

Managing Information and Records



Sponsored by
C-Cube Solutions



ccube solutions

The definitive guide—2013 Edition

Chapter 4

Enterprise Content
Management
from A to Z



- Information and Records Management Best Practice
- IRM Solution Options
- Enterprise Content Management from A to Z
- Designing and Implementing an IRM Solution
- Making the Business Case

Cimtech

Innovation Centre ● University of Hertfordshire ● College Lane ● Hatfield ● Herts ● AL10 9AB ● www.cimtech.co.uk

information OVERLOAD



**Do you want to spend
time searching or finding**

Do you want compliance

Do you want process efficiencies



OITUK Ltd., specializes in providing C-Cube Electronic Document and Content Management & Workflow solutions, based on the C-Cube software suite. Systems scale from small departmental applications to large enterprise -wide solutions and include: the C-Cube Portal, Electronic Forms, Content Searching, and C-Cube Electronic Document & Records Management System (EDRMS), offering specialised solutions, including:

- Legal Compliance
- Health Records Management
- Law Enforcement Applications
- Information Web Portals
- Invoice Capture and Authorisation
- Local Authority Applications
- Human Resource Management

The key to all C-Cube Solutions is integration with your business to ensure that information is delivered on time and to the right place. C-Cube Solutions have met customer requirements in the public and private sectors over the last 15 years using the following underlying technologies:

- Document Management
- Workflow
- Web Portal & XML Integration
- COLD / Microfiche Integration
- Electronic Forms Processing
- Electronic Records Management
- Collaboration Facilities



ccubesolutions



the information people

CCube Solutions
13 Diamond Court
Opal Drive,
Fox Milne
Milton Keynes,
MK15 0DU

Call: +44 (0)1908 677 752
Email: sales@ccubesolutions.com
Web: ccubesolutions.com

CCube Solutions is a trading name of OITUK Ltd.
Registered in England & Wales, No: 04727067

CONTENTS

- Chapter 1 Welcome to the 2013 edition 1**
 - 1.1 The introductory guide 1
 - 1.2 Classified directory 1
 - 1.3 Who is Cimtech and how can we help? 3

- Chapter 2 Improving corporate information and records management: a guide to best practice 5**
 - 2.1 Coverage 5
 - 2.2 The case for improved information and records management 5
 - 2.3 Policies and procedures 7
 - 2.4 Roles and responsibilities 8
 - 2.5 Best practice for information and records management 9

- Chapter 3 A guide to the solution options for improved information and records management 15**
 - 3.1 Terms 15
 - 3.2 A brief history of enterprise content management 16
 - 3.3 Reviewing the solution options for information and records management 22

- Chapter 4 Enterprise content management from A to Z 27**
 - 4.1 Defining enterprise content management 27
 - 4.2 Input 28
 - 4.3 Management 32
 - 4.4 Output 37
 - 4.5 Collaboration and business process management 40

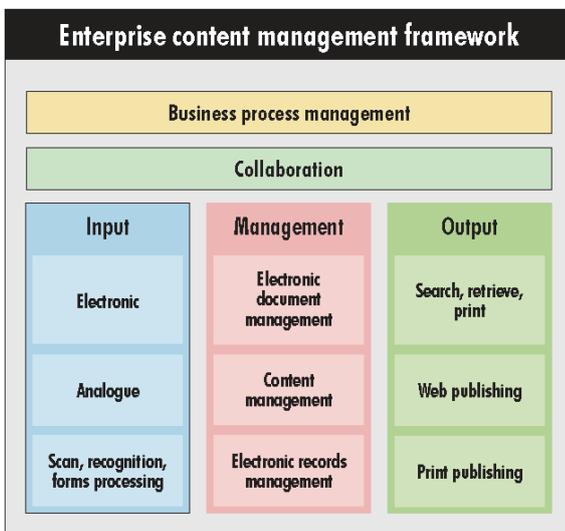
- Chapter 5 Designing and implementing an information and records management solution 43**
 - 5.1 Stage 1—Positioning information and records management solutions 43
 - 5.2 Stage 2—Defining and managing a project 44
 - 5.3 Stage 3—Information gathering and analysis 47
 - 5.4 Stage 4—Feasibility study and options review 50
 - 5.5 Stage 5—Making the business case for the preferred approach 51
 - 5.6 Stage 6—Defining the statement of requirements (SOR) 51
 - 5.7 Stage 7—Procuring the solution 53
 - 5.8 Stage 8—Managing the implementation 53
 - 5.9 Stage 9—Measuring the results 55
 - 5.10 Stage 10—Project closure and solution support 55

- Chapter 6 Making the business case 57**
 - 6.1 Reviewing the tangible benefits 58
 - 6.2 Reviewing the intangible benefits 59
 - 6.3 Reviewing the costs 60
 - 6.4 Cost benefit analysis 60

Enterprise Content Management From A to Z

4.1 Defining Enterprise Content Management (ECM)

A full ECM platform should support information and records management and collaborative working. Drilling down, an ECM solution should provide all the tools needed to input content, manage content, output content and manage all the related collaborative processes (Fig. 4.1).



These four key functions can then be further subdivided into a whole range of facilities, some of which are generic and will be required by most users, and some of which are more specialised and may only be required by a subset of the market (Fig. 4.2).

For input we need facilities to capture and digitise existing analogue content, to create and edit new electronic content, to capture emails and to capture data via electronic forms.

For management we need to store the content, index it for identification and search purposes and classify it. We need to declare a subset of the documents to be records and to manage them as records throughout their life. We need to define the authorised users of the content and provide access control and security facilities. We also need to link the content to related applications and databases and to administer the system.

For output we need to provide authorised users with facilities to search for, retrieve, view, annotate

Core Function	Collaboration
INPUT	Capturing existing content and data capture Content creation Content syndication Electronic forms for data capture
MANAGEMENT	Indexing Classification File and metadata management Application integration Digital asset management Application development and administration
OUTPUT	Searching and retrieval Content distribution Website design and development Web publishing Personalisation
COLLABORATION AND PROCESS MANAGEMENT	Collaboration Workflow and business process management

Fig. 4.2
The four core ECM functions

Fig. 4.1
Enterprise content management framework

and edit all the content and documents held and to provide facilities to identify, assemble and distribute any of the content to any publishing/delivery system that may be appropriate, including websites and printing systems.

For collaboration and process management a full ECM platform should also provide the tools needed to facilitate collaborative working and social networking and the automation of key business processes (Fig. 4.3).

In any organisation there will be scope for improving the way content is managed in all departments and across all major business processes. Where organisations start and where they place their initial investment depends on where the main problems are, where their priorities are and where they consider the best returns will be achieved.

A core principle of ECM is that the content should be created or captured only once and stored in one



Fig. 4.3
The ECM framework should provide a full range of facilities and tools

Fig.4.5
Planetary scanner suitable for bound books

location – a ‘single source of truth’. How the content is retrieved will vary depending on whether the user is an author or editor retrieving content from the main repository for editing, or an end user retrieving published content by navigating a website or a printed publication.

4.2 Input

4.2.1 Capturing existing content and data capture

Most ECM solutions will need to support the capture of existing content. There is an endless range of content and document types that can be captured and managed on an ECM system but all should fall into one of the categories of analogue or digital content, with most content now ‘born digital’.

