facilitate a robust understanding of the likelihood and impact of cybersecurity events." For data security, a key reference for the NIST Framework is FIPS 199—Standards for Security Categorization of Federal Information and Information Systems.



NIST Cyber Security Framework

Identify	Protect	Detect	Respond	Recover
Asset Management	Access Control	Anomalies and Events	Response Planning	Recovery Planning
Business Environment	Awareness and Training	Security Continuous Monitoring	Communications	Improvements
Governance	Data Security	Detection Processes	Analysis	Communications
Risk Assessment	Info Protection Processes and Procedures		Mitigation	
Risk Management Strategy	Maintenance		Improvements	
	Protective Technology			

When states take a risk based approach they improve operational efficiency, assessments are more accurate, attack surfaces are reduced and decision making is improved. Taking an enterprise mentality brings together previously silo-based security and IT tools and allows for ongoing and continuous data monitoring and assessing. States can benefit from an automated and closed loop process based on risk. While these functions were traditionally addressed through access management, data management and data architecture roles, they would gain much more strategic traction collaborating with these roles and functions instead of creating yet another silo. It is important to look at these intersections strategically to incorporate them into cybersecurity and data management.

Likewise, the NIST Framework risk assessment is knowing where data is and knowing how to protect it. NASCIO also named data the "lifeblood" of state government in 2015. So, data is important. Identifying, organizing and classifying data, then, becomes crucial to the quality and integrity of a state's data. So crucial that, in the 2016 CIO survey, 71% of CIOs report that they have established standards for data

State Policy

Ohio's Data Classification Policy states: "data classification is a process that identifies what information needs to be protected against unauthorized access, misuse and the extent to which it needs to be secured and controlled. Each agency shall serve as a classification authority for the data and information that it collects or maintains in fulfilling its mission."

classification and security. But, data classification is no easy task. The goal of this guide—a joint effort between NASCIO's Cybersecurity Committee and Privacy and Data Protection Working Group— is to provide a startup framework for the identification and classification of a state's data. The guide has been broken up into two parts plus a guide with the elements necessary for a data classification policy that states can use.

There is certainly much that this guide cannot cover, and many questions will remain for further exploration and research after this guide has been reviewed. The intent of this guide is to present a point of embarkation into the data classification exercise, and to reveal some of the complexities, as well as the benefits and rewards of data classification. Once data is classified, there are additional steps that can and should be taken to realize all the benefits of classification, so this guide is intended to provide a path into what can become a "life-cycle" type of exercise, repeating at periodic intervals on into the future, rather than a project that becomes final, or "completed." Systems and system data continue to change, so classification of the data must be updated in order to remain accurate and useful.

Perhaps the most important thing to keep in mind, is that data classification must be part of the overall cybersecurity enterprise architecture. Data classification must be done on multiple levels, in multiple agencies and on an ongoing basis. The framework must be in place to make it work.

Before we get started, let's define what data classification means. Ohio's Data Classification Policy states: "data classification is a process that identifies what information needs to be protected against unauthorized access, misuse and the extent to which it needs to be secured and controlled. Each agency shall serve as a classification authority for the data and information that it collects or maintains in fulfilling its mission."

The Value of Data

Information may be the one element in a business that can truly not be "unexposed" once it is breached (loss of confidentiality), replaced once it is lost (preservation of availability), or repaired once it is improperly altered (maintained integrity). **Confidentiality, availability, and integrity** are the three pillars of the information security (C.A.I.) triad. Maintaining C.A.I. is the ultimate and overarching goal of data asset protection endeavors in every state.

Types of Data: Critical and Sensitive (and Everything Else)

We track a wide array of activities, relationships, contacts, appointments, personal information, family information, transactions, events and more. The rate at which the data is being collected continues to grow with increasing velocity each day. Within this aggregation of data collected, there is certainly a considerable amount of

State Policy

The State of Arkansas provides Data and System Classification Grid Guidelines, the purpose of which is "for agencies to examine the data in their information systems, determine its sensitivity or criticality to the agency's functions, and then determine the appropriate level of security to apply to the information technology systems."

