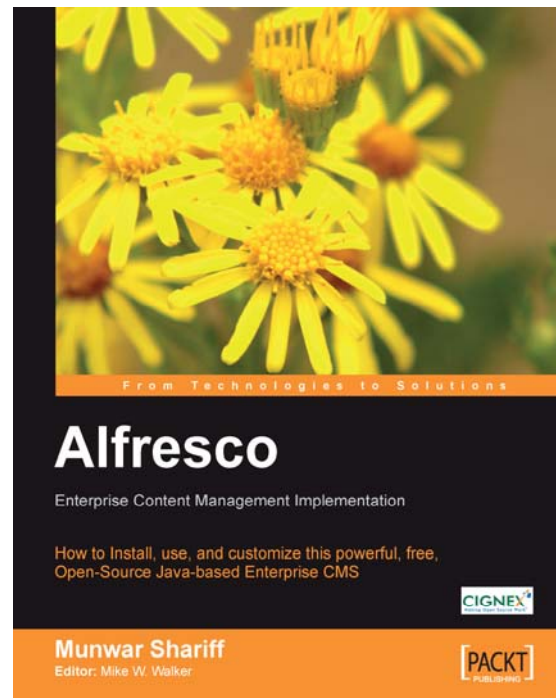# [PACKT] PUBLISHING

Alfresco Enterprise Content Management Implementation

How to Install, use, and customize this powerful, free, Open Source Java-based Enterprise CMS

Munwar Shariff

Chapter 5
"Implementing Document Management"

In this package, you will find:

A Biography of the authors of the book

A preview chapter from the book, Chapter 5 "Implementing Document Management"

A synopsis of the book's content

Information on where to buy this book

About the Author

Munwar Shariff

Munwar Shariff is the CTO and VP of Business Development at CIGNEX.   CIGNEX is the leading provider of open-source Enterprise Content Management (ECM) solutions for businesses and government agencies.

He has worked as the chief architect and manager of engineering teams for 15 years in the areas of system software, Internet applications, and mobile commerce applications for customers in the United States, Japan, Germany, and India.

 He is an expert in Content Management Systems (CMS). Since co-founding CIGNEX in late 2000, he has successfully delivered more than 50 CMS applications using various open-source technologies. He has written a number of articles on open-source CMS; he is an experienced trainer and a frequent speaker at conferences related to this topic.

Munwar earned his MS in Digital Electronics and Advanced Communications from REC Surathkal, India. Munwar is the co-author of the book on the Plone Open Source Content Management System called Plone Live.

About the Reviewer

Mike W. Walker

Mike W. Walker is a Vice President, in-charge of consulting at CIGNEX (http://www.cignex.com). He has over 16 years of progressive consulting experience with world class consulting firms, including KPMG, Accenture, and Ernst & Young Cap Gemini America. Mike's organization at CIGNEX is among the very first to successfully deliver ECM projects using Alfresco. He is currently engaged in very complex ECM projects based on Alfresco. Mike has deep expertise both in Alfresco technical architecture and in creating Content Management solution architectures containing Alfresco. Mike's main areas of interest include consulting services delivery processes and methodologies development.

Alfresco Enterprise Content Management Implementation

Looking at the title of this book, most people may think, "Another book on an open-source CMS?" It's a natural reaction as there are about 1700 content management systems out there in the market. For the past ten years, I have been implementing various content management systems. I started with multi-million dollar implementations of proprietary software, and faced all kinds of challenges including vendor lock-in, rigid code base, and expensive upgrades. At CIGNEX, our focus has been proving value to our customers using open-source alternatives to commercial CMS products. We research on various open-source products, and propose the most suitable product, which satisfies our customers' requirements over a long period of time.

Unlike most other open-source CMSes, which offered only web content management, Alfresco provided a wide range of solutions to Enterprise customers with an impressive roadmap. And most importantly, it is created using completely open standards. This excited us a lot, and we started implementing Alfresco in many enterprises. As part of our implementation, we also train our customers, so that they are equipped with all the information required to manage their systems. I have trained many users, administrators, and developers in Alfresco and many other systems. This book distils the hands-on approach of my training courses into a concise, practical book.

The book focuses on business needs rather than technical syntax. I started by showing the reader how to do something — a step by step example. I explained how that process worked. Then, I explained what other options are available, and how they fit into the overall picture. I hope this helps the reader "generalize" from such examples. I hope that you take advantage of this book by setting up a flexible enterprise content management system for your company and customers.

Your feedback is very valuable to me. You can contribute by reporting any errors you find in the book, making suggestions for new content that you'd like to see in future updates, commenting, and blogging about it.

What This Book Covers

This book will take you through the complete cycle of planning, implementing, and customizing your ECM installation. The topics that this book covers are:

Chapter 1 includes an overview of Alfresco architecture and key features of the software. This chapter also includes using Alfresco for your document management, records management, web content management, and collaboration requirements and also a future roadmap.

Chapter 2 includes tips to choose the right installation for you, and also installation of the software and start using it.

Chapter 3 includes basic planning and configuring your Alfresco installation.

Chapter 4 includes working with users, and membership accounts, including LDAP integration.

Chapter 5 includes using Alfresco as a smart document repository; working with automatic version tracking and control, and accessing the repository from the Web, shared network folders, or FTP.

Chapter 6 includes automating document management tasks with business rules and complete workflows.

Chapter 7 includes designing custom content types.

Chapter 8 includes making content easy to find using search, content categorization, and metadata.

Chapter 9 includes enhancing automated document management tasks with business rules, and complete workflows.

Chapter 10 includes working together using Alfresco's collaboration and syndication features to create effective working groups.

Chapter 11 includes customizing the user interface and creating your own dashboard layouts, presenting content in custom ways relevant to your business.

Chapter 12 includes maintaining your system including exporting data and upgrading your system to newer versions.

Chapter 13 includes collecting paper documents and forms, transforming them into accurate, retrievable information, and delivering the content into an organization's business applications.

5
Implementing Document Management

This chapter introduces you to the basic features of creating and managing content in Alfresco. Using Alfresco, you can manage any type of documents such as HTML, text, XML, Microsoft Office documents, Adobe PDF, Flash, scanned images, media, and video files. You will also learn about the concepts of creating and using categories and smart spaces. This chapter also focuses on the most important aspect of adopting a new enterprise content management system, which is migrating the existing data and using it effectively.

By the end of this chapter you will have learned how to:

- Create spaces and fill them with documents
- Automatically control the document versioning
- Lock, check-in, and check-out the documents
- Categorize content to facilitate searching
- Recover deleted content
- Create and use space templates
- Access documents in the Alfresco repository from your web browser, or a networked drive, FTP, or WebDAV
- Migrate existing documents to Alfresco

Managing Spaces

A space in Alfresco is nothing but a folder, which contains content as well as sub-spaces. Space users are the users invited to a space to perform specific actions such as editing content, adding content, discussing a particular document, etc. You

need to have the administrator, contributor, collaborator, or coordinator role, on a space to create sub-spaces. Similarly you need to have the administrator, editor, collaborator, coordinator role, to edit space properties. For more information about user roles on a space, refer to Chapter 4.

Space is a Smart Folder

Space is a folder with additional features such as security, business rules, workflow, notifications, local search, and special views. These additional features that make a space a smart folder are explained below:
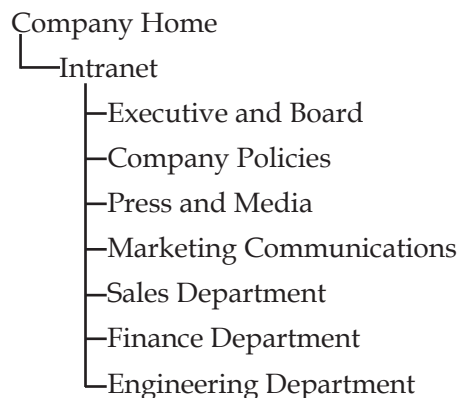
- Space Security: You can define security at the space level. You can specify a user or a group of users who may perform certain actions on content in a space. For example, on the *Marketing Communications* space in *Intranet*, you can specify that only users of the marketing group can add the content and others can only see the content.

- Space Business Rules: Business rules such as transforming content from Microsoft Word to Adobe PDF and sending notifications when content gets into a space can be defined at space level.

- Space Workflow: You can define and manage content workflow on a space. Typically, you will create a space for the content to be reviewed, and a space for approved content. You will create various spaces for dealing with the different stages the work flows through, and Alfresco will manage the movement of the content between those spaces.

- Space Events: Alfresco triggers events when content gets into a space, or when content goes out of a space, or when content is modified within a space. You can capture such events at space level and trigger certain actions such as sending email notifications to certain users.

- Space Aspects: Aspects are additional properties and behavior, which could be added to the content, based on the space in which it resides. For example, you can define a business rule to add customer details to all the customer contract documents in your Intranet's *Sales* space.

- Space Search: Alfresco Search can be limited to a space. For example, if you create a space called *Marketing*, you can limit the search for documents within *Marketing* space, instead of searching the entire site.

- Space Syndication: Space content can be syndicated by applying RSS feed scripts on a space. You can apply RSS feeds on your *News* space, so that other applications and websites can subscribe for news updates.

- Space Content: Content in a space can be versioned, locked, checked-in and checked-out, and managed. You can specify certain documents in a space to be versioned and others not.

- Space Network folder: A space can be mapped to a network drive on your local machine enabling you to work with the content locally. For example, using the CIFS interface a space can be mapped to the Windows network folder.

- Space Dashboard View: Content in a space can be aggregated and presented using special dashboard views. For example, the Company Policies space can list all the latest policy documents that have been updated during the past one month or so. You can create different views for the Sales, Marketing, and Finance departmental spaces.

Why Space Hierarchy is Important

Like regular folders, a space can have spaces (called sub-spaces) and sub-spaces can further have sub-spaces of their own. There is no limitation on the number of hierarchical levels. However, the space hierarchy is very important for all the reasons specified above in the previous section. Any business rule and security defined at a space is applicable to all the content and sub-spaces underlying that space.

