Folders vs Metadata in Electronic Information Management

1. Overview

Most people have used folders throughout their lives. Whether physical or electronic, folders are the way that many of us were taught to organize and store our information. We've used manila folders and filing cabinets to keep important papers (e.g., tax documents, birth certificates, etc.); and we've used electronic folders to manage the data files on our computers (e.g., Word document, Excel sheets, photographs. etc.)

When working within an Enterprise Content Management System or Document Management System, the use of folders, while tempting, can compromise the performance of these tools instead of simplifying and promoting their use as one might expect. For these types of electronic systems, the use of defined metadata tags on electronic files provides a cleaner and easier to use approach to managing data.

This document will provide a brief comparison of both of these approaches, pros and cons for each of them and some suggestions for when each option is more appropriate.

2. Comparison

Folders

For the sake of this paper, when we talk about "folders" we will be referring to electronic folders as you might find within the operating system of your personal computer or in use within an electronic data management system.

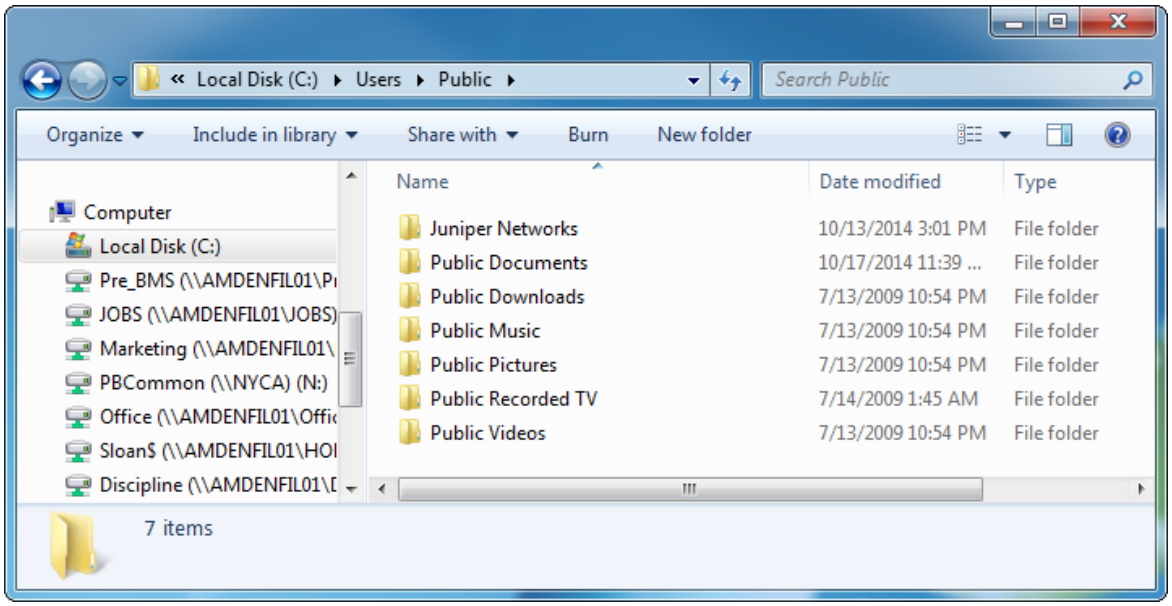

Figure 1 - Sample Folders from a Personal Computer Operating System

Pros

- **Historical Approach** – As mentioned above, most people working on a project site have real-life experiences working with folders to draw upon.
- **User Comfort** – People who have extensive experience working with folders (or even just managing files on their own electronic devices) tend to feel more comfortable working with folders when managing their data.
- **Complex Access Requirements** – For situations with complex access requirements within a single repository, folders can be configured to lock down information while still showing the relationship within the nested folder structure.

