

- How many people here have a big issue with unmanaged documents?
- By **unmanaged files** I mean files that have little or no IG oversight.
- These are typically stored in file shares, email, local drives, or even SharePoint or EFSS repositories like Box.
- By **IG oversight** I mean,
 - the application of policies and procedures to manage
 - file security,
 - permissions,
 - preservation,
 - records management, and file classification.

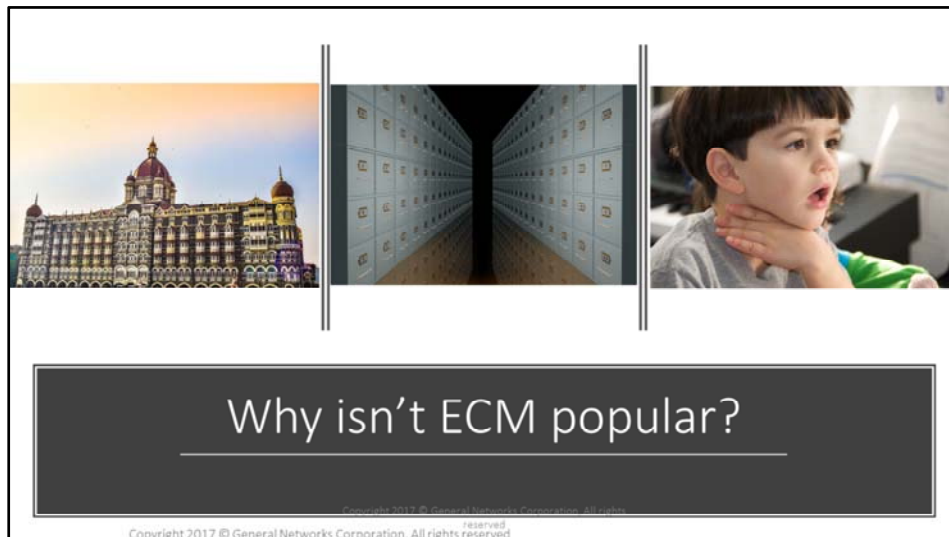
Agenda

- Is there a problem with unmanaged files?
- If there is, what are the options?

Copyright 2017 © General Networks Corporation, All rights reserved



- I'm going to first talk about why there is a problem with unmanaged files.
- Then I'll discuss options for dealing with it.



When we talk about unmanaged files, one of the elephants in the room is, why didn't enterprise content management systems take care of it?

- In my experience, the big reasons are cost, effort, and adoption.
- The Investment is often higher than the perceived return.
- The time and level of corporate involvement required to define a classification structure never gets completed.
- Too much time is needed from business stakeholders to classify content, making this process untenable. And finally,

- The use of an ECM system is seen by users as an intrusion and extra work, rather than a business aid.

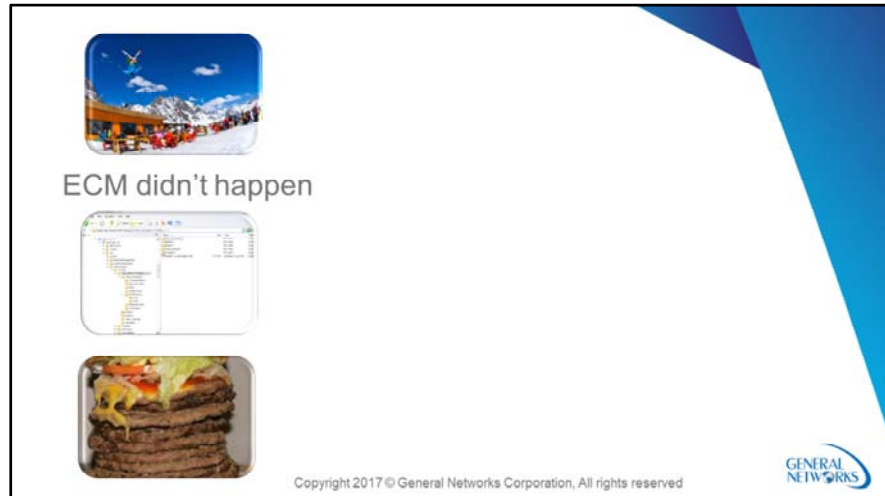


- ECM was an aspiration, if not a mantra, for more than 20 years. Why didn't we figure out earlier that it wasn't going to become the enterprise container for business content?
- **Besides** the high implementation costs, I would say that it was because we underestimated the importance of user adoption.
- It turned out that many business users behave like consumers.
- They gravitate to products that have personal benefit, ease-of-use, convenience, aesthetics.
- Unless they're forced to, many users don't adopt technologies just because they're there

or because they have value to their employer.