In the previous chapter, you created system users, groups, and spaces for various departments as per the example. Your space hierarchy should look like the one given below:

Company Home
└─Intranet
    ├─Executive and Board
    ├─Company Policies
    ├─Press and Media
    ├─Marketing Communications
    ├─Sales Department
    ├─Finance Department
    └─Engineering Department

Space hierarchy enables you to define various business rules, dashboard views, properties, workflow, and security for the content belonging to each department. You can decentralize the content management by giving access to departments at individual space levels.

The Intranet space in our example should contain sub-spaces as shown in the screenshot on the next page. If you have not already created spaces as per the example given in the previous chapter, you must do it now by logging in as the administrator. The examples used in the remaining chapters of this book refer to these spaces. Also,

it is very important to set security (by inviting groups of users to these spaces) as explained in the previous chapter.
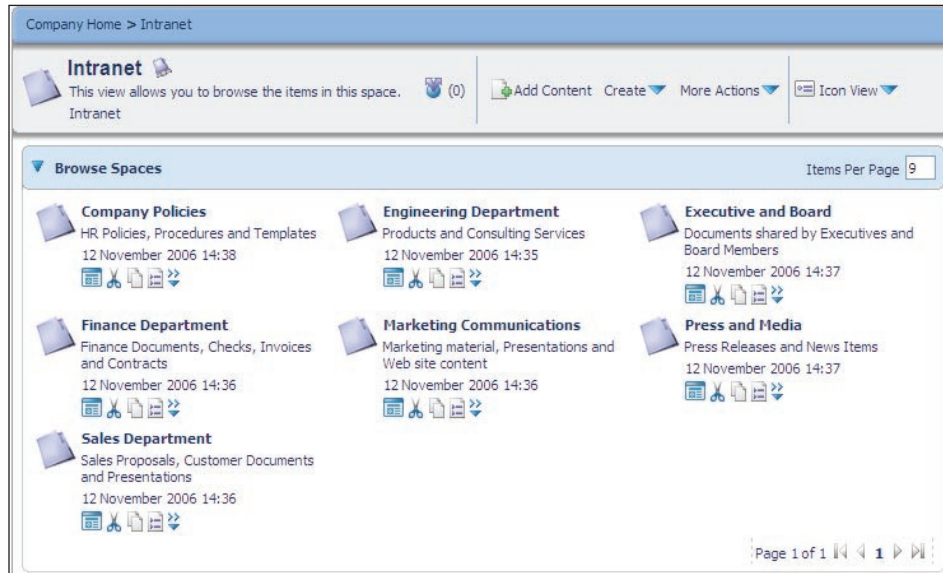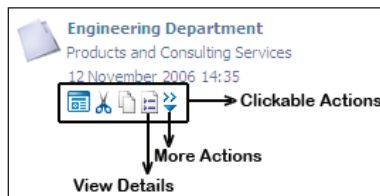

Edit Space

Using the web client, you can edit the spaces you have added previously. Note that you need to have edit permissions on the spaces to edit them as explained in the previous chapter.

Edit Space Properties

Every space listed will have clickable actions as shown in the following screenshot:

These clickable actions will be dynamically generated for each space based on the current user's permissions on that space. If you have copy permission on a space you will notice the copy icon as a clickable action for that space. On clicking the **View Details** action icon the detailed view of a space will be displayed as shown in the next screenshot:

The detailed view page of a space allows you to select a dashboard view, to view and edit existing space properties, to categorize the space, to set business rules, and to run various actions as shown in the above screenshot.

To edit space properties, click on the edit properties icon shown in the above screenshot. You can change the name of the space and other properties as needed.

Delete a Space and Its Contents

From the list of space actions, you can click on the **Delete** action to delete the space. You need to be very careful while deleting a space as all the business rules, sub-spaces, and the entire content within the space will also be deleted.

Move or Copy a Space Using the Clipboard

From the list of space actions, you can click on the **Cut** action to move a space to the clipboard. Now you can navigate to any space hierarchy, given that you have the permission, and paste this particular space as required.

Similarly, you can use the **Copy** action to copy the space to some other space hierarchy. This is useful if you have an existing space structure (such as a marketing project or engineering project) and you would like to replicate it along with the data.

The copied or moved space will be identical in all aspects to the original (source) space. When you copy a space its properties, categorization, business rules, space users, entire content within the space, and all sub-spaces along with their content will also be copied.

Create a Shortcut to a Space for Quick Access

If you need to frequently access a space, you can create a shortcut (similar to a browser favorite) to that space to reach the space in one click. From the list of space actions, you can click on the **Create Shortcut** action to create a shortcut to the existing space. Shortcuts are listed in the left-hand side **shelf**.

Choose a Default View for Your Space

Out of the box, four different views (as shown in the screenshot on the next page) are supported to display the space information.

Details View provides listings of sub-spaces and content in horizontal rows.

Icon View provides title, description, timestamp, and action menus for each sub-space and content item present in the current space.

Similarly, **Browse View** provides title, description, and a list of sub-spaces for each space.

The **Custom View** is disabled and appears in grey. This is because you have not enabled a dashboard view for this space. In order to enable a dashboard view for a space, you need to select a dashboard view (refer to the icon in the earlier screenshot).
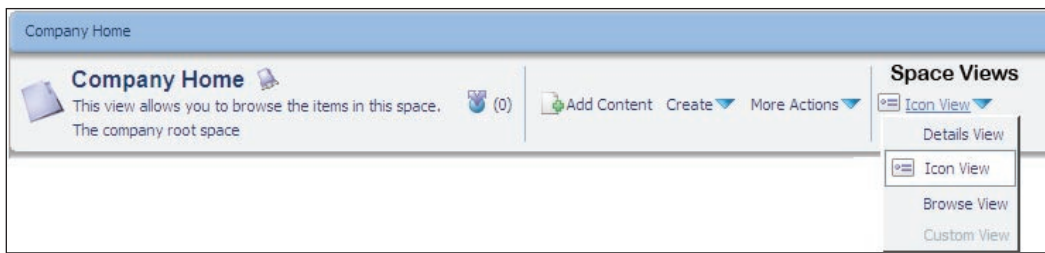
Sample Marketing-Project Space Structure

Let us say you are launching a new marketing project called 'Switch to open source ECM'. You can create your own space structure within the marketing project space to manage content. For example, you can have a space called *Drafts* to keep all the draft marketing documents and so on. **Go to the Company Home | Intranet | Marketing Communications** space and create a new space called *Switch to open source ECM* and create various sub-spaces under that as shown in the screenshot below.

The new marketing project space and the sub-spaces created are used in the remaining examples listed in this chapter to manage content.
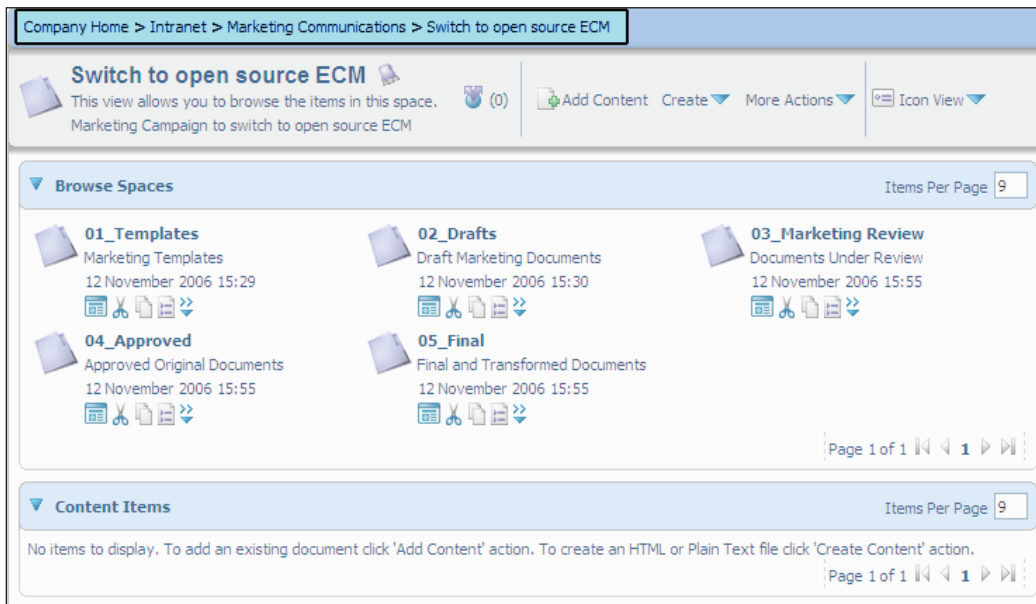
Managing Content

Content could be of any type as mentioned at the start of this chapter. Using the Alfresco web client application, you can add and modify content and its properties. You can categorize content, lock content for safe editing, and you can maintain several versions of the content. You can delete content and you can recover the deleted content as well.

This section uses the space you have already created as a part of your *Intranet* sample application. As a part of sample application, you will manage content in the **Intranet | Marketing Communications** space. As you have secured this space earlier, only the administrator (*admin*) and users belonging to the Marketing group (*Peter Marketing* and *Harish Marketing*) can add content in this space. You can log in as *Peter Marketing* to manage content in this space.

Create Content

The web client provides two different interfaces for adding content: one to create inline editable content such as HTML, Text, and XML and the other to add binary content such Microsoft office files and scanned images.

You need to have the Administrator, Contributor, Collaborator, Coordinator role on a space to create content within that space. For more information about user roles on a space, refer to Chapter 4.

Creating Text Documents—HTML, Text, and XML

To create an HTML file in a space, follow the steps given below:

1. Ensure that you are in the **Intranet | Marketing Communications | Switch to open source ECM | 02_Drafts** space.

2. On the header, click **Create | Create Content**. The first pane of the **Create Content** wizard appears as shown in the screenshot on the next page. In this wizard, and in any Alfresco wizard, you can track your progress through the wizard from the list of steps at the left of the pane.

3. Provide the name of the HTML file, select HTML as **Content Type**, and click the **Next** button. The **Enter Content** pane of the wizard appears as shown in the next screenshot. Note that **Enter Content** is now highlighted in the list of steps at the left of the pane.

