Microsoft®
Windows®
SharePoint™ Services

Windows SharePoint Services:
Content Migration Strategies and Solutions

Microsoft Corporation
Published: September 2005

Abstract

Once an organization has deployed Microsoft® Windows® SharePoint® Services, one of the most requested tasks is to help individuals and project teams migrate their existing Office documents and other content to their SharePoint sites and document libraries. The most common content sources are existing file servers, Web sites, file systems on individual users' personal computers, and public folders on Microsoft Exchange public folder servers.

This white paper is written for IT Professionals who are responsible for helping users plan and execute the migration of content from a source content store to a Windows SharePoint Services environment.

Michael Herman
SharePoint Solutions Architect
Parallelspace Corporation

Microsoft®

Contents

Introduction

Microsoft Windows SharePoint Services (WSS) is built on the Microsoft Windows Server™ 2003 platform, and makes it easy for IT professionals to implement a dependable, scaleable infrastructure that enables teams to create Web sites for information sharing and document collaboration, using straightforward administrative tools and services.

WSS provides Information Workers more effective information sharing, collaboration and discoverability capabilities. When deployed within business units and across organizations, WSS benefits IT departments and administrators through:

- A reliable and scalable platform

- Powerful administrative control and security

- Better management and consolidation of corporate file shares

- Ease of extensibility and integration

- A rich partner ecosystem

- Rapid user adoption.

When an organization has deployed Windows SharePoint Services, one of the most common customer requests is to help individuals and project teams migrate their existing Office documents and other content to their new SharePoint sites and document libraries. The most common content sources are existing file servers, Web sites, file systems on individual users' personal computers, and public folders on Microsoft Exchange public folder servers.

This white paper is written for IT Professionals who are responsible for helping users plan and migrate their content from a source content store to a SharePoint environment. The migration may be centrally planned and managed by the IT organization, but more often requires close collaboration with the individual line-of-business (LOB) organization and project teams. The latter are the people who understand the purpose, value, structure, and access controls that need to be created in the target SharePoint environment

A key requirement of a successful content migration project is a deep understanding of the documents and other content in the source content store, the options for representing the documents and content in the target store, and developing or using a framework or process to plan and execute the overall process.

What is Content?

The most common types of content include Microsoft Office documents and Adobe PDF files. In addition, the source content to be migrated can include Extensible Markup Language (XML) documents, images, audio and video files, as well as data from line-of-business applications or third-party databases such as Lotus Notes.

From a content migration perspective, source content may have additional descriptive information associated with the file, in addition to the document file stream – the actual file content. The additional information may include:

- *Metadata*. Additional descriptive information stored in a document, in a companion file; or a record in a database or line-of-business application.

- *Structure*. Information implied by the folders and containers used to store or group the content files.

- *Access Controls*. The security mechanisms (usually associated with the folder structures) that control the rights granted to a particular user for reading, writing, creating, and deleting content.

- *Event-Based Workflow*. Application logic that may be associated with a content store or folder that is executed when a new document is created or a change to an existing document is made.

This collection of attributes associated with a collection of documents is described in more detail in the section 5-Level Content Model (5LCM).

What is Content Migration?

A simple definition of content migration is the copying of documents (and data) from a source content store to a target content store. In reality, the process of migrating content from one content store to another requires an appropriate level of planning along with process execution. Small content migration projects of a few hundred documents may not require sophisticated planning, processes, or tools. On the other hand, projects involving several hundreds, thousands, or millions of documents require a process framework to ensure successful planning and execution of the content migration project. Effective content migration frameworks are composed of models for the source and target content stores; the processes used to plan, execute and track the migration process; the tools to support the chosen processes; and, best practices.

What If I Want to Migrate Right Now with No Planning or Extra Tools

This approach is applicable whether a small number of documents need to be migrated, or a large project can be divided into several smaller departmental projects. The design of the SharePoint environment can be modeled using tools like Microsoft Excel or Microsoft Visio®, and implemented manually using the SharePoint HTML administration pages. The design needs to include the hierarchy of SharePoint sites, document libraries, and folders in the target SharePoint environment that are to receive the content from the source content store.

Content can be uploaded manually using either the single-document upload or multiple-document upload features of a SharePoint site. (The multiple-document upload feature is available if Microsoft Office 2003 is installed on the Windows client or server computer that is being used to perform the uploading). In addition, SharePoint document libraries support a Windows Explorer views that let you drag the file system content directly to a SharePoint document library or document library folder.

A content migration framework is recommended for a content migration project that involves more than a few hundred documents.

Using a Content Migration Process Framework

The key to a successful, large-scale, content migration project is effective organization, planning, and execution. These can be attained using a content migration process framework.

A complete content migration framework includes:

- A reference model for describing the different types of source content and source content stores, as well as the characteristics of the target content store.

- A set of processes, activities, and tasks for planning and executing the content migration project.

- A set of recommended content migration tools for each step of the project.

- Best practices accumulated over several content migration projects.

Each of these elements is described in the remainder of this document.

Understanding Your Content

What could be complicated about migrating a few hundred Microsoft Office documents from a file server or Exchange public folder to a SharePoint site? This type of project can be a simple task: all of the documents on the filer server can be copied using Windows Explorer and pasted into the Explorer view of a SharePoint document library; or uploaded using the document library multiple document upload feature that is available when Office 2003 is installed on the user's personal computer.

However, there are often other considerations that need to be analyzed. For example, are all of the documents stored in the same folder (or in multiple folders) on the file server or Exchange server? What was the reason for placing the document in multiple folders? Was the design for the folder hierarchy driven by a particular corporate taxonomy or information architecture? Or, alternatively, was the folder hierarchy designed to control access to the documents in a specific folder by a particular group of people person?

One approach to consider when migrating the documents from a source content store to WSS is to use 5LCM developed by Parallelspace Corporation.

