



雛 ZNYSFORUM

Session Overview

Objective of the session is to build a foundational understanding of best practices in data classification.

So many questions to answer, including:

- What is data classification?
- Why does data classification matter to your organization and to whom?
- What should you be aware of before embarking on a data classification effort?
- And more...

Data Classification: Definition & Benefits

Data Classification Definition

The process of organizing data by relevant categories so that it may be used and protected more efficiently

Involves applying a variety of tags to data

Eliminates multiple duplications of data

Data Classification Benefits

Simplifies location and retrieval of data within your organization

Supports your risk management, compliance, and data security efforts

Reduces storage and backup costs while expediting the search process

Though the classification process may seem technical, it is an important topic that should be <u>understood by your organization and its leadership</u>. A simple data classification scheme that is well-implemented and provides real benefits to the organization.

雛 ZNYSFORUM

Data Classification as a Practice

Data classification often involves applying a multitude of tags and labels that may define the type of data, its sensitivity, its integrity, and its availability; those factors affect the security measures put in place to protect each classification level.

Three (3) main types of data classification that are considered industry standards:

Context-based

1.

Inspects and interprets files looking for sensitive information.

2.

Content-based

looks at application, location, or creator among other variables as indirect indicators of sensitive information.

3. User-based

depends on a manual, end-user selection of each document. Userbased classification relies on user knowledge and discretion at creation, edit, review, or dissemination to flag sensitive documents.

Content-, context-, and user-based approaches can be both right or wrong depending on the business need and data type.

NYSICA 2019 Spring Conference

Importance of Data Classification

Appropriate classification of data can meaningfully contribute to your efforts in the areas of regulatory and policy compliance, IT security, and overall organizational risk management. In the absence of data classification, your agency may not:

- Be able to identify data whether data you hold affects your compliance obligations
- What IT security measures are most appropriate based on your data, how it is classified, and where different types of data reside; don't under or overdo it
- Have appropriate entitlements set up in your systems and processes to control data access as needed
- Understand how to limit risk exposure associated with data you are handling

NYSICA 2019 Spring Conference

雛 ≝NYSFORUM

Organizational Roles: IT Executive

CISOs and information security leaders must:

- Create an organization-wide data classification scheme that is simple to understand, yields consistent classification, and is applicable to structured and unstructured data.
- Support data classification projects with the policies, processes, and technical tools that are necessary to meet the project's stated objectives.
- Ensure that data classification processes and tools take into account the full lifecycle of your data, and that controls remain effective and appropriate if and when classification changes.
- Start with focused initiatives that address funded business needs (i.e. regulatory compliance) and have well-defined project scopes.

Organizational Roles: Others

Users

翻 NYSFORUM

• If users don't understand the classification scheme, they will either not use it or use it inconsistently; both are bad and will undermine the program.

Management

 Management support and guidance on priorities for a data classification program are key elements of its success. Understand how the classification will support their objectives for the organization on the whole.

Partners and customers

 Any external entity you conduct a form of business or data exchange with may be impacted by your approach to data classification, from needing to add tags to data, reformat files being exchanged, etc. – to on the positive side knowing you are taking steps to protect their confidential appropriately.

Vendors

 Like other partners and customers, there may also be an impact of a data classification program on existing vendors handling your data – sometimes with a cost – to be mindful of from the outset. Vendors may also play a role in implementing and maintaining your data classification through products and services.

NYSICA 2019 Spring Conference

翻HNYSFORUM

Data Classification Programs Overview

Through 2020... 75% of organizations implementing data classification will report limited deployments and tangible benefits; regulated industries and governments will continue to represent the majority of the data classification success stories.

- Data classification projects are the evergreen projects of most organizations perceived as important but often not successfully completed or applied.
- They often fail or stall because they are overly complicated. If the data classification schemes are
 not intuitive to the layperson, they will be inconsistently applied (if at all), waste time and
 resources, and provide a false sense of security.
- Organizations that have practical data classification programs focus on a few key subtle but extremely important – points based on a shared understanding of program goal(s).
- Data classification programs should create large categories of similar data and establish policy regarding the proper use and applicability of various protection profiles for each category.
 - The classification schemes, policies, and procedures inherently establish the scope of the overall initiative, and the effort and resources required to support it.
 - Successful classification initiatives almost always include senior management support, business users' involvement, training, technology, and funding; this applies to the initial setup, day-to-day operations, and the long-term maintenance of the program.

NYSICA 2019 Spring Conference

Challenges & Misconceptions

Making it overly complex

雛 ^世NYSFORUM

• Complex and ambiguous data classification schemes will be poorly understood, difficult to apply, and unlikely to be sustained over time.

Equating data classification with data security

• Organizations may have a false sense of security because they assume that having a classification policy or applying a tag/label to a data item will protect them.

Lack of maintenance

• Information security leaders often apply data classification in a static set-and-forget approach that does not reflect the data life cycle.

Neglecting operational impact

• Organizations often define classification requirements without taking into account the ongoing business impact and the volume of historical information involved.

Underestimating the effort

 As with any significant undertaking, a classification program will only be effective with adequate effort and priority behind it. Organizations typically underfund training and education related to data classification, resulting in poorly understood and implemented programs.