"junk." However, there is also a vast amount of extremely important data. Data must be protected primarily because of these characteristics: its necessity, confidentiality, or sensitivity. That is why it is best to start by identifying what information needs to be protected (Secude).

Data can be identified in several ways:

- **Critical data** is so necessary that in its absence important business cannot continue normally, e.g. property records for county governments or voter registrations for state governments.
- **Sensitive data** implies that if it is obtained by or exposed to the wrong people, the outcome can be harmful to persons, e.g. tax records or bank statements.
- **Protected Health Information (PHI)** includes a large amount of personal medical information that could lead to discrimination if it is revealed publicly or to a malicious person.
- **Personally Identifiable Information (PII)** is generally information collected by financial and similar institutions which, if compromised, can lead to financial harm like identity theft and other associated harmful outcomes.

While these are some of the ways data can be organized, it is important to note that in addition to the above categories, states collect additional data that doesn't fit into any of the categories above. Nevertheless, this data requires storage space, and is maintained at a cost. Identifying the life-cycle for this data, and understanding when it can be purged, is a goal of the data classification exercise.

Documenting the Characteristics of Your State's Data: Data Classification

To adequately protect a state's data, the state must understand what data it possesses, and then take appropriate measures to protect sensitive data based upon its level of sensitivity. Extremely sensitive data, of course, deserves commensurate levels of protection.

The State of Arkansas provides Data and System Classification Grid Guidelines, the purpose of which is "for agencies to examine the data in their information systems, determine its sensitivity or criticality to the agency's functions, and then determine the appropriate level of security to apply to the information technology systems."

Data classification is the exercise required to categorize data according to its value, and sensitivity. Until a state has its data classified, there is no way to adequately protect it, or even to understand how much protection is adequate. Further, in the aftermath of a disaster, unless the data asset is well understood, the ability to recover computing systems in a sequence that is most beneficial to the state is not possible. Business continuity planning, the discipline of organizing the re-assembly of a business function in the aftermath of a business disruption, cannot hope to succeed if the data involved is mysterious in any way. Disaster recovery, the re-creation of a computing environment to support the business application and function, cannot proceed in a correctly prioritized manner, if the highest value business functions, and the data associated with them, are not identified and backed up, or mirrored, in such a way that their restoration can be accomplished quickly, systematically and effectively.

Assigning the Correct Level of Protection Requires Information About the Data

Data classification enables the state to align security controls and levels of protection to data in accordance with its business value, as measured by business criticality (needed for the business to function and survive) and sensitivity (the extent to which the loss or compromise of the data can impact the state and/or its customer/constituents adversely). This may sound unimportant, but improper access to data that is sensitive can cost millions of dollars, and even more in terms of intangible or tangible business or political reputation.

Knowing what data is collected and stored, and knowing where the data is, can allow a state to properly care for it with a level of “due care” that is matched to its criticality or sensitivity (“due care” is a legal term that describes the effort expended to protect something that “due diligence” has led the state to understand the full value of). Who would ask someone to handle dynamite, unless the person had been trained adequately on the use of explosives? Due care requirements may be placed at all levels of a state.

Risk Mitigation

Data classification is a foundational pre-requisite to comprehensive risk mitigation, because data classification gives the state tangible information that is critical to meaningful risk assessment activities. The link between data classification and risk assessment is irrefutable.

Enterprise Foundation

A pre-requisite to data classification is that it must be founded at the enterprise level. There must be joint decision making and collective action at all levels and agencies. Data classification in state government must be guarded by an enterprise policy. Different agencies cannot have different policies.

Proper Data Disposal

Data classification can identify sensitive data that no longer needs to be stored. By properly purging a state’s sensitive or confidential, but unneeded, data, that source of potential breach is eliminated, reducing risk. Elimination of all unneeded data reduces steadily increasing storage costs. However, it is important to comply with records management and the state’s archival policy. Also keep in mind data sharing agreements may be active and not visible to the data classification process.

Classify on Mobile Devices – Location is a Risk Factor

Data classification can include the identification of data that is known to be collected by, or ported to mobile devices, thus allowing a state to make informed decisions about how to manage that mobile device and the data in motion that it contains. Who should carry a device with highly sensitive data on it, how strong does the device password need to be? Is the data encrypted? Can the state afford for the data to be unencrypted? Can the device be remotely wiped? Will it self-wipe if too many wrong passwords are attempted on it? What provisions need to be made in advance for the device if would be lost or stolen? Should the data on it have time limits for residence on the device? Is the data backed up often enough to recover the most current data, etc.?