5-Level Content Model (5LCM)

5LCM is used to analyze content – both documents and data – from the perspective of a content migration project. The model consists of five levels representing increasingly sophisticated views of what might initially appear to be a simple collection of documents stored in a hierarchical folder system. These five levels, illustrated in Figure 1, are best understood as "layers of an onion."

5. Event-Based Logic
4. Access Controls
3. Folder Structure
2. Metadata
1. Documents
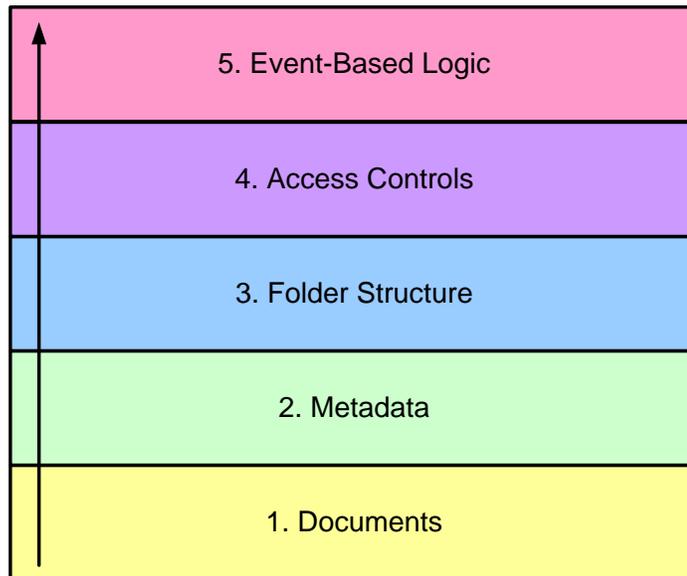
Figure 1. 5-Level Content Model (5LCM) Levels

The lowest and simplest level consists of the documents themselves (the *document content,* or *file streams*) that have a file name and file path used to name and locate the file in the computer's hierarchical folder system.

The next level of information associated with the content is metadata. Basic metadata is the additional information that may include the name or e-mail address of the person who last modified the document, or the person who initially created the document; the dates on which the document was created and last modified; or the length of the document – that is, the basic properties that are found in either the document itself (Office OLE document properties) or in the information used by the file system to store the document in a particular folder. Sources of additional metadata for a particular document may include a line-of-business application (for example, grant application processing), or a companion database that holds additional per-document information (for example, an electronic document or records management solution).

Virtually every file storage system, whether it is an operating system file system, document management system, or related source content store, has some sort of folder or container hierarchy for organizing and storing individual collections of documents. This is the third level in the 5LCM model – sitting strategically between Metadata at level two and Access Controls at level four. The folder hierarchies are often an integral part of an organization's taxonomy or information architecture. On the other hand, for organizations that are largely dependent on file servers for document sharing, the chosen folder hierarchies are often created on demand, by the particular individual or project team as the need arises. Folder hierarchies are virtually the only way of using metadata principles to organize

and structure a collection of documents into meaningful groups, particularly if the only available content storage system is a file server.

Access controls are strategically placed above the folder hierarchy level because while folders represent a common way of encoding metadata (level two) folders, hierarchies are also the most common way of encoding access controls. That is, access controls are administered on a folder-by-folder basis to manage, create, read, write, and delete permissions for the documents contained in each folder. Some content stores also enable document- or item-level access controls, but this capability is becoming less common.

The top level in the 5LCM model corresponds to any event-based logic or application workflow that needs to be supported in the target SharePoint environment. This level is often overlooked, particularly in the case of Exchange public folder content migration projects. Although it is uncommon for Exchange public folders (or file server folders) to have any event-based logic connected to them, a content migration project needs to determine whether this type of application logic is being used. The content migration project plan will need to allow for the time it will take to recreate the corresponding functionality in the target SharePoint environment.

Deploying 5LCM using SharePoint Sites and Document Libraries

As important as it is to understand the in-place content model for the source content store, it is equally important to understand the content model supported by the target content store: the Windows SharePoint Services sites, document libraries, and lists.

SharePoint Sites are multi-purpose, easy to create and manage Web sites supported by Windows SharePoint Services. WSS enables users to quickly incorporate additional capabilities into a SharePoint site, such as additional document or picture libraries with check-in / check-out and version history capabilities; XML form libraries; and customizable lists for including features such as Announcements, Contacts, Events, Issues, Links, and Tasks.

SharePoint document libraries are used to store Office documents, or any type of content that can be stored in a file. SharePoint picture libraries and XML form libraries are similar to document libraries, but have special features for organizing, viewing, and working with images and XML forms, respectively. SharePoint document libraries provide the 5LCM model level one storage requirements for document file streams.

At level two, the 5LCM model enables support for rich metadata in SharePoint documents, pictures, and XML libraries (in addition to SharePoint lists). The column properties in a specific SharePoint library or list can be easily extended using the HTML administration pages for the library or list. Column properties can be given specific data types such as string, date, number, and currency. A property can also be configured as a drop-down list box, option button, or check box field, based on a predefined list of values or as a lookup against another SharePoint list of values.

There are three ways to structure and organize content at the 5LCM model level three: sites, document libraries, and folders. One of the easiest means of providing structural support is to use document, picture, and XML forms libraries. The second way, given that a SharePoint site can have multiple document libraries and that libraries are easy and quick to create, is to use multiple document libraries in a SharePoint site.

The third way to organize and structure content is to use SharePoint sites themselves, which can have a hierarchical organization similar to file system folders. SharePoint sites are lightweight users of

storage and other resources compared to conventional file system-based Web sites, since they take up only a small number of rows in a small number of SQL Server™ database tables.

