



Controlling Content Growth

The amount of unstructured content held by organizations continues to grow exponentially. Unstructured content is difficult to dispose of because it is not well described, commonly sitting in repositories such as shared network drives (aka file shares). In these situations, content is held indefinitely and it is common to see the following:

- More storage being added each year and storage costs increasing.
- Many files are duplicates, have become obsolete over time, or contain no content of business value.
- Disposition decisions cannot be made because documents cannot be identified, resulting in content being held in perpetuity.
- New Enterprise Content Management (ECM) systems are put into place but a lack of appreciation for the underlying content issues prevents large scale file share cleanup, classification and migration.
- Document retrieval. Required in the course of regular business activities is costly and time consuming
- eDiscovery costs continuing to increase with the growth of content.

The investment in projects designed to improve how content is organized is often higher than the perceived return. The perception that content classification is optional, is the root of the content issues facing most organizations. Internal classification projects have low success rates due to a lack of expertise. Common issue that come up include:

- Too much time is needed from business stakeholders to classify content making the process untenable.
- The time and level of corporate involvement and cooperation required to define a classification structure and taxonomy is usually too great for it to be completed.
- New ECM structures are a mystery; after content has been migrated users are stymied by the new structure and feel discouraged.
- The use of an ECM system is seen as an intrusion and extra work, rather than a business aid.

Success Story – BHP Billiton

BHP Billiton had 1,550,584 legacy files located in shared network drives, many of which were critical documents that needed to be moved to the corporate Enterprise Content Management system. Inconsistent folder structures, folder names, and filing practices made identification of these critical documents difficult and time consuming.

FileFacets was able to marry the requirements for a consistent corporate folder structure at the top two levels, previously defined by BHP Billiton to drive the records management functionality of ECM, with the requirement for business areas to use folder naming and organization which best aligned with what the business area's wanted to work with content.

The project, which spanned 18 business areas, saw one business representative per area spending on average 6.7 hours on the project. The simplicity of FileFacets as a means of reviewing folder structures and file mapping was the key to keeping staff involvement to a minimum. The tools allowed business representatives to provide meaningful input on how files should be classified.

In the course of a three month project, 783,722 files originally targeted for migration to ECM were identified as content to be stored in a secondary information system. Further analysis of this content identified 117,554 duplicate files and 96,401 files of no business value, all of which were removed from the holdings prior to the migration. Overall, just 35% of the original holdings were deemed suitable and required for migration to ECM.

The Current State of Content Management

In many organizations it is common to see little or no overarching practice, general guidelines, or strategy for content management. Ineffective management of content has many tangible and intangible impacts on organizations:

Compliance

- Electronic copies of documents are retained past destruction dispositions.
- Poor description of content prevents retention and disposition rules from being applied.
- No integrated retention and disposition system means there is no simple facility for file classification.

Cost

- 50% or more of current storage costs involve content that could be disposed of.
- Notes Databases or Exchange servers are used to store content, thereby inflating infrastructure costs.
- Continued reliance on paper records produces operating inefficiencies and higher costs from printing, filing, and retrieval.

Standards

- Reorganization or consolidation costs are higher because there are no standards for record keeping practices and business policies and procedures.
- Corporate compliance policies are left to divisional or departmental interpretation.
- Employee retirement or departure accompanied by IP loss.

It is common to see a diverse set of applications in use for the management of documents along with large file system repositories. Documents and records can be found in:

- Content repositories that sit under BPM solutions
- Back-ups or mirrored storage
- Personal content stored on local drives
- Duplicated electronic content which has been copied and put on a hold status in support of legal discovery
- Web Content Management repositories
- Help inside social and collaborative systems
- Content copied onto local drives

Without controls placed on content growth it is common to see four-fold increases in volumes and three-fold increases in cost.

Significant quantities of content are sitting unused. There is a significant amount of duplicative content being help, most of it resulting from backups, document sharing where groups of users all save the same document, or groups copying content for convenience. It is common to see a file share where over a period of 12 months or more, 80-85% of the files remain unchanged and 70% have not been accessed in 12 months or more.

