Information Management Resource Kit

Module on Management of Electronic Documents

UNIT 5. DATABASE MANAGEMENT SYSTEMS

LESSON 3. USING A DATABASE FOR DOCUMENT MANAGEMENT

NOTE

Please note that this PDF version does not have the interactive features offered through the IMARK courseware such as exercises with feedback, pop-ups, animations etc.

We recommend that you take the lesson using the interactive courseware environment, and use the PDF version for printing the lesson and to use as a reference after you have completed the course.

At the end of this lesson, you will be able to:

• understand the **requirements** for **information management**, and

• comprehend the **role of database** in an information management system.

Introduction

How can we manage our electronic documents?

A meeting is in course in the headquarters of the Organization for Agricultural Policy to decide how to manage the electronic documents that will be published on the new website.

Requirements for document management are not the same as those for document delivery, but it is important to consider that there is a common need: as well as final users, also users of a document management system (e.g. authors, editor, publisher) will need to access the documents **quickly and easily**.

Requirements for document management

Managing your documents implies two main areas of activities:

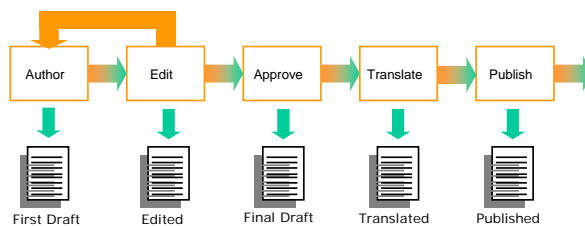1) managing **the process** of creating and revising documents (workflow management), and

2) managing the **document content** itself; this means administrating different instances of each document (resource administration) and managing its content meaning (semantics management).

Let's analyse these activities in more detail...

Requirements for document management

A **workflow** is a sequence of tasks that need to be performed to complete a business process. In order to manage the workflow, you can assign **a role** to each user (e.g. Author, Editor, etc.), so that you control which users can perform a specific task.

Author	Edit	Approve	Translate	Publish

First Draft	Edited	Final Draft	Translated	Published

Metadata associated with each document indicate its status in the process, and it changes as each task is completed: first draft, edited, final draft, translated, published.

In this example the sequence of tasks is linear, even if the document edited must be revised by the Author before going to the next step. Workflow processes can be more complex; for example, they can include parallel flows, multiple feedback or control loops.

Content management is the other area of document management .
This implies the following activities:

Resource administration
This consists of identifying each resource (i.e. document) and managing different **versions and variants** of the same basic document.
Versions are different instances of the document with small modifications (e.g. some changes in the text). Variants are different instances with fundamental changes (e.g. different renditions or languages).
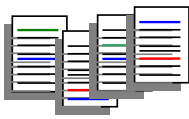
Content semantics management
This relates to the meaning of the content.
For example: grouping documents into **collections**, managing the **metadata** associated with each document, managing the **links** between documents and between elements within the document content.

Requirements for document management

The Organization for Agricultural Policy carried out a short analysis, generating some requirements.

Some of them are listed below: can you tell which area they fall into?

	Workflow management	Content management	
		Resources administration	Semantic management
Users with 6 different roles will access the system.	○	○	○
Documents will contain internal links.	○	○	○
The language version will be specified for each document.	○	○	○

Please click on the answer of your choice.

Requirements for document management

The question is: for our document management system, is it sufficient to create a file system or we need a database?

You have two main methods at your disposal for managing electronic documents:

• managing on the **file system**;

• managing using a **database**.

Let's look at these methods in more detail.

Using a file system for management

Using a file system may save us a lot of money…

If your resources (money, time, people) are limited, you could manage electronic documents on the **file system** without using a database.

The main advantage to managing documents on the file system is that it is **cheap** and **simple**.

It requires **no additional software**, other than:

• the **operating system** of your computer, and
• the **network** (if you are sharing documents with others).

Using a file system for management
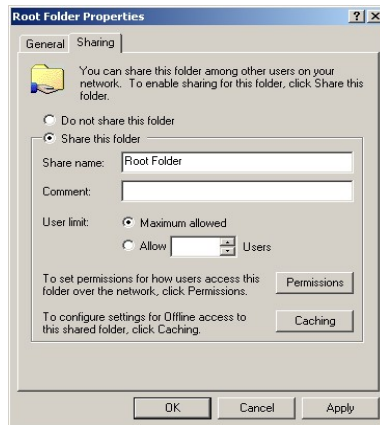
The first thing you need to do to manage documents on the file system is to **create a root folder** which will be the 'base' of your document management system.

You can set up user access **permissions** on that folder using the operating system administration tools.

Using a file system for management
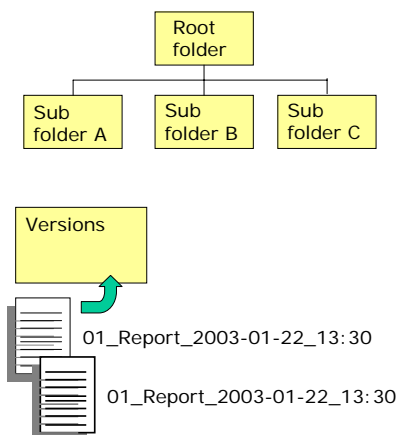
Inside the root folder you can create sub-folders to store the documents. You will need to keep to a **strict naming convention** for your documents and it's a good idea to make sure **each document has a unique name**.