A level four content store allows control over who has access to a content store for creating, reading, writing, or deleting documents. SharePoint sites and document libraries are level four stores that support access controls at the level of a SharePoint site, library, or list. Windows SharePoint Services does not support access controls at the folder or document level, but this function is expected to be included in a future release. SharePoint sites are different from traditional file servers. By default, they do not allow a user to access the site's contents (such as libraries and lists) unless the site administrator explicitly grants the user role-based access. This makes SharePoint sites more secure than traditional file servers.

Level five event handler and workflow support is often overlooked. Common examples are Exchange public folders with script event handlers that trigger routing and approval or other application functionality; or file server folders monitored by a Windows for file systems events. There is corresponding event handling functionality in SharePoint libraries and lists for each of these scenario types . Alerts and moderator approval are two available features that are configurable through easy-to-use HTML administration pages. Both features use e-mail to provide notification capabilities. Additional event handling can be programmed using the SharePoint Products and Technologies Software Development Kit (SDK). Figure 2 illustrates how the 5LCM model is applied to Windows SharePoint Services.

5. Workflow	Event handlers, Alerts and Approvals
4. Access	Site, Library and List Access Controls
3. Structure	Site Hierarchies, Multiple Libraries, Folders
2. Metadata	Custom Document Library and List Column Properties
1. Documents	Document, Picture and XML Forms Libraries
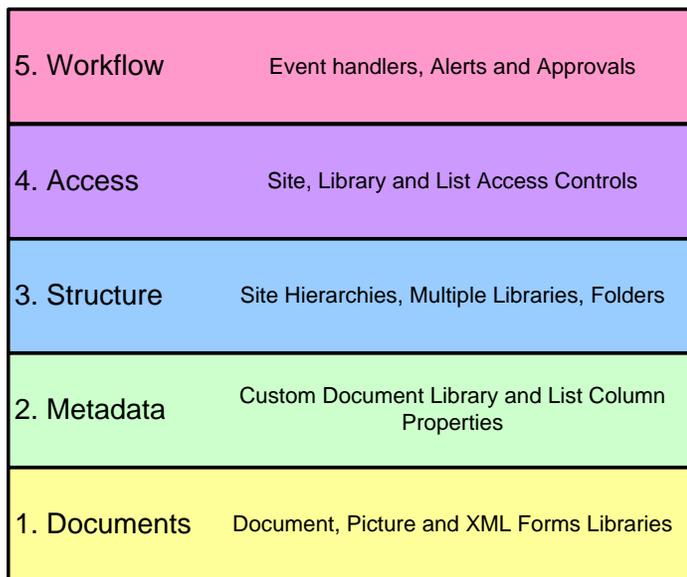
Figure 2. The 5LCM model applied to Windows SharePoint Services

Applying 5LCM to Source Content Stores

The 5-Level Content Model (5LCM) must be applied to each source content store to be migrated to Windows SharePoint Services. In this white paper, the 5LCM model is applied to the following source content stores:

- File Servers and Local File Systems

- Web Sites

- Exchange Public Folders

- SharePoint® Portal Server 2001 Workspace Libraries

- Lotus Notes Databases and line-of-business applications

File Servers and Local File Systems

Documents are stored in one of the file systems supported by the operating system. The underlying file system determines the limitations on the length of file and folder names and the set of valid characters that can be used to name them. Figure 3 maps the capabilities of a traditional file system to the 5LCM model.

5. Workflow	File System Change Notification APIs
4. Access	File System Access Controls
3. Structure	File Shares, File Systems, Folder Hierarchy (Nested)
2. Metadata	File System Directory Information, Document Properties, LOB Application Data
1. Documents	Traditional Document Store
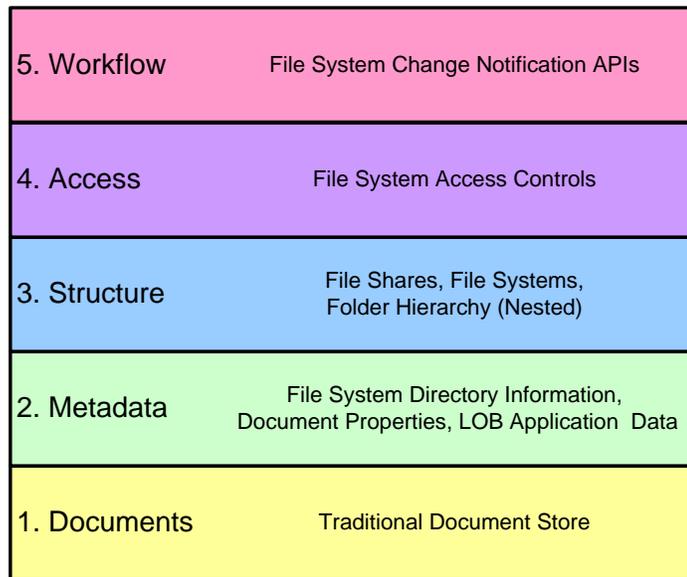
Figure 3. The 5LCM model applied to File Servers and Local File Systems

Web sites

Web Sites are typically used to store HTML documents. They are also used to store Office documents, particularly if the Web site is enabled for WebDAV (Web Distributed Authoring and Versioning) or Microsoft FrontPage Extensions protocols. Web sites typically do not support storing additional metadata for each HTML document. However, HTML <meta> tags can be used to store metadata information internally in an HTML document. The SharePoint full-text indexing service uses the metadata, stored internally in HTML documents and as OLE document properties in Office documents, to build catalogs that support both rapid full-text searches and property searches of the content in one or more Web sites.

Although Web pages in a Web site appear to be arranged hierarchically, documents stored in a Web site are often viewed as a simple collection of documents with the hierarchical information presented in the document's URL (uniform resource locator). From the perspective of a content migration project, the goal is to migrate all of the documents that a particular user account has access to. Web sites themselves typically do not have event-based logic connected to content in the Web site. Figure 4 illustrates the 5LCM model applied to Web site content.