Getting Started

State Buy-In and Top-Down Support

Data classification in a state is virtually impossible to accomplish without the support and assistance of others. Again, this is where the importance of enterprise architecture governance comes into play. If data classification isn't part of a state's cybersecurity enterprise architecture, it will not be successful.

As with many initiatives with broad scope, if there is lack of buy-in and support from the top of the state, data classification stands little chance of success. True for the small states, and even more true in larger states. The reason for this is that all but the smallest of states consist of multiple business functional areas, and each of these areas collects, creates, stores and processes its own type of data. The inputs and outputs (and life-cycle) of the data may be known only to those who work in that functional area, and perhaps only to a small subset of those who work in the area. It is safe to generalize that there is no single person, in any state, who knows all about all the data. So, to accomplish data classification, cooperation from a number of people is essential. However, these people have competing priorities, so the initiative will need that leader, or someone with broad authority to enact the data enterprise architecture, and to instruct all the people in the state who have knowledge of that data – information that the initiative needs – to cooperate in responding to queries about their data. Another pragmatic approach would be to leverage the data management and data governance discipline to implement data (protection) classification.

The Game Plan – Identify the "Players"

The initiative will be, in effect, taking a survey of the state's data. The surveyor will need to get to know who the data "owners" or "managers" are, or will need to work with people who know these data people, and can elicit the needed information from them. Database administrators, programmers or other technical experts are often those who know most about the data. This person can provide information about the data's characteristics, and how it is stored, shared, protected, duplicated (backed-up for protection against loss), and will know what a data record "looks like," how many records of each record-type are stored, how many records are added, deleted or change every day, and if the data is encrypted.

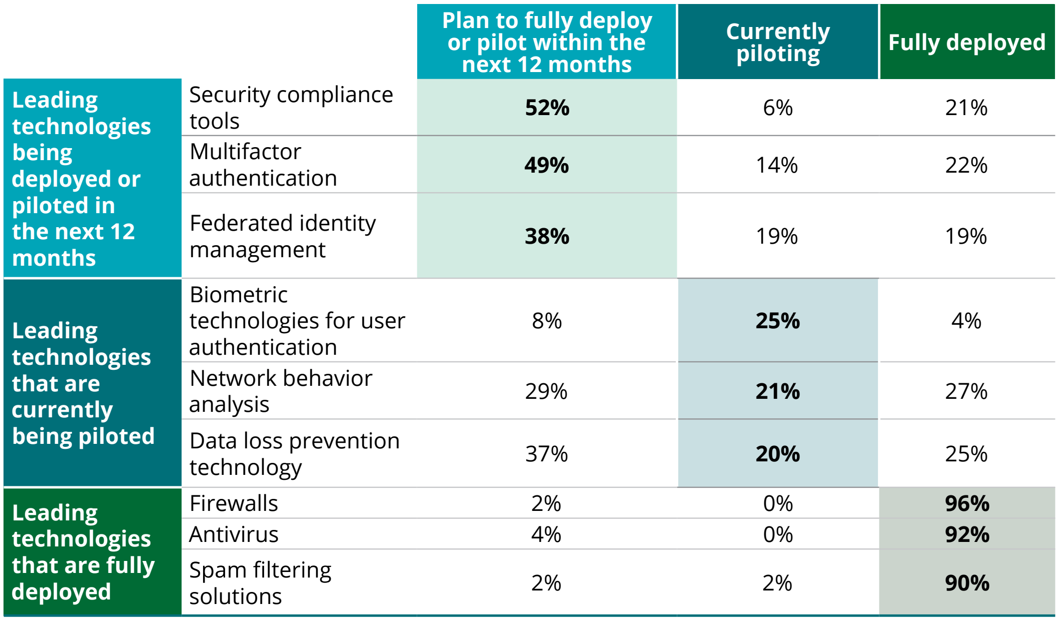
It is also necessary for the policy experts and owners of the data to be involved. Someone such as the director of the business function is the person who really knows how critical certain data is to the overall business function. While the director may know nothing about a record, or the details of its storage and organization in the computer system, the director may be the person who can say how long the business can survive if this, or any specific information, is not available for a prolonged interval. Additionally, states will want to involve records managers, legal experts, archivists and other policy professionals.