The easiest way to do this is to assign to each new document a progressive number.

You can implement simple version control by creating a folder (e.g. called 'versions') inside the root folder.

Each time you start to edit a document:

• make a **copy** of it,
• append a **consistent format of date and time** to the name of the document, and
• move the copy **to the versions folder**.

Root
folder

Sub
folder A

Sub
folder B

Sub
folder C

Versions

01_Report_2003-01-22_13:30

01_Report_2003-01-22_13:30

Using a file system for management

If you are to be successful in managing documents on the file system you will need to be **organized and disciplined**.

That is to say, you need to make sure that you **follow the rules** you have set for how you will manage your documents.

If you are the **only person** managing documents (on the hard drive of your own computer, for example) **it is much easier** to keep to the rules than if you have many different users.

Using a file system for management
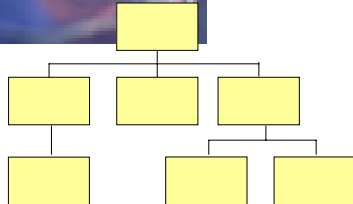
Remember: a file system is a viable option if...

• requirements for document management are **simple**;

• there is only a **limited number of users**; and

• you are confident that issues such as **security and backup** are adequately handled by administration of your file system and servers.

Using a file system for management

From the analysis of requirements, it emerges that complex workflow management is needed to support collaborative authoring and review.

With which of the following opinions do you agree?

• As far as the costs are concerned, choosing a file system to manage documents is better than using a database!

• There are a lot of users: a file system would give us some problems that could be avoided by using a database.

• Let's be realistic... a file system is not powerful enough to manage a website: we need a database!

Please click on the answer of your choice
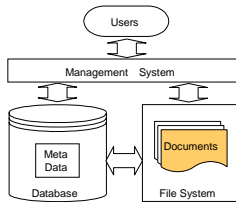

Using a file system for management

If it is important for you to support any of the following functions, then you will struggle to manage your documents successfully just using the file system:

• Associate **complex metadata** with documents.

• **Long term check out** of documents so that they can be exported and worked on (edited) only by designated users.

• **Reuse of document content** in different documents, supported by mechanisms for configuring 'compound' documents.

Using a database for management

If you decide to use a database, you have to know that there are two main ways in which it can manage metadata:
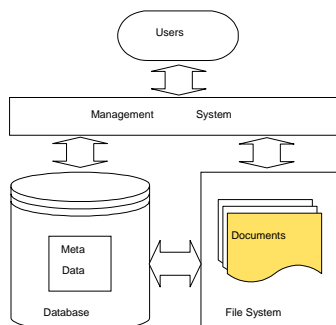

1) managing metadata in the database, with **document content on the file system**, or
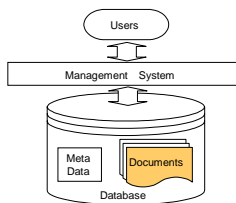
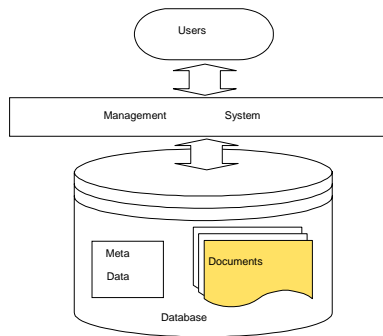2) managing document content and metadata **inside the same database**.

Using a database for management

It is often easier to create a system which manages metadata in the database and points to document content stored on the file system, so this can be a **good 'next step' if you were previously managing documents on the file system** and want to build your own document management system.

Likewise, **products** which manage documents in this way are **likely to be cheaper, easier to configure and simpler to administer** than systems which do everything in the database.

Users

Management   System

Meta
Data

Documents

Database

There are advantages to using systems that manage both document content and metadata **in the same database**:

• the database manages the **transactional integrity** of the content and metadata, meaning that the two types of information are keep in sync with each other at all times, and

• content is **more secure** when stored in the database, rather than the file system.

Using a database for management

Databases are used in **document management systems** and in **web content management systems**.

The choice between these two kinds of system is normally quite easy to make.

DOCUMENT MANAGEMENT SYSTEM
If you need to do **general document management**, with **varied workflows** and **complex version and access controls**, then you will find a document management system fits the bill.

WEB CONTENT MANAGEMENT SYSTEM
If you are managing content specifically for **publication through a website**, then you will find that web content management systems have built-in features that make your life quite easy.

Document Management System

The main features of **Document Management Systems** are:

• import and export of the resources
• version control
• access control
• document scanning and imaging
• metadata management

Leading commercial document management systems include products from **Documentum**, **OpenText** and **FileNet**.

If version control and access control are the two most important features for you, then you may be able to use the **open source CVS**.