4.2.1.1 Analogue content capture

Analogue content is usually held on paper and photographic media and has to be scanned and digitised before it can be managed on the ECM solution. If it is necessary to capture the text content of a digital image—such as the text on a page or a paper form—then recognition software will be required. Other analogue content that can be digitised and held on ECM systems includes analogue video and audio.

Scanning

A combination of a scanner hardware and image processing software allows users to scan analogue content, digitise it and store the digital image. If your organisation needs document scanning the next decision is whether it just needs to capture new documents on an ongoing basis or whether it also needs to capture an existing collection of docu-

Fig.4.4
Kodak rotary scanner



ments. For such large back file conversion projects many users opt to use a bureau service.

Most paper documents need to be manually prepared prior to scanning. Scanning bureaux typically apply an hourly charge for document preparation on top of the per-page cost for scanning.

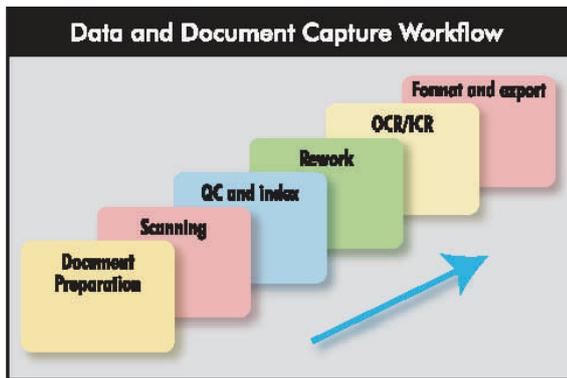
There are four main types of scanners for scanning paper documents—rotary, flatbed, large format, and planetary.

- Rotary scanners use a mechanical transport to feed single sheets through the machine (Fig. 4.4).
- Flatbed scanners are like photocopiers—users can manually place each sheet on the platen of the scanner or stack multiple sheets in an automatic document feeder.
- Large format scanners are useful for maps and plans and scan by passing up to A0 size paper over rollers.
- Planetary scanners use mounted cameras to scan mostly delicate material or bound books (Fig. 4.5).

There are also specialist scanners for scanning all formats of microfilm and 35mm colour slides, etc.

Most users will need an A4 or A3 flatbed or rotary scanner. Some scanners capture only one side of the page at a time (simplex) but most rotary and flatbed scanners can scan both at once (duplex). Document throughput rates range from a few pages per minute to 100–200 pages per minute. A high throughput rating is accompanied by greater resilience and increases the cost of the scanner. Average prices for workgroup scanners are £100-£200, departmental £1,000-£2,000 and production £5,000-£250,000.

The higher the resolution of the scanner, the better the quality of the image but the larger the file size. 200dpi (dots per inch) is the preferred standard for document capture with 300dpi preferred for optical character recognition or 400dpi for poor quality originals. Most scanners can output black-and-white or greyscale images. Black and white is good



*Fig.4.6
Data and
document
capture
workflow*

for clean text but greyscale helps where the background is coloured or marked. Many scanners can capture colour although this greatly increases the file size and should be used with caution. Some are supplied with built-in image enhancement software to improve image quality from poor quality originals.

Quality control is very important when scanning, not only to make sure all images are legible but to make sure every side of every page is captured so that the originals can be safely destroyed. The British standard for evidential weight and legal admissibility of electronic information BS 10008 and the accompanying Code of Practice BIP 0008 place much focus on the quality control around scanning. Some degree of automatic quality assurance software can be purchased on high-end systems but quality is mostly a matter of procedures and training.

Prior to scanning you need to review the documents and decide how you want them to be indexed to support your management and retrieval requirements. The range of indexing and classification techniques is reviewed in section 4.3.2 below. Cimtech and other consultants help users to define their requirements for large back file conversion projects in detail and this saves the client money and improves the quality of the results.

The next step is to load the images and index data into the ECM solution. For ongoing capture this is a simple 'release' process invisible to the user. For back file capture a bulk import process needs to be planned to ensure the images and index data are provided in the format and sequence required for loading (Fig 4.6).

Suppliers of document capture software, scanners and scanning bureau services can be found in the various sections of the Cimtech Directory.

Data capture

In addition to capturing images of a document, it is often important to capture either the full text or selected data from the document so it can be processed, searched and reused. While the cost of computer hardware and software continues to fall, the cost of manual data entry increases at an alarming

rate and there are significant accuracy issues with manual data capture. There are two key capture applications linked with scanning, full-text capture and forms processing, both of which can significantly reduce, if not eliminate, the volume of manual data entry required.

Full-text capture

Full-text capture requires character recognition software. It takes the scanning and capture process a stage further. There are three main options:

(i) The solution takes the document image, applies recognition software and loads the images into the file server and the text into a search engine. The user can search the text content of the document and retrieve the images for viewing. The user cannot edit the document content but the image is a legally admissible facsimile of the original.

(ii) The solution uses an intelligent capture software such as Acrobat from Adobe. In full PDF mode this divides an image into text and graphic zones and recognises the text and the type fonts to create an intelligent, formatted PDF document that can be searched and viewed but not easily edited. Acrobat can be used without an ECM system. The text is held within the document but can also be captured by an ECM search engine.

(iii) The solution processes the image, divides it into zones, recognises the text, captures the graphic and image zones, marks up the text using a mark-up language and tags the graphic and image files to the marked-up text. The text can be loaded into a full-text search engine for searching and the user can edit the retrieved content. This approach is used in publishing applications to fully exploit the content.

The cost of these three options is in reverse order of functionality as options (ii) and (iii) can involve significant manual processing depending on the quality/complexity of the documents.

Forms processing

If organisations need to capture data from customers or contractors outside the company then the traditional answer has been to send them a paper form. Keying data from these forms into computer systems is a costly and error-prone business that most organisations would like to avoid. As indicated in section 4.2.4 below, many organisations are now using electronic forms on the Internet. However, many organisations continue to need both paper and electronic forms processing to capture all their data.

Paper forms processing begins with the design of the paper form, which needs to be optimised for automated data capture. When the completed forms come into an organisation they need to be

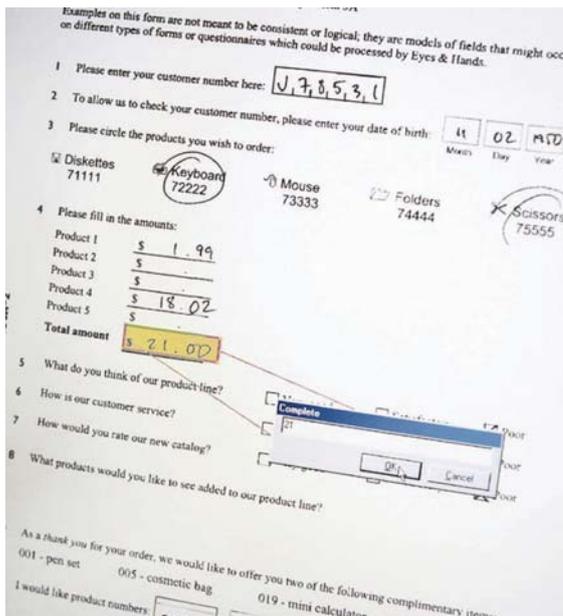
Fig. 4.7
Forms processing
workstation



scanned and converted into digital images (Fig. 4.7). The images are then processed, the form templates recognised and the boxes located. The data is recognised and extracted from each box, validated and accepted or rejected based on confidence levels and business rules. Where the data is rejected the software presents an image of the box and prompts the operator to enter the correct data via key-from-image (KFI) software (Fig. 4.8). When all the data is captured the images of the forms are saved and the data exported to the business application for loading and additional validation.