5. Workflow	N/A
4. Access	"None" (exposed over HTTP)
3. Structure	Flat List of Structured URLs
2. Metadata	Document Properties
1. Documents	Web Pages, Images, CSS, Traditional Documents

Figure 4. The 5LCM model applied to Web Site Content

Exchange Public Folders

Exchange public folders support storage for rich collections of metadata associated with each document. Exchange public folders inherently support deep heterogeneous folder hierarchies. A rich set of access controls is supported both at the folder and subfolder level. Synchronous and asynchronous script and COM (Component Object Model)-based event handlers are supported at the level of each public folder and public folder subtree. Figure 5 illustrates the mapping between the Exchange public folder architecture and the 5LCM model.

5. Workflow	Folder Event Service, Event Scripts
4. Access	Folder Level Role-based Access Controls
3. Structure	Single Folder Hierarchy
2. Metadata	Extensible Item Schemas, Document Properties
1. Documents	Documents, Contacts, Discussion Items, Calendar Appointments, Tasks

Figure 5. The 5LCM model applied to Exchange Public Folders

SharePoint Portal Server 2001 Workspace Libraries

In SharePoint Portal Server 2001, each workspace can have one document library or workspace library. A workspace is similar to a Windows SharePoint Services top-level site, and a workspace library is similar to a Windows SharePoint Services document library.

Workspace libraries support hierarchical folders with role-based access controls supported at the workspace, workspace library, and folder level. Limited support for access controls at the document level is provided, using a "deny"-based permissions model (and as a result, document-level access controls were not often used because of the complexity of the deny permissions model).

Workspace libraries support extensible metadata based on document profiles. For example, a folder can contain documents where each document potentially uses a different document profile depending on the business purpose of the document. In addition, the Microsoft Web Storage System in SharePoint Portal Server 2001 supports both synchronous and asynchronous event handlers for implementing workflow or other business logic.

In SharePoint document libraries, all documents in a particular library share the same set of column properties, or document metadata. If the documents in a SharePoint Portal Server 2001 folder use multiple document profiles, then each set of files that use a particular document profile will need to be stored in either its own SharePoint library, or a single SharePoint library that includes a column property for each property in each of the document profiles – that is, a set of column properties that represents a union of the properties in the multiple document profiles.

The 5LCM model capabilities of SharePoint Portal Server 2001 workspace libraries are illustrated in Figure 6.

5. Workflow	Asynchronous/Synchronous Events
4. Access	Workspace and Folder Level Access Controls (Item-level Exclusions)
3. Structure	Document Workspaces, Document Libraries, Folder Hierarchies
2. Metadata	Extensible Document Profiles (Multiple), Category Hierarchies, Document Properties
1. Documents	Traditional Documents

Figure 6. The 5LCM model applied to SharePoint Portal Server 2001

Lotus Notes Databases and other LOB Database Applications

Office documents can be stored in Lotus Notes databases as attachments to Notes documents. A Notes document is more similar to a database record than a conventional document. It is often best to

think of a Notes document as a collection of field name-value pairs where a particular document may have certain fields defined, but not others. This is different from a conventional database record where there are typically a fixed number of fields (database columns) and each field has a value or the special null value. Access controls are supported at the database and document levels. Lotus Notes databases support a rich agent-based event infrastructure. The 5LCM model for a Lotus Notes database appears in Figure 7.

5. Workflow	Agents and Scripts, Workflow
4. Access	Server, Database and Document Level Access Controls
3. Structure	Notes Servers, Notes Databases, Notes Documents (Nested)
2. Metadata	Notes Document Items (Columns)
1. Documents	Notes "Documents" (Nested Database Records), Traditional Documents (as Attachments)
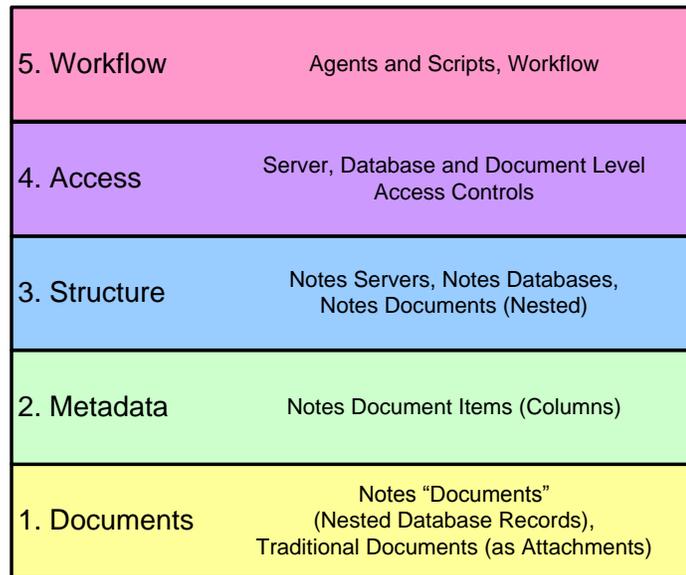
Figure 7. 5LCM applied to Lotus Notes Applications

Office document attachments will most often be migrated to SharePoint document libraries, along with the associated metadata. Lotus Notes documents will be migrated to a SharePoint list, based on one of the standard templates or a custom list design, or as an XML document stored in a SharePoint XML forms library. In the case of XML, Microsoft Office InfoPath® 2003 is useful for mapping selected fields from the XML document to SharePoint list column properties via a process known as property promotion. In addition, InfoPath can be used to create specific forms templates for displaying the XML data in a document.

An approach similar to migrating Lotus Notes documents can also be used to migrate data from a line-of-business application, collection of XML documents, or relational database. That is, the data can be imported or injected into a custom SharePoint list as column properties on top of which custom views can be easily defined for filtering, sorting, and grouping the data to match the needs of the users. Alternatively, the data can be formatted as XML documents and stored in SharePoint XML forms libraries similar to the Notes data.

