bringing order to chaos

a guide for information governance with netgovern solutions
INTRODUCTION

Organizations have been accumulating data for years, sometimes decades. Corporate networks are storing literally millions of files, folders, and emails. As the demand for more storage grew over the years, there were two options to consider:

1. Clean up what was no longer needed and reclaim the space.
2. Add more storage capacity to the network systems.

Storage vendors reduced prices drastically, and we know that users are not very good at “cleaning up”, and so option #2 was most often the solution. Humans are data hoarders; they will always say that they need to keep everything, “just in case”. However, for organizations, this should not be a viable option. There are all sorts of issues to deal with around information governance, or IG – how to protect, optimize, and use the collective information as an asset.

Unfortunately, most organizations have lost control of their unstructured data repositories; over time, they have simply thrown more capacity at the problem without necessarily putting in place the policies, practices, or tools to bring order to the chaos.

The information that resides in that massive amount of data:

⇒ may have business value, but we don’t know where it is, or that it even exists
⇒ may be requested in the case of litigation against the organization
⇒ may contain non-compliant information and put the organization at risk

External factors are putting more and more pressure on organizations to do something about the chaos: GDPR, HIPAA, FINRA, and more recently, the California Consumer Privacy Act. Organizations are realizing that the status quo – keep everything and have no idea what’s in there – cannot be maintained. There are key questions that need to be addressed:

1. Who can access our data?
2. What information do we save?
3. When should we delete it?
4. Where is my data being stored?
5. How do we audit everything?

These questions can be answered with a clear Information Governance strategy and realized with the right technology.

NetGovern proposes several solutions that help customers regain control of their unstructured data, which in turn will help them mitigate risk, benefit from cost savings, and put them in a much better stance when it comes to compliance.
In this guide, we will discuss the following points:

1. The Framework: this section will discuss NetGovern’s framework and approach to help customers regain control of their unstructured data.
2. The Solutions: this section provides a brief overview of each product in the NetGovern portfolio.
3. Mapping Out the Solutions: here you will learn how each product maps to each step of the framework; in other words, what to use when.

It is hoped that this guide will provide clarity as to what the NetGovern framework is around Information Governance, specifically around unstructured data, as well as how the products fit within that framework.
THE FRAMEWORK

The ARMA framework of Information Governance, or Information Management, as they call it, includes a lot of moving parts. Essentially, it points out that Information Governance is a cross-functional discipline that involves many different stakeholders. It is a complex affair that requires much collaboration.

At the center of it all are the standards and principles that are used to guide the Information Governance initiative.

This model provides a solid framework for developing an Information Governance strategy, as well as best practices. However, in the practical, everyday world of actually dealing with the data, what are the steps and solutions that can be used to help organizations regain control of their information? How should they tackle the mountain that is their unstructured data?

NetGovern’s framework of Information Governance, as it pertains to unstructured data, can be represented this way:

Starting from the outer circle, it is important to understand that while Information Governance encompasses many things in an organization, one of the most neglected aspect is the unstructured data.

Within this sphere, customers need to focus on 3 different steps to regain control of this massive amount of data; we will examine each of these steps in the next section.
Analysis

The first step towards an Information Governance initiative around unstructured data is to understand what is in the data and who has access to it.

What types of files are stored in the file systems? How old are the files? How many duplicate files are out there? Who’s hogging all the storage space? Which folders are the biggest?

From an access governance perspective, who has access to what files? What permissions do they have to any given location? How did they inherit those rights?

File analysis serves to identify the ROT – the Redundant, Obsolete, and Trivial data. Redundant refers to the duplicate files scattered across the network or any instances of duplicate information. Obsolete data refers to files or information that is no longer valid; for example, old product brochures about products that no longer exist. Trivial files are those that have no business value whatsoever: music & video files, old Christmas party invitations, etc.

It is widely estimated that ROT can represent anywhere from 40-70% of the unstructured data in any organization; we’ve actually had organizations tell us that it’s probably closer to 80% in their case! However, all this ROT is treated the same way as business-critical or sensitive information: it is protected, backed-up, and located on expensive storage.