The accuracy of forms processing applications depends on how complex the data entry requirements are, how well designed the forms are, and how much scope there is for validating the data. Generally speaking, such applications aim to automate between 70 and 90 per cent of the data capture process. In large-scale applications the payback can be very significant indeed. Where possible, however, paper forms should be replaced with electronic forms that can be completed by customers online.

Fig. 4.8
Verifying data
accuracy is an
important part
of the forms
processing
operation



4.2.1.2 Digital content capture

Digital content can be created on applications such as office suites, CAD packages, digital cameras, graphics and digital multimedia packages. If it is created outside the ECM system it needs to be captured.

When a new ECM solution is implemented, there will be legacy digital content held that may need to be loaded into the new system. This will include emails and attachments, content created on office applications and content created on obsolete applications. If the content needs retention then organisations will need to load it into the new system, otherwise best practice would be to delete it.

Digital content created outside the system and legacy digital content created on old application packages may need to be converted into more common formats before being loaded into the ECM solution. The full text from digital content will be automatically captured into the search engine database.

When planning the implementation of a new ECM solution it is important to formulate a data migration plan. Folder structures in good order can be uploaded in bulk using migration packages but it is much more common to let users upload their content of choice from previous shared and personal drives. Although ECM can handle almost any file type you might wish to convert formats into something longer-lasting or more widely accessible. If you are converting formats then, depending on the importance of the material, it is advisable to run tests and check that no content or formatting data is lost.

4.2.2 Content creation

The bulk of the documents and content to be managed on ECM solutions are created electronically and, on average, 50-60 per cent are created internally in the organisation. Hence it is vital that ECM solutions are fully integrated with the applications used to create content in the organisation.

The majority of corporate documents are still created in office suites, email software, graphics, CAD, digital video and audio packages, etc. Office applications have been interfaced to document management systems for many years so all the content created on those packages can be saved directly into the system and managed from creation to disposal. These systems can be set up to make it impossible to save the content anywhere other than in the new system.

The majority of our information now comes in the form of e-mails and attachments. Increasingly,

ECM systems are designed to integrate closely with popular email packages, embedding favourite ECM folders in the email client so that users can save emails and attachments directly into the ECM system and automatically capture the sender, recipient, timestamp and subject line into document meta-data.

Web content

An important subset of content is content published to the web site and intranet. When the Internet was very new and organisations were setting up intranet sites and starting to use web content management for the first time, there was a clear distinction between back-office documents created on office applications and content being created for loading onto a website. Users who wished to use office applications to create content destined for publication on a website would typically be required to send it to the staff responsible for the website who would review it, transform the content into a web-friendly format, format it and then load it on the website. Alternatively, regular authors of content for publication on websites would be provided with editing packages designed to produce HTML and/or XML output and templates designed to reduce the scope for error and impose house standards.

Today, most desktop applications can output content in web-friendly formats including HTML and XML and there is less of a need to make a rigid distinction. In most organisations users creating simple content for intranet websites are allowed to load the content directly or via a simple approval process. Increasingly, templates are set up to ensure content is created to a house style and in a format that can be automatically transformed without the need for checks.

Today, any integrated ECM solution will need to provide support for the following:

- authors who create content in an office or third-party application and save it in native format or transform it into HTML/ XML, etc. to publish it on web servers.
- authors who create content that can be loaded directly onto web servers.

Design of web page layout is a separate activity that is usually confined to the web team and is covered in Section 4.4.3 below. Web pages can be defined and held as templates in an ECM system.

Web 2.0 content

Content is created in the wikis, blogs and discussion boards that are now well-used collaboration features in Web Content Management systems and some EDRM systems.

- A wiki enables many non-technical contributors to add and update a web page without an approval step or a conversion process. The most notable example of a wiki is Wikipedia. Wikis are useful for project collaboration and accumulating a knowledge base.
- The web log, or blog, is another important Web 2.0 application. Blogs are mainly used by individuals to publish news and ideas although readers can leave comments. They often serve as a personal home page for the individual owner.
- Discussion boards, also known as forums, message boards or bulletin boards, enable users to post questions and ideas and receive responses.

Mostly there is no retention and disposal functionality in Web 2.0 features and long term content might need to be archived back to the main ECM libraries for records management purposes.

4.2.3 Content syndication

Syndication is a content capture process offered by ECM and Web Content Management (WCM) suppliers to enable users to gather selected content from third-party sources over the internet. Content syndication enables users to set up systems to go out and gather dynamic content that is updated on a regular basis on selected web sites. This may include news feeds from news agencies such as Reuters, stock exchange data, streaming video content from broadcasters, images from the providers of image library services, audio from the providers of audio libraries, etc. In order for such a sophisticated process to work successfully content interchange standards have to be widely agreed and supported. Many of the suppliers of these content feeds will be commercial publishers looking to charge for the provision of the service. Hence they will require digital asset and rights management software, see Section 4.3.9 below.

As Tom Jenkins points out¹, as a website continues to develop, dynamic content plays an increasingly important part and the integration capabilities of the WCM software acquire central importance. The more dynamic a web site becomes the more flexible the WCM system must be to support external applications, products or interfaces.

4.2.4 Electronic forms for data capture

In Section 4.2.1 above, we described traditional paper forms processing solutions for data capture. Increasingly, organisations are replacing paper forms with electronic forms that allow their clients to enter data online, which can then be validated and loaded into a database (Fig. 4.9 overleaf). Such electronic forms software is provided bundled with business administration systems. It is also provided as part of workflow and collaboration packages.

Fig.4.9
An example of
an electronic
form

Finally, there are dedicated suppliers of electronic forms software.

Organisations are placing electronic forms on their websites and customers complete them via a web browser. Such an approach forms the basis for most business-to-consumer (B2C) electronic commerce applications today. Electronic forms solutions have to provide:

- A design module to design the most appropriate form for each application.
- A data-entry module with built-in validation and security.
- A print module for applications where users still need to print the form before completion.
- A database connectivity module for online validation via databases.
- A workflow module for routing and back-end/exception processing.

It is important to design forms that engage the user and are pleasant to complete but that also validate the data and screen out as many errors as possible. Where required for legal or audit purposes, the data entered by an individual can be captured and placed on the form template and an electronic image of that form stored so, just as with paper forms processing, users can be provided with an electronic image of the completed form and with the data for loading into a database.

Fig.4.10
ECM repository
management

4.3 Management

After inputting content, we need the functionality to store it and manage it.

4.3.1 Storage

The most basic function, which any comprehensive ECM solution should support, is to organise, store and control access to the content components in one or more repositories (Fig. 4.10).