8-Step Content Migration Process

With a deep understanding of the 5LCM model, how should the actual content migration project be organized? Based on partners' experience with several content migration projects, the following eight-step content migration (8SCM) process has proved the most successful:

1. Analyzing
2. Planning
3. Modeling
4. Extraction
5. Mapping
6. Injection
7. Verification
8. Workflow Migration

Each of these steps is described below.

1. Analyzing

The primary focus of the analyzing step is to develop a detailed understanding of the content to be migrated from the source content store. Data is gathered to answer the following types of questions in the planning step:

- How many items are in the source content store?
- How many of these items need to be migrated?
- Which document types are represented?
- What metadata is available?
- How much storage?
- Access controls?
- Folder hierarchy?
- How old is the content?
- Events and workflow?
- Who owns which subsets of the source content?

2. Planning

The detailed understanding of the source content to be migrated is the "what". The next step – the "how" – is to build a plan to prepare the target Windows SharePoint Services environment, as well as the approach, processes, and tools to be used to migrate the content from the source content store to the target content store. The scope of the migration plan may include creating the target Windows SharePoint Services infrastructure in addition to the actual migration of the source content. Refer to the section Related Links for information about planning, deploying, and operating a Windows SharePoint

Services single server or multiple-server server farm environment. Topics to consider include capacity planning, system performance, availability, network performance, and scalability planning.

For the actual content migration tasks, the key consideration is understanding the source content to be migrated. This includes:

- Understanding the business value of migrating each collection and subcollection of content – what content needs to be migrated and what content can be simply be backed up, rather than migrated.

- Understanding which metadata properties need to be migrated with the document content.

- Determine if the access controls on the source content are adequate and can be simply replicated. Establish whether to use item-level access controls; and how to improve access control management using the Windows SharePoint Services role-based access control model.

- Determine whether application logic needs to be replicated in the target SharePoint environment. Decide whether out-of-the-box SharePoint features can be used directly; and which capabilities need to be developed as an event handler in Windows SharePoint Services.

Determine two things in planning the time and effort for a content migration, both of which are greatly influenced by the size of the source content store:

1. The length of the planning phase. This frequently involves hands-on analysis of the content in the source store; as well as the time it takes to work with the owners of the content to determine what needs to be migrated and how it should be done.

2. The actual amount of source content to be migrated. The extraction and injection phases of the content migration process often take the most time and are dependent on the number of documents to be migrated and the quantity of content, in megabytes, to be migrated.

3. Modeling

Once the analysis and planning steps are complete, the project team will have developed a deep understanding of the source content to be migrated. The next step is modeling; that is, the design and implementation of the target SharePoint environment. Figure 2 illustrates how the 5LCM model is applied to a Windows SharePoint environment.

If the project team has enough members, the modeling step can be approached top-down and bottom-up. The top-down approach begins with modeling the hierarchy of top-level sites, subsites, and document libraries to be used to store the source content. The primary consideration is the access controls that need to be implemented in Windows SharePoint Services to control access to the imported content as well as the different teams and roles that need access to the content.

The bottom-up design considerations focus on the structure of the SharePoint document libraries to be used. This includes the folder hierarchy for structuring the collections of documents as well as the column properties that need to be used to store the metadata for each collection of documents. Different collections of documents may require a different set of metadata column properties based on their business purpose. Different document libraries should be used for document collections that have a distinct, different business purpose or, for other reasons, have a different set of metadata properties that needs be stored with each document. Once a SharePoint document library (or list) has been designed, it can be modeled using Windows SharePoint Services and saved as a document library or list template. The template can be used directly in the site where it was created to create multiple

identical document libraries or lists in any number of sites. Also, it can be downloaded from or uploaded into the template library of another Windows SharePoint Services server, and then used to create new document libraries or lists on the new WSS server.

Once the design of the target SharePoint environment is complete, the last task of the modeling step is to physically implement the design on the target Windows SharePoint Services server. This can be accomplished using the Windows SharePoint Services HTML administration pages, or with third-party software. Some of these tools are described in the section *Choosing a Toolset*.

4. Extraction

Extraction is the process of physically copying or moving content from the source content store to an intermediate file format or database. Depending on the sophistication of the software tools used, extraction may involve copying data from all five levels of the 5LCM model or at a minimum, level one (content) and possibly level two (metadata). Depending on the content access protocols supported by the source store, the document content (file streams) may be left in the source store and accessed based on the documents' URLs, or the file streams may be extracted and temporarily stored on a file system for use by the subsequent mapping and injection steps.

5. Mapping

Mapping is an optional intermediate step between extraction and injection that handles situations where the metadata needs to be mapped or transformed before the extracted source content is injected into the target SharePoint environment. Scenarios where mapping is required include:

- Differences in metadata property names. This scenario occurs when, during planning, the names chosen for the metadata properties in the target SharePoint environment are not the same as the corresponding names from the source content store. For example, the source content store has a metadata property named Surname, while an organization's internal conventions dictate that the corresponding property in the target SharePoint environment is called LastName.

- Differences in metadata property values. This scenario occurs when the values of a property in the target SharePoint environment are different from those used for the corresponding property in the source content store. A frequent example of this is when a source property uses a numerically coded value and an organization chooses to use an easier-to-read set of string values for the target SharePoint environment.

The existence of these scenarios in a particular project strongly indicates the need to consider a commercial content migration solution.

6. Injection

Content injection is the final step in physically copying the extracted content into the target SharePoint environment. At this point in the content migration process, the source content has been extracted and the model for the target SharePoint environment has been implemented on the physical server(s). What remains to be done is the application of the mappings from step five of the process to the extracted metadata; and, together with the document file streams, the writing of the content to the appropriate SharePoint site and document library (or list).

A key requirement of the injection step is to maintain a log of the documents that have and have not been successfully written to the target SharePoint environment. Being able to collect and track the list

of documents that failed to migrate is particularly important as most content migration projects require every document to be accounted for.