4. You can see that there is a comprehensive set of tools to help you format your HTML document. Enter some text, using some of the formatting features.

5. If you know HTML, you can also use an HTML editor by clicking on the HTML icon given. The HTML source editor is displayed. Once you have updated the HTML content, click on the update button to return to the **Enter Content** pane in the wizard, with the contents updated.

6. After the content is entered and edited in the **Enter Content** pane, click **Finish**. You will see the **Modify Content Properties** screen to update metadata associated with the content as shown in the screenshot below:
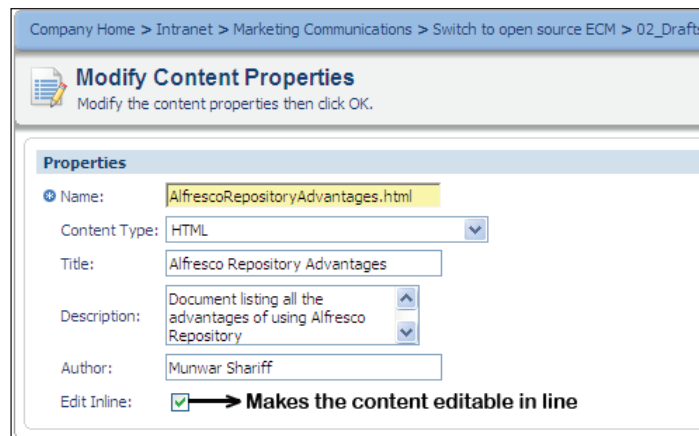
7. If you are satisfied with the properties, click the **OK** button to return to the **02_Drafts** space, with your newly created file inserted.

You can launch the newly created HTML file by clicking on it. Your browser launches most of the common files such as HTML, text, and PDF. If the browser could not recognize the file, you will be prompted with the Windows dialog box containing the list of applications, from which you must choose an application. This is the normal behavior if you try to launch a file on any Internet page.

Uploading Binary Files—Word, PDF, Flash, Image, and Media

Using the web client, you can upload content from your hard drive. Choose a file from your hard disk that is not an HTML or text file. I chose *Alfresco_CIGNEX.doc* from my hard disk for the sample application. E**nsure that you are in the Intranet | Marketing Communications | Switch to open source ECM | 02_Drafts** space.

To upload a binary file in a space, follow the steps given below:

1. In the space header, click the **Add Content** link.
2. The **Add Content** dialog appears.
3. To specify the file that you want to upload, click **Browse**. In the **File Upload** dialog box, browse to the file that you want to upload. Click **Open**. Alfresco inserts the full path name of the selected file in the **Location** text box.
4. Click the **Upload** button to upload the file from your hard disk to the Alfresco repository. A message informs you that your upload was successful as shown in the following screenshot.
5. Click **OK** to confirm.
6. The **Modify Content Properties** dialog appears. Verify the pre-populated properties and add information in the text boxes. Click **OK** to save and return to the **02_Drafts** space.
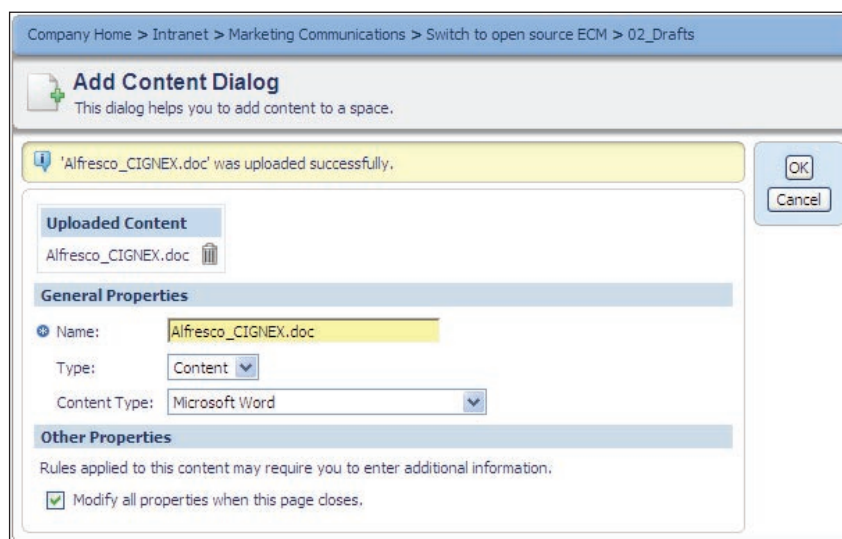
The file that you uploaded appears in the **Content Items** pane. Alfresco extracts the file size from the properties of the disk file, and includes the value in the size row. Now that you have two files, you can edit them as you like.

Edit Content

Using the web client you can edit the files that you have added previously. Note that you need to have edit permissions on the content to edit them as explained in the previous chapter.

Inline Editing of HTML, Text, and XML

HTML files and plain text files can be created and edited inline. Each file type is edited in its own WYSIWYG editor. If you have edit access to a file, you will notice a small pencil (edit) icon as shown in the screenshot below. Clicking on the pencil icon will open the file in its editor.
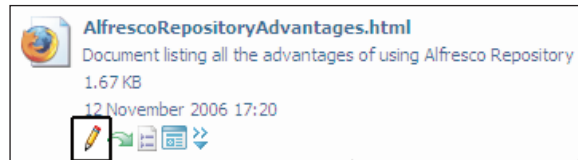
Upload an Updated Binary File

If you have edit access to a binary file, you will notice the **Update** action icon in the more actions link as shown in the screenshot below. On clicking on the update icon, the update pane opens. Click on the **Browse** button to upload the updated version of the document from your hard disk.
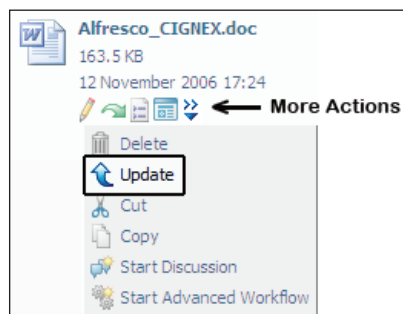
Content Actions

Content will have clickable actions as shown in the screenshot on the next page. These clickable actions (icons) will be dynamically generated for a content based on the current user's permissions for that content. For example, if you have copy permission for the content, you will notice a copy icon as a clickable action for that content.
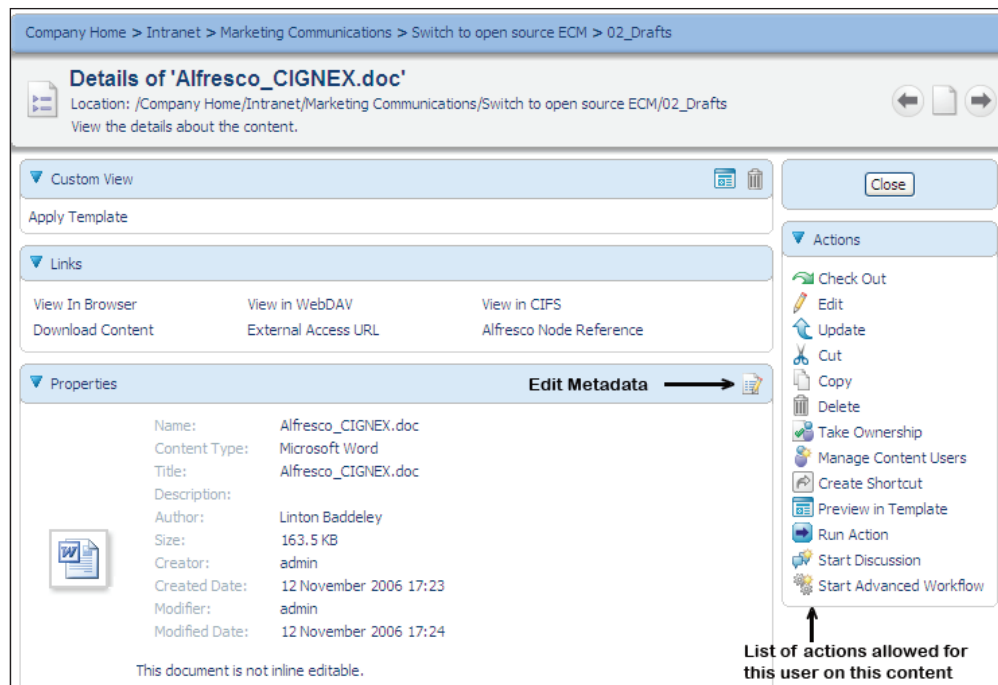
Delete Content

Click on the **Delete** action, from the list of content actions, to delete the content.

Move or Copy Content Using the Clipboard

From the list of content actions as shown in the previous screenshot, you can click on the **Cut** action to move the content to the clipboard. Now you can navigate to any space hierarchy and paste this particular content as required.

Similarly, you can use the **Copy** action to copy the content to another space.

Create a Shortcut to Content for Quick Access

If you have to access particular content very frequently, you can create a shortcut (similar to a browser favorite) to that content to reach the content in one click. **From** the list of content actions as shown in the **View Details** page, you can click on the **Create Shortcut** action to create a short cut to the existing content. Shortcuts are listed in the left-hand side **Shelf**.

Managing Content Properties

Every content item in Alfresco will have properties associated with it. Refer to the screenshot on the tenth page to see the list of properties such as **Title**, **Description**, **Author**, **Size**, and **Creation Date**. These properties are associated with the actual content file `Alfresco_CIGNEX.doc`.

The content properties are stored in the relational database and are searchable using advanced search options.

What is Content Metadata?

Content properties are also known as *Content Metadata*. Metadata is structured data, which describes the characteristics of content. It shares many similar characteristics with the cataloguing that takes place in libraries. The term Meta derives from the Greek word denoting a nature of a higher order or more fundamental kind. A metadata record consists of a number of pre-defined elements representing specific attributes of content, and each element can have one or more values.

Metadata is a systematic method for describing resources, and thereby improving access to them. If access to the content will be required, then it should be described using metadata, so as to maximize the ability to locate it. Metadata provides the essential link between the information creator and the information user.