ECM systems store content either as datafiles on a server (which may be separate from the software

and data servers) or as BLOBs (Binary Large Objects) within the ECM database. The user is not aware of where or how the storage is located. SharePoint by default stores datafiles within the database but add-on software (Remote Blob Storage) can be purchased to enable content to be stored elsewhere. Other ECM systems vary and some systems allow different subsets of the content to be stored on different hardware.

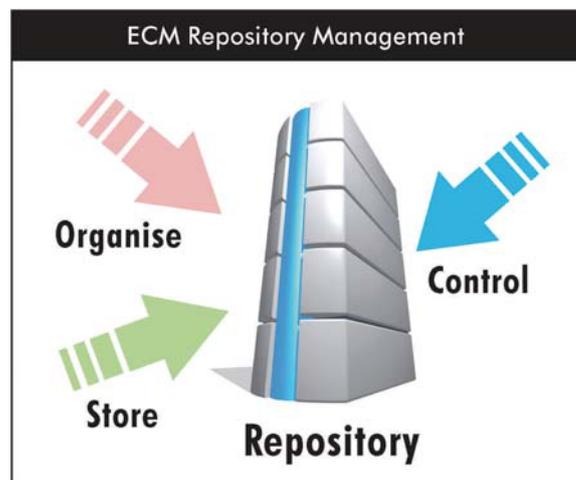
When discussing storage we should think to the future and how we are going to maintain the ECM content over time. Long record retention schedules will require the transfer of content between several generations of successive systems. The systems we buy should support bulk export facilities so that all, or a defined subset, of content components can be exported together with the associated metadata, audit trail, version history etc. Our best chance of future-proofing our content is to avoid proprietary hardware and implement MoReq2010 export formats.

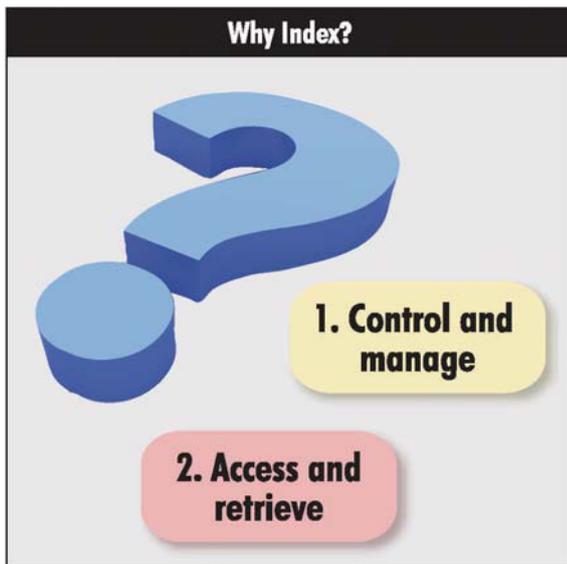
4.3.2 Indexing

What primarily distinguishes an ECM solution from storage on a Windows or another file server is the fact that content is registered on a central database and you can use the database to attach metadata (index fields) to the content.

There are two main reasons why organisations need to attach metadata to their content. The first is to control and manage the content. The second is to describe the content and facilitate search and retrieval. In ECM solutions both purposes are served (Fig. 4.11 overleaf).

Most traditional electronic document management (EDM) systems have been used in back-office functions to control and manage documents in core business processes. This requires structured indexing and structured searching.





Most ECM and all WCM systems however are geared to the front office and to delivering content to knowledge workers. Users want to retrieve and examine a number of documents containing information about the subjects they are interested in. For this purpose we can add user-selected keywords to our structured indexing but we also need full-text indexing (see below) and we may need hypertext linking (Fig. 4.12).

Structured indexing

Of the structured metadata some is captured automatically by the system (Date Created, Date Modified, File size, File type etc). Other metadata fields are set up by the organisation during implementation and completed by users when creating, uploading or scanning documents. A good system will automate field completion as much as possible by integrating with application databases and providing drop-down options and validation methods.

An ECM solution should be able to use a corporate thesaurus as an aid to consistency in describing content. A corporate thesaurus is a controlled vocabulary of terms and is designed to ensure that authors and users employ the same terms to represent the same concepts. This is especially useful where there are lists of countries, products, parts, etc. which may be known by different names that need to be standardised before they can be used to index content.

Full-text indexing

In addition to structured indexing we also want our ECM system or WCM system to find documents by the text they contain. In ECM and WCM the user is often less interested in locating a specific document and more interested in finding content about a particular subject. An ECM system will offer a search engine that 'crawls' the system, captures all the text in new documents and stores it in

a database that is reverse-indexed so that the keyword points to all the documents that contain it.

If the document is a scanned image it will need to pass through character recognition for the text to be captured by the search engine.

Hypertext linking

The third option, hypertext linking, is mainly used to create links between documents. Authors are able to link concepts directly, creating trails of associations between ideas, either within a single document or across documents. The links may be hierarchical, leading the user to material at ever-increasing degrees of detail, or lateral, using some form of cross-reference. Hypertext links (hyperlinks) are essential for websites.

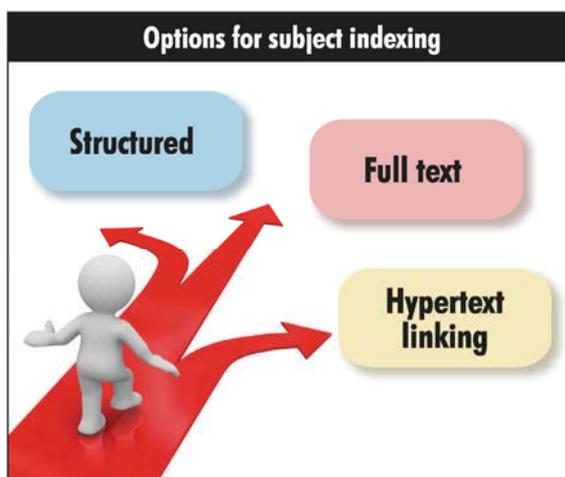
Hyperlinks are an important feature of wikis, blogs and tweets, adding to their power and usefulness.

Hyperlinks can be used between documents in web-based ECM systems, but the links will be broken if the destination document is exported or moved to another ECM system.

4.3.3 Classification

Organisations need to structure their information carefully in order to make it easy to share, navigate and manage. In EDRM systems classification schemes can be used either to create the structure of classes and folders that the user sees and browses or to create a structure behind the scenes that the records manager uses to control the management and retention of contents. Either way the classification is attached to the document, folder or class as metadata.

Best practice for most organisations is to classify their records on the basis of functions, activities and transactions, which creates a more sustainable structure than one based on organisational hierarchies and one that maps more readily to retention schedules. The conduct of an audit and the devel-



*Fig.4.12
Options for
subject indexing*

*Fig.4.11
Index data is
used to control
access data*

Fig. 4.13
Document
image with
annotations

opment of a classification scheme are covered in Chapter 5 (Stage 3 Step 2).

The National Archives and the European Union (MoReq) have published guidelines to assist organisations developing such record classification schemes (see Chapter 2). The Information and Records Management Society (IRMS) has developed a Local Government Classification and Retention Scheme (LGCRS) which offers members a largely function-based classification scheme designed for local authorities but applicable in large part to any organisation. JISC has commissioned and made freely available on its web site a classification and retention scheme for further and higher education institutions which is also transferrable to other organisations.