- And there are also other factors.
- Many organizations gave up on enforcing RM disciplines across the enterprise.
- Individual departments and business process owners solved the problem with Shadow IT. They purchased point solutions to solve their particular problems.
- The result is, a lot of business data is unmanaged or siloed.

Just curious... Do any of these realities apply to your organization?



- The results:
- Large enterprises have failed to execute on the dream of one repository to rule them all
- Multiple systems from multiple vendors is typical
- According to Forrester Research, 37% of firms they surveyed use more than three ECM systems
 - How many of you have an enterprise DM system?
 - How many of you have three or more systems for managing content?
 - How many of you are using EFSS repositories?
- Many technology and business decision-

makers demand more flexible, faster to deploy options, including cloud services. ECM vendors are not dominating the marketplace.

The Burden of Ungoverned File Shares

Enterprises store large volumes of "unstructured data" on in-house storage systems

- 37% store 1000 terabytes or more
- Regulated industries are top holders of unstructured data

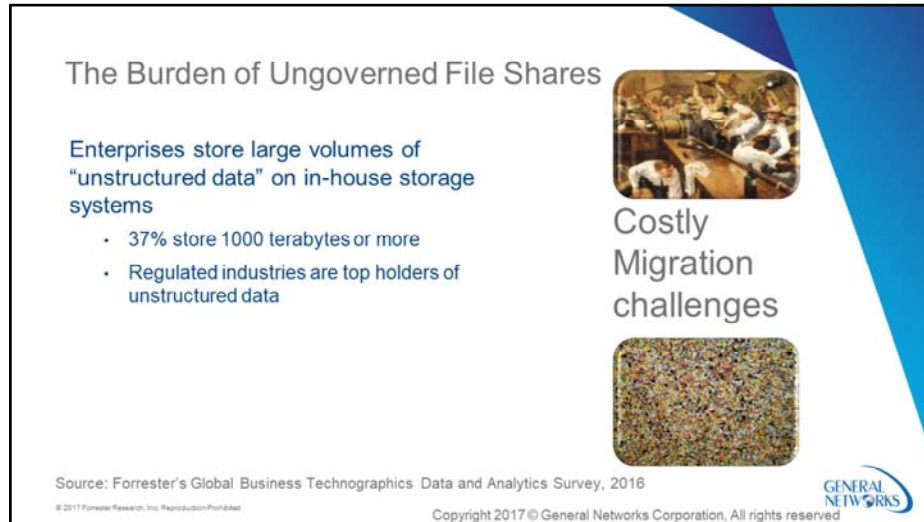
Costly Migration challenges

Source: Forrester's Global Business Technographics Data and Analytics Survey, 2016

© 2017 Forrester Research, Inc. Reproduction Prohibited

Copyright 2017 © General Networks Corporation. All rights reserved

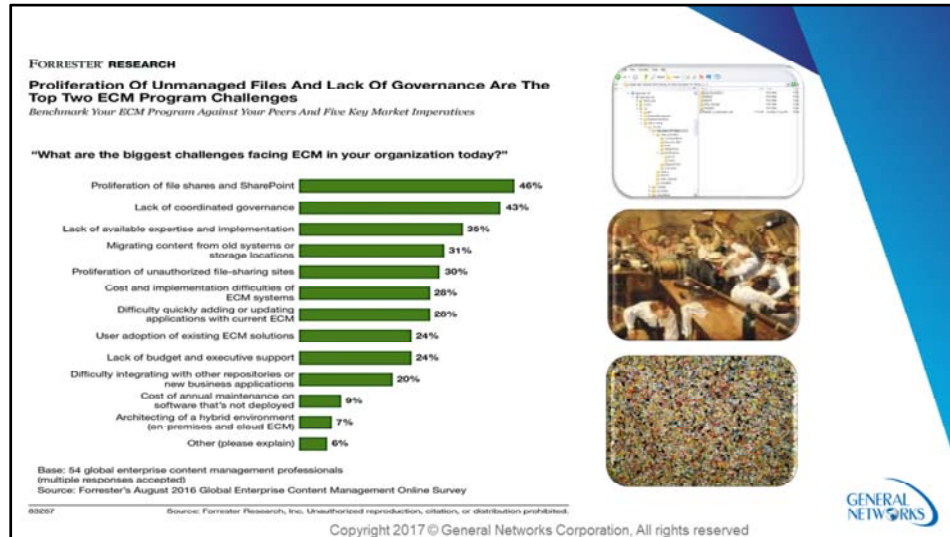
GENERAL NETWORKS



- So this brings us back to the problem we're addressing – managing unmanaged files.
- The problem of unmanaged files isn't the result of some outlaw departments or users. It's systemic.
- And the problem is growing, not shrinking.
- Forrester says that 37% of surveyed businesses store 1,000 terabytes or more.
- 1,000 TB (which is called a Petabyte) is a daunting size.
- That's about **20 million 4-drawer** filing cabinets.