While the primary aim of metadata is to improve resource discovery, metadata sets are also being developed for other reasons, including:

- Administrative control
- Security
- Management information
- Content rating
- Rights management

Metadata Extractors

Typically, in most content management systems, once you upload the content file, you need to add the metadata (properties) such as title, description, and keywords to the content manually. Most of the content such as Microsoft Office documents, media files, and PDF documents contain properties within the file itself. Hence, it is a duplicated effort having to enter those values again in the content management system along with the document.

Alfresco provides built-in metadata extractors for popular document types to extract the standard metadata values from a document and populate the values automatically.

This is very useful if you are uploading the documents through FTP, CIFS, or WebDAV interface, where you do not have to enter the properties manually, as Alfresco will transfer the document properties automatically.

Editing Metadata

To edit metadata, you need to click the **edit metadata icon** in the content details view. Refer to the edit metadata icon shown in the screenshot that shows the detailed view of the file `Alfresco_CIGNEX.doc`. You can update the metadata values such as **Name** and **Description** for your content items. However, certain metadata values such as **Creator**, **Created Date**, **Modifier** and **Modified Date** are read-only and you cannot change them. Certain properties such as **Modifier** and **Modified Date** will be updated by Alfresco automatically, whenever the content is updated.

Adding Additional Properties

Additional properties can be added to the content in two ways. One way is to extend the data model and define more properties in a content type. More information is provided in the chapter titled *Extending the Alfresco Content Model*.

The other way is to dynamically attach the properties and behavior through **Aspects**. Using aspects, you can add additional properties such as *Effectivity*, *Dublin Core Metadata*, and *Thumbnailable* to the content. More information is provided in the chapter titled *Implementing Business Rules*.

Library Services

The library services are the common document management functions for controlling users with permissions to create multiple instances of a document (versioning) and users with access a document to make changes (checking in/out).

Versioning

So far you have learned about creating spaces, adding files, and editing them. You might have more than one person who can edit a document. What if somebody edits a document, and removes a useful piece of information? Well, you can use versioning features to avoid such issues.

Versioning allows the history of previous versions of a content item to be kept. A content item needs to be *versionable* for versions to be kept. You can enable versioning in four different ways.

- Individually: To enable versioning for an individual content item, go to the **View Details** page and click on the **Allow Versioning** link. The screenshot on the next page illustrates the way to enable versioning on an individual content item.

- Using Smart Spaces: A business rule can be set for a space to allow versioning of all the content or selective content within that space. More information about this is provided in the chapter titled *Implementing Business Rules*.

- By Type: By default, versioning is disabled for all content types in the Alfresco content model. Versioning can be enabled for a specific content type in the Alfresco content model, irrespective of the location of the content. More information about this **is provided in the chapter titled** *Extending the Alfresco Content Model*.

- Globally: Alfresco can be configured globally to enable versioning for all the content throughout the site. More information about this is provided in the chapter titled *Extending the Alfresco Content Model*.

Enable versioning for the sample file you have already uploaded to the system. Go to the **Intranet | Marketing Communications | Switch to open source ECM | 02_Drafts** space and view details of `Alfresco_CIGNEX.doc`. Click on the **Allow Versioning** link to enable versioning as shown in the screenshot below, and you will immediately notice that a version with 1.0 is created.
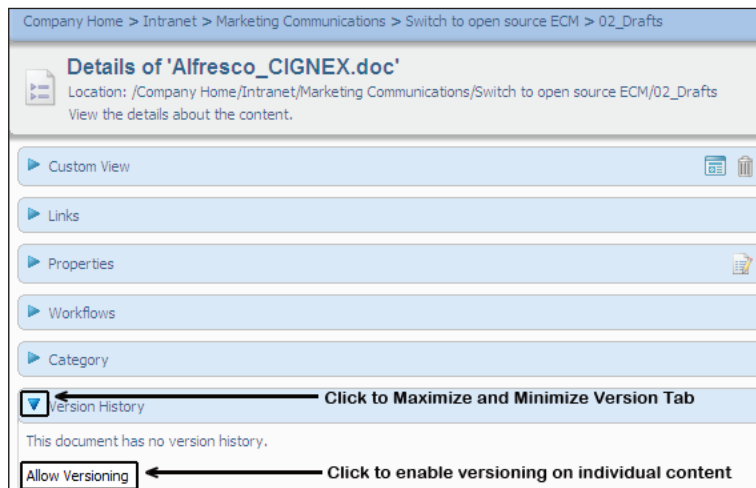
At the time of writing this book (Alfresco version 1.4), reverting back to an older version of the content is not supported. There is plan to support this feature in the future releases of Alfresco. The work-around is to download the older version and upload it again as the current version.

For a checked-out content item, the version is updated when the content is checked in. The version number is incremented from the content version number that was checked out.

Auto Versioning

Auto versioning can be enabled by editing the content properties and selecting the **Auto Version** checkbox.

If auto versioning is enabled, each **Save** of the content results in an incremented version number when it is edited directly from the repository. Each **Update** (upload) of the content also results in an incremented version number.

If auto versioning is not enabled, the version number is incremented only when content is checked in.

Check-in and Check-out

Using the versioning feature, you can ensure that all the changes made to a document are saved. You might have more than one person who can edit a document. What if two people edit a document at once, and you get into a mess with two new versions. To resolve this issue, you'll need the library services.

The library services provide the ability to check out a document, reserving it for one user to edit while others can only access the document in a read-only mode. Once the necessary changes are made to the document, the user checks in the document and can either replace the original or create a version of the original.

Check-out locks the item and creates a working copy, which can be edited (content and details). Check-in replaces the original item with the working copy and releases the lock.

Checking out Documents

Ensure that you are in the **Intranet | Marketing Communications | Switch to open source ECM | 02_Drafts** space. Click on the check-out action of Alfresco_CIGNEX. doc as shown in the screenshot below:
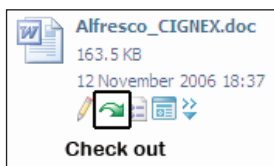
The **Check Out** pane is displayed as shown in the screenshot below. You can either check out the file in the current space or to any other pre-defined space. Typically, in a production environment the file resides in the **Published** space and you can check it out to a drafts space to make modifications to the file.

For the current example, choose the file check-out in the current space option and click on the **Check Out** button. You will return to the **02_Drafts** space.
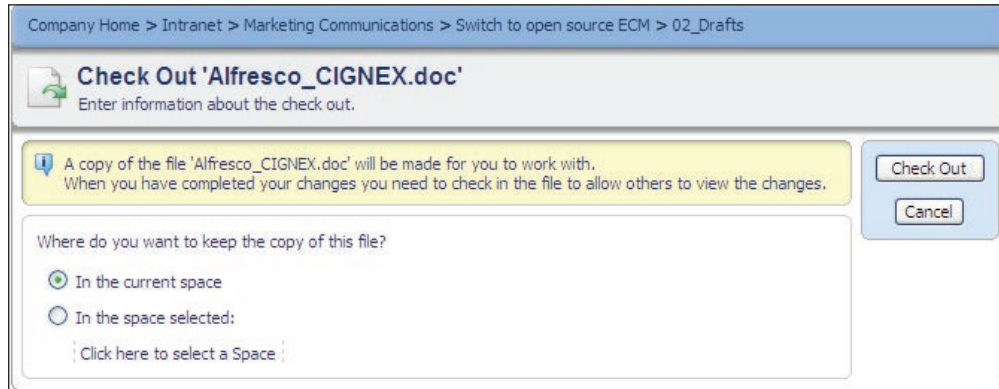
You will see two copies of the same document as shown the following screenshot. The original version of the file has a lock. This indicates that no one else can check out this file until you have checked it in again. The original version of the file cannot be edited (no pencil icon) or checked out (no check-out icon). The desired effect of all these features is that you cannot edit a checked-out file deliberately or accidentally.

You can only update the working copy. The checked-out file has **Working Copy** inserted in the file name. The working copy can be edited and checked in.

Checking in the Working Copy

Update the working copy **Alfresco_CIGNEX (Working Copy).doc** file. Updating the document is explained earlier in this chapter in the *Upload the Updated Binary File* section. After you update the working copy, you can check it in by clicking the

check-in button as shown in the above screenshot. Once you click on the check-in action, you will see the **Check In** dialog window as shown in the following screenshot:

If you have only made minor changes to the file, you will check the **Minor Change** checkbox. By selecting the *Minor Change* checkbox, you will be able to increment only the number after the decimal (from 1.0 to 1.1); else you will increment the number before the decimal (from 1.0 to 2.0).

For this example, select the **Minor Change** option and provide some meaningful version notes. Version notes are very important documentation to help understand the differences between various versions of the same document.
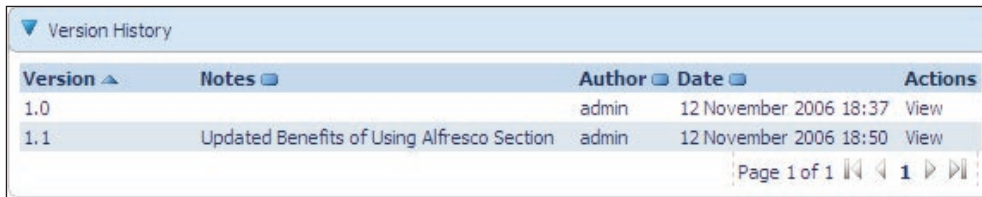
The option **Check in changes and keep file checked out** is an useful option if you intend to keep the file locked for several days. You can continually mirror your changes from the working copy to the locked copy. This means that if another user wants to look at the file, they will see a more up-to-date copy than if they had to wait until you checked in the file.

You now have two options to check in the file:

- **Use copy in current space**: You would use this option if you had previously used **Update** on the working copy.
- U**se copy uploaded from my computer**: If you use this option, you do not need to have previously used the **Update** option.

There will be situations where either of these options is preferable. For this example, select the **Use copy in current space** option and click on the **Check In** button. You are returned to the *02_Drafts* space.