4.3.4 Access control

ECM solutions need to provide sophisticated levels of access control and privileges. If your organisation operates an access-control package such as Microsoft Active Directory then the ECM system should be able to use it for single sign-on. The directory holds user logins and passwords and assigns users to one or more groups or roles. The ECM system itself will usually manage ECM user access rights and privileges, including rights to read, annotate, edit, create, delete, enter metadata, amend or delete metadata and define new users etc.

4.3.5 Document management

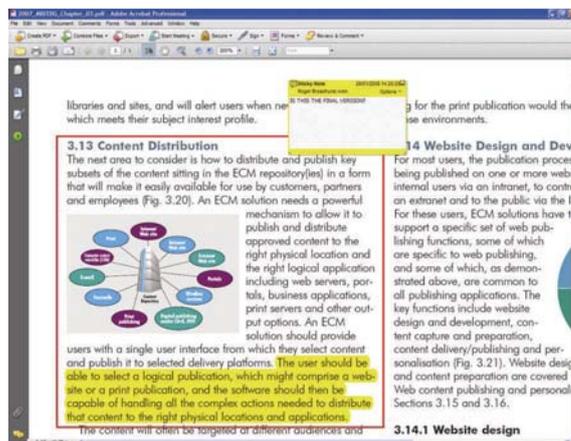
To look further into some of the management facilities required it is helpful to split our overall ECM solution up and look at some of the specific document, content and records management facilities that may be required in an overall ECM solution.

In Section 3.2 we traced the separate development of document image processing systems and electronic document management systems and how they merged to form integrated document management solutions.

Document image management

Digital images are not easily searchable so it is vital to attach metadata to the image content components as soon as they are scanned (Section 4.2.1). This metadata may specify how many images belong to one document, the date when the document was received, the author, the recipient, the document type, the document status etc.

A viewer is needed to look at the images. The viewers generally provided with office systems may be too basic for professional use. A viewer can be purchased along with the ECM system that will enable users not only to view the images, but also to rotate, enlarge or reduce the images, annotate them



and print them. Digital image files are compressed for storage and transmission so image viewers also provide image decompression. A viewer of the right type might also display other less common formats such as CAD files.

Annotation allows authorised users to attach notes, comments and mark-ups without changing the original image—as distinct from editing (Fig. 4.13). Annotations are treated as separate layers linked to the source component. Authorised users can see the original image or the annotated image as required.

Electronic document management

The ability to edit documents and manage versions is a basic feature of document management. To avoid conflicts between users many systems impose a library model where content must be checked out before it can be edited and then checked back in again. You can also configure the systems to enforce the updating of the version number, with options to update either the major (before the decimal point) or minor (after the decimal point) parts of the version number. EDM solutions are often provided with workflow management facilities so the process of checking out a document, editing, review, approval and issuing it as a new version could be automated.

4.3.6 Content management

Advanced content management is needed for publishing and for web content management. This enables content to be managed in a neutral format and rendered into different styles for different delivery channels through document type definitions (DTD), mark up languages and style sheets. A CM solution is designed to keep the content separate from the delivery mechanism or the presentation format.

With a simple EDM system you have no choice about how you view the document—you see an image of a page in an image viewer or you see the content in its native application. With a CM system

the content can be marked up using SGML, HTML or XML and then a document type definition and one or more style sheets determine how that content will be rendered on one or many delivery devices. The delivery device can be a PC, a smartphone screen, a tablet screen etc.

Hence a CM system treats content more like data in a database. A CM system manages semi-structured information to create a re-usable resource that enables an organisation to fully exploit that content just like a database enables them to fully exploit their structured data.

CM systems should:

- manage all content from creation or capture through to eventual destruction.
- provide facilities for capturing, managing and providing access to the content.
- provide facilities for processing the content and publishing it on websites, in printed form, on DVD or via tablets and mobile phones.
- manage dynamic (constantly changing) content.

CM systems support assembly: the ability to assemble a document from all its content components

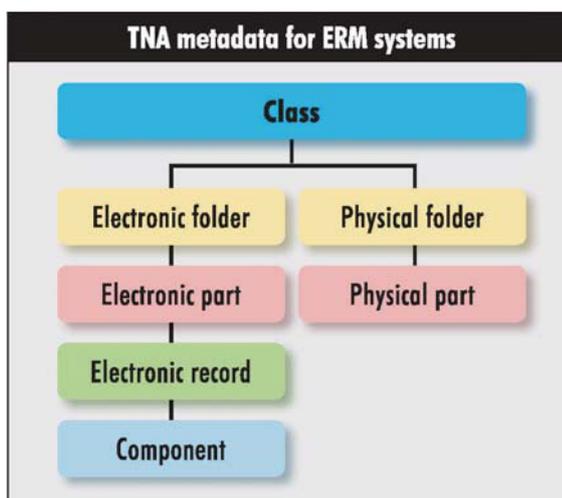


and render it on screen, on paper or however it is required to be rendered in future.

4.3.7 Electronic records management

As already described in Chapter 2, requirements for ERM software were specified by The National Archives in the UK, by MoReq for the European Union and by the US Department of Defense (DoD 5015.2). Software was initially developed to operate independently but now is usually integrated into an EDRM system in which editable documents and read-only records coexist in the same folders.

Compared to EDM, the main additional records management requirements include the following:



*Fig.4.14
The National Archives
hierarchy for
ERM systems*

- Individual documents can be declared records and from that point on their content and metadata is controlled and cannot be changed.
- The system must support a classification scheme. The National Archives file plan defines a hierarchy of classes, folders, folder parts, documents/records and components (Fig. 4.14).
- Metadata can be used not only to describe documents but to describe classes, folders and other aggregations.
- The system must support the maintenance of records over time, to retain only those that should be kept and manage the controlled disposal of those no longer needed. Retention and review schedules have to be assigned at the folder or record level and processes in place to support the controlled review and destruction or export of records at the end of the retention period.
- Most ERM systems will be required to manage electronic, paper and hybrid records. The metadata held at the class and folder and part level is similar but in most cases paper documents will not be individually indexed on the system unless they are of considerable value. The system will track the physical location of the paper records in boxes, on shelves in specific stores.
- The requirements call for electronic preservation and archiving methods. These would include tagging with preservation attributes and support for the rendition of content into new formats to avoid obsolescence.

4.3.8 Application integration

Today, most organisations will not be looking to install an ECM solution in isolation but to integrate it with core business administration systems. Tom Jenkins¹ sums up the requirement very effectively: "To maximise its effectiveness, organisations need to connect content to the appropriate business

processes—and make it automatically accessible to anyone involved in the process".

The business systems hold the key data and the ECM system holds the supporting content so, clearly, the two need to be designed to work together.

There is a wide range of business administration systems in organisations today. These include bespoke systems developed in-house, best-of-breed package solutions implemented in one area of the business, or enterprise-wide systems—most notably the enterprise resource planning (ERP) suites provided by companies such as SAP and Oracle. The ECM solution should be provided with open, documented application programming interfaces (APIs) to all the ECM functions so that any data application can in theory connect with ECM.