((click))

- It seems paradoxical that regulated industries would be the biggest holders of unstructured data? Maybe their users are afraid to throw anything out.



Here's a bit of validation.

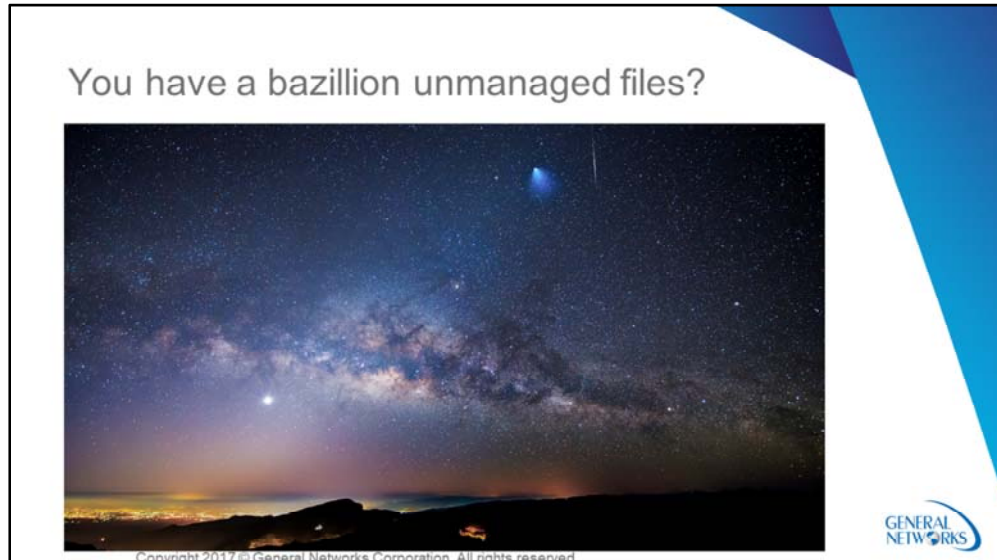
Forester research did a survey to identify what are the biggest challenges facing ECM in organizations today;

- No. 1 was Proliferation of file shares and SharePoint
- No. 2: Lack of coordinated governance
- Note No. 4: Migrating content from old systems or storage locations.



So, should we leave things the way they are?

- Do we say: “We can live with the risk.” –
- Is it too hard and too costly to embark on a large-scale file share cleanup?
- Do we say: “Let’s go after the low hanging fruit.”
- Does the low hanging fruit contain the documents we need to keep?
- Do we decide to boil the ocean and go after it all?



How do we know how many of our unmanaged files have business value?

There's lots of ROT.

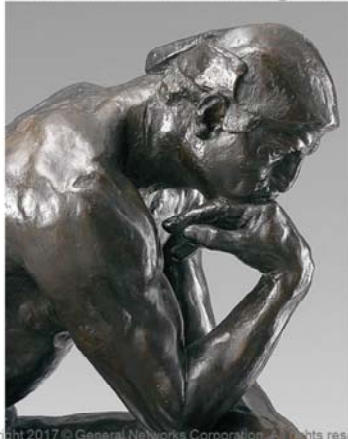
Do you know this term?

R, O, T? WE HAVE...

- Redundant files are duplicates
- Obsolete are older versions of newer files, older than x years, files from applications that have been sunset,
- Trivial – are .tmp, .lnk, .db, .exe, etc. They have no business value

Maybe we have a bazillion unmanaged files, but all we care about is 50,000

We have a bazillion unmanaged files?



Copyright 2017 © General Networks Corporation. All rights reserved.

GENERAL
NETWORKS

- That raises more questions:
- How much time do we spend figuring out what we have?
- How much engagement from the file owners should we expect?

You have a bazillion unmanaged files?



GENERAL
NETWORKS

Copyright 2017 © General Networks Corporation. All rights reserved.

- Then what do we do with the unmanaged files we want to keep?
- Leave them in place?
- Does that solve the problem?

You have a bazillion unmanaged files?



Copyright 2017 © General Networks Corporation. All rights reserved.

GENERAL
NETWORKS

- Are we going to migrate the unmanaged files to a single or to multiple repositories?
- Who's going to do that?
- Who knows the content and how it should be classified?



- Then, there's that **"you break it; you own it"** thing...
- What would our job be like if we start moving people's files around and they can't find them?

What are the options for dealing with unmanaged files?

Copyright 2017 © General Networks Corporation, All rights reserved



- Before we can answer these questions, we need to figure out the magnitude of the problem with unmanaged files, and
- What the options are for dealing with them.



Dilbert cartoon break.