There will no longer be a working copy of the document. Notice the latest modification timestamp of the original document. If you click on the **View Details** action and scroll down to **Version History**, you will see that the history has been updated as shown in the following screenshot:


Undo Check-Out to Unlock a Document

Now that you can use library services, you might still have questions such as; How long does a file remain checked out? Can we see who checked it out and when? And who can cancel the lock?

A document remains in the checked-out state (locked) forever till the working copy is checked in or till somebody cancels the checked-out status from the working copy. To cancel check-out, locate the working copy of the document, click on the **More Actions** icon and select the **Undo Check Out** option as shown in the following screenshot. **Undo Check Out** will delete the working copy and release the lock, as if the check-out had not happened.
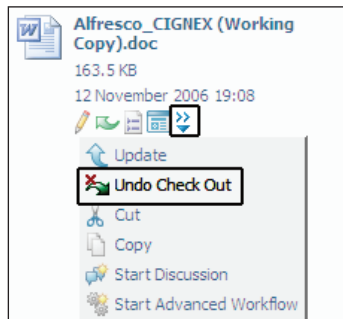
You can enable auditing on the Alfresco repository, and find out audit trail information such as who locked the content and when. More information about the auditing is covered in Chapter 7.

The owner of the document, or a coordinator, or an administrator can unlock the document by executing the **Undo Check Out** action on the working copy. Other users, who have read access to the space, can still see the working copy of the document but they can never edit or check in the document. The possible actions on a working copy by other users are shown in the following screenshot:
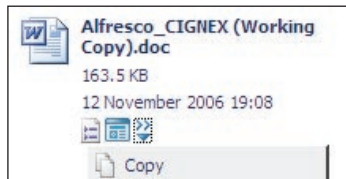
Categorizing Content

Categorization helps to classify information in a number of ways. Various technologies use various terminologies such as hierarchies, taxonomies, and ontology for the same concept.

In Alfresco, all content can be linked to one or more categories. Categories are defined and managed by administrators only. Categories can have sub-categories and there is no limitation on the number of categories or the depth of the hierarchy.

Categorization aids in searching and the advanced search form in Alfresco allows you to search the content filtered by various categories.

Managing Categories

Follow the steps below to create two new categories called Technology and Products for your example application.

1. In any space, click on the **Administration Console** icon in the top tool bar. The **Administration Console** pane appears.

2. Click on the **Category Management** link. The **Category Management** pane appears as shown in the following screenshot. Notice the existing categories such as **Software Document Classification**, **Regions**, and **Languages**.

3. In the header, click the **Create | Add Category** link to create a new category.

4. As an example, create a new category called *Open Source Products*. Under that new category, create few sub-categories such as *Alfresco* and *Plone*.

You can add additional categories and sub-categories and you can edit the existing categories.
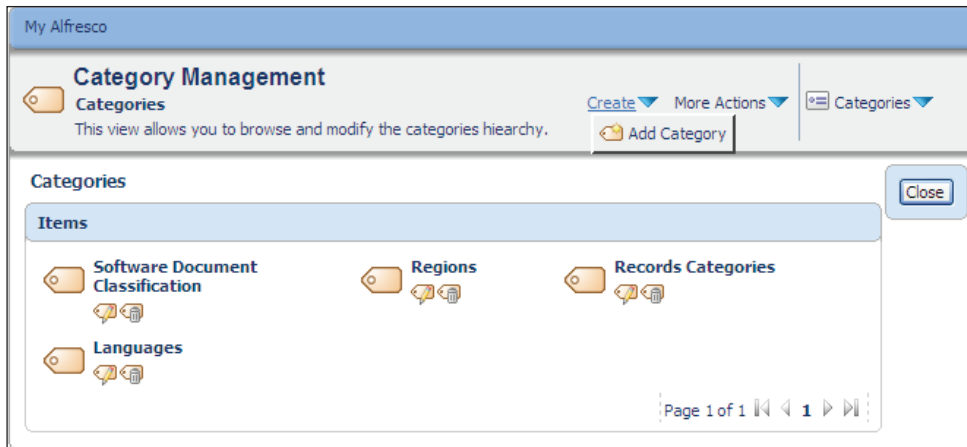
Adding Categories to Content

In order to categorize content, you need to have an Administrator, Editor, Collaborator, Coordinator role on that content.

To enable categorization for an individual content item (say `Alfresco_CIGNEX.doc`), go to the **View Details** page, and click on the **Allow Categorization** link. Click on the **Change Category** icon as shown in the following screenshot overleaf to apply categorization to the content.
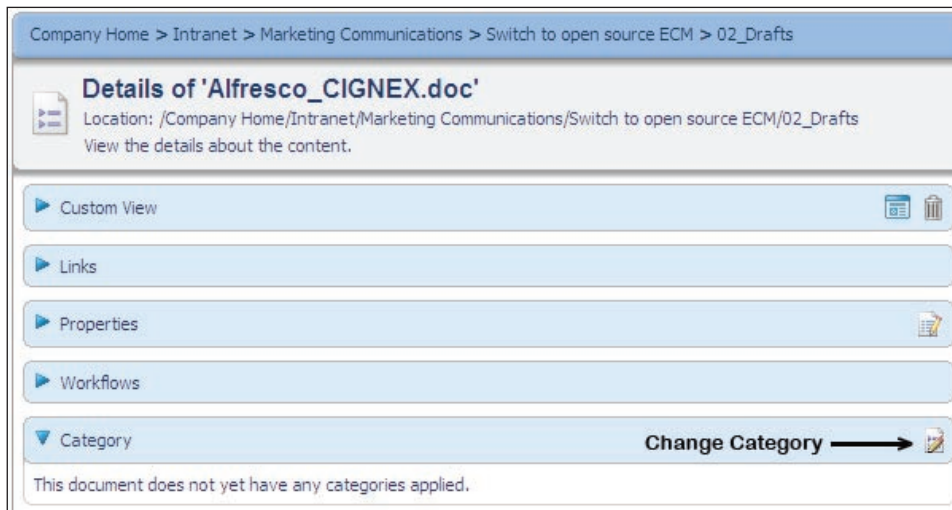
You will see the **Modify categories** dialog as shown in the screenshot below page. Apply two different categories to the content. Click on the **OK** button to confirm.

You will notice these two categories associated with your document in the **View Details** page.
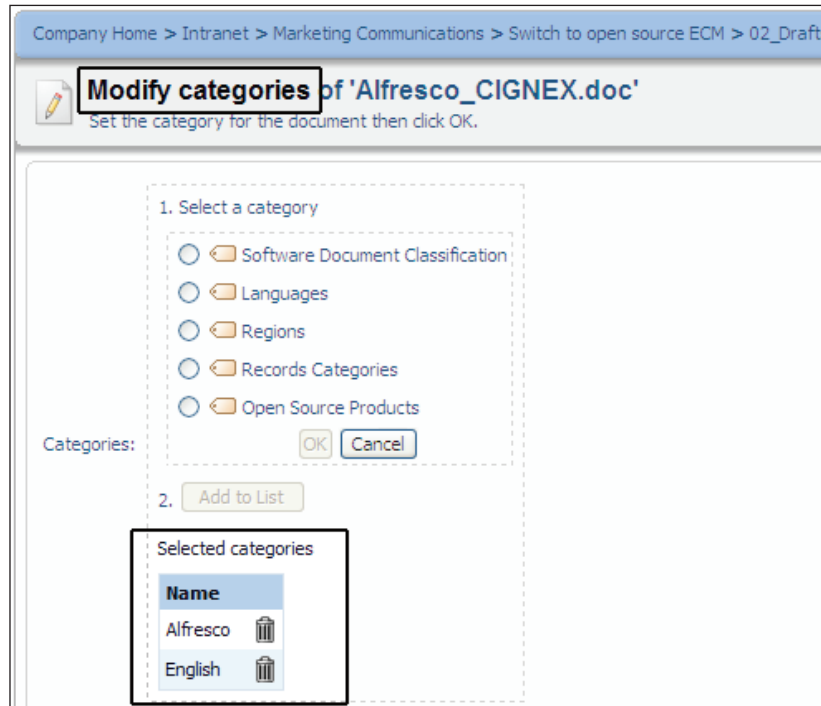
Search Content by Category

Categorization helps to narrow down a search or filter the search results. From the earlier example, we will search for the marketing documents, with a text called *Content* and categorized under **Open Source Products | Alfresco**.

Click on the **Advanced Search** link from the drop-down list of the search menu. From the form options, click on the **Show me results in the categories** pane. Click on the **Click here to select a category** link. Select **Open Source Products | Alfresco** as an option and provide text in the **Look For** text box and then click on the **Search** button.

Recovering Deleted Content

When you delete an item (content or space) in Alfresco, the item will not be deleted from the server; it will be moved to a temporary store called the **Archive Space Store**. This gives you a chance to recover items that were deleted earlier.

A deleted item will be in the temporary store forever, or till you decide to either recover or purge the deleted item. This feature is available to administrators through the **Manage Deleted Item** action.