Another aspect of what needs to be analyzed around all those files are the permissions. In many cases, over the years, IT has granted ad-hoc permissions to users and groups to fulfill collaboration requests. These permissions are rarely “cleaned up” and are very difficult to inventory. Customers need to gain clarity into these permissions when taking into consideration the security of their data.

Once an organization has gained some clarity into the metadata surrounding their file systems, they must now look inside their data to deep dive into the content. This is where customers will discover the dark secrets that lurk in their unstructured data that can put them at risk.

There are many types of data to look for, and it will depend on the customer’s industry. For example, in any organization that handles credit card payments, they would be striving to be PCI compliant. One of the constraints of PCI compliance is that credit card numbers should not be stored unencrypted in files or emails. How do we find those instances of non-compliance? How do we know that Johnny isn’t storing a spreadsheet in his OneDrive that contains customer credit card numbers? Or that Joan hasn’t been receiving credit card numbers through her email?
Customers concerned with external regulations (like GDPR, HIPAA, FINRA, etc.) would need to know if certain information is being stored on their networks. As previously mentioned, most have no way of assessing whether this is the case.

Customers must also understand that any data stored on their network is subject to eDiscovery in the case of a litigation request. This means that all this ROT could potentially put the organization at risk! In the case of compliance & audit concerns, customers should be worried about what lurks in all that data. If the organization is susceptible to litigation (e.g.: healthcare, construction, engineering, etc.) or Freedom of Access to Information requests, then it would greatly benefit them to be easily able to find that information, review it, and export it for the requesting party. Ideally, this would be handled internally by the customer’s own legal department, rather than having to use expensive, outsourced firms.

As indicated on the NetGovern framework diagram, the outcome of the Analysis phase is a set of “Recommended Actions” to be taken. These can include:

- Steps to remediate non-compliant data
- Clarity into data that can be disposed of (ROT)
- Suggestions for organizing/archiving/storing unstructured data more efficiently
- Suggestions for correcting and structuring network permissions
- Training and policies that could be elaborated/clarified for end users
- Suggestions for optimizing current storage
By now a customer should have gained clarity into their unstructured data: their files, their emails, email attachments, network permissions, etc. They can now define Information Governance policies based on what they’ve found and what the regulations for their industry require. Based on their findings, their policies should cover retention schedules, access governance, and the defensible deletion of any data that no longer has business value.

Armed with this knowledge, they can take all the necessary actions to “clean house” and therefore mitigate any risk to the organization. Any ROT should be deleted, such as duplicate files, files that are no longer relevant to the business because they are so outdated, and non-business-related files. Old emails that go beyond the customer’s retention schedule should be deleted as well, and emails belonging to users who are no longer with the organization but still within the retention period should be archived, where automated policies can take over to apply the necessary retention lifecycle.

This is where an organization would also decide – again, based on their policies – what to do with information that could pose a risk: those credit card numbers in emails and files, or Patient Health Information that could be in those, as well. Do they delete them, or is there truly any business value to having them where they are?

The organization should now also have a complete inventory of all the permissions assigned on the network, and there is a very high likelihood that a lot of remediation will be needed to reign those back in! This is a critical step in an Information Governance initiative, as it ensures proper protection of company data. Additionally, it can help prevent the spread of ransomware attacks, as infected workstations can normally only infect files that the user has access to; limiting users’ rights strictly to what they need access to can go a long way towards mitigating the risk of having all data encrypted.
If an organization has gotten this far in the Information Governance cycle, they should – in theory! – now have much less data to worry about. It is not, however, the end of the road – Information Governance is not a “project” that is worked on and then checked off as “Done”. Ever!

Certainly, the act of starting an Information Governance initiative, developing a strategy, and developing policies, could be seen as a project, but once in place, it is an ongoing effort. This can be likened to doing the spring cleaning in a home: if regular maintenance hasn’t been taking place, it’s a big job! Once completed and the house is impeccable, though, one is faced with a choice: let it get dirty, dusty, and filthy for a whole year, again, and do another big spring cleanup in a year, or do regular, weekly cleaning and always have a beautiful home! The latter also ensures that one is always ready for guests! It is also much easier to find things you need in a clean, well-organized home!

You’ll recall that our diagram was in the shape of a wheel; that is entirely intentional and indicates that these 3 activities must be conducted on an ongoing basis. Akin, if you will, to the weekly house cleaning that one should do to make sure the home is always clean.