Use Cases: “Find and Mine” Unstructured Files



- ☐ Convert unmanaged to managed files
- ☐ Clean up the ROT
- ☐ Compliance
- ☐ Security
- ☐ Enrich metadata
- ☐ Reconcile repositories

Copyright 2017 © General Networks Corporation. All rights reserved



We know we have a problem with unmanaged documents, but what are some typical reasons for doing something about it?

Here are some generic use cases for managing your unmanaged files:

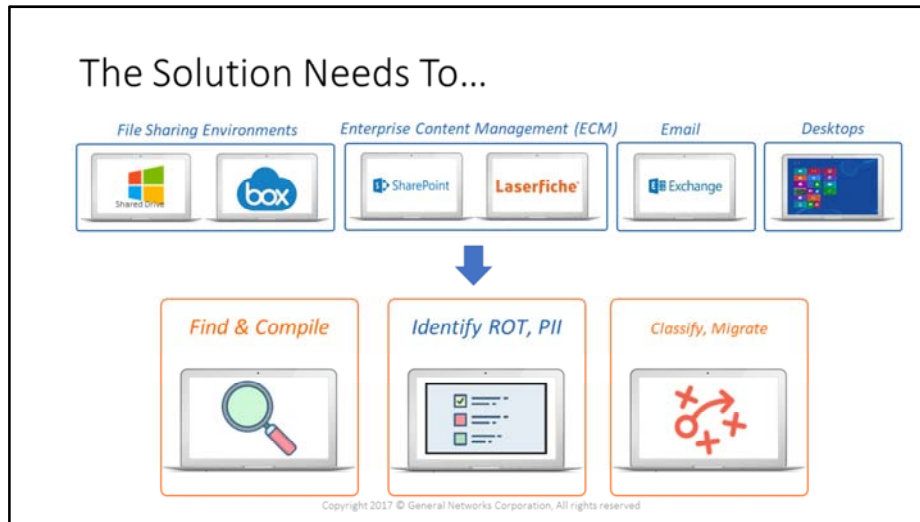
- Moving files with business value from unmanaged to managed repositories. Users should only need to check one repository to find what they’re looking for.

- Cleaning up the ROT.
- Maintaining compliance with file retention and privacy requirements.
- Privacy compliance is getting complicated with lots of domestic and international regulatory requirements.
- Unmanaged files are also a security issue.
Unmanaged data is easier to hack.
- Adding metadata to documents is an essential step to transition a file from unmanaged to managed.
- You can't, for example, enable RM disposition if

documents aren't classified.

- Reconciling documents stored in file shares or ECM systems with those in LOB apps. Again, users should only need to refer to one repository to find what they're looking for.

Do you have one or more of these use cases in your organization?



For these use cases, what kind of functionality do we need in the solution?

We will need to

- *Locate and uncover* unstructured content from multiple sources,
- *And Compile the data* into a single enterprise-wide view of all content
- Identify and *remove* Redundant, Obsolete & Trivial data (ROT),
- And Locate, identify & *flag* Personally Identifiable

information (PII).

- *Classify* data and
- move it to a secure location, or
- *Migrate* it to ECMs.

That's the overview.

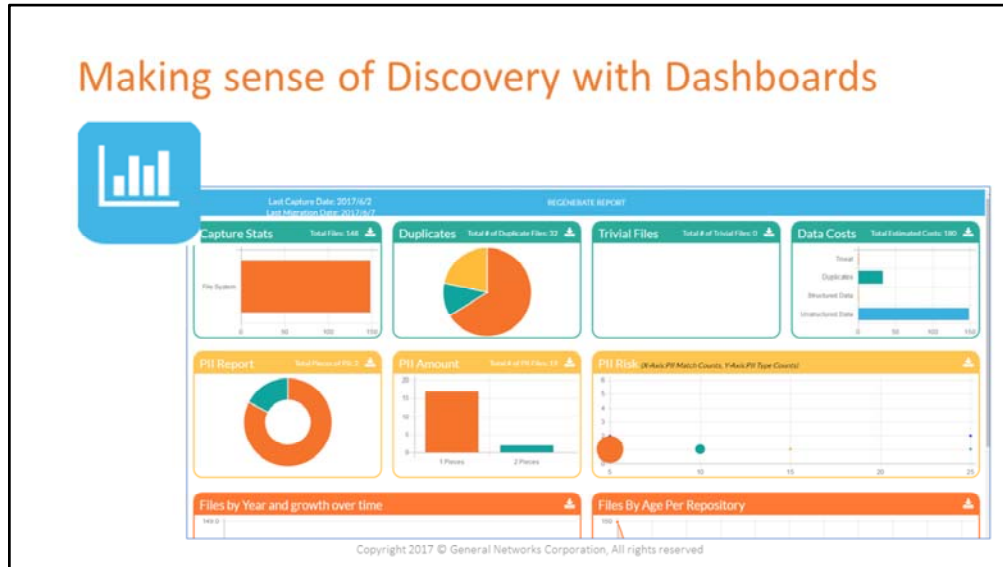


What are the tools we'll need?