The takeaway is that the initiative's success depends on the cooperation of people at many levels of the state, and for this reason it is important to the success of a data classification effort to know who these people are, and have people with adequate authority supporting and/or sponsoring the data classification initiative. It is also important that a formal project management discipline be overlaid on this exercise to make sure that the endeavor is defined in manageable increments, that all the right people needed for the success of the venture are identified and advised that their participation and cooperation are vital to the success of the exercise, and that their support is expected from the state's leadership.

Other Drivers – Understand the Cascading Benefits of Data Classification

It is a fact that many derivative benefits come from knowing what and where specific data is. Risk assessments often depend on the metadata gathered in a data classification initiative. Often, compliance initiatives rely on risk analysis that depends on data classification. It can be quite beneficial to understand the initiatives that will derive benefit from the data classification exercise. It may well be that once this is fully understood, the buy-in and support at all levels becomes much easier to secure.

Once a data classification architecture is in place and is being implemented, data is better protected. Per the 2016 Deloitte-NASCIO Cybersecurity Study, 45% of states have fully deployed or are implementing data loss prevention (DLP) technology. Another 37% plan to fully deploy or pilot this technology in the next 12 months. It doesn't make sense, nor is it possible for states to spend limited resources protecting all data equally. Classification allows states to prioritize where to use DLP technology and where to implement other protections.



		Plan to fully deploy or pilot within the next 12 months	Currently piloting	Fully deployed
Leading technologies being deployed or piloted in the next 12 months	Security compliance tools	**52%**	6%	21%
	Multifactor authentication	**49%**	14%	22%
	Federated identity management	**38%**	19%	19%
Leading technologies that are currently being piloted	Biometric technologies for user authentication	8%	**25%**	4%
	Network behavior analysis	29%	**21%**	27%
	Data loss prevention technology	37%	**20%**	25%
Leading technologies that are fully deployed	Firewalls	2%	0%	**96%**
	Antivirus	4%	0%	**92%**
	Spam filtering solutions	2%	2%	**90%**

Source: 2016 Deloitte-NASCIO Cybersecurity Study.

Graphic: Deloitte University Press | DUPress.com

Looking Ahead

As previously stated: data is important, and, data classification is not any easy undertaking. If a state simply cannot immediately take on the exercise, there are some things that can be done right away:

- Use the **NIST Cybersecurity Framework** as a roadmap
- Incorporate **Data Governance and Data Architecture** activities into the state's cybersecurity data classification. Metadata management is already part of data management and this also seems to be a good place to embed data classification.
- Incorporate a cybersecurity data classification practice through new projects to show value.
- Leverage your **Project Management Office (PMO).**
- Use data classification when enhancing and upgrading systems.
- Leverage ITIL's **Change Management**
- Create a **Data Sharing Agreement** (NASCIO's data sharing agreement brief).
- Adopt a **Records Management** archival process.
- Set forth clear **Access Management**: Defined through roles, group access, and Service Oriented Architecture (SOA).
- Leverage and incorporate the National Information and Exchange Model (NIEM), a data exchange used by federal agencies, states and local governments.

States and agencies are at different maturity levels regarding data classification. But, in short, any state or agency can find a pathway that makes sense and get started.

We addressed several issues here regarding getting started in the process of identifying, organizing and classifying data. But what are the policy implications of such? How do you engage and train others? Those topics and others, including guidelines for creating your own data classification policy, will be addressed in part two of this brief.

Contributors

NASCIO would like to thank the following individuals for their contribution to this white paper series.

Tom Baden, Commissioner and Chief Information Officer, State of Minnesota

Sallie Milam, Chief Privacy Officer, State of West Virginia

Rita Reynolds, Chief Information Officer, County Commissioners Association of Pennsylvania

Ellena Schoop, Enterprise Data Architect, State of Minnesota

Josh Spence, Chief Information Security Officer, State of West Virginia