The amount of data versus the available resources in any given organization is disproportionate – that’s how most of them got in the mess they’re in in the first place! There’s just too much data for humans to manage easily or properly.

Organizations should be using tools to automate most of these tasks and help make sense of it all. Delegating menial tasks through policy driven automation can keep the ROT and non-compliant data in check.

Setting up access governance policies and tracking those automatically against roles and responsibilities allows for significant reduction in risk. Monitoring sensitive information automatically will reduce risk and minimize the impact of security breaches. Alerts on HIPAA, PCI, or other compliance information allows the security department to ensure that storage and communications policies for sensitive information are applied properly by the users.
THE SOLUTIONS

In this section, we will look at the individual solutions offered by NetGovern, after which we will map out the solutions to the framework we previously elaborated. This should help customers clarify what products fit where, and what are their functions.

NetGovern Analyze

Analyze is an enterprise-class file analysis solution. Driven by the Analyze engine, communication is established with servers hosting network files and a complete inventory of all files and folders is taken. This includes all metadata, such as creation, modified, and last accessed dates, owner, file type, etc. Permissions on the files and folders can also be inventoried for analysis.

The important thing to note is that Analyze will inventory all files across all servers, across all locations. The data is then stored in a database so that administrators have a centralized view from which to access the reports and queries. When looking for duplicate files, for example, these can be found across the entire, distributed network, not just on specific servers.

Permissions reports shed light on rights assigned across the network, thereby aiding with data access governance initiatives. Permissions can be reported on from an identity perspective (what do these users have access to?) or from a location perspective (who has access to these files?).

All of the reports are created and accessed through a web browser, and there is a Windows Data Analytics client which allows administrators to deep dive into specific servers to conduct further analysis.

NetGovern eDiscovery

Our eDiscovery solution allows customers to index all their emails, email archives, and files to conduct searches required due to litigation requests. The files can be located on premises or in cloud repositories, such as SharePoint, OneDrive, Box, Egnyte, Citrix ShareFile, and others. The latest release of NetGovern eDiscovery has a built-in OCR engine, which means that any scanned documents that have been saved to the file system as TIFF or PDF images can now be made searchable.

The Case Management console allows organizations to create as many cases as they need. Cases can have different Case Managers, Auditors, and Reviewers, as well as different locations and criteria to search on. Multiple search queries can be saved in each case, allowing members to view different sets of results for review and analysis.

Unlike most eDiscovery solutions and the eDiscovery Reference Model, NetGovern eDiscovery does not go out and collect the data first – we index the live data, at the source, and allow discovery, review and analysis of this data. This means that case searches are quite dynamic: for example, search criteria yielding a certain set of results one week
can show more results when the case members return to it a week later due to the fact that new emails and/or files meet the criteria and show up in the results. Once the analysis and review has been completed, then the data is collected and made available for export.

Although the most frequent use case for eDiscovery revolves around legal matters, many public organizations, such as state and local governments, may also be subject to access to information laws and must be able to produce the requested data. NetGovern eDiscovery is also frequently used for this type of request.

NetGovern Audit & Remediation
Using the same engine and architecture as NetGovern eDiscovery, our Audit & Remediation console is targeted to Compliance, Security, and Audit personnel within organizations. Similar to the eDiscovery console, this allows organizations to search live data for any kind of information that could put them at risk.

For example, organizations that need to comply with the PCI Security Standards Council regulations would want to know if unencrypted credit card numbers are being stored in emails or files; health care providers would want to ensure that no PHI data is being stored where it shouldn’t be; and finally, any organization concerned with GDPR regulations would want to search their data for any information that could put them at risk, such as social security numbers, phone numbers, etc.

NetGovern Audit & Remediation offers a wide variety of search criteria and also allows RegEx expressions to be created to search for very specific patterns.

NetGovern Archive
NetGovern Archive is our flagship solution, out of which grew the eDiscovery and Audit & Remediation features. For customers hosting their email systems on premises, archiving emails has been a way to tame the size of the systems, allowing for better performance and maintenance. Policies can be configured that automatically archive emails after a certain period – 90 days, for example, or a year – so that the messages can be purged from the mailboxes and the mail database size kept manageable.