To test these features, log in as an administrator, create a couple of dummy files in Alfresco and delete them. In any space, click **More Actions | Manage Deleted Items**. The **Manage Deleted Items** pane appears as shown in the following screenshot:
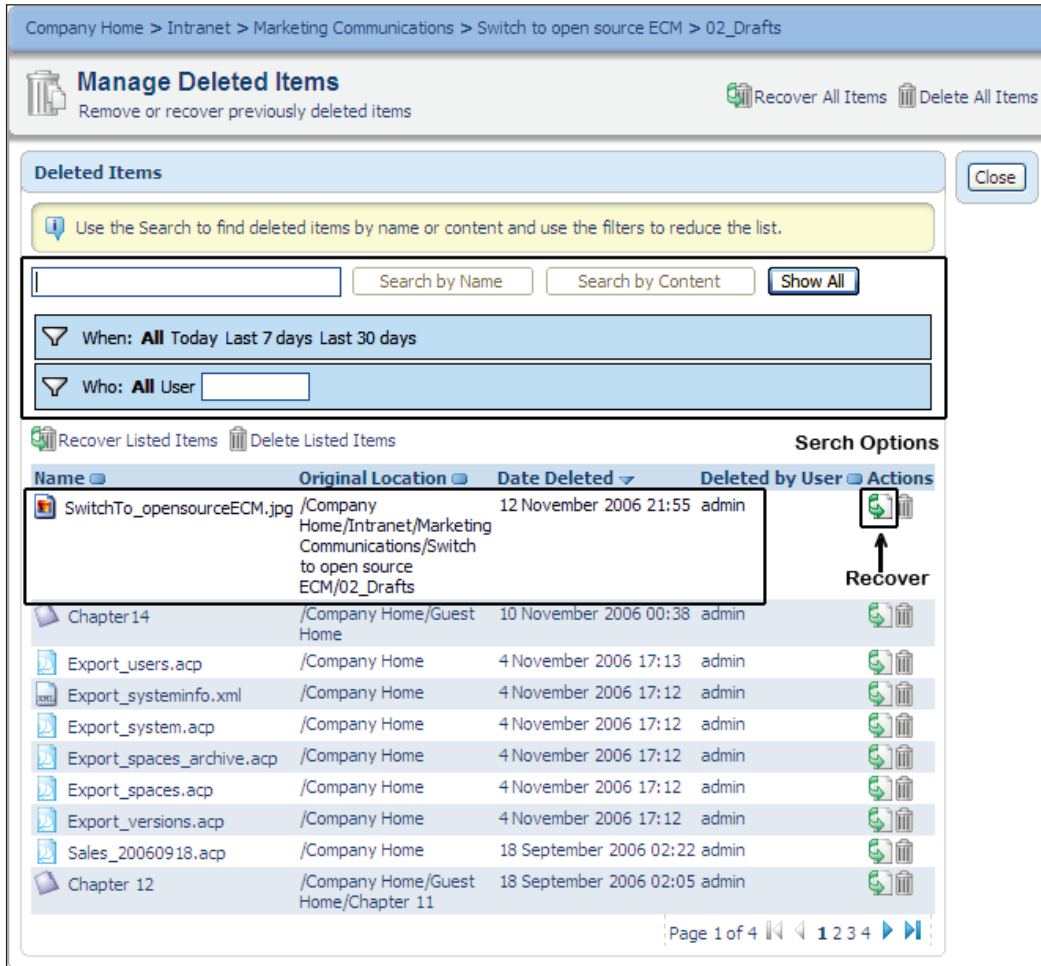
You can list all the deleted content by clicking on the **Show All** button as highlighted in the above screenshot. You can also search the deleted items by name, content, date, or the person who deleted them using the search options provided. Select the item that you deleted earlier and click on the **Recover Listed Items** icon as shown in the screenshot above. You will notice that the item has been recovered to the original space.

When an item is recovered, it will be removed from the archive space store and moved to the original space from where it was deleted.

Purged items are deleted forever and can not be recovered. Since the deleted items will be in the temporary store forever, it is best practice to purge them periodically. It is also recommended to take regular backups of your data. More information about maintenance and backup is provided in the *Maintaining System* chapter of this book.

Use Network Drives to Manage Content

The out-of-the-box installation comes with a web client, where you can connect to the Alfresco repository through a web-based application. Apart from the web client, the Alfresco out-of-the-box installation supports various client applications to access Alfresco content using protocols such as FTP, WebDAV, and CIFS.

CIFS

CIFS, an enhanced version of **Microsoft Server Message Block** (**SMB**), is the standard way for the computer users to share files across intranets and the Internet. CIFS enables collaboration on the Internet by defining a remote file-access protocol that is compatible with the way applications already share data on local disks and network file servers.

CIFS supports the usual set of file operations like open, close, read, write, and seek. CIFS also supports file and record lock and unlocking. CIFS allows multiple clients to access and update the same file while preventing conflicts by providing file sharing and file locking. CIFS servers support both anonymous transfers and secure, authenticated access to named files.

CIFS helps you to map an Alfresco space as a folder in your local file system thus giving you flexibility in working with files in the repository as if they are in your local file system. You will be able to bulk upload files to the server and edit them directly using your desktop applications.

Mapping the Drive

As an example, you will now map one of your spaces, say **Intranet | Marketing Communications**, as a local folder.
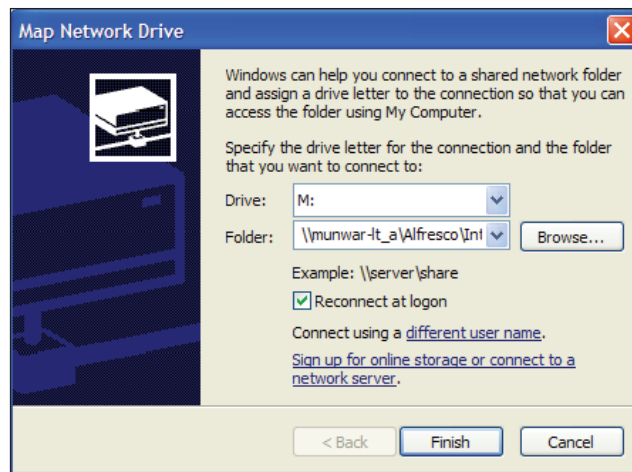
To map a space in Windows Explorer as a network drive, follow the steps given below:

1. In Windows Explorer, click the **Tools | Map Network Drive** link. The **Map Network Drive** dialog appears as shown in the above screenshot.

2. Select an unused drive letter (say **M** for Marketing Communications space).

3. In the **Folder** text box, type **\\AlfrescoServerName_a\Alfresco\Intranet\Marketing Communications**. The syntax is:
   \\YourMachineName_a\alfresco\YourSpaceName

4. Check the **Reconnect at logon** checkbox.

5. Click **Finish**. As the space is secured the system will prompt for your authentication.

6. Type your Alfresco user name and password when prompted.

Another easy way of mapping your space as a network folder in your local file system were using the web client. Go to the space and click on the **Open Network Folder** icon as shown in the following screenshot:
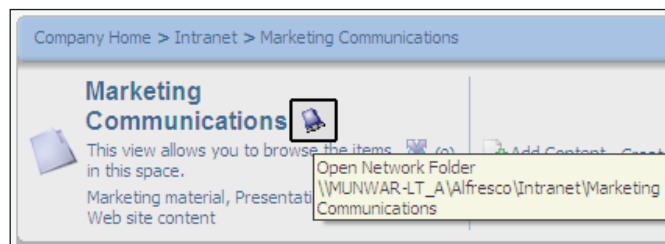
Once a space is mapped as a network folder, you can browse the space as if it were your local folder. The above screenshot illustrates that you could browse your Alfresco content on your local file system and the server name is *YourMachineName_A*.
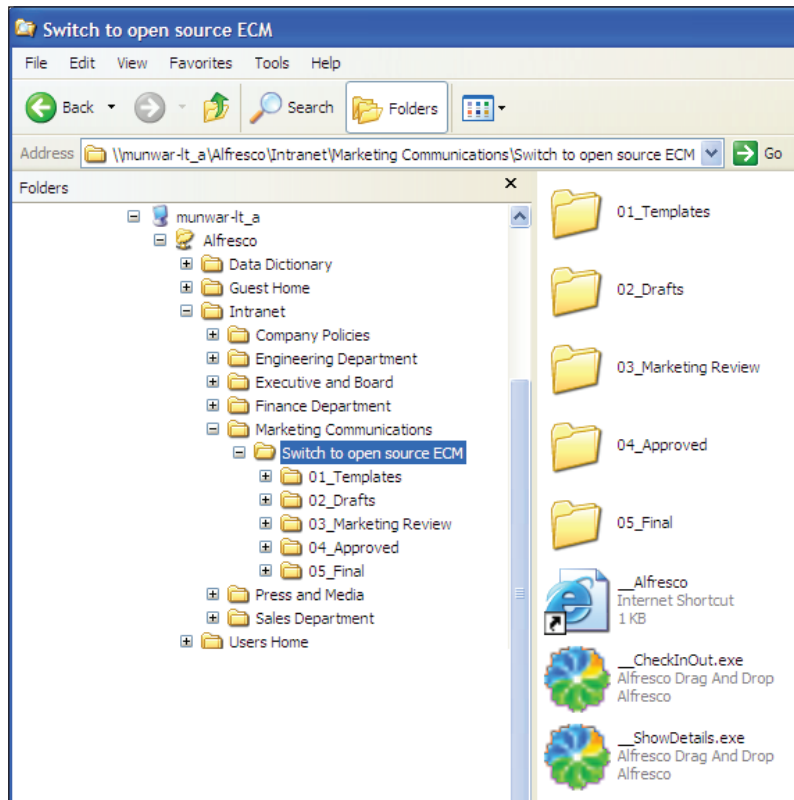
Drag-and-Drop Documents in Alfresco through CIFS

Once a space is mapped as a network drive, you can drag-and-drop files from your local hard disk to the Alfresco server. Similarly you can copy files from the Alfresco server to your local hard disk.

As an example, drag-and-drop an image file from your hard disk into your **Intranet | Marketing Communications | Switch to open source ECM | 02_Drafts** space.

People belonging to a department can map their department-specific space and transfer files from their local file system.

As an administrator, you can map the root space (Intranet) to your local drive and bulk transfer documents between the server and the local machine.

Check out and Check in Documents in CIFS

For the check-out and check-in feature to work with CIFS, you need to make some configuration settings. These settings are already explained in the *Configure CIFS Desktop Actions* section of Chapter 3. For more information, refer the custom configuration file `file-servers-custom.xml` in `<extension>` folder and view the configuration settings for the following XML block:

```
<desktopActions>
… xml configuration settings for desktop actions …
</desktopActions>
```

Note that any configuration changes to XML files require restarting of server, otherwise the changes will not be effective.

Follow the following steps to try check-out and check-in of a document in CIFS:

1. In Windows Explorer locate the **02_Drafts** folder, drag-and-drop the `Alfresco_CIGNEX.doc` file (or any other file) onto `__CheckInOut.exe` to check out the file.

2. Observe that a working copy of the document is made and the original document is locked as shown in the following screenshot.