Why do some documents fail to inject into the target SharePoint environment? The most common reasons are:

1. The file type (suffix) of a source document is on the list of Windows SharePoint Services blocked file extensions list for the server or server farm. The default list of blocked file extensions is managed using the SharePoint Central Administration HTML administration pages.

2. File and path name restrictions. These include the set of characters allowed for use in a file or folder name, the maximum length of a file name or individual folder name, as well as the overall length of the path name (including the folder, subfolders and files).

Specific details on the file and path name restrictions can be found in the section Appendix A – File and Folder Naming Conventions and Limits.

7. Verification

The verification step reports on the progress of the overall content migration process. The primary goal of the verification step is to report on the differences between the contents of the target SharePoint environment and the source content store resulting that, primarily, result from issues encountered during the Injection step.

8. Workflow Migration

Although workflow or application migration appears as the last step in the 8SCM process, this step usually begins immediately after planning is complete and runs in parallel with the core migration processes (steps 3-6). This is because this migration step usually requires some level of custom software development to implement the event handling logic or the deployment and configuration of a third-party workflow solution. The amount of effort required for this step can vary considerably depending on the original workflow requirements, if any, of the source content and any new workflow requirements for the migrated content.

Migration vs. Co-existence

Many organizations begin their project planning with the goal of undertaking a one-time migration of content from a particular source content store into their SharePoint environment. This approach often turns out to be unrealistic for the following reasons:

- There is more source content than can be migrated in, the time originally specified, for example, one day, overnight, or a weekend.

- All of the users cannot be migrated at the same time due to internal considerations such as the migration of applications that use the source content store to an updated set of applications that use the target content store.

- More time is needed to train each user department on the best use of their SharePoint environment, for example.

This creates the need for the content migration project to include a co-existence strategy: the simultaneous use of the source and target content stores (and associated applications and business processes) until a time when the content, applications, and users have been successfully migrated to the new environment.

During the co-existence period, based on business requirements, there are additional technical process requirements for keeping the content synchronized between the original source content store and target environment. This may involve ongoing one-way synchronization (incremental migrations) from the source content store to the target store. In addition, there may be a requirement to support bi-directional content synchronization and change conflict resolution to support simultaneous use of both environments while they co-exist.

Following a successful co-existence, the original source content store and applications can be backed up and then decommissioned. It is very important to plan for this event; otherwise, there will be a few users who continue to use the old environment. A co-existence period that runs longer than necessary means increased operations and support costs.

Choosing a Toolset

Develop Your Own Tools or License a Solution?

Some organizations have opted to develop their own content migration tools, while a number of Microsoft independent software vendor (ISV) partners have created software solutions for all or part of the overall content migration process.

The decision to create a set of internally developed tools versus licensing a third-party solution is primarily determined by the following factors:

i)   Ability of third-party tools to address an organization's specific requirements

ii)   Licensing cost

iii)   Software development expertise using the SharePoint Products and Technologies object models and web services

iv)   Experience managing content in the source content store and its application programming interfaces (APIs)

v)   Content migration domain expertise

Considering each of these factors, the following describes selected content migration scenarios and a Microsoft ISV partner solution that can addresses the requirements of each scenario. Most of the partner-developed content migration solutions described below can also be applied to a number of other scenarios.

File Server and Local File System Content

One of the first and most advanced solutions for migrating content based on file servers and local file systems into SharePoint Products and Technologies is the Tzunami Deployer from Tzunami Incorporated (http://www.tzunami.net). The Tzunami Deployer supports all of the core content migration steps and is a very good tool for modeling the target SharePoint environment.

Web Sites

The management and migration of Web site content into a SharePoint environment is a growing area of interest amongst Microsoft customers. Metalogix Migration Manager from Metalogix Corporation (http://www.metalogix.net), a long-time Microsoft Content Management Server ISV partner, supports all of the core content migration steps with tools to inventory and analyze content and metadata from Web sites, file servers and third party content stores, and remotely migrate HTML content, documents, files and data to Window SharePoint Services and SharePoint Portal Server 2003 repositories.

Lotus Notes Databases and other Line-of-Business Database Applications

Casahl Incorporated (http://www.casahl.com) is an ISV partner that has focused on the migration and two-way synchronization between LOB data in Lotus Notes databases, SQL Server, and other relational databases and Microsoft Exchange public folders. Casahl ecKnowledge support the integration, extension, and migration of content and data from Lotus Notes and LOB database applications for SharePoint Products and Technologies, as well as other target content stores.

SharePoint Portal Server 2001 Workspace Document Libraries

A number of Microsoft customers deployed SharePoint® Portal Server 2001, the predecessor product to SharePoint® Portal Server 2003. To provide support for updating and migrating content from SharePoint Portal Server 2001 to SharePoint Portal Server 2003 document libraries, Microsoft created the SharePoint Portal Server 2003 Document Library Migration Tools. These tools consist of two programs: Spout.exe (SharePoint output from SharePoint Portal Server 2001) and Spin.exe (SharePoint input from SharePoint Portal Server 2003). More information on the SharePoint Document Library Migration Tools can be found in the Microsoft Download Center (http://www.microsoft.com/downloads).

Exchange Public Folders

Windows SharePoint Services and SharePoint Portal Server 2003 are the strategic server platforms at Microsoft for IT professionals who are developing and deploying collaborative business applications. Over time, it is expected that Microsoft customers will prefer Windows SharePoint Services for deploying new collaborative applications instead of using Exchange public folders. For Exchange Server 5.5 and Exchange Server 2000 customers upgrading to Exchange Server 2003 or later, many are considering migrating their existing public folder content to Windows SharePoint Services.