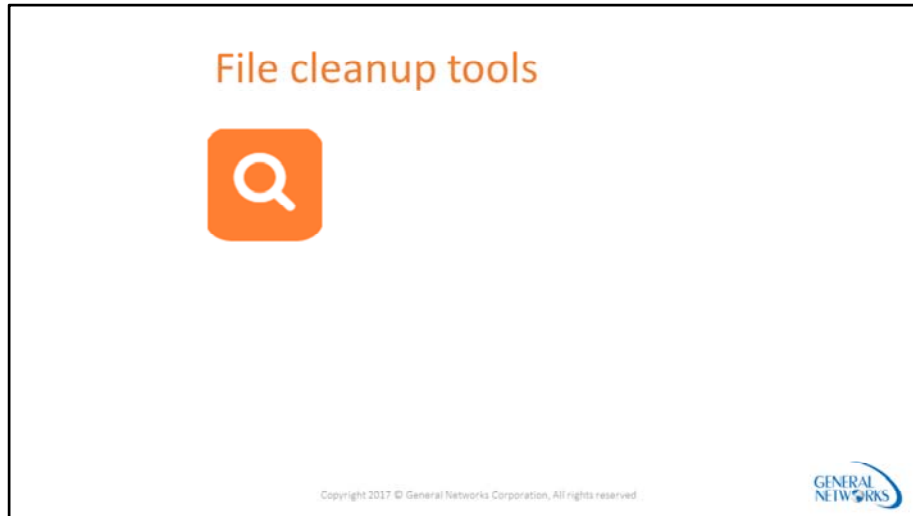
- We'll need **Discovery tools** that gather the data from our unmanaged directories, collect their properties, like file size, creation date, and last access date, and provide us with a high-level view of what's in our unmanaged repositories.
- After that we'll need **File cleanup tools**, because we only want to keep files that have business value. We'll need analytics to discover the files that contains important content, and the ROT.

- We'll need **Migration tools** to classify our file content, apply metadata, and move our unmanaged files to our ECM systems or other target repositories.
- And, unless we permanently fix the problem of creating unmanaged content, we need **Ongoing compliance management tools** that apply the rules that we created to maintain compliance.

Making sense of Discovery with Dashboards





- Gathering the high-level data is the first step. What file shares, local drives, SharePoint sites, Exchange servers do we want to examine?
- Analytics is what gives us the ability to efficiently make decisions and take action. Dashboard views typically show high-level results like **total files by repository, files by age, by size, by authoring application, by specific content (PII, etc)**



- Most of our unmanaged file systems contain a large percentage of unnecessary files.
- File cleanup tools allow us to create rules to identify these files and remove the files we no longer need. This can reduce the size of our managed files by half or more.

Developing Queries and Rules



Redundant: true duplicates.

Obsolete: last accessed 10 years ago or later

Trivial: Excluded File Types e.g., .tmp, .lnk, .db, .exe, .dll, thumbs.db, shortcuts, 0 byte


Other rules:

Stale: Last accessed \geq 6 months

Large: File size is more than 2 Gb

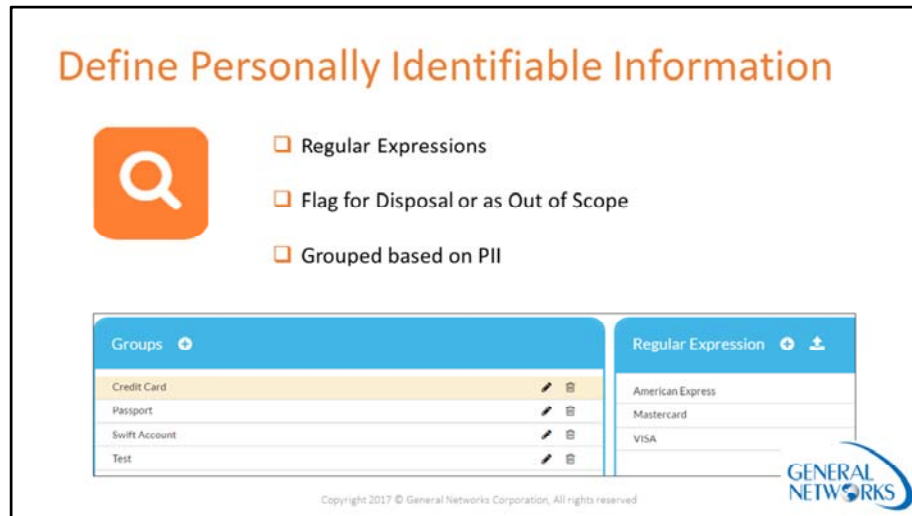
Out of Scope: Files unrelated to a classification