3. Make changes to the working copy by directly editing it using your favorite editor on your hard disk.
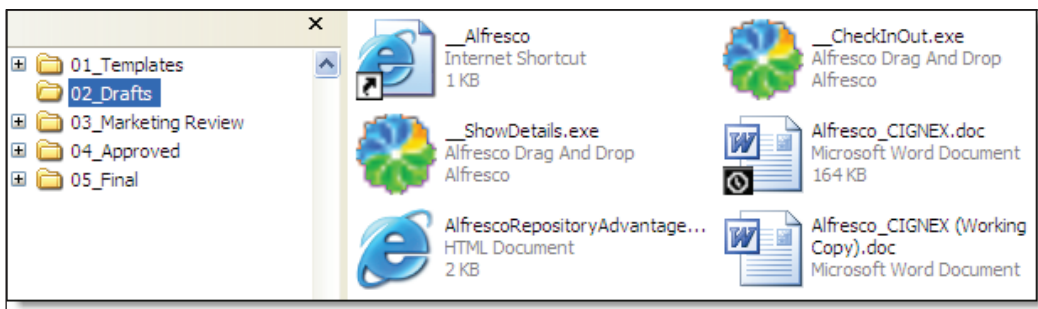
4. Click the **__Alfresco** icon (shown in the above screenshot) given in CIFS to open the web client interface directly from your CIFS folder. You can verify that the file has been checked out in web client interface as well. You will notice that the changes made to the Alfresco repository through the CIFS interface are visible in the web client user interface as well.

5. Again drag-and-drop the working copy onto `__CheckInOut.exe` to check in the file. You will see that the working copy file is checked in and deleted. The original file is updated and unlocked.

If auto versioning is enabled on the document, you will also notice that the changes made to the document are versioned as shown in the following screenshot. This is a great advantage to the content authors as they can edit the documents in their local file system using their choice of editors, and maintain various versions in the Alfresco repository automatically.
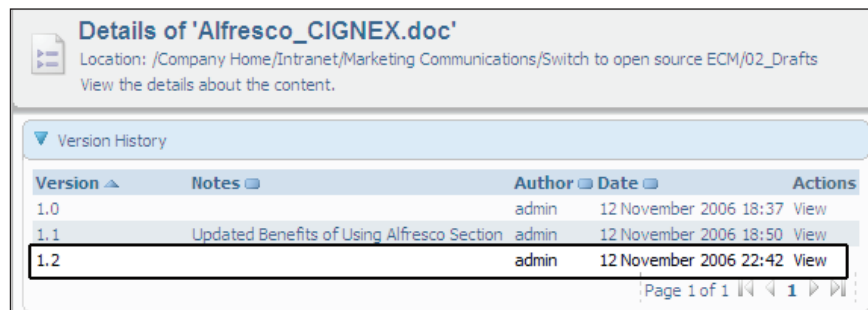
The sample works fine with Windows Internet Explorer and other browsers that already contain the CIFS plug-in. If you are using an older version of Firefox or any web browser that does not contain a CIFS plug-in, you need to install the plug-in to take advantage of this feature.

For Firefox (or Mozilla) browser, a plug-in is available on the sourceforge.net site. To install the Firefox extension, follow these steps:

1. Browse to the URL: `http://sourceforge.net/projects/alfresco`.
2. Click on the green box, **Download Alfresco Content Management**.
3. Scroll down to **Firefox extension**, and click on the **Download** link.
4. Click on the link **alfrescoext-0.9.xpi**.
5. In the list of download sites, click **Download** for the site nearest to you.
6. Allow permission to install the extension.

FTP

FTP is useful to transfer files from your local file system to the remote server. Using any FTP client, you can connect to Alfresco server as if it were a FTP site and upload and retrieve files.

If you are connecting to a space all of the rules are applied and all of the permissions are enforced. If versioning is enabled, then content will be versioned regardless of how it is updated.

Follow the steps below to use FTP on the Windows DOS prompt to upload a file from the local hard disk to your **Intranet | Marketing Communications | Switch to open source ECM | 02_Drafts** space.

```
> ftp localhost
Login as: admin/admin
> ls
> cd Alfresco
> ls
> cd Intranet
> ls
> cd Marketing*
> cd Switch*
> cd 02_Drafts
> put c:\press2A.txt
> ls
```

Verify that the text file is now in the Alfresco repository. Similarly, you can use any FTP client application to connect to the Alfresco repository and access files.

WebDAV

WebDAV is primarily designed for editing and managing files on remote web servers, in a structured way. For example, an application like Adobe Photoshop can directly open a file in the Alfresco content repository and edit it. This gives you the flexibility of using your own favorite editor to edit the content on the Alfresco server.

If you are connecting to a space all of the rules are applied and all of the permissions are enforced. If versioning is enabled, then content will be versioned regardless of how it is updated.

If you have a WebDAV client, then you can access the Alfresco server using the URL:

`http://localhost:8080/alfresco/webdav/`

To open a space in WebDAV, log in to the web client, go to the space, and click on the **Detailed View** link of the space and then click on the **View in WebDAV** link under the links section as shown in the next screenshot:
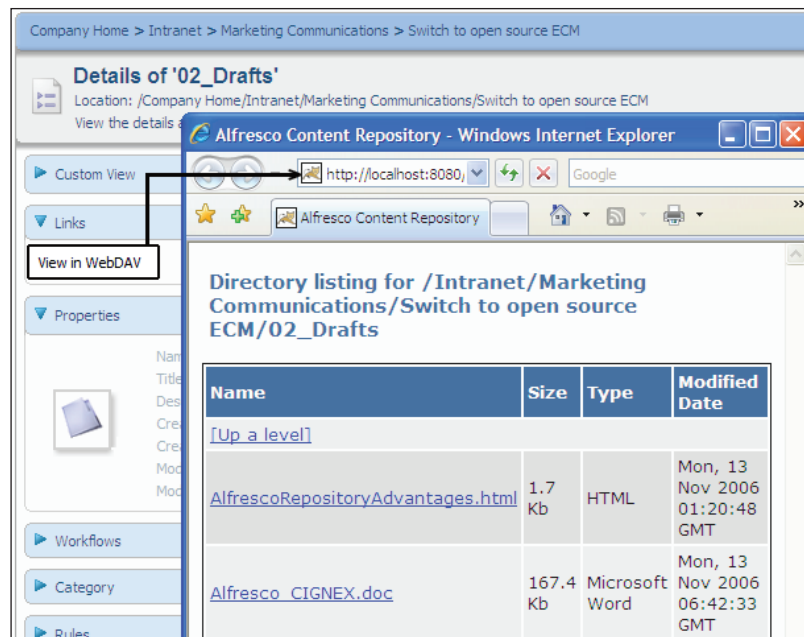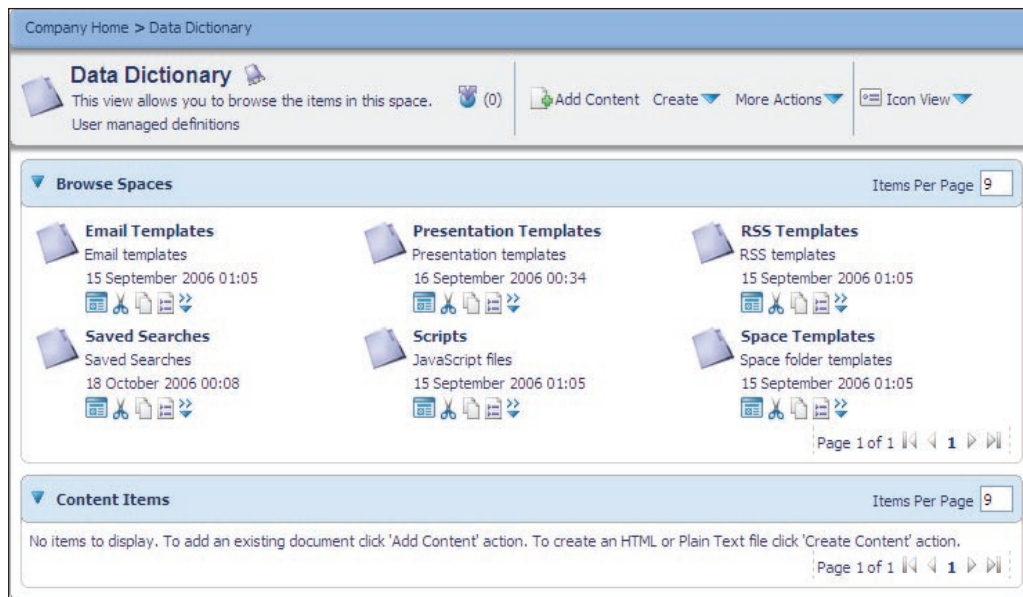
Data Dictionary and Space Templates

The **Company Home** space is the root space that contains the sub-spaces such as **Data Dictionary**, **Guest Home**, and **Users Home** spaces.

Data Dictionary Space

The **Data Dictionary** space contains all the user-managed definitions as shown in the screenshot on the next page:

The **Data Dictionary** space contains the following sub-spaces. The scripts and the templates contained in these spaces are covered in detail in the further chapters of this book. For the time being, note that **Data Dictionary** is a shared resource. All the scripts and templates provided in the data dictionary can be defined by the administrators and used by the users.

- Email Templates: Contains email templates for notifying users of an invitation to a space or document and for sending notifications to users from a rule or an action. Email templates are written in the *FreeMarker* template language and have the .ftl extension.

- Presentation Templates: Contains presentation templates, which are useful to consolidate and view content in different ways. Presentation templates are written in the FreeMarker template language and have the .ftl extension.

- RSS Templates: Contains RSS templates, which are useful to provide RSS feeds on spaces. More information about RSS templates is provided in Chapter 10.

- Saved Searches: Contains pre-built queries, which are saved by a user from the search results page. Each user will have their private saved searches. This space will also contain all the saved searches that are publicly shared by the users.

- Scripts: Contains JavaScript files, which are used to perform certain operations on content.

- Space Templates: Space templates contain space structures that can be used as templates to create new spaces. Any space can be saved as a space template for future reuse of the space structure and data. More about space templates is covered in the following section.

