SPHERE
TECHNOLOGY SOLUTIONS

Data Governance: Unstructured Data

Whitepaper: Implementing a Functional Strategy

This whitepaper outlines in detail the methodology behind creating a robust Data Governance strategy, with insights on what should be included, how it should be implemented and other key areas.

Table of Contents

Executive Summary

Data governance is one of the most business-critical disciplines to have emerged within enterprises over the last decade. Through data governance, organizations are looking to exercise positive control over the processes and methods used to manage their data repositories.

Data Governance is a fairly broad term that can relate to a variety of disciplines. One work stream of Data Governance is a focus on unstructured data. Unstructured data can be anything that resides on your system, from basic Excel spreadsheets on file systems and e-mail in mailboxes to large training videos stored in public folders. Due to the sheer volume of this data, maintaining control becomes a daunting task.

Over the years, many organizations have turned a blind eye to the massive amounts of unstructured data that resides within their infrastructure. But, more recently organizations are being asked to provide insights into the state of their "unstructured data". This includes providing metrics around how much data exists and how it's used, identifying any security issues with access to this data and quantifying the extent of any vulnerabilities. In addition to this, organizations are being tasked with acting in a more proactive manner, looking at their existing policies and processes and learning if these meet the needs of effective governance and control. Also, companies are beginning to understand that this is a difficult problem to solve but you have to start somewhere; high risk data is a good place so reporting on sensitive data is typically a starting point for many.`

Much of this is due to the results of a failed audit, a breach they've experienced, or even more simply, a precautionary measure to avoid one of these scenarios. With the public "shaming" that occurs when a major firm experiences a breach, companies are aware of the repercussions due to reputational damage.
The recent and high-profile events that have struck major organizations include inappropriate access and exfiltration of PII, ransomware attacks and attempts by hacktivists and nation states to disrupt business operations. These will only increase in number and severity.

In order to mitigate any of these potential issues there needs to be a deep and thorough analysis of the environment to uncover any major risks, as well as, to create a baseline of how much is required to remediate the environment. A Target Operating Model needs to be designed and all the necessary policies and processes must be implemented to ensure there are controls in place for ongoing management and governance. Once the proper controls are being implemented, a work stream dedicated towards remediation activities must be incorporated to stop the bleeding and standardize the data and access controls.

To start, the following are key questions IT departments are being tasked to answer:

- ***What data exists and where is sensitive data exposed?***
- ***Is there data that can potentially be eliminated or removed from the managed system?***
- ***Who has access to access to what and where are controls problematic?***
- ***How is the provisioning and managing of access currently done?***
- ***Who owns what data?***
- ***Where are the risks and how do i prioritze them?***
- ***Are there areas where is there a lack of compliance?***

While many firms are scrambling to kick off preliminary investigations to understand their current state, many are discovering the amount of critical issues they are facing in terms of compliance, security and overall lack of oversight of the vast and vastly growing data.

This is a concern at the most senior levels within corporations. “87% of Board Members and C-level executives have said they lack confidence in their organizations’ level of cybersecurity.”[1]

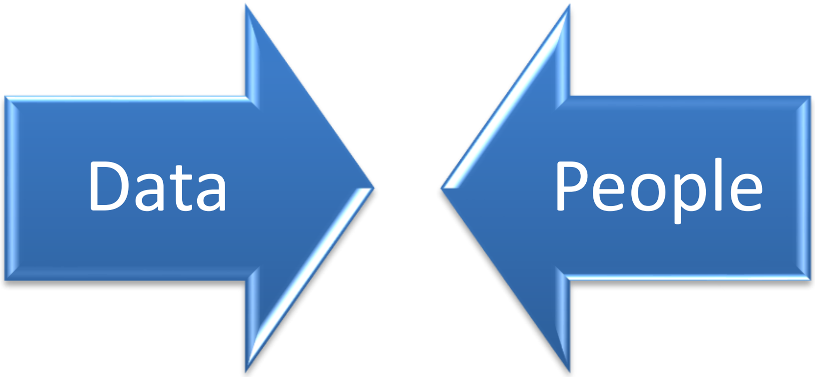
Not surprisingly, there is an increased need to have a handle on where information exists and how it is accessed throughout an organization. The dangers from both external and internal threats are only increasing. Without a well-defined and well-executed governance plan, the ability to respond to audits, ensuring that data is secure causes aggravation, and in many cases elevated fines for too many organizations.

Governance Framework

Creating a robust governance strategy can seem like an incredibly overwhelming activity. Instead of starting with complexity, focus instead on taking a more simplistic approach and building on top of that.

Process is the key to effective governance.