Copyright 2017 © General Networks Corporation. All rights reserved



- The low hanging fruit in file cleanup is ROT.
- We need to adapt rules that identify what we consider to be redundant, obsolete, and trivial files.
- For example, matching the hash values of files can determine whether they are duplicates – even if they have different file names. We might want to set a rule to determine which duplicate file to keep - the oldest or the newest duplicate. That decision can effect our file retention clock.
- We'll need search queries to define obsolete files by when they were last accessed, or

- Trivial files by file suffix (like .tmp files), size (like zero byte files), or expression (like shortcuts),
(click)
- and rules, which apply logical conditions, like defining a stale file.
- These queries and rules give us effective criteria for deleting, saving, and migrating our files.



- Criteria for defining PII for our organizations is as much subjective as objective.
- PII can include the usual suspects, like SS and credit card numbers, but our organization might define specific general ledger account numbers or confidential project names as information we want to protect.
- We need to be able to identify which files containing PII we want to keep or delete.
- That means we need to be able to define our own Searches for PII using expressions.
- Finally, we need to be able to store PII files in separate repositories by using the content inside those files to identify whether they belong HR, accounting, or some other repository.

Use Regular Expressions to...



- ❑ Locate, classify, flag, dispose
- ❑ Identify PII - define search patterns: e.g., Social Security #: [0-9]{3}-[0-9]{2}-[0-9]{4}, phone #: [0-9]{3}-[0-9]{3}-[0-9]{4}
- ❑ Search content: e.g., Livingston Properties
- ❑ Tag files: e.g., Livingston Properties

Copyright 2017 © General Networks Corporation. All rights reserved.

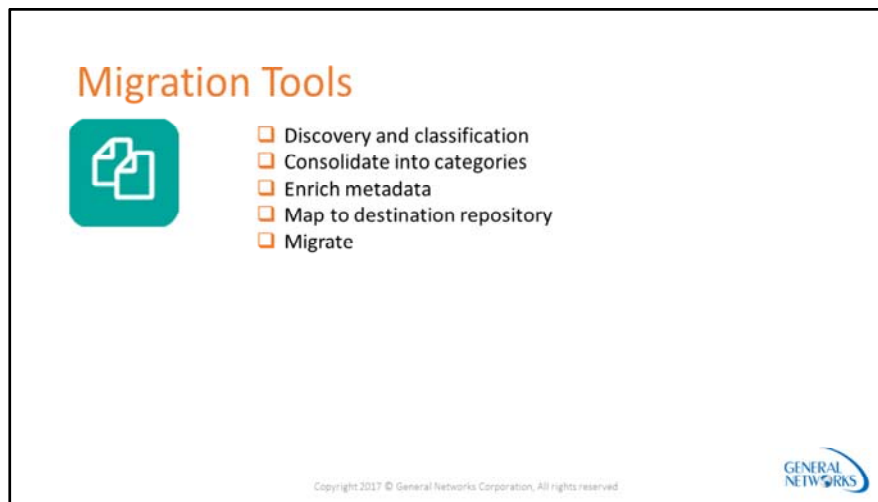


We can use expressions

- to locate, classify, flag, or dispose of records that contain keywords or any other defined pattern of characters.


You can see some examples.

- A formatted expression defines a search pattern for a SS no. or phone no. This can be used as a switch to classify a file as having PII.
- We can also use these expressions to search on content within the documents themselves, like project, vendor, case, or product names.
- The content that matches these expressions can be tagged to the files and later used as metadata attributes in a content management system.

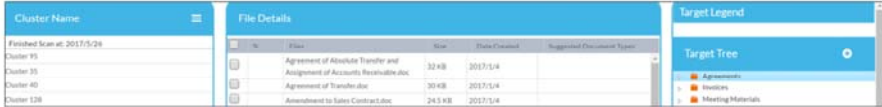


- File migration is a workflow process. The main steps are
- Classifying files – that's using selected file content and attributes to provide a basis for organizing files into sets.
- Consolidating the files into clusters or collections based on the classifications,
- Enriching the files with metadata that will be used by the destination ECM system,
- Mapping the files and their metadata to the destination ECM systems, and
- Performing iterative migrations.

Clustering




- Based on the documents contents
- View keywords in the Cluster
- Map whole clusters with drag and drop



The screenshot shows a software interface with three main panels. The left panel, titled 'Cluster Name', lists clusters: 'Cluster 95', 'Cluster 35', 'Cluster 40', and 'Cluster 128'. The middle panel, titled 'File Details', contains a table with columns for 'File', 'Size', 'Date Created', and 'Suggested Document Type'. The right panel, titled 'Target Legend', shows a 'Target Tree' with categories like 'Agreements', 'Invoices', and 'Meeting Materials'.