The main points of integration required between ECM and data applications are:

- A hotlink (icon or option) can be put on the data application screen to pull up and view an associated document or documents from the ECM system.
- Documents generated on the data application (e.g. mail merge letters) can be automatically captured and indexed into the ECM system.
- Data from the data application can be used to validate metadata in the ECM system during indexing.
- Data from e-forms and scanned forms can be entered into the data application to avoid manual data entry. Sometimes workflow (BPM/WFM software) is used to act as a bridge between the user and the different systems needed.

Where an organisation is rolling out an ERP system and an ECM solution the two can be closely integrated. The same integration approach can be replicated in a number of applications such as finance and human resources. Not surprisingly, a number of the major ECM suppliers have strategic alliances with the ERP suppliers.

Fig.4.15
Screenshot of a
digital asset
management
(DAM) system



4.3.9 Digital asset management

Another module offered in many integrated ECM solutions is digital asset management (DAM) and digital rights management (DRM) software. DAM and DRM are aimed at organisations that own or manage rich or valuable digital content assets including images, audio tracks, video, etc.—all of which is or could be marketed to other organisations on a commercial basis. The owners of the content need software to manage these assets, help potential customers to identify the content assets and make the content assets available in the format required by the customer (Fig. 4.15). The more flexible the output options the greater the chance of a sale.

Initially, DAM was used primarily by video production, printing and publishing departments. The Internet revolution has made DAM much more widely applicable. DAM now allows organisations to manage large collections of diverse content centrally and prepare and stage that content for delivery via digital distribution channels. The Digital Media Resources (DMR) Group at Penn State University provides a useful definition. 'DAM ingests, indexes, categorises, secures, searches, transforms, assembles and exports content that has monetary or cultural value. The fact that an asset is represented digitally presents many opportunities for revenue generation and operating efficiency. This is what defines DAM and distinguishes it from ECM in general'.

The DMR Group² define seven core steps in the DAM process:

- At creation an artist, writer, editor or filmmaker creates new content.
- The content is indexed and the metadata describes when it was created and can be used to interpret and catalogue content.
- The content is stored.
- When required the content is delivered in the required format.
- The content is resold.
- The content is reused as required.
- The content is reviewed.

So DAM is used to manage rich multimedia formats and in most cases the owners want to be able to charge the user for access and to prevent the user from gaining unlawful access. The objective of DRM is to give customers what they want, when they want it (one-time use, subscription, redistribution rights, etc.), how they want it (format), where they want it (channel), and at a price they want to pay (pay per view, multi-subscription rates, try before buy, etc.) in an easy way.

From a publisher or asset owner's perspective DRM needs to cover content reuse, security of rights protection against piracy, etc. There are lots



of different views about the benefits and viability of DRM. Someone in your organisation will probably need to make use of such software. If so, it will be worth checking if the ECM suppliers on your shortlist of all provide a DAM module.

Recent examples of the power of DAM include the transformation of the music industry by Apple and others so we can download digital audio tracks online for a fee; the development of electronic book readers so we can download the content of selected books and the emerging transformation of the television broadcast industry as multimedia becomes available for download as well (Fig. 4.16).

4.3.10 Application development and administration

Although many organisations will use a systems integrator or the main supplier to implement their preferred solution and integrate all the component parts, they will need to administer the solution themselves once it is fully implemented and they may need to change the configuration and develop workflow applications.

For a truly scalable solution, a good central administration module is needed to standardise and control the configuration settings across the whole system.

Administrative functions to be supported will include:

- defining users and their roles, access rights and privileges,
- defining the classification scheme to the system and ongoing configuration control/maintenance,
- administering workflows and changing process definitions and business rules,
- defining, running, maintaining and modifying key system reports,
- supporting multi-server architectures,
- roll-back/recovery in the event of faults when content is being distributed to servers, etc.

- monitoring the usage of content on websites and archiving old or infrequently-used content,
- backing up the content objects and metadata in the repository,
- archiving the content objects and metadata,
- controlling the mass import and export of content and metadata as required,
- maintaining a complete log and audit trail of user transactions on the managed content,
- keeping the repository and the web servers in synchronisation,
- preserving electronic records via migration and other preservation strategies.

*Fig.4.16
Multimedia
downloads to
portable devices*

4.4 Output

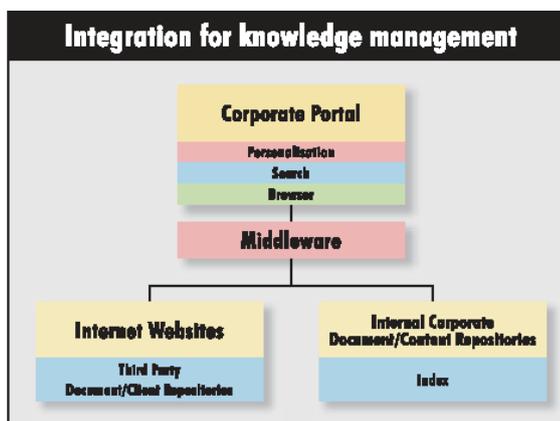
4.4.1 Searching and retrieval

Search

A core attribute of any ECM solution is to improve access to the content held. Core technologies include structured and free text searching (Fig. 4.17).

When searching on structured attributes (metadata fields/indexes) users will require a form-based or query-by-example interface where they are prompted to enter the attributes to be matched in specific fields. Structured searching is ideal when back office staff are looking for a specific document or content component that they know exists in the library or for all the documents by a specific author or on a specific date.

Content or free-text searching is ideal when users are looking generally for information on a particular subject. They can enter any word or combination of words and retrieve all documents that contain that word or combination of words. Search engines support natural language searching so users do not have to learn Boolean operators. Many search engines also contain fuzzy logic facilities to overcome poor spellings and employ algorithms that can expand the search terms used to locate all related concepts. Some search engines can parse and search against non-textual assets, such as im-



*Fig.4.17
Integration for
knowledge
management*

ages and video, and profile users based on which hits users opt to retrieve.

ECM solutions support combined text searching and structured searching. Mostly they offer the combined search as standard and offer structured metadata searching as an 'advanced search' option.

Search engines are now familiar from internet search engines, which are impressively fast and accurate. Users often ask for 'Google-like' search functionality. However, one reason internet search engines are so accurate is that they work on huge volumes of content, hyperlinks and users, which enable them to use past searches to accurately predict what a future user wants. The search engines used in corporate systems have a smaller body of content to work on and must use different algorithms to rank their results. These are not always so impressive, so the ability to view and refine the results in different ways is a useful feature to ask for.

Navigation

As well as searching, users also need to find information by browsing around the fileplan or classification scheme. A good ECM system presents its content in a tree structure similar to Windows Explorer and other file system applications. The fileplan needs to be designed to make browsing as friendly and intuitive as possible. Too many levels makes it harder to find information. Folders within folders are not advisable unless it is necessary to separate content in order to apply different access control or retention schedules.

Views

Some ECM systems enable metadata to be used to present different views. This allows the content to be sliced in a variety of ways and enables the fileplan to be shallower. An example would be a library of invoices which could be viewed by date,

Fig. 4.18
Web publishing
functions



by supplier, or by amount, avoiding the need to create subfolders for date ranges or suppliers.

4.4.2 Website design and development

ECM solutions that include Web Content Management functionality support a specific set of web publishing functions. The key functions include website design and development, content capture and preparation, content delivery/publishing, approval and personalisation (Fig. 4.18).