NYSICA 2019 Spring Conference

羅 署NYSFORUM

Best Practices

- **Understand your data**: It is important to have a grasp on the data from the outset; absent that a classification program may ineffective or even disruptive.
 - · What will be included in the program and why
 - · What characteristics does that data have
 - What are the usage patterns associated with the data and for what purpose(s) and audience(s)
 - What data is in use versus at rest
- Predetermine program goals: A data classification program must be tied to concrete goals understood by management, IT, and relevant departments; if not there is a risk that the program will fail to achieve them and not be of any use at all.
- Consider leveraging technology: Successful data classification projects are well-supported with a blend of
 policies, processes and technical tools. While ongoing training and education initiatives are often the most
 important elements in implementing a data classification program, tools can assist with automating data
 classification, making the program more sustainable over time.
 - Will leverage characteristics about the data itself or a context related to its life cycle to determine and assign the appropriate data classification
 - Reduces classification errors through consistent application of rules built into the tool, such as through the use of qualifiers and disqualifiers to assign a classification to a dataset

NYSICA 2019 Spring Conference

雛≌NYSFORUM

Best Practices (cont.)

 Follow a methodical process: starting checklist; assigned resources (e.g. PM); identify the data and its sources; discover they relevant details; classify the data; secure the data; monitor ongoing maintenance of the program – including tools, users, and application of classification as new data sets or systems are identified. 12

- Keep it simple: If the terms used to identify different classifications are in any way near-synonyms, then change them. If the delineation between the different classifications is ambiguous, redefine each classification in terms of business impact. Resist the urge to expand the classification scheme.
- **Examine a simple three-classification approach**: If the focus is compliance and/or intellectual property, a variation of this has been used for many successful programs as it is easy to understand and train on.
 - "Public" This is data that is published on your publicly facing website or other official external communications, such as social media feeds and various product collaterals.
 - "Internal" This data is for internal use only, but consists of your routine business communications and documents created as part of your normal, day-to-day activities. This includes the majority of internal email.
 - "Confidential" This data is your sensitive data that typically requires special handling procedures. This can be data subject to regulations, intellectual property, or information that is not publicly known or available internally, such as HR data, etc.

WESTORIUM Additional Detail: Decoupying data classification initiatives, providing automation and controls that can increase the overall fidelity of a classification initiatives, ule and offer efficiency required to make the program effective and sustainable. Examples of potentially useful technological tools

E-discoverv

software

Encryption solutions

Note these tools can significantly improve the viability of an ongoing program, but in all cases human involvement is required for overall success, especially in the initial phases when the organization has yet to achieve classification process maturity.

Data-centric audit

and protection

solutions

Enterprise content

management

systems

NYSICA 2019 Spring Conference

Enterprise digital

rights management

翻HNYSFORUM

Additional Detail: Applying qualifiers

- Qualifiers are characteristics that increase the accuracy in the processes of identifying the targeted data by verifying the content matches, known syntax or values.
- As an example, if the structure you are looking for comprises known credit card values that have been entrusted to you by your clients, qualifiers could include:
 - The occurrence of 16-digit numbers
 - The 16-digit sequence adds to a specific verification value
 - The 16-digit number occurrence is within close proximity of text that matches names and addresses, syntax and values, and so on
- Disqualifiers are characteristics that increase the accuracy by verifying that the content does not
 match known syntax or values that have been identified as outside the actual dataset.
- Continuing the above example:
 - 16-digit numbers that are found in the headers and footers of documents have been found to be corporate document tracking numbers, and should be disqualified
 - 16-digit numbers that are in the engineering files have been found to be part numbers, and should be disqualified
 - 16-digit numbers that match the previously listed qualifiers, but have a total of fewer than five unique occurrences, are not subject to regulatory compliance and would otherwise cause many false positives based on known use cases

雛 ZNYSFORUM

Conclusion

- Data classification projects can address just about any use case imaginable but only with the right mix of business support, technology integration, time and resources.
- Organizations must assess their own individual level of maturity, senior support, available resources and patience before attempting to implement more advanced data classification schemes.
- Focus data classification deployments on specific and funded business issues; resist the urge to try and accomplish too much and too early.
- Data classification efforts require additional staffing and budgets for technology, process implementation, training and support – and thus need senior management support.
- Ultimately, the key to a successful data classification project is to look at the big picture, but also to start slowly and build experience with different data owners, gathering internal references and momentum over time.

The NYS Forum, Inc. Demystifying Data Classification



Sajid Qureshi Spruce Technology (703) 989-0652 squreshi@sprucetech.com

翻 NYSFORUM



16

17 The NYS Forum, Inc. **Artificial Intelligence: How Data Analytics** Manages Risk



John Maynard, CPA Fraud & Risk Specialist

Agenda

Artificial Intelligence – What is it? Form & Function

ronn & runction

Data Analytics Lifecycle From data to insight Where AI plays a role

Al Government Use Cases Where and how is Al used





器 ≝NYSFORUM

What is Artificial Intelligence (AI)?