Once messages are stored in the Archive, lifecycle policies can be applied to stored messages. For example, an organization’s retention policy might state that emails need to be kept for 7 years and then be subject to disposition. Different policies can be applied to different groups of users.

Another use case would be to archive emails belonging to users who have left the organization. This not only affects the size of the email system but can also help reduce the number of mailbox licenses required, driving down cost.

Although solutions like Microsoft 365 offer practically unlimited mailbox archiving for users (depending on the plan subscribed to), organizations still have a need for 3rd-party archiving. There is a white paper from Osterman Research titled “Supplementing the Limitations in Office 365” that covers some of the reasons organizations might want to do this.
NetGovern Enforce

There is a slew of features in NetGovern Enforce that will make management and protection of network file systems easier.

To begin with, organizations can start with identity-based storage policies. For example, one of the most common problems that practically all organizations face is cleaning up Home directories that belong to users who are no longer with the company. Most of the time, these are left orphaned and no one knows if there is anything of any business value stored in them. A NetGovern Enforce Home folder policy can take care of provisioning and deprovisioning these folders. When a user’s account is deleted, the policy could move the user’s folder to a vault area where someone else could be granted access to view the contents; a lifecycle policy can also be applied, so that when a user is deleted, the Home folder is vaulted, then deleted after 90 days.

It is also possible to create location-based policies, one of which is a Security Notification policy. Assuming that an organization has cleaned up all the permissions on a given network file system (using NetGovern Analyze to gain clarity into those!), it would then be possible to create a policy for any sensitive areas on the file system that would monitor any changes in permissions. For example, after all the rights have been cleaned up, administrators might want to be alerted if any permission changes occur on the “Finance” folder. If any user, or group, gains access to the folder through any kind of assignment, Enforce would notify the administrator. Corrective action could then be taken to remove those rights if the assignment is not appropriate.

Another great feature of NetGovern Enforce is Epoch Data Protection. Organizations can create Epoch Data Protection policies as an additional layer of security for what we call High-Value Targets; think “Finance”, “HR”, or “Payroll” folders. The policy can then take “snaps” of all the files and folders at regular intervals, from once an hour to once a day, or as required. A user can then be assigned as a Data Owner on this High-Value Target and he or she can use the Data Owner Client to quickly restore files. If someone overwrote or deleted important files, or if the files were encrypted due to a ransomware attack, the files and folders could easily be restored within minutes.

And finally, Enforce can be used to import “Workloads” to take action on files. For example, administrators can use NetGovern Analyze to create a report of all *.TMP files older than 1 year across the network and then export the results into a CSV file. This CSV file can then be imported into Enforce and appropriate actions taken; these include Copy, Move, or Delete. Obviously, in this example, it is highly likely that the administrator would choose to Delete these TMP files, thereby reclaiming some storage.
MAPPING OUT THE SOLUTIONS

We have discussed the 4 steps required to regain control of unstructured data, and we have seen an overview of the solutions. The last piece is mapping the solutions to the steps – what do we use, and when? What should the whole process look like?

Going back to our wheel, we can map out the solutions in the following way:

Next, let’s examine how each solution helps in each step of our Information Governance vision.

ANALYZE

Use:

- NetGovern Analyze
- NetGovern Audit & Remediation
- NetGovern eDiscovery

In this phase, customers would use NetGovern Analyze to inventory all of their file systems. Customers can start accessing reports that will show duplicate files, how old that data is, who it belongs to, what types of files are being stored, etc. They can also analyze the file permissions across the different shares; this might reveal many accesses that are non-compliant.

Using Analyze to inventory all file metadata and permissions, a customer should then be able clarify what policies need to be developed, or if policies are already in place, whether they are being respected. Either way, we have seen that ROT can represent 40-70% of all stored data, and so a customer who is serious about reducing ROT can go a long way towards that goal by leveraging Analyze.
Then, customers will need to look *inside* that content and see if there’s anything lurking in there that could put the organization at risk.

This will all depend on which, if any, regulations the organization may be subject to, but essentially, once we have indexed all the content of the emails, attachments, and file systems (on premises or in the cloud), customers can look for any kind of information or pattern: social security numbers, patient health records, keywords, etc. *Any* info they want!