Cons

- **Nested Physical Structure** – Folders are usually created within a nested "tree" structure to provide deep layers of related information, separated out into logical collections of related information. When utilizing folders as a method to store information, items are then stored within in a specific folder somewhere within the "tree", as deemed the most appropriate location by the person placing that item.
- **Navigation issues** – Due to the nature of a nested folder structure, users typically require additional navigational assistance to understand the structural relationship between folders within the tree and how best to navigate that structure.
- **Structure Confusion for File Upload and Access** - Additionally, when data is stored within a deep, layered structure determining where data should be placed, and finding that data once it has been stored within the system becomes a challenge. This challenge is exasperated as the number of folders and nested layers increase.
- **Structure represent only one possible grouping approach** – One of the reasons that the nested folder "tree" approach is so confusing for most users, is the fact that it represents only one possible way of organizing the data, and usually is reflective of one person's take on that organizational approach.
- **Situational Data Duplication** – If a particular item logically falls into two or more categories within defined data structure, the user faces a conundrum regarding how to proceed. If they place the file within one of the possible folders that match the organizational structure, they risk the file not being found by users who are looking in the other areas when that file naturally belongs. To combat this, they may choose to make multiple copies of the file, and place one in each of the areas logically associated with the data provided by the file, but this then leads to the loss of a single data source and the possibility of data corruption as the multiple copies are not controlled as a single item.
- **File Access Security** – In a traditional folder tree structure, security is managed at the folder access level; while affective, this approaches is time consuming and can be an administrative headache especially as the number of folders and branches on the tree increase in number.
- **File URL Tied To Folder Location** – When utilizing folders within an electronic filing system, the address for accessing a file is directly tied to the location where that file lives within the system. The deeper the file is stored within the nested folder structure, the longer the address is for accessing the file.
- **259 Char Name/URL Limitation** – When naming files stored within a folder structure, it is important to remember that the address information is automatically appended to the file name within the system. This can be an area of concern when using long folder names and deep folder structures, because Microsoft has a built-in length limitation of 259 characters for file names.
- **No User Data Control** – When accessing data stored within a folder, the user is presented the data as it exists within that folder and is fairly tied to the items in the manner in which they are presented. To limit the items shown within the folder, it is often necessary to change the contents of that folder.

- **Difficult to Reuse** – In order to reuse a folder structure tree that has been proven successful in the past, it requires significant effort as the tree must be duplicated within the new environment, and each folder must be recreated individually in order to recreate the tree structure.
- **Structured Labeling Management** – When utilizing folders to manage data across a large organization, or across repetitive branches, it is not uncommon for the naming structure to be modified by users to better meet the needs of the project where those folders are being utilized. This can cause data access issues for users who may be thrown by the new labels and may not realize that the new name reflects the information they are looking for.
- **No workflow / automation** – Because workflows are usually tied to specific items, it is very difficult to associate them with folders.
- **Sort / Filter Limitations** – Traditional folders do not provide users with a standardized way to sort or filter the data stored within the folder. This difficulty is increased when trying to access data spread across multiple folders within a structure.

Metadata

In order to address the difficulties inherent with the use of folders within an electronic data management environment, the developers of these environments (e.g., Microsoft) suggest the use of metadata for data storage and access management instead of folders.

"Metadata" is the process of adding identifying tags to files within an electronic data management system in order to identify key information about that data and how it applies to the project and work being performed.

Example

As an example of how metadata might be used in a real world situation, let's assume we a contract document (A-001-PBA-XYZ.docx) that represents the working agreement between WSP and Contractor A, associated with contract "A-001" under project "XYZ" for the intersection of "First Avenue" and "Main Street" in Denver, Colorado, that needs to be added to our electronic file management system. Based on the information provided here, we could tag this document with multiple pieces of metadata, including:

- **File Name** – This is the name of the file and is captured automatically by the system.
- **Created** – This is the date that the file was added to the system and is captured automatically.
- **Created By** – This is the person who uploaded the file to the system and is captured automatically.
- **Modified** – This is the date that the file was last modified within the system and is captured automatically.
- **Modified By** – This is the person who last modified the file within the system and is captured automatically.
- **Document ID** – This is a unique number that can be associated with the file to identify it where ever it lives within the system and can be provided automatically by the system.
- **Content Type** – This identifies the type of content represented by the file, and is provided automatically.
- **Version** – This reflects the current version of the file and is provided automatically.
- **Contract Number** – This is a manually entered tag that identifies the contract number associated with the file.
- **Party 1** – This is a manually entered tag that identifies the first party of the contract associated with the file.
- **Party 2** – This is a manually entered tag that identifies the second party of the contract associated with the file.
- **Project Name** – This is a manually entered tag that identifies the project that the file is associated with.
- **City** – This is a manually entered tag that identifies the city where the project is located.

- **State** – This is a manually entered tag that identifies the state where the project is located.
- **Location** – This is a manually entered tag that identifies the location on the project that is associated with the contract file.

What the user would see would be similar to Figure 2:

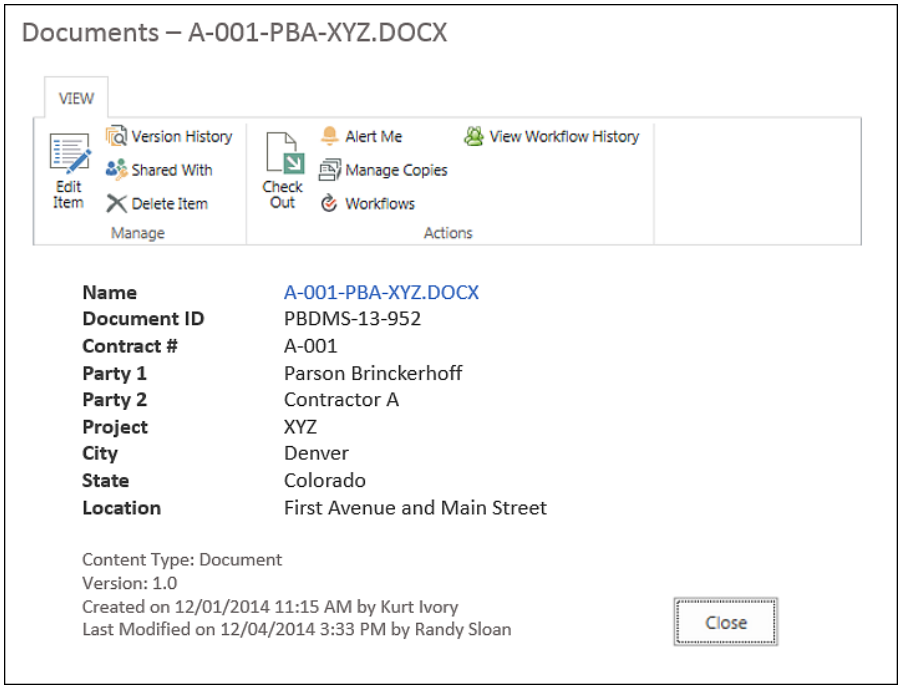

Figure 2 - Sample Metadata Properties Screen

Each of these tags can in turn be used to customize the navigation, the search capabilities and the method of access for files stored within the system.

Pros

- **Centralized Access** – The metadata data management approach is typically applied to files within a single source library, which aids in the search and access of files in a timely manner as it provides one central location for accessing related project information.
- **Single Data Source Approach** – The metadata data management approach encourages the use of one single source of every data file, accessible in multiple ways, which ensures that the user is always looking at the correct and most up to date file.
- **Custom Data Views** – One of the primary benefits of the metadata data management approach is that it allows users to customize their view of the information within a source library to show them only a subset of the items related to what they are requesting without requiring the other items to be moved or deleted from the container.
- **Consistent Labeling** – By using pre-populated selection lists to guide the tagging process, the metadata data management approach helps to ensure labeling consistency across all items entered into the system.
- **Workflow & Automation** – Because the metadata data management approach is built around individual files within an electronic management system, and not specific folders, it makes the use of workflows and process automation more effective and available for use across a wider audience.