Space Templates for Reusable Space Structure

In the earlier sections, you have created a marketing project in the **Company Home | Intranet | Marketing Communications | Switch to open source ECM** space. Let us assume that you are going to launch many such marketing campaign projects in your marketing department. Each marketing project space will have a similar structure to hold project information. For example, your marketing project space has marketing templates, draft documentation, and approved marketing collateral. Also, each project will have a hierarchy of spaces, security settings, business rules, notifications, and workflows.

Instead of having to repeatedly create the same structure for each project, you can maintain a *Marketing Project template* and keep on replicating it for every new project. Thus, all the work you do manually could be done in a few seconds simply by using such a template.

To give you an idea, **Company Home | Data Dictionary | Space Templates** contains a built-in space template called **Software Engineering Project** as shown in the next screenshot:
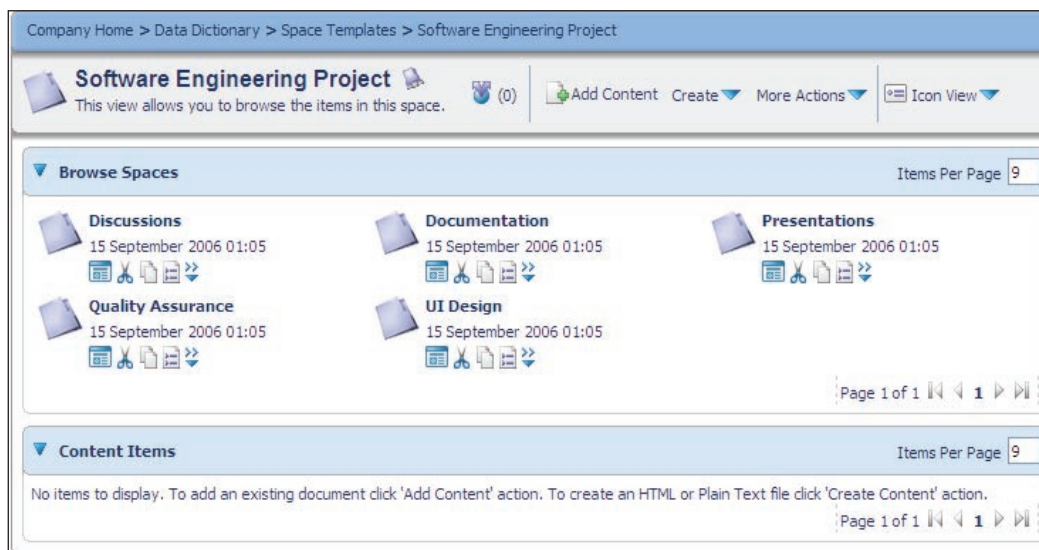
This represents a typical engineering project space structure with sub-spaces for documentation, discussions, quality assurance, user-interface design, and project presentations. You can use this *Sowftware Engineering Project template* and keep on replicating it for every new engineering project. If the template does not match your needs exactly, you can add or delete some spaces in the template itself.

Note that the **Software Engineering Project** does not contain rules. This is because rules will vary from one company to another. For example, workflow processes will vary. For your own company projects, adding rules and standard content will increase the value of the space template.

Create a New Space Template for Reuse

Follow the steps given to create your own space template for your **Marketing Projects** using the existing space structure.

1. Log in as the admin and go to the **Company Home | Data Dictionary | Space Templates** space.

2. Click on the **Create | Advanced Space Wizard** link. The **Create Space Wizard** pane appears.

3. Click on the **Based on an existing space** radio button option and then on the **Next** button to go to the second pane titled **Space Options**.

4. Browse and select the **Company Home | Intranet | Marketing Communications | Switch to open source ECM** space as shown in the screenshot below. Click on the **Next** button to go to the **Space Details** pane.

5. In the **Space Details** pane, give an appropriate title (such as *Marketing Project*) to your space template and click on the **Finish** button to confirm.
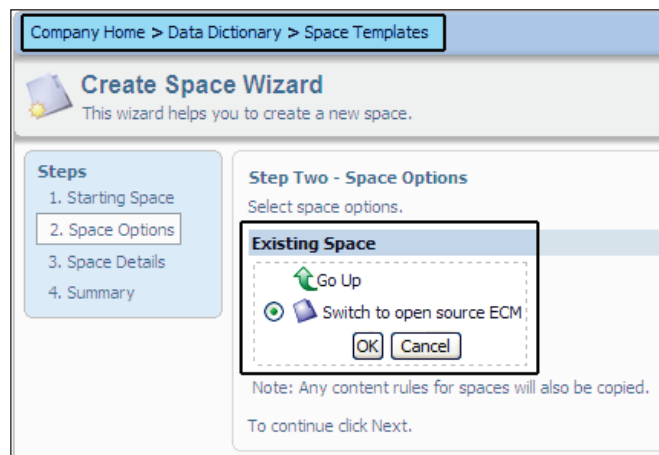
- You will notice a new space template called **Marketing Project** in the **Company Home | Data Dictionary | Space Templates** space. Examine the space structure.

Use an Existing Space Template to Create a New Space

You can reuse the *Marketing Project* space template to create new marketing projects. To create a new *Marketing Project*, ensure that you are in the **Company Home | Intranet | Marketing Communications** space and follow the steps given below.

1. In the space header, click **Create | Advanced Space Wizard**.

2. The first pane of the **Create Space Wizard** pane appears. Click on the **Using a template** radio button option and click on the **Next** button.

3. The **Space Options** pane of the wizard appears as shown in the following screenshot:
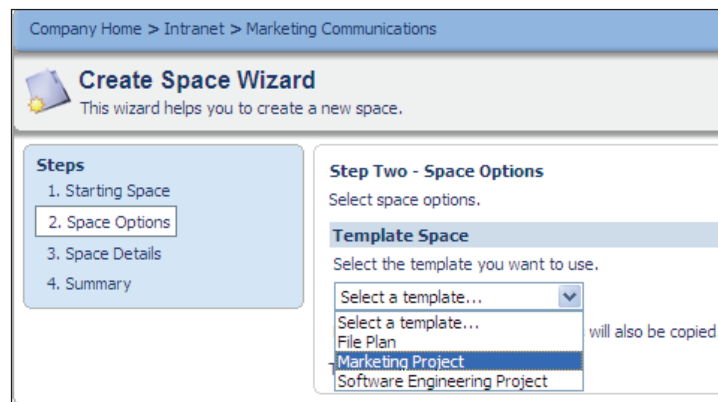
4. Choose the **Marketing Project** template and click the **Next** button.

5. The **Space Details** pane of the wizard appears as shown in the next screenshot.

6. Specify your new project name and click on the **Finish** button to confirm.
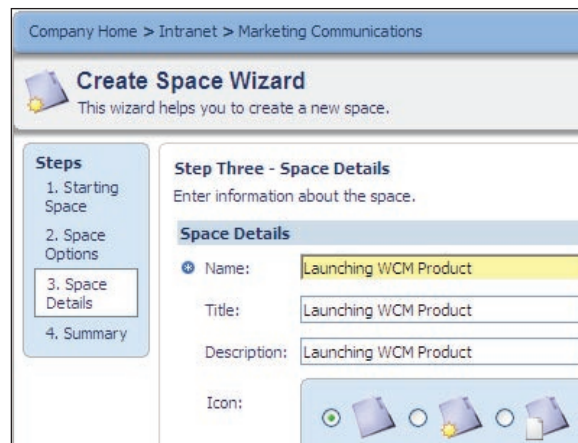
7. You can now browse around your new space and compare it with the *Marketing Project* space template. You will notice that the space contents are identical.

Migrating Existing Content into Alfresco

If you want your Enterprise Content Management initiative to be successful, you need to make sure that you can move the existing content into the new system. Most enterprises will have content in the form of files (in local or shared hard disks), email attachments, faxes (invoices) and scanned images. It is very important to move the content to a centralized and highly scalable content repository such as Alfresco.

Alfresco, being a powerful content management system for the enterprise, supports various ways to migrate existing content in the enterprise.

Drag-and-Drop Content to a Network Drive

You can drag-and-drop (bulk upload) content from your local hard disks to the Alfresco server using options such as CIFS, FTP, or WebDAV. Refer the *Use Network Drives to Manage Content* section in this chapter to know how to move content from your hard disk to the Alfresco server.

The only issue with this approach is that you will have to manually update the metadata (properties) of the content.

Using the Web Services API to Migrate Content

Alfresco provides a very rich web services API, using which you can transfer your files as well as metadata to the Alfresco server.

ACP Generator Bulk Upload Utility

The ACP Generator project provides a tool for bulk uploading of content into any Alfresco repository. It reads custom content models and a comma-separated variable (CSV) list of required property and category values for each content item and creates an ACP file.

The syntax is as follows:

```
acpGeneratr arg1 arg2 arg3 arg4 arg5
Where:
Arg1: The path and filename of the custom model to be used.
Arg2: The path to the directory where the content to be imported is
                                                    located
Arg3: The destination path
Arg4: The content type we're dealing with
Arg5: The csv file of property values
```

Example:

```
acpGeneratr "D:\PIER\ACPGeneratr\sampleModel.xml"

            \ACPGeneratr\content  ..\ACPGeneratr\testDestDir dm:gqa

                                  "D:\PIER\ACPGeneratr\faqlist.csv"
```

This will create the defined destination directory, and within it a directory with all the content and the required `.xml` file. All you need to do is zip it up into a `.acp` file and import to Alfresco. More details about this tool are provided in Alfresco Forge at `http://forge.alfresco.com/projects/acpgeneratr/`.

Summary

You can customize Alfresco features such as smart spaces, library services, and security to implement your enterprise document management requirements. Various interfaces such as web client, CIFS, FTP, and WebDAV can be used to manage content in the Alfresco repository. The content can be edited on your local desktop using your choice of favorite content editors. You can also use tools to migrate your existing content and to do bulk upload and retrieval of content.

Where to buy this book

You can buy Alfresco Enterprise Content Management Implementation from the Packt Publishing website: `http://www.packtpub.com/alfresco/book`

Free shipping to the US, UK, Europe, Australia, New Zealand and India.

Alternatively, you can buy the book from Amazon, BN.com, Computer Manuals and most internet book retailers.

[PACKT]
PUBLISHING

www.PacktPub.com