Website design

The first task is to define the structure of the website or the overall publication. Website designers usually start with a site map that illustrates the site's structure and storyboards that detail the pages and their relationships. It is essential in web publishing to separate the content from the overall structure or logic of the publication. Web content needs to be held in a form which allows it to be reused and displayed in different formats.

Web pages are defined and held as templates and need to be tightly managed. They define the layout of the page and where specific content objects will be displayed and where the menu items will appear. They define the containers into which the content is placed and the display formats. Good website design involves the use of a limited number of page templates. The WCM system should manage the page templates as content components. The leading ECM suppliers support multiple presentation templates so content can be created once and then displayed in different devices. For marketing purposes this technique can be used to take content and re-brand it for different applications.

The HTML format supports the creation of links between pages and content items but cannot guarantee the integrity of all the links. Changes to the website can break links. A good WCM tool can be used to check the links and guarantee the integrity of the links within a website. The links that need to be supported are the links needed to cater for all the ways in which users may need to access, cross-reference and browse through content on a website.

Content preparation and approval

The content that organisations want to publish on a website may need to be created or captured using a range of techniques as outlined in Section 4.2 above. Once obtained the content may need to be reviewed and approved for publication and third-party content may need to be prepared for publication by converting it into a supported format.

The leading WCM systems provide content collection or content aggregation server software that can collect diverse types of information from other

websites, databases, flat files and e-mails and transform the data for presentation on the website. All WCM software should support collaborative authoring and editing, allowing organisations to devise automated processes for reviewing and approving content prior to publication.

Web publishing

The third function that any WCM system should support is the publishing or distribution process (Fig. 4.19). This involves taking the editorial content components and the relevant web page templates for a website from the CM repository and distributing them all to the selected web site server or a staging server. The preferred web server and application server software will then normally be responsible for running the live web site and delivering the content to the visitors. The WCM software should be able to identify the web page templates and content components that need to be delivered from the content library to the web site by reviewing the attribute data held for the content. The WCM software will also be required to remove selected content from the live web site if that content has expired or has been replaced. Finally, the software is required to maintain all the links between pages and content components on the web site.

WCM software should provide tools that address the needs of multiple website deployments. These simplify issues such as synchronisation, localisation, branding and content delivery. An author may be creating content that is to be delivered on the corporate website and the corporate intranet. The content may be customised and reformatted for each site. All this should be automated as far as possible as soon as the content is approved for publication.

The web publishing process, unlike conventional publishing, is not always a one-time process. The content delivered to visitors often needs to be assembled from different content sources. The WCM solution should be able to support a range of options from fully static assembly with pre-prepared HTML pages to a fully dynamic assembly model where the web or application server does the assembly of web pages. In a static delivery case, all the content assembly should be conducted by the WCM software prior to content distribution. Where this is feasible it will optimise performance.

In a dynamic situation, the assembly should be conducted by the web or application server. Dynamic assembly is well suited to applications where real-time data needs to be inserted from a newsfeed or back-office application, or where the content needs to be personalised based on the profiles of a specific visitor.

Increasingly, on active websites, the web content needs to be delivered on demand in response to specific queries. For example, many organisations' website home pages contain lists of news items, and when a user clicks to retrieve those news items, the list of news items is created by issuing a query on the CM library.

Personalisation

The interactive delivery of web content means that the content can be personalised to meet the needs of the individual who is searching the website. There is a range of personalisation options, from translating content into different languages, selecting content that suits the age and level of expertise of a user, to reacting to users' expressed preferences.

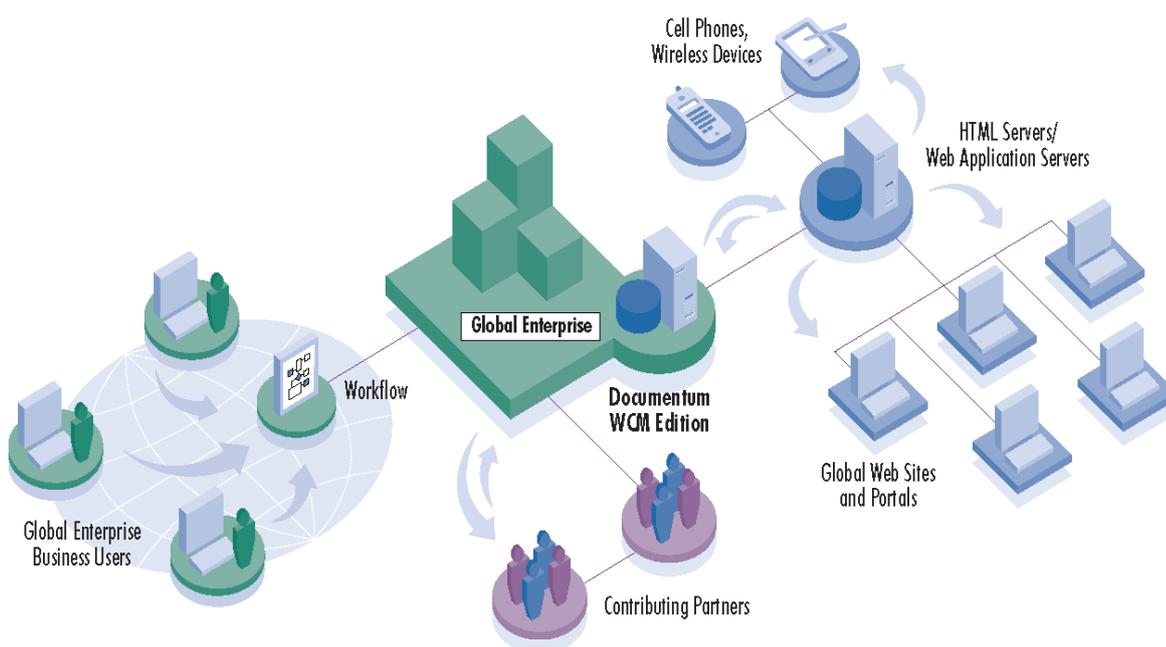
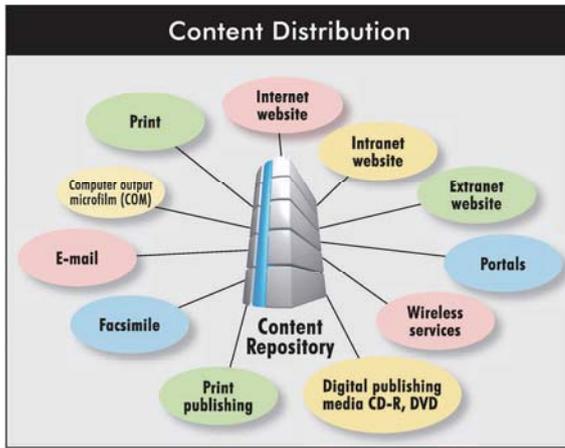


Fig. 4.19
Enterprise web content creation and publishing to multiple devices and sites (courtesy EMC Documentum)

Fig.4.20
Content
distribution



WCM systems must be capable of generating on the fly content that is directly relevant to the user. Examples of the facilities available include automated system adaptation for various browsers, language and location detection, behaviour adaptation, implicit optimisation of site structure and content to suit the needs of the visitor and the use of a recommendation engine.