In short, the growth of documents saved in file shares and attached to email are the biggest consumers of disk space. Industry standards show that when large repositories are classified, the age of 40% of content is found to be past its destruction disposition point. When 80% of files are not being accessed, one can expect that a large number of these files are old and no longer need to be retained. This can account for hundreds of Terabytes of content and millions of dollars in annual cost.

From a compliance perspective, this 40% of content represents a significant risk. Failure to apply a well-thought-out and effective records management program within the organization prevents the removal of all content that has exceeded its required retention period. Not only does this make the organization less legally defensible, but increases the cost of eDiscovery in case of litigation.

A Better State of Content Management

Improving an organization's ability to meet compliance requirements and reduce storage costs while defining a new set of standards for how content is managed requires a vision for how information will be governed and a roadmap of initiatives that focus on leveraging immediate cost reduction to fund the work required. The objective: clean up current content, remove outdated content, and put in place new practices which significantly slow the growth of content in the future.

Recent Gartner studies of organizations which have successfully completed similar processes highlight some key elements of what has made these programs successful, including:

- Building an information governance business case around saving money by reducing storage and e-discovery costs, enabling regulatory compliance, and improving IT efficiency and responsiveness.
- Using a records management program in the context of ECM introduction, in order to establish clear roles and responsibilities, an explicit information life cycle, and create a vehicle for enforcement.

- Deleting information that does not need to be retained while minimizing the time required of the business user or creator of information to help with the process.

These key elements reflect a set of proven practices that, when followed, ensure higher levels of success for such changes. In devising a plan based on these practices the following actions are required:

- Classify content and then remove legacy documents from file shares and email that has reached its disposition point.
- Reorganize the remaining content so that information management and records management controls can be applied.
- Establish a set of standards and capabilities for content governance.
- Elimination or strict limitation of desktop file storage and ability to create personal email repositories (NSFs and local PSTs).
- Application of classification and other descriptive practices (use of metadata and well-formed taxonomies) on an ongoing basis to all new file systems and email content to support search, retrieval, and application of retention rules.
- Begin a process of replacing paper records with electronics records.

Success Story – eDiscovery

A government organization, located in Canada, had seen a dramatic increase in the number of litigation events on an annual basis. As a result, eDiscovery costs had skyrocketed. Relying largely on shared drives to store information, the organization had little insight into what information was valuable or how information was organized. The Legal and IT departments both recognized that there was too much content and that the lack of organization of the content was compounding information growth and hindering eDiscovery efforts.

The client performed a simple ROT (Redundant, Obsolete, and Trivial files) analysis to determine what percentage of content was eligible for destruction. Upon realizing that more than 50% of their content was eligible for destruction they took on a project to both remove those files from the holdings and restructure the remaining content into a folder structure that mirrored the corporate classification scheme.

Analysis of subsequent eDiscovery costs for comparable projects show that the new folder structure, and the removal of ROT led to a nearly 50% reduction in price for eDiscovery.

The FileFacets Process

The FileFacets process is tested, repeatable and scalable. As a cloud-based Information Governance solution for migrating unstructured content in Shared Drives, SharePoint, and all ECM repositories, FileFacets provides numerous opportunities for quality checks by all affected stakeholders prior to migration of files into a new file structure. It is an inclusive methodology that guaranteed buy-in by all business units within an organization by using efficient collaborative technologies and interactions that support good communication.

Phase	Description
Discovery and Analysis	<ul style="list-style-type: none"> Using a representation of the current file structure, it assesses and identifies volumes, anomalies, file types and patterns. FileFacets is used to create a replica of the current file structure and identify duplicate documents.
File Plan Development	<ul style="list-style-type: none"> Implement folder structure and metadata taxonomies previously established by your organization into FileFacets. Immediately show stakeholder how their input has a direct effect on the way in which the file plan is built.
Mapping	<ul style="list-style-type: none"> A user begins to map files into the new plan Content is mapped to the appropriate disposition classification previously applied to their folder structure taxonomy within FileFacets.
Migration & Disposition	<ul style="list-style-type: none"> Records eligible for disposition are segregated into a secondary folder structure. Any files identified as being ROT are segregated so they can be disposed of. FileFacets Migrate is used to automate the movement of content from the legacy structure to the new repository.
Post-Migration Support	<ul style="list-style-type: none"> The Content Map provides all client staff a quick reference to the location of files based on the legacy file structure. This assists in the change management aspect of your shared drive cleanup.