Things to plan for with the 5LCM model for Exchange public folders include:

- A single Exchange public folder can store many types of items including e-mail messages, calendar appointments, tasks, and contacts, in addition to traditional documents. Each type of Exchange item will have its own corresponding set of metadata fields associated with each item.

- Like file servers, public folders many contain significant amounts of old information that can be backed up and deleted instead of being migrated.

- Selected public folders can be enabled for event scripting. Public folder event scripts enable business logic to be executed whenever a new item is created, modified, or deleted in a folder.

- Public folders that are e-mail enabled. These are public folders that have simple mail transport protocol (SMTP) e-mail addresses associated with them. This enables e-mail users to send e-mail messages (including attachments) to an Exchange server and have the messages automatically stored in the public folder associated with the message recipient's e-mail address.

Parallelspace Metadata Management Services (MMS) is a content migration solution from Parallelspace Corporation (http://www.parallelspace.net) for inventorying, analyzing, and migrating Exchange public folder content to Window SharePoint Services and SharePoint Portal Server 2003. It supports all eight steps in the 8SCM process.

Some sample code for uploading a document into a SharePoint document library can be found in *Appendix B – File Upload Sample Code*.

Other Content Stores

Many customers have other types of content stores that they will want to migrate to WSS. The 5LCM model can be used to analyze the content and help you develop a strategy for migrating the content to Windows SharePoint Services.

In addition, contact the Microsoft partners mentioned above. Their content migration solutions support source content stores in addition to the ones highlighted in this whitepaper. The following table highlights the partner solutions that can be used to migrate each of the key content stores.

Table 1. Partner Content Migration Solutions

Partner	File Servers	Web Content	Lotus Notes Databases	Line-of-Business Databases	SharePoint Portal Server 2001	Exchange Public Folders
Casahl Incorporated	●	●	●	●	●	●
Metalogix Corporation	●	●				
Parallelspace Corporation	●		●		●	●
Tzunami Incorporated	●				●	●

Summary

While Windows SharePoint Services provides basic capabilities for uploading single documents and small collections of documents into a SharePoint document library, enterprise content migration projects will be more successful if they use a content migration framework consisting of reference models, processes, and a supporting set of tools and best practices. This white paper has described the 5LCM (five-level content model) and the 8SCM (eight-step content migration) process, as well as a set of content migration tools available from Microsoft and its solution partners. Equipped with this information, IT professionals can approach their content migration projects for Windows SharePoint Services and SharePoint Portal Server 2003 with greater understanding and confidence.

Related Links

For more information on Microsoft SharePoint Products and Technologies, please visit the following Web sites.

Technical Resources

- Coexistence and Interoperability Guide for SharePoint Products and Technologies
  http://office.microsoft.com/en-us/assistance/HA011607771033.aspx

- Microsoft Developers Network (MSDN) Developer Center for Microsoft SharePoint Products and Technologies
  http://msdn.microsoft.com/sharepoint

- Microsoft TechNet SharePoint Portal Server 2003 Technical Resources
  http://www.microsoft.com/technet/prodtechnol/office/sps2003/default.mspx

- Microsoft TechNet Windows SharePoint Services Technical Resources
  http://www.microsoft.com/technet/prodtechnol/windowsserver2003/technologies/sharepoint/default.mspx

- Microsoft TechNet Downloadable Resources for Microsoft SharePoint Products and Technologies
  http://www.microsoft.com/technet/downloads/sharepnt.mspx

- Microsoft IT Showcase Deployment White Papers
  http://www.microsoft.com/itshowcase

- WSS FAQ Web site
  http://wss.collutions.com/default.aspx

Product Information

- SharePoint Products and Technologies product Web site
  http://www.microsoft.com/sharepoint

- Windows SharePoint Services and SharePoint Portal Server 2003: Product Information
  http://www.microsoft.com/office/sharepoint/prodinfo/relationship.mspx

- Benefits of Microsoft Office SharePoint Portal Server 2003
  http://www.microsoft.com/office/sharepoint/prodinfo/benefits.mspx

- SharePoint Portal Server 2003 Customer Evaluation Guide
  http://www.microsoft.com/office/sharepoint/prodinfo/guide.mspx

- Implementing Rich Collaboration Infrastructure Using Windows SharePoint Services and SharePoint Portal Server 2003
  http://www.microsoft.com/sharepoint/evaluationoverview.asp

- Microsoft Web Enterprise Portal
  http://www.microsoft.com/downloads/details.aspx?FamilyID=ac26898b-6893-48b9-8ec0-667f1ba22d6b&DisplayLang=en

Analyst and Research Reports

- Process Goldmine: Microsoft Office System Integrated Solutions Deliver Business Value, Navigant Consulting Inc., September 2003.
http://www.microsoft.com/office/business/value.mspx

- The Secret(s) Behind SharePoint Portal Server 2003, Microsoft Watch, December 2004.
http://www.microsoft.com/office/sharepoint/prodinfo/microsoftwatch.mspx

For the latest information about Windows Server 2003, see the Windows Server 2003 Web site.

Appendix A – File and Folder Naming Conventions and Limits

To ensure a smooth migration of a collection of files stored in folders and subfolders on a local file system or remote file server, one best practice is to make sure your folder, file, and path names conform to the following limits and other restrictions.

Folder path and fully qualified file name lengths

A fully qualified file name consists of two parts:

- File name including the file base name and file suffix.

- Folder path for the fully qualified file name.

The filename is the last component of a fully qualified file name. The folder path for a fully qualified file name starts with the partition's drive letter and includes everything up to the filename and the last folder character (e.g. '\' or '/').

Restrictions – Microsoft Windows

The fully qualified file name can have a maximum length of 260 characters. The maximum length of a folder path portion of a fully qualified file name is 248 characters.

Restrictions – Microsoft SharePoint Portal Server 2003

The fully qualified file name can have a maximum length of 255 characters in Windows. The maximum length of a folder component is 30 characters.