An integrated ECM solution needs to include a powerful suite of personalisation software to ensure that organisations can meet all the current and future requirements of their various website users.

4.4.3 Content distribution

An ECM solution needs a powerful mechanism to allow it to publish and distribute approved content to the right physical location and logical application (Fig. 4.20) including web servers, business applications and other output options. An ECM solution should provide users with a single user interface from which they can select content and publish it to selected delivery platforms. The user should be able to select a logical publication, which might comprise a website or a print publication, and the software should handle the complex actions needed to distribute that content to the right physical locations and applications.

Fig.4.21
Email software
facilitates simple
collaboration

To guarantee delivery of content via an ECM solution, the transport of the content across the network should be transactional, where the transaction is successful only when confirmation is received that the content has been successfully delivered to all the different locations. The actual delivery of the published content to the end users will be performed online by a range of standard web, application, portal and e-commerce servers and offline by print systems as required.

Content distribution is very important in the construction industry, as an example, where 'transmittals' of updated documents, typically a drawing, specification or procedure, can be automated and audited to ensure that everyone on a project is working off the latest versions.

4.4.4 Mobile access

The increasing use of smartphones and tablets has led ECM suppliers to develop applications (apps) that provide remote access to their products from a variety of mobile devices. ECM apps enable users to browse, search, view, manage and email documents while on the move. Special plug-ins enable the reformatting of document types such as spreadsheets and PowerPoints for smaller screens. Users can contribute to corporate discussion threads and micro-blogging sites while travelling and use mobile workflow to fill in forms, trigger processes and approve documents and decisions without returning to the office.

4.5 Collaboration and Business Process Management

4.5.1 Collaboration

Alongside information and records management the other main function which an ECM solution must support is collaboration. People and systems need to work together and access and create information in order to conduct essential business processes. Tools that facilitate the sharing of information and help people work together and interact with computer systems more effectively can save organisations considerable amounts of money and improve customer service significantly.

ECM by offering a corporate and open fileplan can enable collaboration not only within divisions and teams but across the whole organisation. This is particularly useful for cross-cutting projects and initiatives and for organisations that manage their work in work streams rather than team structures.

In addition to the basic ability to share information offered by ECM we have a number of other features that assist with collaboration: workflow (see next section), email and social networking software.





E-mail

The most common form of collaboration is by email (Fig. 4.21). Your ECM system should integrate with the email system for the capture of emails sent and received into ECM folders (see section 4.2.1). It should also be able to send out an ECM document by email. If the recipient has access to the ECM system you should be able to email a link to the document, otherwise it will be an attachment. One of the big benefits of EDRM is the reduction in email attachments. That is one reason why it is often better to roll ECM out to all staff quickly with a 'vanilla' configuration, then add the complexity later.

The problem with basic email as a form of collaboration is that there is not much feedback to let you know whether the user has received the message and acted on it, and so if you are forwarding the same kind of information to the same people on a regular basis you will be looking at content distribution techniques (Section 4.4.3).

Collaboration and social networking software

In addition to basic ECM collaboration some ECM suppliers provide their own collaboration software suites and almost all interface with what has become the most popular collaboration suite, Microsoft SharePoint. SharePoint is good for collaboration because it enables content to be presented in sites each with a home page to which can be attached web parts. The web parts might be used for news feeds, charts, project calendars, pictures or dashboards showing the state of the project or the company. Wikis, blogs and discussion threads can also be linked to the home page.

A number of the major ECM providers offer advanced collaboration technology. EMC Documentum has released CenterStage as a replacement for their eRoom collaboration software. Benefits to the customer are said to include better performance, easier customisation and quick access to new user interface frameworks such as wikis, blogs, social networking and RSS feeds.

OpenText has also unveiled a major new release of its Web Solutions that gives organisations new capabilities to rapidly deploy intranets, extranets and Enterprise 2.0 solutions, and meet the expanded demands of new digital strategies. The new release delivers a complete set of Web 2.0 tools tightly integrated to give customers far greater security and control over social media than possible with a set of point solutions.

Other forms of collaboration are provided by software usually outside the ECM suite, such as instant messaging, presence detection, shared calendars and telephone and web conferencing. Some of these packages will capture content that needs to be archived into the ECM system for longer retention.

Your organisation needs to decide how many of these tools it needs to support collaboration and social networking and then decide whether it wants to implement a collaboration suite as a separate project and integrate it with the ECM solution or whether it purchases the collaboration/social networking module of the preferred ECM platform.

4.5.2 Business process management

Business process management is also known as workflow management. Workflow management means literally managing the flow of work through a business process—a process that will increasingly have been modelled and reengineered to meet business objectives (Fig. 4.22). The word to emphasise is 'managing'. As indicated above, BPM software is usually implemented by organisations that must have tight control over the flow of work through a particular structured business process. The reasons why they need tight control usually include some of the following:

- high volumes of cases being processed at any one time,
- many people and computer systems involved at various stages in the process,
- to impose control on the process,
- to truncate the process-elapsed time,

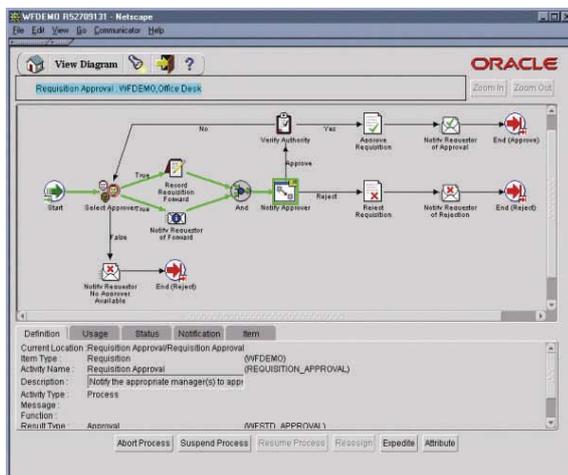


Fig. 4.22 Example of graphical workflow with business process management software

- to ensure each case is processed to quality standards,
- to ensure each case is processed according to a set of rules,
- to capture management information and audit trail data,
- to improve customer service.

When looking at ECM solutions it is useful to distinguish between the types of basic BPM software that will be bundled with the ECM solution and the features requiring third-party production BPM software. Most ECM solution suppliers will provide some bundled BPM facilities. They will typically offer document-based workflow: content capture and approval, content editing and approval, version control, content routing and distribution and content publishing on a website.

If more complex BPM capabilities are required to carry out sophisticated process modelling and workflow design then the integrated ECM solution should include a powerful BPM software package.

Production BPM is not document-based: it is process-based and may involve multiple documents or none. The kind of functionality available with production BPM includes the ability of processes to fork and 'rendezvous', to put processes on hold and reactivate them, to set timeouts and to interact with line-of-business systems such as ERP applications. Managers can use BPM to automate load-balancing between staff, to prioritise work items and to obtain performance data.

A list of WFM/BPM suppliers is provided in the Cimtech Directory.

References

1. Jenkins, Tom. Enterprise Content Management Technology: what you need to know. Open Text. Third printing, 2005. ISBN 0-9730662-5-3 www.opentext.com.
2. www.drm.uk.com, www.digital-rights-management-review.com.



The Cimtech Directory

Classified listings of information and records management products and services.

Available online at
www.doconsite.co.uk
(free access)