File	Size	Date Created	Suggested Document Type
Agreement of Absolute Transfer and Assignment of Accounts Receivable.doc	32 KB	2017/1/4	
Agreement of Transfer.doc	30 KB	2017/1/4	
Amendment to Sales Contract.doc	24.5 KB	2017/1/4	

Copyright 2017 © General Networks Corporation, All rights reserved.




- The discovery tools for migration are very sophisticated. Let's talk about machine learning, for example.
- Using one example document or a set of documents, machine learning can build rules to automatically classify similar documents into collections known as clusters.
- Think of clustering as organizing your documents based on content into a bunch of boxes.
- You can then label these clusters (e.g., purchase agreements, NDAs, engineering drawings), and migrate these documents as a group, instead of one document at a time.
- What about creating metadata for these documents?

- Clustering supports automatic categorization which applies tags to documents.
- Again, machine learning benefits from starting with a model, so you would need to tag a typical contract or several contracts, for example, and the clustering module will look for similar content in other documents and tag them.
- Clustering can be configured to look for conceptual similarities, e.g., whether or not the documents are about the same topic, OR
- literal similarities, which measures whether or not the documents contain identical passages of text. (That's useful for finding different version of the same document – called near duplicates).
- There are lots of permutations to machine learning. For example, you can have literal containment similarity, which is used to identify email threads where the last email in the thread contains the text from all the preceding emails.


- BTW, it's very helpful if these tools automatically perform OCR on non-searchable PDF files, so clustering will work on them as well.

Define Metadata Relationship



☐ Link source metadata to target metadata


Source Metadata	Target Metadata
Name	--Select Any--
Extension	--Select Any--
Date Created	--Select Any--
Date Modified	--Select Any--
Author	--Select Any--
VISA	--Select Any--
Mastercard	--Select Any--
Canada	--Select Any--



Copyright 2017 © General Networks Corporation. All rights reserved.

- Once we have identified and classified the documents that need to be migrated, we will need a data mapping tool.
- This will map the metadata in a collection of unmanaged documents to the relevant metadata fields in the target ECM database.
- If we don't classify the unmanaged documents and map their metadata to the ECM system, we will not be able to conduct metadata searches.
- The only way we will be able to find files without metadata is using full-text searches. We would be left with the Google experience.

Foldering



- Drag and Drop files/Folders
- View Metadata Elements in the Target Legend and Metadata values in Target Tree
- Add new Document types to the Target Tree
- Utilize the Filter option to view all unclassified/unmapped files

Source Tree

- Demo data
 - Agreements
 - Budgets
 - Employee Agreement
 - Facility Management
 - Images
 - Incidents
 - Meeting Minutes
 - Meetings


File Details

%	File	Size	Date Created	Suggested Document Type
100%	Agreement of Absolute Transfer and Assignment of Assets Rescindable.doc	32 KB	2017/1/4	
100%	Agreement of Transfer.doc	30 KB	2017/1/4	
100%	Amendment to Sales Contract.doc	24.5 KB	2017/1/4	
100%	Conditional Sale Agreement.doc	23.8 KB	2017/1/4	
100%	contract-reduces.doc	34 KB	2017/1/4	

Target Legend

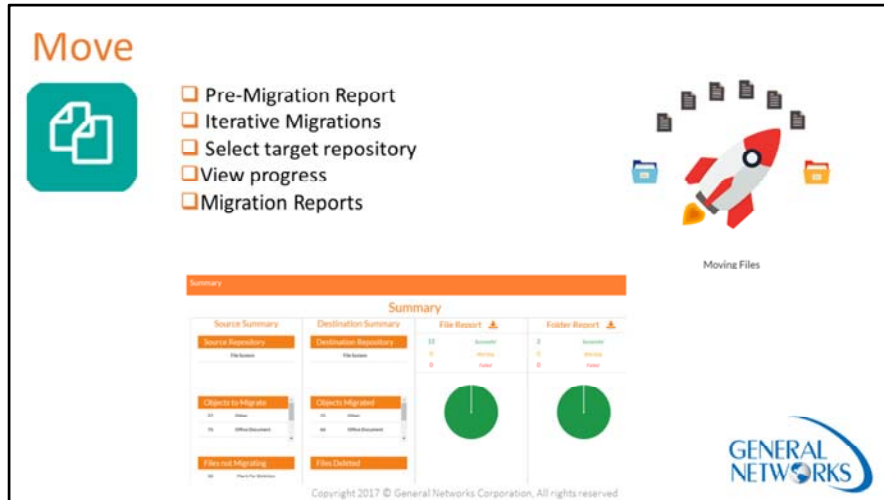
Target Tree

- Agreements
- Meetings
- Meeting Materials
- Purchase Orders
- Training Materials



Copyright 2017 © General Networks Corporation. All rights reserved.