The first stage of creating an effective process is to define the policies your organization must adhere to. This set of “rules” creates the outline of how processes and procedures are to operate and under what premise.
As it relates to Data Governance, developing policy, and then processes, is understanding that there are two basic building blocks that need to be clearly understood:



There are many types of “data” that exist as assets of an organization. Unstructured data, as opposed to semi-structured and structured, is typically most out of control. Whether it’s email in a user’s mailbox, alerts that are sent to an Exchange Public Folder, PowerPoint presentations on a SharePoint site, or the data stored in the most commonly used file shares, all of these repositories can prove to be a risk to the firm if not properly governed

The first step in defining your framework is to define the scope of your data governance policies.

As we start taking our deep dive into the data, we focus on the second building block around “people”. We always find that there is a surprising amount of data that is easily accessible to people who shouldn’t have access to it. It is all too common to easily find the CEO’s helicopter schedule and learn that it is open to the entire company or compensation data that thousands of people can easily read and sometimes even modify! With Enterprise Search

[1] EY ’s 19th Global Information Security Survey 2016 - 17

solutions being leveraged, employees can find information without even having to know where to explicitly look for it. What if a disgruntled employee or an external vendor used this information for an improper purpose?

Thinking about governance in terms of Data and People will allow you to create policies that are unique to your organization, but also ensure that you are focusing on the right areas of governance.

Still, there is no cookie cutter approach. No one size fits all.

If it's going to work, it must be tailored around how your business operates. If creating a Data Governance framework was something that could be easily created or a simple template to be downloaded, companies would not invest this much time and effort into these programs. The reality is that true Data Governance takes time, understanding that as you embark on your mission you may need to adjust and reprioritze, ches as you understand more about how your organization can operate.

The remainder of this document focuses on how we take the basic concepts of Data and People and use them as stepping stones for building a robust governance program.

Assessment

The first, and almost always overlooked, stage for creating an effective Governance program, is knowing what you have on hand. Having the necessary visibility, understanding the Data and the People and how they relate, will allow the creation of policies and processes that make sense for your unique organization. Corporate culture and the desire of your business community to modify their day-to-day activities is something that must be considered. The strategy must be adjusted based on how cooperative different areas of the business are expected to be.

The Data

Data is the crux of an organization, probably one of the most critical assets your company owns. And, this is not new; it's always been the case. Over time, these critical assets have grown in size and number, and understanding all the metrics and statistics around your data is a pertinent component of the Assessment we undertake in order to move forward with the appropriate strategy.

Critical Questions To Answer

Where does the data reside? How is the data distributed across servers or regions or business units?

How do I organize and categorize the data? Where do I start grouping data into "collections"?

How do I make sense of the millions of folders and present them in a logical way? Who will need to see this information and how will they need to see it?

Are there mechanisms to identify sensitive data? Can I start to isolate areas where there may exist data that needs enhanced controls?

Is the data stale or active? Is it being accessed or modified? Is it a risk to maintain stale data longer than required by the regulations?

Should it be moved to an alternative repository? Is a certain repository required to be phased out i.e. Exchange public folders?

What metadata should I extract? What tools do I have to collect and report on this data about my data?

Being able to accurately answer these questions is vital for the creation of a substantive policy and a way to enforce certain behaviors. While attempting to answer these questions, you will naturally discover many organizational aspects that you may not have known in the past. For example, you may start to see that there are tooling requirements that you may or may not have deployed. Or, you may need insights from different business units to create definitions before you can accurately answer the above questions i.e. what type of data is considered sensitive?

The end goal is to protect your most important assets, reduce the risk of a data breach, and minimize your exposure. Starting with these foundational components will set the stage for ongoing governance needs long term. It will also allow you to bring together all the institutional knowledge related to technology and business needs that are fundamental in beginning the analysis, and more importantly, putting in a place a system to manage governance requirements long term.

As seen in the Ponemon Study “Global Insights on Documents Study”[2] companies focus on protecting information, not the IT stack. Sixty-seven percent of respondents (33 percent + 34 percent) agreed that the protection of information is more critical than the IT stack. This finding indicates that organizations are recognizing the value of their information assets.

The People

Understanding the “People” is the second half of assessing your unstructured data in an effort to create an effective Governance strategy. This area is usually more dynamic, but a baseline is still required. Focusing your assessment now on understanding the “Who” as opposed to the “What” will help deliver a full assessment and provide all the required information to move forward with creating an appropriate strategy.

[2] Ponemon Institute Research Report “Global Insights on Documents Study”, June 2014

Critical Questions to Answer

Who has access to what? How was that access provided? Is it a unique permission or inherited?