NG eDiscovery may also be relevant if the customer deals with either litigation or access to information requests. Although both of these products have very similar feature sets, there are some slight differences that could make one or the other more appropriate. If the customer is concerned with audit and compliance and being able to retract emails or documents from the source, then the Audit & Remediation console will work best. On an ongoing basis, this product can also monitor the data for any new instances of non-compliance in the unstructured data. If the customer is more concerned about legal aspects of finding information in the data – such as in the case of litigation issues – then the eDiscovery console is the right solution. It is very likely that some organizations will want to use both, too!

**REMEDIEATE**

*Use:*

- NetGovern Audit & Remediation
- NetGovern Archive
- NetGovern Enforce

At this point, the expectation is that if the customer is really serious about “cleaning house”, then there are significant steps to take.

To begin with, the reports produced in NetGovern Analyze should have helped identify a lot of the ROT; however, it would be impractical to go and manually delete all the files that have been identified as such. This is why a customer would want to export the reports to CSV; these files could then be imported into Enforce as a “Workload” and actions automatically taken on them: copy, move, or delete. Duplicate or obsolete files, or files that have no business value whatsoever, could easily and quickly be disposed of, thereby reclaiming all that storage space.

Using Enforce, customers can also gain control of orphaned user Home directories (those that belong to users who are no longer with the organization) and either move them to secondary storage or simply delete them.

Next, the Audit & Remediation console should have identified data that does not fit within the compliance policies of the organization; these records – emails, files – can also be retracted and either stored to quarantine or deleted.

And finally, customers may want to use NetGovern Archive to reduce the size of their email systems. This can include archiving old mailboxes belonging to users who have left the organization or putting in place a policy that automatically archives all emails older than a certain period, and then deletes older emails from the users’ mailboxes. For example, users may only have 1 year of emails in their mailbox, and should they need to access any older emails they would login to Archive. These steps can have a significant impact on the email system by
reducing storage, making the databases easier to maintain, and shortening backup windows. Additionally, a leaner email system normally also helps improve performance.

**ENFORCE**

Use:

- NetGovern Enforce
- NetGovern Archive
- NetGovern Audit & Remediation

Let’s begin with Archive: once emails have been stored in the archive, it is very easy to setup lifecycle policies on these records. For example, if the organization has determined that emails need to be kept for 5 years, then any older emails can be automatically disposed of from the archive. This helps ensure the “defensible deletion” of any data that is no longer required. (Customers need to keep in mind that ALL data that is stored on their systems is subject to eDiscovery in the case of litigation; even though their policy might say that email records need to be kept for only 5 years, if they’ve kept everything for the last 10 years, it all becomes discoverable in a lawsuit. This could potentially put the organization at risk!)

Using NetGovern Enforce, policies can be put in place to automatically provision/deprovision user Home folders, making sure that these are never “orphaned” again. Group and project folders can also be managed the same way. When a user leaves the organization, the Home folder can be “vaulted” to secondary storage and deleted after the required retention period.

We’ve already discussed that Enforce can also be used to notify administrators of any changes in permissions; once these have been inventoried with Analyze, and all non-compliant assignments remediated, then policies can be put in place to monitor any changes. This would allow for ongoing remediation of any unauthorized permissions.

Finally, NetGovern Audit & Remediation can be used to monitor for any instances of non-compliant data hiding in files and emails: PCI, PHI, PII, etc. Compliance/Safety personnel can have a report delivered daily to their inbox, from which they can even take action, such as Delete or Quarantine. The ability to be proactive like this can go a long way towards reducing risk for organizations.

**REMEMBER:** it’s a wheel!! A good Information Governance strategy is a continuous cycle of analysis, remediation, and enforcement.

**About NetGovern**

NetGovern’s software enables regulated organizations to cost-effectively define and deploy vertical market ready Information Governance strategies in under 30 days, eliminating the “analysis-paralysis” phase that negatively impacts most enterprise data projects. Connect, Collect & Control petabytes of unstructured data stored in your file sharing, instant messaging, email and collaboration platforms, whether on-premise, on-cloud, or across hybrid systems. By providing comprehensive File Analysis (Audit), eDiscovery (Search), and Enforcement & Remediation capabilities, our clients can proactively organize, preserve, secure, and gain insight from what is arguably their most valuable asset – Information.