- **Enhanced Search** – By tagging files within an electronic file management system with the appropriate associated metadata tags, the capabilities of the search functionality are enhanced and increase the likelihood of users finding the information they are searching for.
- **Document Templates** – Most electronic document management and content management systems allow for the use of document templates; additionally, the metadata associated with these templates can be predefined and used to guide users into providing the data that is needed to enhance the value to the organization of files they add to the system.
- **Multiple Levels of Metadata** – The use of metadata for tagging items within an electronic content management system can be provided at both the individual file level and the Document Set (file collection) level. This allows search results to find files both individually and as part of a collection of related files.
- **Built in Sort and Filter features** – The metadata data management approach includes multiple ways to sort and filter the data being displayed at the individual user level without having those custom views of the data affect any other user of that data collection in any way.

Cons

- **"New" Concept** – The use of metadata and file tagging is a relatively new concept for most people, even though the act of tagging data has been around for centuries in both publishing and data management fields.
- **Initially Confusing** – The concept of tagging files with metadata can be difficult to understand at first, which can keep people from attempting it.
- **Additional Steps** – The act of tagging files as you add them to an electronic data management system adds additional steps to the file upload process and can lead to incomplete uploads when people feel that they are too busy to perform the action in the required manner.

3. Findings

In order to capitalize on the multiple benefits associate with the use of metadata for tagging, navigation and data search and retrieval, it is highly recommended that users of electronic data management system make full use of metadata tagging when it is available to them.

Does this recommendation mean that users should never utilize folders? No. We will now address when folders are still a data management option that should be considered.

Situations that Warrant the Use of Folders

When faced with any of the following situations, the use of folders should be considered prior to moving forward:

- **Client/Project Requirement** – some clients are known to include very rigid, structured document environments that all projects are required to follow. In these cases, creating a matching folder structure may be warranted in order to meet the stated business requirements.
- **Risk of Low User Adoption** – if it appears that there is a high chance that users will reject the metadata approach, it may be worthwhile to implement a folder-based solution in order to ensure that the data is captured as opposed to them simply not using the systems provided.
- **Varying Permission Levels within a Single Repository** – if all data provided is required to be contained within a single repository / data source, and there are varying levels of permissions required for accessing that data, it may make sense to create a matching nested folder structure that can be utilized to lock down the data as required by the project. Preferably, this requirement would be handled through the use of multiple libraries of various security levels.

- **Flat Data Structure** – If the data being captured is of a "flat" nature and is naturally broken into predefined collections of information (e.g. a Photo Album with various collections of related pictures), a set of corresponding folders may be warranted to provide the desired structure.

4. Conclusion

In conclusion, the use of metadata for tagging, navigating and retrieving data files is highly recommended in order to provide the various benefits mentioned above. The majority of electronic data management, enterprise content management and document management systems on the market today, recognize the value of metadata tagging and utilize that within their systems to enhance data management and retrieval.

About the Author

Kurt Sloan is the Client Services Manager for the WSP USA's Project Information Management group, within the Program Management Services Technical Excellence Center. 2014 represents Kurt's 26th year providing information technology support and guidance to companies such as WSP, Parsons Brinckerhoff, Accenture / Andersen Consulting and Bank of America.

Data Sources

The data presented here was based on the writer's personal experience supplemented with input from the following data sources:

- 12 reasons folders in SharePoint are a bad idea (http://sharepointmaven.com/12-reasons-folders-sharepoint-bad-idea/)
- Ten Reasons to Avoid Folders in SharePoint (http://www.slideshare.net/bobbyschang/10-reasons-to-avoid-folders-in-share-point-spsvb)