- Clustering is not the only way to perform migrations. Sometimes we might have files or folders that don't belong to a cluster that we want to migrate.
- We can map our unmanaged files or folders from a file share or email into a folder in our ECM system.



- Migrations are tricky to reverse, so we want to them right the first time.
- Before pulling the trigger, we can use a pre-migration report to validate the plan.
- In any large migration, we'll need to perform iterative migrations because we will be using different criteria for different clusters or sets of documents.
- We will need to select which target repository we would like to move our files into.
- We can view the progress of our move in real-time.
- We'll need to download Migration Reports and Logs

to document what was done.

100% OF YOUR EMAIL
MESSAGES THIS MONTH
INVOLVED LINKS TO
FUNNY VIDEOS.









These are some vendors who have tools to help manage unmanaged files.

- FileFacets is a self service tool. It's in the cloud. It has all the capabilities we've covered here. It's designed for use by stakeholders.
- Active Navigation is a powerful and expensive tool. It's designed for IT mavens.
- Varionis focuses on data security and permissions management.
- Nuix is primarily an eDiscovery tool
- DocsCorp contentCrawler converts TIF, PDF, JPG images to full-text for searching. It can update ECM systems with a new version of the original, a new file, or replace the original.

- I'm leaving out platform specific vendors.

FF: Runs on Azure in a SOC2/SSAE16 secured data center. Only metadata goes to the cloud. Single enterprise-wide view of content analytics across systems and jurisdictions

Can scan network, cloud, desktop, laptop.

About General Networks

Year Founded 1986

ECM Practice: 1990
OpenText: 2006
SharePoint: 2007
DocuSign: 2014
M-Files: 2015
webMethods: 2015
ServiceNow: 2016



Team Members 83+

Headquartered in
Glendale, CA with
Staff in six US states
And Shanghai.



Over

Two Million

Users Supported

Our solutions support over two million users at 150 organizations, including state and local government, ports, airports, utilities, entertainment, technology, and manufacturing.

14+

Strategic Partners



Our partner network includes major platform vendors and 3rd party products we love!

Over

60%

Revenue from Repeat Customers



Our people and processes help us build long-term relationships with our clients.

General Networks Professional Services



Copyright 2017 © General Networks Corporation, All rights reserved

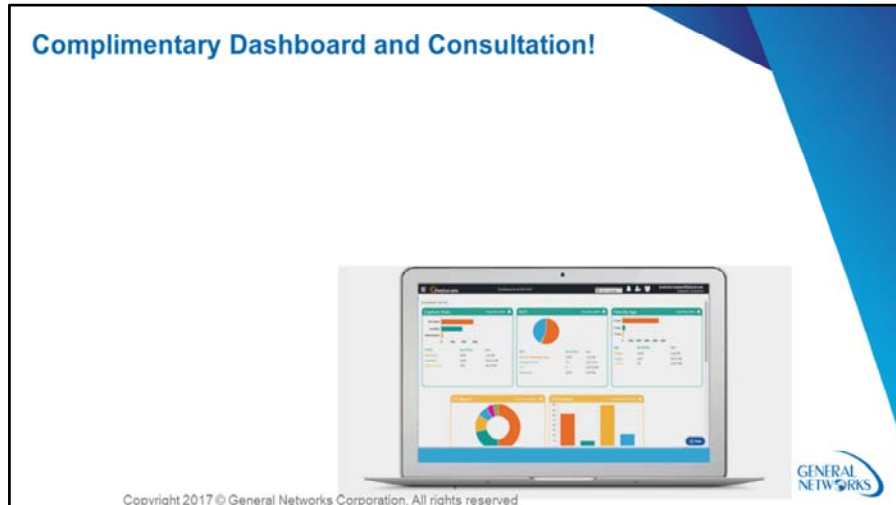


Should I stay or should I go?

Copyright 2017 © General Networks Corporation, All rights reserved



If you and your organization need to know more before answering this question, you might be interested in seeing a high level view of your unmanaged repositories.



If you want to get a dashboard view of the problem,

Leave us your business card and we'll

Contact you to schedule a **no-charge** on-site visit where we'll install a FileFacets dashboard and show you how to do your own analysis of your unmanaged repositories.

Our Services and How to Engage with Us

<http://gennet.com/solutions/>

- **Services include:**

- Discovery
- Strategy
- Needs assessment
- Implementation
- Migration
- Process automation
- Custom solutions
- Integration
- Support

Contacts:

David Horwatt

V.P. Professional Services

dhorwatt@gennet.com

(818) 330-7513

Cathy Ashbaugh

Sr. Account Manager

cashbaugh@gennet.com

(818) 330-7517

Copyright 2017 © General Networks Corporation, All rights reserved