Is access open to all users? Is access open to an excessive number of users? Is there inappropriate or non-standard access?

Was the last access warranted? Is the access to data a common activity for a specific user or is this behavior abnormal?

Who owns the data? Is it the person who uses it the most or the manager of the people that use the data? Who is the true authoritive source for changes?

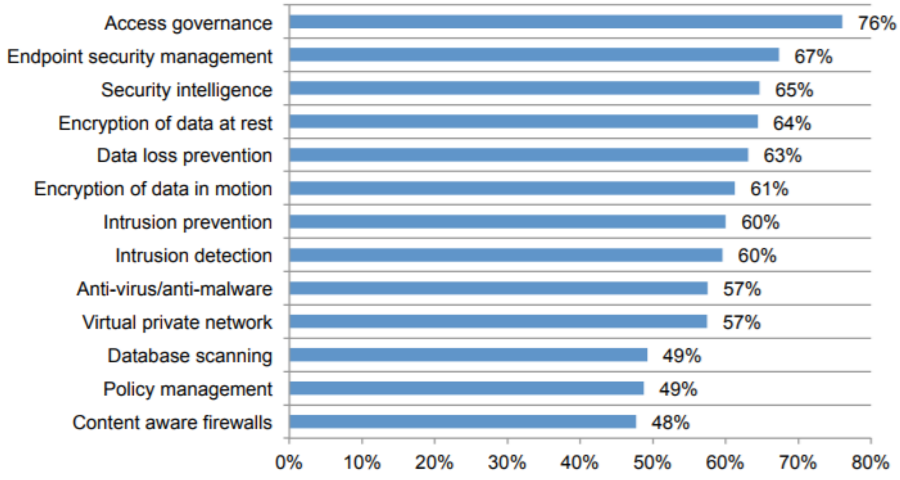
Do I have the controls in place to determine policy violations? Are your policies only reactive? How can I prevent policy violations?

Who would be affected by our strategy or a remediation of existing permissions? What regions are these users in? Any executives we need to avoid?

Having a clear and concise understanding of how access is provided and utilized, who owns the data and what end-users would be affected by any changes, is of utmost importance. These are fundamental in being able to prepare for dealing with the dynamic aspects surrounding access controls.

It isn't uncommon for permissions to require updates; individuals join, change and leave different functional areas regularly. So, having a baseline of what access controls look like now and a way to determine where permissions are inappropriate is fundamental. At the same time, you must also have a way to constantly refresh this information so that you can see improvement over time.

And as it relates to Data Governance, implementing effective access controls is a fundamental step in data security and a proactive approach to protecting sensitive or confidential business information. As per the Ponemon study titled, "The Human Factor in Data Protection"[3], access governance is most important in preventing a data breach:



[3] Ponemon Institute Research Report "The Human Factor in Data Protection", January 2012

Additionally, the following table was included in the study showcasing specific security and governance procedures that are most important. Notice managing entitlements and privileges is most important and therefore most effective.