Permanently Containing Content Growth

With the use of FileFacets, organizations can tackle the exponential growth in content. But making these improvements permanent requires that an organization puts into place and follows a well-defined set of content management practices after the project is completed.

1. Define a strategic framework in which principles are established for managing all information that provide guidance to:
 - In-flight content initiatives.
 - How to manage the lifecycles of documents, records, web content, and the eDiscovery process.
 - Make changing the status of content a transparent part of your organizational business processes.
 - Have content management be an enabler to how employees work.
2. Identify opportunities to optimize the use of existing technologies where:
 - In-place initiatives could be expanded to provide more value to the organization.
 - Rationalization of competing technologies can reduce IT costs.
 - Current governance can be enhanced to improve ECM use or create more standardization.
3. Be ready for business-driven ECM related initiatives:
 - Put an ECM Program in place.
 - Identify where current ECM technology solutions need to be improved.
 - Define a vision for ECM within your organization.
 - Develop an ECM roadmap based on your vision.

Quantifying your Unstructured Content Costs

There are many parameters that contribute to the costs incurred by your growing holdings of unstructured information. Quantifying these costs is the first step in arming your organization with the information required to take action.

Hardware costs and personnel time are straightforward costs to quantify, however the impact of unchecked information growth on business productivity is much more complex. A study by IDC found that workers spend 9.5 hours per week, or an average of \$14,000 per year searching for information. One organization in the industry determined that 30% of this time spent looking for information could be eliminated if the information was cleansed, classified and managed in an Enterprise Content Management System.

For this analysis, we will reduce this impact of information cleansing, classification, and management on business productivity to just a 5% reduction in the time spent looking for information. This will reduce the \$14,000 annual cost per employee down to \$700.

The following factors contribute to the overall cost that your unstructured content has to your organization.

Example of unstructured information costs for mid-sized Oil and Gas Company with 10,000 employees and 5 TB of File Shares.

Expense	Average Cost	Solution	Average Cost After Cleanup with FileFacets
Hardware maintenance costs, averaging \$9.55 per GB per year, for 50 TB	\$477,500	Deletion of ROT files	\$238,750
Productivity lost due to lack of IM (1500 workers)	\$10,500,000	Classification of documents	\$5,250,000
Regulatory fines related to information requests (Industry average)	\$3,500,000	Metadata Tagging for easy searching	\$0
e-Discovery for litigation	\$1,800,000	Obey legal retention dates	\$900,000
Totals	\$16,277,500		\$6,388,750

Over a one year period the costs savings would be \$9,888,750

Don't take our word for it! Use this the workbook below to see how the numbers work for your organization.

Expense	Average Cost	Average cost after Cleanup with FileFacets
Hardware maintenance costs	$\frac{\text{_____} \# \text{ of GB}}{\text{X } \$9.95 \text{ Cost/GB}}$ $= \text{_____ Cost/year}$	$\frac{\text{_____ Cost/year}}{\div 2 \text{ Without ROT}}$ $= \text{_____ Cost/year}$
Productivity lost due to lack of IM	$\frac{\text{_____} \# \text{ of GB}}{\text{X } \$700 \text{ Cost/GB}}$ $= \text{_____ Cost/year}$	$\frac{\text{_____ Cost/year}}{\div 2 \text{ Without ROT}}$ $= \text{_____ Cost/year}$
Possible Regulatory fines related to information requests	$\frac{\text{_____ Mean Fine in your industry}}{\text{X } \text{_____ Mean Fines /Year}}$ $= \text{_____ Cost / year}$	\$0
e-Discovery for litigation	$\frac{\text{_____} \# \text{ of GB}}{\text{X } \text{_____ Cost/GB}}$ $= \text{_____ Cost/year}$	$\frac{\text{_____ Cost/year}}{\div 2 \text{ Without ROT}}$ $= \text{_____ Cost/year}$
Totals	_____ Total Cost/year	_____ Total Cost/year

Over a one year period the cost saving would be \$_____