<u>AI</u>

- Artificial Intelligence
- AKA "Narrow AI"
- Programmed defined parameters
- Cannot apply known to unknown
- Good at simple goals, big data
- More like machines



<u>AGI</u>

- Artificial General Intelligence
- Sentient & self-aware
- Understands self and others
- Can apply learned info to new area
- Conquers more complex goals
- More like humans





Key AI and Analytics Concepts

What is Deep Learning?

翻 ^{||} **NYSFORUM**

Deep Learning is a type of machine learning used in recognizing speech, identifying objects in images and more.



NYSICA 2019 Spring Conference



Key Al and Analytics Concepts

What is Computer Vision?

翻 ^{||} **NYSFORUM**

Computer Vision enables systems to see, identify and process images or videos in the same way that human vision does



NYSICA 2019 Spring Conference

Key AI and Analytics Concepts

What is Predictive Analytics?

翻 NYSFORUM

Predictive Analytics is the use of data, statistical algorithms, and machine learning to identify the likelihood of future outcomes based on historical data





Key Al and Analytics Concepts

What is Forecasting & Optimization?

翻 ^{||} **NYSFORUM**

Forecasting is predicting future needs for a product or service, while **Optimization** is maximizing results within a set of constraints





- Virtual Assistants
- Translators
- Chatbots

鐵 置NYSFORUM

- Personalized Shopping
- Facial Recognition
- Autonomous Cars & Drones
- Medical Genome Mapping
- Assisted Medical Diagnosis

NYSICA 2019 Spring Conference







30



翻 ^{||} **NYSFORUM**











Procurement Integrity

Risk Management - Traditional

• Ethics training

ﷺ≝NYSFORUM

- Internal controls
- Annual audits
- Whistleblowers



Procurement Integrity

< rillers	Letity Type	Enity Name	Entity No.	Entity id	Actionablescore	No Of Alerts	Faded .	Setwork Score Guster	id No Of Network Nodes	C Direct Link
	Vendor	Biglitted	204091	4430574062	183,25930711	1,982	1.1	3 23	120	3
Earlity No.	Vendor	Farmer Jack	200388	4520225013	18.967559052	171	13	2 33	120	
• Emility Id	Vendor	Plunkett Home Furnishings	200661	4120130541	52,854476574	672	- 13	3 64	11	3
+ Emility Type	Vendor	Eliman S Catalog Showroome	202430	4060558435	35.952929644	381	12	2 23	120	
All (1713)	Vandor	Buckeye Funniture	205308	4110242767	35.170400049	232	12	3 3	37	
Employee (852)	Vendor	Lional Playworld	202708	4320123030	38.000963084	243	11	2 25	120	2
D User (24)	Vendor	Exect Solutions	201763	4300556872	8.6048807905	22	(R. R	3 11	10	
Weder (2000)	Vendor	Mr Ag S	200655	4790584534	18.382107368	64	10	1 1.8	1 5	
	Sundor	Carden Master	200579	4800642580	58.55477492	677	100	3 64	25	5
 Enity Name 	U vendor	KainbowLfe	203010	4900039030	14.003532358	25	10.	2 1.9	4 4	2
 Direct Links 	Sendor	Creat Western	208293	4700263035	15.294947299	333		0 0	0	
No Of McDarock Hodes	i 🛄 Nendor	Bumper To Bumper Auto Parts	203340	4205511390	18.000150357	55	1] 0 0		4
ChaterM IncreaseN Score Falled Texts Ho Of Alexts	undor	Your Putters is Now	205683	42005559855	8.5584079057	63		3 33	120	1
	uwedor	The Arraw Pottery	200.063	4520260506	6.3259795686	13	2	2 1.0	5 5	2
	Vendor	Mariberix	203458	4090536004	14,803139455	35		2 52		2
	Vendor	Custom Sound	202026	43005455332	27.490053422	139	5	0 0		
	D stendor	Sold seaf Sarder Hana	202307	4202544494	6.7549793596	221	100	1 1 1	6 3	1
	A stendor	Kenny Ropers Rosebers	206119	4100195819	3.3814985542	18	BINK	1 50		-1
	Vandor	Strong Life	203225	4320085256	13.133301807	229		2 25	2 4	2
	Vendor	Cohman S	225249	4550264853	42.080072558	18	Contract of the local division of the local	2 1.0	1 5	2
	-									

Analyze



Detect



Investigate



ﷺ≝NYSFORUM

Artificial Intelligence





Risk Management w/ AI + PA

- Continuous monitoring
- Predictive analytics scoring
- Social Network Analysis (SNA)
- Data at auditor's fingertips







雛 営NYSFORUM Child Safety Detect Investigate Analyze Risk Management w/ AI + PA Continuous monitoring Predictive analytics scoring Highest risk cases handles first Artificial Social Network Analysis (SNA) Intelligence Data at caseworker fingertips NYSICA 2019 Spring Conference









翻 NYSFORUM

John Maynard, CPA SAS Institute 919-531-5184 john.maynard@sas.com

LinkedIn: https://www.linkedin.com/in/johnmaynardcpa/

Twitter: https://twitter.com/JohnMaynardCPA, @JohnMaynardCPA



44