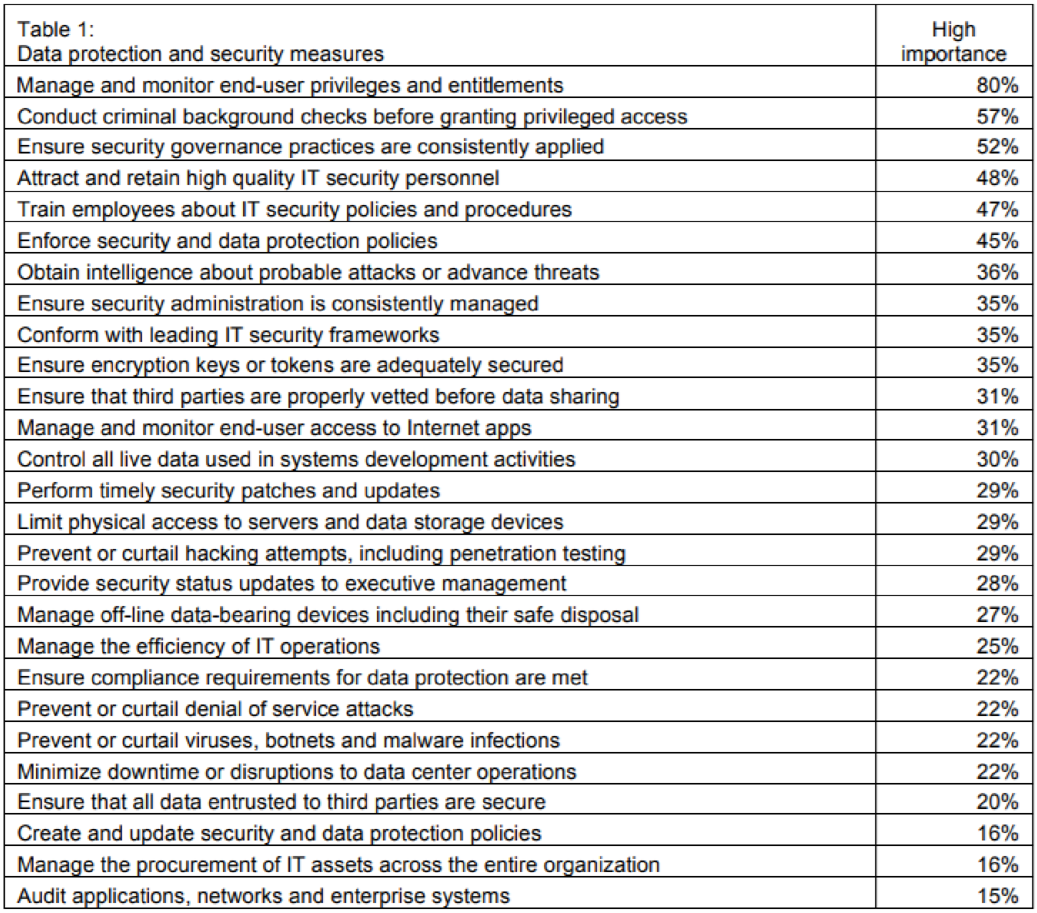


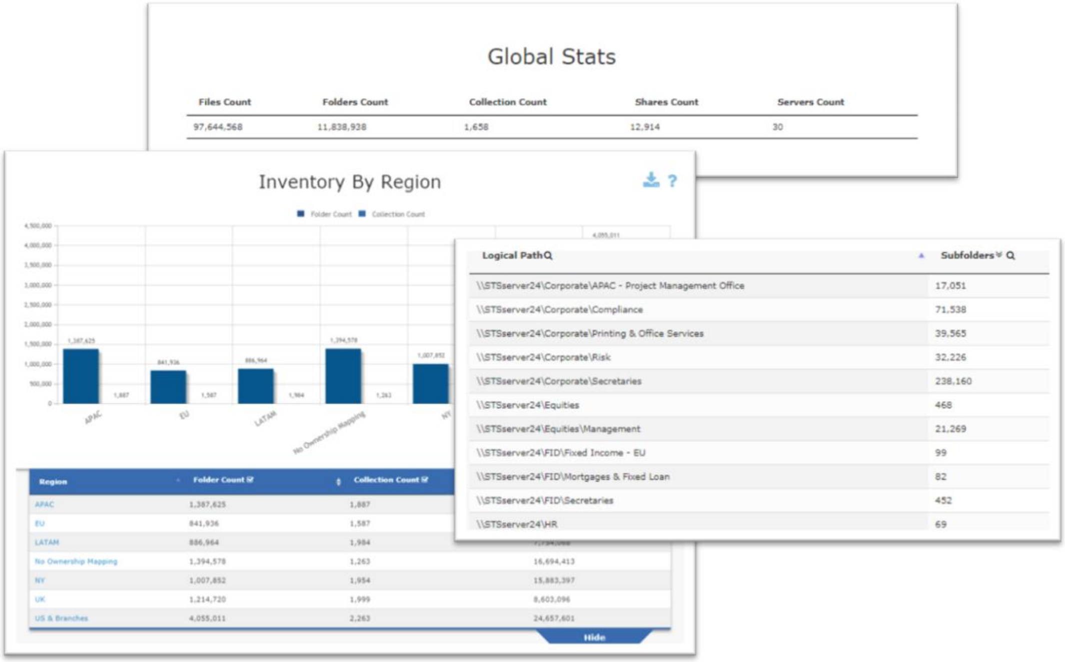
Table 1: Data protection and security measures	High importance
Manage and monitor end-user privileges and entitlements	80%
Conduct criminal background checks before granting privileged access	57%
Ensure security governance practices are consistently applied	52%
Attract and retain high quality IT security personnel	48%
Train employees about IT security policies and procedures	47%
Enforce security and data protection policies	45%
Obtain intelligence about probable attacks or advance threats	36%
Ensure security administration is consistently managed	35%
Conform with leading IT security frameworks	35%
Ensure encryption keys or tokens are adequately secured	35%
Ensure that third parties are properly vetted before data sharing	31%
Manage and monitor end-user access to Internet apps	31%
Control all live data used in systems development activities	30%
Perform timely security patches and updates	29%
Limit physical access to servers and data storage devices	29%
Prevent or curtail hacking attempts, including penetration testing	29%
Provide security status updates to executive management	28%
Manage off-line data-bearing devices including their safe disposal	27%
Manage the efficiency of IT operations	25%
Ensure compliance requirements for data protection are met	22%
Prevent or curtail denial of service attacks	22%
Prevent or curtail viruses, botnets and malware infections	22%
Minimize downtime or disruptions to data center operations	22%
Ensure that all data entrusted to third parties are secure	20%
Create and update security and data protection policies	16%
Manage the procurement of IT assets across the entire organization	16%
Audit applications, networks and enterprise systems	15%

