Managing File Server Content with Folder Metadata

File servers are almost always part of an organization's IT infrastructure. They are used as repositories to share files and documents and as archives for legacy data.

Managing file server content is recognized as one of the greatest IT challenges. The lack of guidance when storing new files and documents often results in information being misplaced or duplicated. Information workers moving files and folders around also add to the file server chaos.

Metadata can be used to solve the issue. They can help classify all files and documents enabling their management and retrieval. Unfortunately, manually adding metadata to years of legacy data can be a long and expensive process. Moreover, modifications to the classification can obsolete already added metadata.

Folder metadata are a powerful complement to file metadata. They leverage the existing hierarchy to tag all files and documents within a folder's scope. They can also define what a folder should or should not contain and how to manage it. They are easy to maintain and administer. Folder metadata are often the only solution to add context to a large number of files and documents.
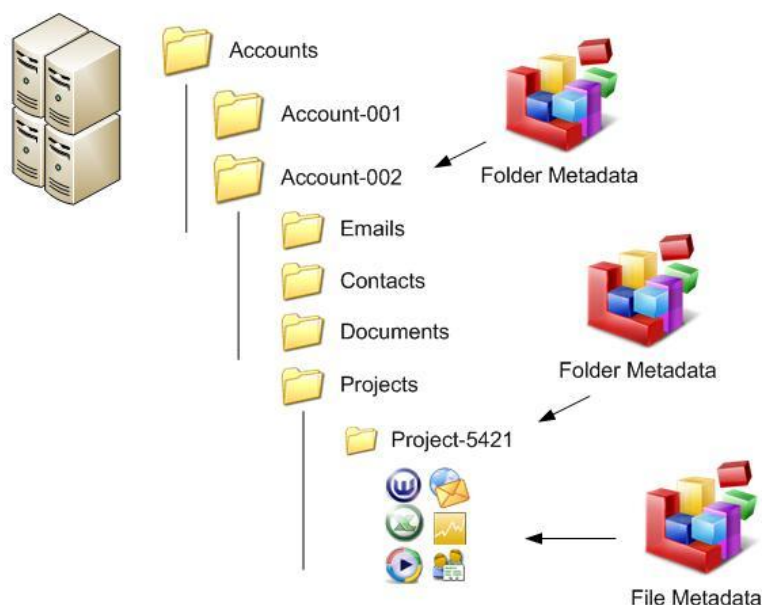
Enterprise Content Management

Implementing an Enterprise Content Management (ECM) system is often seen as the ultimate way to manage all files and documents produced by an organization.

Unfortunately, ECM systems often fall short of meeting all expectations for multiple reasons:
- High cost and complexity
- Low user adoption
- Legacy data migration
- File size and type limitations
- Performances

In most cases, organizations have to continue to support file servers:
- As ECM companions
- As legacy data repositories

Managing File Server Content

Information workers have very little guidance to help them finding the appropriate location to store and retrieve documents. This lack of guidance results in information being misplaced and duplicated. On the other hand, IT workers don't have the required information to support the Information workers. With no context surrounding most of the files and documents, it is becomes quite complex to apply retention and disposition policies.

Metadata solve the issue by giving a context to each file and document. Metadata is critical to know:

- What a document is about
- Who does it concern and who should have access
- When it should be dispose

File metadata alone is often not sufficient to provide the required information to manage file server content. Tagging all files and documents is a long and time consuming process that can take months to complete. File metadata is also not supported by all document formats.

Folder metadata is a powerful complement to file metadata. It leverages the existing hierarchy to tag all files and documents within a folder's scope. Moreover, it can define what a folder should or should not contain and how to manage it.

Folder Metadata

Folder metadata are a powerful approach to add context to large volume of data because they leverage the existing folder hierarchy. Adding metadata to a single folder will give a context to all its child files and documents. Moreover, all new documents created within a folder's scope will automatically inherit its context without any user intervention.

Folder metadata can be divided in 3 categories:

1. **Context metadata**: Define a context to help the Information workers to retrieve and to know the purpose of a document.
2. **Management metadata**: Define how a file or a folder should be managed. Migration tags, retention policy tags and disaster-recovery-plan tags are all management metadata.
3. **Business application metadata**: Record valuable business application information. By example, a folder representing a project can record: the timeline of the project, the project manager, the project status...

Technologies

Most file systems don't natively support folder metadata. There are two main approaches to overcome this limitation:

1. **The Central Database**: The first solution is to create a centralized database to record all folder metadata. While offering a few advantages in controlling metadata access, this approach faces many challenges. Users moving, copying or deleting folders will break and corrupt links thus losing valuable information. Also, failure to support the database in the long run will result in the complete loss of all metadata.
2. **The Context File**: The second solution is to store the metadata in a file under the folder. This technique guaranties that the context will stay associated with the folder and will always be available without any maintenance or support. XML is the preferred file format to ensure high compatibility and longevity of the metadata.

NOVABRAIN
FILE SERVER CONTENT MANAGEMENT

Example #1: Dealing with Legacy Data

An organization is implementing a new Enterprise Content Management system. All new documents will be created in the new system but legacy data stored on existing file servers need to be addressed.

1. In a first step, the organization defines a legacy-data classification to know:

   • What should be migrated
   • What should be deleted
   • What should be archived
   • What should remain on the file servers and for how long
   • What should be tagged to be indexed

2. In a second step, the organization mandates its Information workers to add the required metadata to all legacy files and documents. By using folder metadata, the Information workers will leverage the existing hierarchy to efficiently add the required information to most files and documents in a short period of time.

3. In a third step, the organization mandates its IT group to start the migration process. Knowing what each folder contains, the IT group can run periodic scripts and reports to quickly complete the process.

Folder metadata will also be used to add searchable facets to all legacy documents.

Example #2: Maintaining a Disaster Recovery Plan

An organization is developing a disaster recovery plan. One part of the plan is the recovery of the file servers. With millions of files on multiple servers, the organization can't recover everything at once. A priority list is required.

The organization needs to know:

• What needs to be recovered first
• How important is the information
• How and where to recover the information

Using folder metadata, Information workers and IT workers can easily add and keep updated the information on the folders of interest. The disaster-recovery-plan administrator will be able to run periodic reports to always have an up-to-date list of all mission-critical folders and the methodology to recover them.

Conclusion

Folder metadata is a powerful approach to manage the content of file servers. It leverages the existing hierarchy to quickly and efficiently give context to all files and documents. Folder metadata can be used to:

- Help Information workers standardize where they store and retrieve documents
- Help the IT workers manage the documents
- Recursively tag all files and documents to enable faceted searches
- Run queries and reports on folders
- Layout the groundwork of a migration or a disaster recovery plan
- Store business application information

Folder metadata combined with the auto discovery of specific document metadata is the most powerful solution to tag large numbers of files and documents.

Contact information

D1006-001 Rev 1.0