Discussion

- Usually, it is not possible using Windows Explorer or a Windows application to create folder paths and fully qualified file names that are too long. However, it is possible under the following scenarios:

- A subfolder one or more levels below the root of the file system is shared as UNC (Universal Naming Convention) file share. Then it is possible to create an entire new folder hierarchy under the root of the file share that can store files using fully qualified file name that are a further 260 characters long.

- Similarly, the `subst` command can be used to anchor a drive letter at a subfolder one or more levels below the root of the file system.

- Using Unicode fully qualified file names prefixed with the letters "\\?\". When this approach is used, fully qualified path names can be 32,000 characters in length.

Unsupported characters in folder and file names

The following characters are not supported in Windows SharePoint Services folder and file names:

> / \ : * ? " < > | # { } % & ~ or tab characters and multiple periods.

If a file, folder, or URL name in your original site contains one of these characters, it is usually replaced with an underscore (_).

Two exceptions are:

- Character '#' (pound sign) which can be replaced with the word "No"

- Character '&' which can be replaced with the word "And"

Multiple periods are replaced with a single period. Additional digits may be appended to the file or folder name if there are conflicting renaming changes. Files or folders with trailing spaces at the very end of a filename ("filename.doc ") are not supported in Windows SharePoint Services. You must change the file name and then add the file manually to your site.

Blocked file types

Windows SharePoint Services (and in turn, SharePoint Portal Server 2003) have the ability to block certain file types from being uploaded or otherwise stored in a document library.

The following is the default list of file types (based on a file's suffix) that are blocked from being uploaded or otherwise stored into a SharePoint document library.

Table 2. SharePoint default blocked file types

File extension	File type
.ade	Microsoft Access project extension
.adp	Microsoft Access project
.app	Application file
.bas	Microsoft Visual Basic class module
.bat	Batch file
.chm	Compiled HTML Help file
.class	Java class file
.cmd	Microsoft Windows NT Command Script
.com	Microsoft MS-DOS program
.cpl	Control Panel extension
.crt	Security certificate
.dll	Windows dynamic link library
.exe	Program
.fxp	Microsoft Visual FoxPro compiled program
.hlp	Help file
.hta	HTML program
.ins	Internet Naming Service
.isp	Internet Communication settings
.jse	Jscript® Encoded Script file
.lnk	Shortcut
.mda	Microsoft Access add-in program
.mdb	Microsoft Access program
.mde	Microsoft Access MDE database
.mdt	Microsoft Access data file
.mdw	Microsoft Access workgroup
.mdz	Microsoft Access wizard program

.msc	Microsoft Common Console Document
.msi	Microsoft Windows Installer package
.msp	Windows Installer patch
.mst	Visual Test source files
.ops	Microsoft Office profile settings file
.pcd	Photo CD image or Microsoft Visual Test compiled script
.pif	Shortcut to MS-DOS® program
.prf	System file
.prg	Program source file
.reg	Registration entries
.scf	Windows Explorer command file
.scr	Screen saver
.sct	Windows Script Component
.shb	Windows shortcut
.shs	Shell Scrap Object
.url	Uniform Resource Locator (Internet shortcut)
.vb	Microsoft Visual Basic® Scripting Edition (VBScript) file
.vbe	Visual Basic Script Encoded Script file
.vbs	Visual Basic Script file
.wsc	Windows Script Component
.wsf	Windows Script file
.wsh	Windows Script Host Settings file

The following sections contain the character limitations that apply to various user-specified parameters in Microsoft Office SharePoint Portal Server 2003.

Other Naming Considerations

Portal Site

Portal site name

The portal site name is limited to 80 Unicode characters. The portal site name can consist of all alphanumeric characters except for the following: **\ / : * ? " < > |**.

Portal site description

The portal site description is limited to 200 Unicode characters. The portal site description can consist of all alphanumeric characters, as well as the following: **# % * + \ | " ? $ @ [ ] { } ¦**.

Portal site logo

The URL for the portal site logo is limited to 256 ASCII characters. It can consist of all alphanumeric characters except for the following: **# % * + \ | " ? > < $ @ [ ] { } ¦**.

The following file formats are supported for the graphics file: .gif, .bmp, .jpg, and .png. TIFF files are not supported.

Portal site URL

The URL for the portal site is limited to 255 characters for the entire URL, including virtual directories.

Account name for portal site owner

The account name is limited to 128 characters.

E-mail address for portal site owner

The e-mail address is limited to 255 characters.

Location for creating SharePoint sites

The URL for creating sites from the Site Directory is limited to 2,048 ASCII characters. In addition, no component of the URL, such as the virtual directory or virtual server, can exceed 128 characters.

The URL must end with /_layouts/language/scsignup.aspx.

Backward-Compatible Document Library (Web Storage System-Based)

Document library friendly name

The friendly (display) name cannot exceed 100 characters in length.

Alternate Portal Site Access Mappings

Mapping name

The mapping name has a maximum length of 64 Unicode characters and must not be null. Each URL has a maximum length of 255 Unicode characters and a minimum length of 5 characters. The user can enter the URL in any combination of uppercase and lowercase letters, but the URL is always saved in lowercase. Each URL must be different from all other URLs.

Content Index

Content index name

The content index name must be less than 50 characters. The content index name cannot contain the following characters: + ~ # ' % * ( ) = [ ] { } | \ " < > . ? / @ & or the euro symbol or a space.

Appendix B – File Upload Sample Code

```
SPWeb site = new SPSite("http://server01:81/SampleES").OpenWeb();

FileStream fStream = File.OpenRead(txtUrl.Text);
byte[] contents = new byte[fStream.Length];
fStream.Read(contents, 0, (int)fStream.Length);
fStream.Close();
site.Files.Add("http://server01:81/SampleES", contents);
```

Other sample code for importing content into SharePoint document libraries can be found at
http://workspaces.gotdotnet.com/spimport.