Coupling all this information we've learned around the "People" with what we've learned about the "Data", gives you the full 360 degree view of your unstructured data landscape.

Sample Reports

As we start to review the above questions, we can identify the various tools that can provide different aspects of the answers, and then we begin to compile all this raw data. Next are a variety of reports showcasing different sets of analytics.

Reports – What do I have? The first set of reports should provide a high-level overview of what exists within the in-scope data set. This should take the concept a few steps further by organizing your data into meaningful "collections", assigning ownership to these "collections" and being able to slice-and-dice the data into categories mimicking how your organization manages the data i.e. region or department.

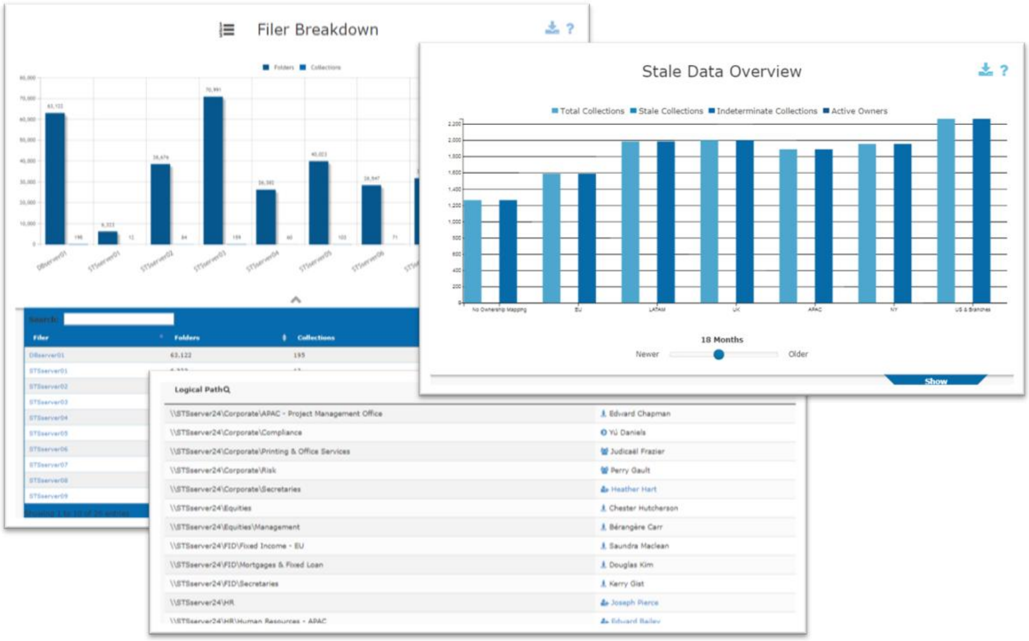

Global Stats

Files Count	Folders Count	Collection Count	Shares Count	Servers Count
97,644,568	11,838,938	1,658	12,914	30

Inventory By Region

Region	Folder Count	Collection Count	
APAC	1,387,625	1,887	
EU	841,936	1,587	
LATAM	886,964	1,984	[illegible]
No Ownership Mapping	1,394,578	1,263	16,694,413
NY	1,007,852	1,954	15,883,397
UK	1,214,720	1,999	8,603,096
US & Branches	4,055,011	2,263	24,657,601

Logical Path	Subfolders
\\STSserver24\Corporate\APAC - Project Management Office	17,051
\\STSserver24\Corporate\Compliance	71,538
\\STSserver24\Corporate\Printing & Office Services	39,565
\\STSserver24\Corporate\Risk	32,226
\\STSserver24\Corporate\Secretaries	238,160
\\STSserver24\Equities	468
\\STSserver24\Equities\Management	21,269
\\STSserver24\FID\Fixed Income - EU	99
\\STSserver24\FID\Mortgages & Fixed Loan	82
\\STSserver24\FID\Secretaries	452
\\STSserver24\HR	69


Filer Breakdown

Search:

Filer	Folders	Collections
DBserver01	63,122	195
STSserver01	[illegible]	[illegible]
STSserver02		
STSserver03		
STSserver04		
STSserver05		
STSserver06		
STSserver07		
STSserver08		
STSserver09		

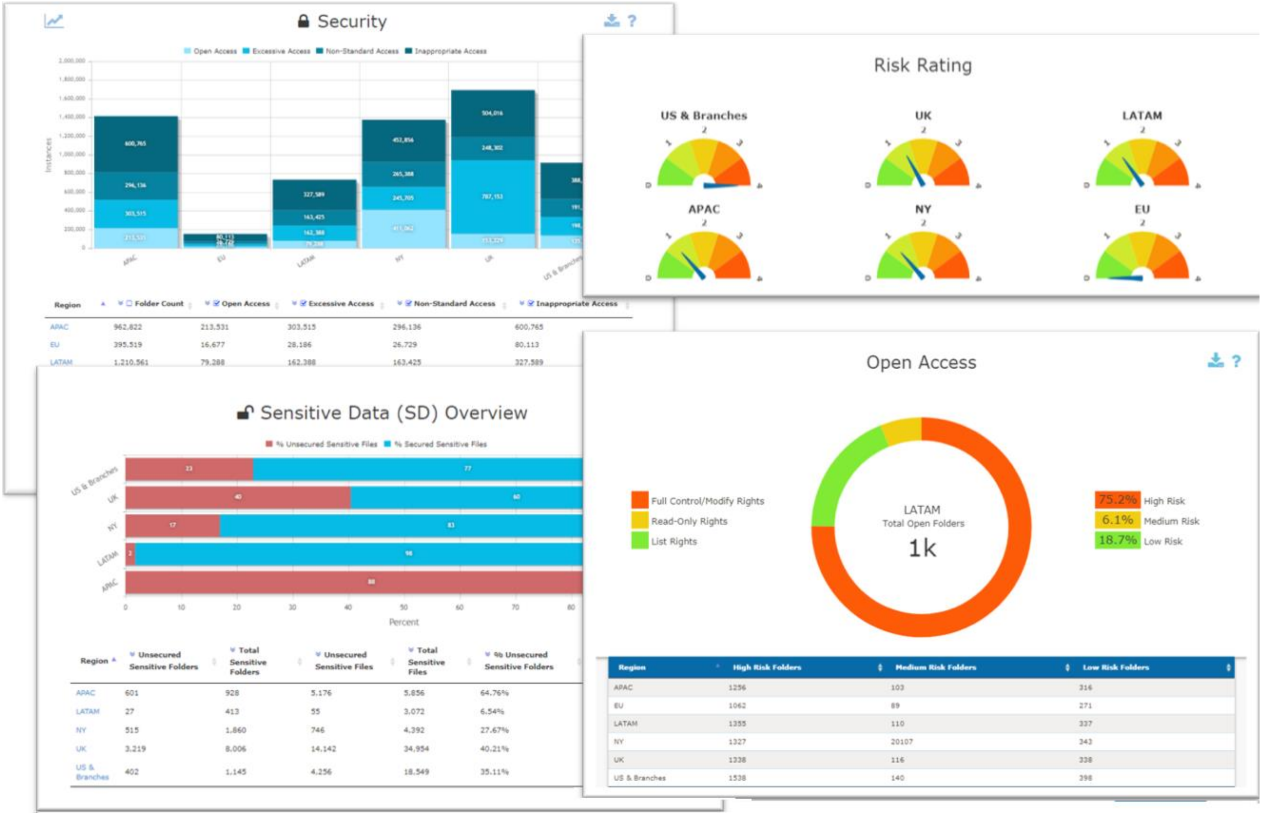
Showing 1 to 10 of 25 entries

Stale Data Overview

Logical Path	
\\STSserver24\Corporate\APAC - Project Management Office	Edvard Chapman
\\STSserver24\Corporate\Compliance	Yú Daniels
\\STSserver24\Corporate\Printing & Office Services	Judicaël Frazier
\\STSserver24\Corporate\Risk	Perry Gault
\\STSserver24\Corporate\Secretaries	Heather Hart
\\STSserver24\Equities	Chester Hutcherson
\\STSserver24\Equities\Management	Bérangère Carr
\\STSserver24\FID\Fixed Income - EU	Saundra Maclean
\\STSserver24\FID\Mortgages & Fixed Loan	Douglas Kim
\\STSserver24\FID\Secretaries	Kerry Gist
\\STSserver24\HR	Joseph Pierce
\\STSserver24\HR\Human Resources - APAC	Edvard Bailey

Reports – Where are my issues?