CVS is the **Concurrent Versions System** (http://www.cvshome.org) a widely used open-source network-transparent version control system.

CVS is useful for everyone from individual developers to large, distributed teams.

Although it is really aimed at managing the source code in software development projects, you can use it for managing any type of electronic documents.

Its client-server access method lets users access documents from anywhere there's an Internet connection. It includes a sophisticated check-out model and version control mechanism and its client tools are available as open source implementations on most platforms.

Web Content Management System

Web Content Management Systems generally incorporate some of the features of Document Management Systems, but are geared specifically to manage content (in **multiple formats**) that is to be published on a website.


The **main features** of this system are:

• management of web content
• workflow to support collaborative authoring, review and sign-off
• site publication workflow
• automated site update
• whole website versioning and release
• Quality Assurance of web content (e.g. checking the integrity of URLs in hyperlinks)

Commercial products are available from vendors such as **Microsoft**, **MediaSurface**, **Interwoven** and **Tridion**.

WebDav
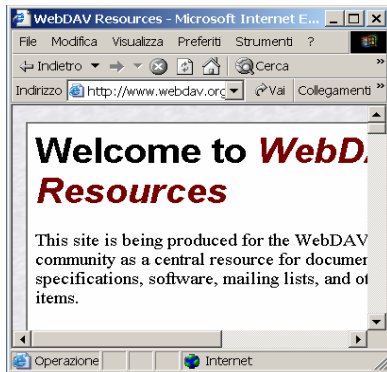
There is also a new standard for the management of resources: the Web-based Distributed Authoring and Versioning (**WebDAV**).

Published by the Internet Engineering Task Force (IETF) in 1999, it allows for the management of resources on **remote servers** so that the web can be used as a distributed document management system.

WebDAV provides **facilities for**:

• Collaboration
• Metadata
• Namespace management
• Ordered collections
• Versioning
• Access control
• Searching

🌐 **www.webdav.org**

Tools

From here you can download and print a guideline document to list the requirements to manage information.

Click on the icon to open the document.

 Guidelines for requirements analysis.

Summary

• There are two main areas to consider when thinking about **document management**: managing the **workflow** and managing the **document content** itself.

• Using a **file system** is a viable option if requirements for document management are **simple**, there is only a **limited number of users,** and you don't have any particular problem with **security and backup**.

• Databases can manage metadata with **document content on the file system**, or can contain both document content and metadata.

• Databases are used in **document management systems** and in **web content management systems**.

• There is also a new standard for the management of resources: the Web-based Distributed Authoring and Versioning (**WebDAV**).

Exercises

The following four exercises will allow you to test your understanding of the concepts described up to now.

Good luck!

Exercise 1

Can you tell in which area the following management requirements fall?

	Workflow management	Content management	
		Resources administration	Semantic management
Links between documents must be managed.	○	○	○
Collaborative authoring and review must be supported.	○	○	○
Information packaged in electronic files must be uniquely identified.	○	○	○

Please click on the answers of your choice

Exercise 2

What are the advantages of each information management mode?

FILE SYSTEM

DATABASE SYSTEM

It is easier to use by multiple users.

It allows the reuse of document content in other documents.

It requires no additional software, other than the operating system of your computer and the network.

It is cheaper to implement.

Click on each option and drag it in the corresponding box.

Exercise 3

You want to change from your current file system to a document management system:
which of the following structures is easier to build for you?

○

Users

Management System

Meta Data | Documents

Database

○

Users

Management System

Meta Data

Database

Documents

File System
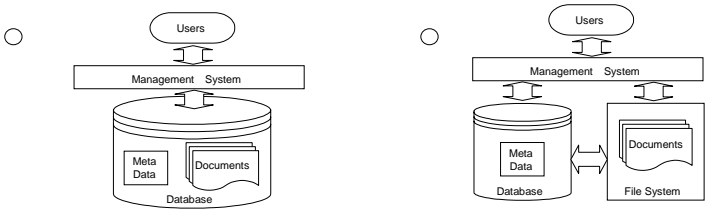
Please click on the answer of your choice

Exercise 4

You've got various options for managing documents using databases: choose the description
that best matches each of the three systems listed.

1   a

| Document Management System | A system that provides version and access control, metadata management and workflow. |

| Web content Management System | Facilitates the management of resources on remote servers allowing for web-enabled distributed document management. |

| WebDAV | Geared specifically to manage content (in multiple formats) that is to be published on a website. |

Click on each option and drag it to the
correspondent box.

If you want to know more...

CVS is the Concurrent Versions System a widely used open-source network-transparent version control system. (http://www.cvshome.org)

WebDAV - Web-based Distributed Authoring and Versioning (www.webdav.org)

Commercial document management systems include products from Documentum (www.documentum.com), OpenText (www.opentext.com) and FileNet (www.filenet.com)

Commercial web content management products are available from vendors such as Microsoft (www.microsoft.com), MediaSurface (www.mediasurface.com), Interwoven (www.interwoven.com) and Tridion (www.tridion.com)

www.B2Business.net - an online portal with information on products for electronic business, including listings of document and content management systems.