The next set of reports should provide details on where your security issues exist within the unstructured data platforms. This can include open access, excessive access, inappropriate access and non-standard access. Additionally, information regarding unsecured sensitive data and risk rating is critical as well.



Region	Folder Count	Open Access	Excessive Access	Non-Standard Access	Inappropriate Access
APAC	962,822	213,531	303,515	296,136	600,765
EU	395,519	16,677	28,186	26,729	80,113
LATAM	1,210,561	79,288	162,388	163,425	327,589

Region	Unsecured Sensitive Folders	Total Sensitive Folders	Unsecured Sensitive Files	Total Sensitive Files	% Unsecured Sensitive Folders
APAC	601	928	5,176	5,856	64.76%
LATAM	27	413	55	3,072	6.54%
NY	515	1,860	746	4,392	27.67%
UK	3,219	8,006	14,142	34,954	40.21%
US & Branches	402	1,145	4,256	18,549	35.11%

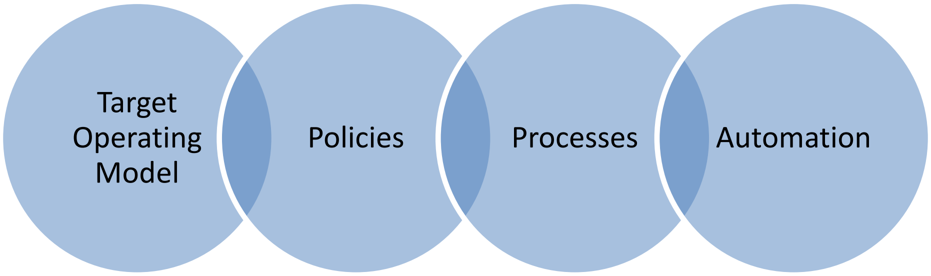
Region	High Risk Folders	Medium Risk Folders	Low Risk Folders
APAC	1256	103	316
EU	1062	89	271
LATAM	1355	110	337
NY	1327	20107	343
UK	1338	116	338
US & Branches	1538	140	398

Strategy Creation

Now that we understand the landscape and where our risks lie, we can start thinking about how to tackle these problems in two ways:

- Developing a desired end state
- Remediating all the issues that already exist

When developing the strategy it's important to focus on many key decisions that must be made early on. Scope, use of existing tools vs. evaluating new tools, who will service what part of the project, etc. are the areas that need to be considered with an end goal to have some completeness in the following areas:


Target Operating Model

Policies

Processes

Automation

Target Operating Model Recommendations

Each component plays a critical role in the overall governance framework. Below are these components:

- Control Standards: The standards that dictate how data is to be managed
- Administrative Roles and Responsibilities: Who can manage what, what tools and processes are used
- Lifecycle: What is the creation, update and deletion process
- Exception Management: How are exceptions handled
- Process Integration: How are ancillary processes integrated with management
- Enforcement: How are deviations to the standard reported, reviewed and resolved

Policy and Process Creation

A set of standard business-as-usual functions need to be outlined, including both the expectations of the personnel who will take part in the processes and the methods they should be using to accomplish specific tasks

- Provisioning of file share access
- Management of access lifecycle
- Scheduled attestation or reviews
- Enforcement of policies
- Manage ownership changes
- Retiring access, stale data and unused file shares
- Integration of automated file share management system

Automation Opportunities

Identify potential solutions to automate the management and maintenance of data. This includes, but will not be limited to, suggestions for the implementation of policies, procedures and technology.

Remediation

Now that you've identified the risks in the environment, the Target Operating Model, you need a plan to go from point A (messy) to point B (clean). How will you get there? It is a good idea to perform a pilot remediation to understand the complexities your organization presents and how to best handle them. Also, this will give you a sense of how long a full remediation will take

This can be a short-term project, on a subset of data, but will represent the organization as a whole. Take one area that is identified as a high risk, such as your HR's group shares, and assess the data, identify the strategy to resolve the issues found and use the remediation options as outlined in the next section, to prove the plan as a justifiable course of action.

This tactical remediation should include:

- Developing a workflow required for the broader remediation
- Identifying the responsibilities across operational teams regarding the remediation process
- Documenting the steps of the process using a swim lane diagram
- Providing recommended communications to be used during entitlement reviews and remediation
- Documenting the entire process for use across the broader environment
- Communication to the associated ownership with remediation requirements
- Completion of these remediation activities on an agreed upon subset of the environment

- Implementing the controls to ensure the environment does not go back to the pre-remediated state

It is also imperative that real change is implemented to show immediate risk reduction. While these changes may be minor in scale, the teams aligned with this program must see the technical steps necessary to complete remediation work, such as:

- Modify inheritance model where applicable
- Apply new groups to affected directories; where required
- Ensure only authorized users have access
- Remodel permissions to follow standards
- Resolve Broken ACLs and broken inheritance

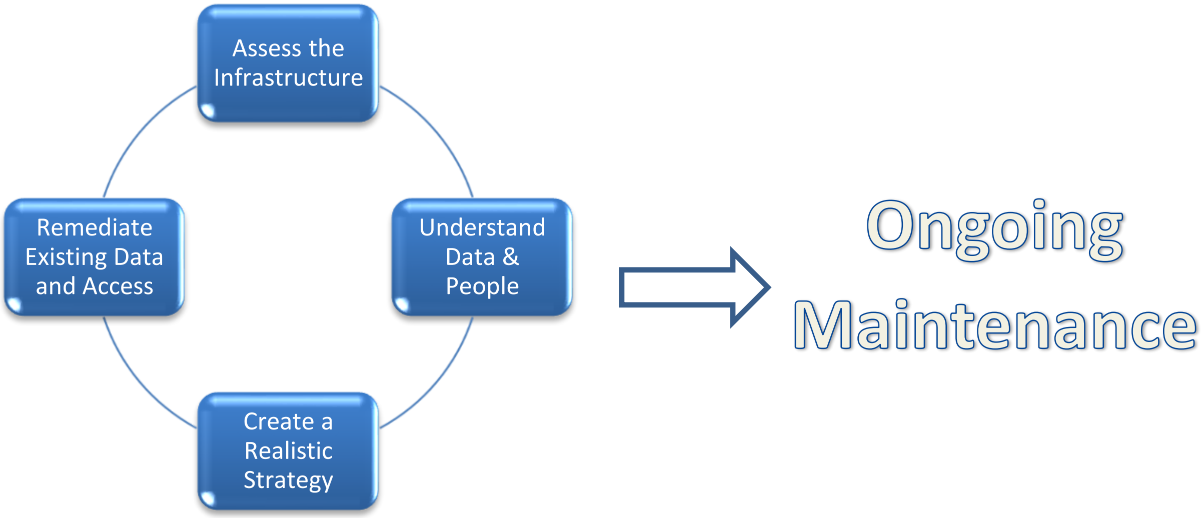
This tactical remediation will essentially provide the "script" for how the larger remediation effort will be undertaken, as well as, provide a guide for ad-hoc remediations that will be required on an ongoing basis as new issues pop up. Most often, there are many different people involved in any one remediation ranging from data owners, folks who generate reports and provide guidance, as well as, administrators who physically make the changes. It must be a well-coordinated effort in order to be successful at scale.

Summary

Not all companies have the same needs for compliance, but all companies have a need for security, therefore a need for a governance policy. We are in a world where data is only going to continue to grow; knowing where it is, who has access and what is being done with it needs to be understood. Whether for compliance or security or both, companies must have a plan in place to deal with their information. Data is a critical asset and needs to be protected.

These projects can be daunting, especially for companies that have never taken the steps as presented here. But, it is imperative that companies, large and small, start now before the issue gets completely out of control. There is no such thing as a perfectly governed environment. But, having the appropriate policies in place and adhering to them goes a long way to mitigating any issues that may arise through a data breach or loss.

Most importantly, there needs to be processes in place for ensuring that all the remediation you've done does not go to waste. Make sure there are clear processes for ongoing maintenance, including entitlement reviews, access authorization workflows, infrastructure reporting, etc.


Assess the Infrastructure
Understand Data & People
Create a Realistic Strategy
Remediate Existing Data and Access
Ongoing Maintenance

About SPHERE Technology Solutions

SPHERE Technology Solutions is an award-winning woman-owned cybersecurity business focusing on improving security and enhancing compliance. We put controls in place to secure your most sensitive data, create the right governance processes for your systems and assets, and make sure companies are compliant with the alphabet soup of regulations surrounding this space.

Our Services – Ranging from strategic security advisory to SWAT-team remediation projects, SPHERE provides solutions to help companies understand their risks, create policies for a target end state, and remediate major vulnerabilities.

Our Products – Leveraging years of experience in providing visibility around data, systems and assets, SPHERE productized its services and released SPHEREboard; allowing customers to have the deep analytics and rich metrics to help them better govern their environment.

Our Partners – SPHERE recognizes the need for third party products and has partnered with the industry's best-of-breed solutions. SPHERE provides guidance, along with technology integration capabilities to both the client's technologies, along with SPHERE*board.*