2012

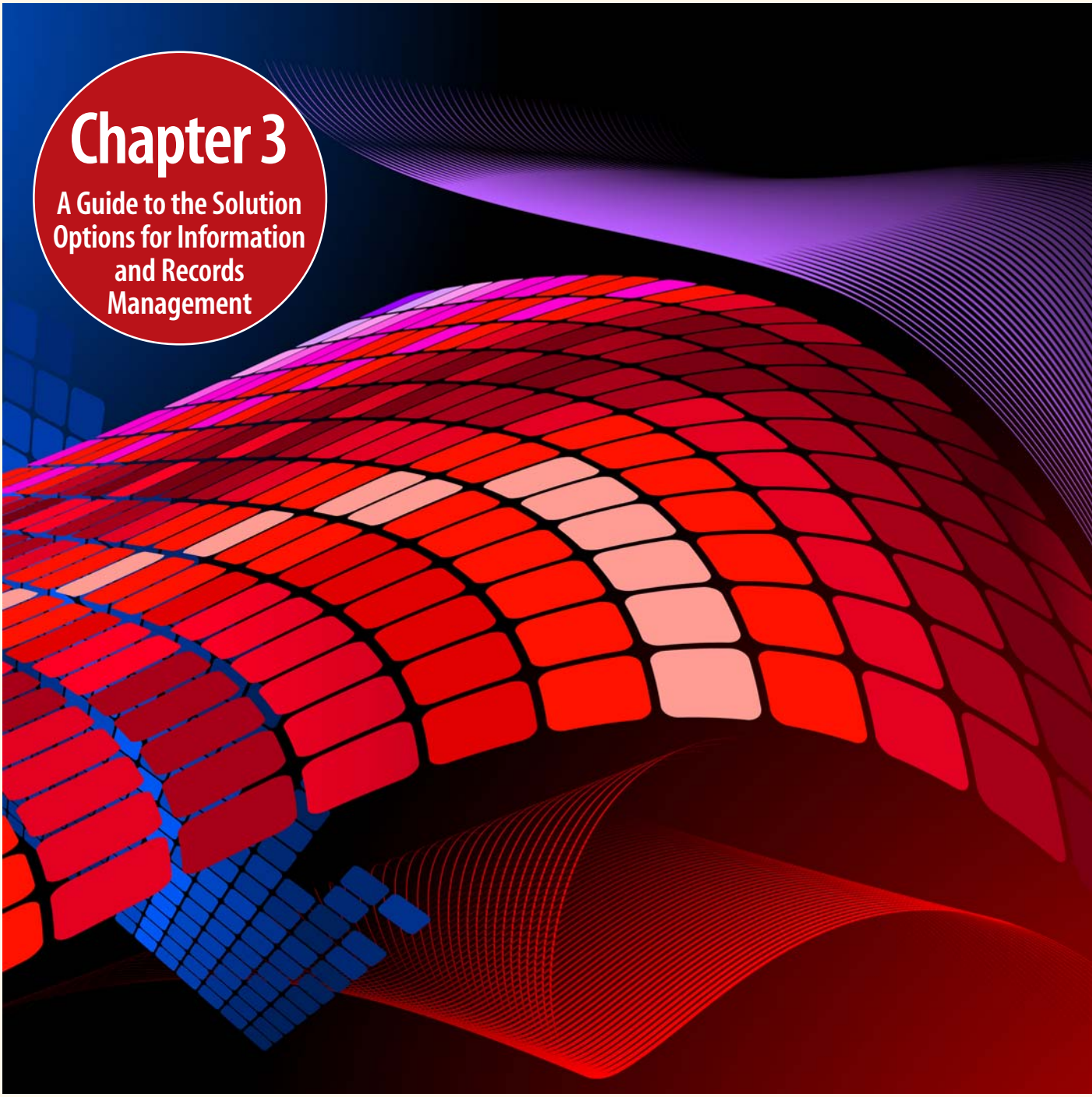
Managing Information and Records

The definitive guide—2012 Edition

Cimtech

Sponsored by
C-Cube Solutions

ccubesolutions



Document Management • Records Management • Content Management • Document and Data Capture • Workflow Management • Services

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

Looking to use SharePoint for Document and Records Management?

Cimtech offers the following one-day course
on dates throughout the year

Information Architecture for SharePoint Document and Records Management

£25 Early Booking Discount

Topics covered include:

- SharePoint information governance
- Information architecture design
- Analysing and organising content
- Applying retention schedules
- Tagging content
- Migration issues

Full details at www.cimtech.co.uk

Reserve Your Place Online

Cimtech, Innovation Centre, University of Hertfordshire, College Lane, Hatfield, Herts. AL10 9AB
Tel: 01707 281060 • Fax: 01707 281061 • e-mail: c.cimtech@herts.ac.uk • www.cimtech.co.uk

CONTENTS

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**

DM DOCUMENT MANAGER AWARDS 2010 WINNER
Project of the year - Public Sector
CCubesolutions from OITUK Ltd St Helens & Knowsley NHS Trust

DM DOCUMENT MANAGER AWARDS 2010 RUNNER-UP
Document Manager Product of the Year
CCubesolutions from OITUK Ltd C-Cube Portal

the information people

ccubesolutions

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

Chapter 3

A Guide to the Solution Options for Improved Information and Records Management

3.1 Definitions

- **Content**

Content is used to define the range of data types that organisations need to manage in order to manage the full range of corporate information. Text, graphics, images, audio, video, etc., are all content types.

- **Content Models/Containers**

The computer industry uses 'files' to hold content data typically created using one application—so text created on a word processing package is stored in a text file while image data created on an image processing application is held in an image file. The file name suffix indicates the application used to create it. So we could say that 'file' is the basic element in the content model or hierarchy.

Confusingly, in the world of records management and paper filing, the term 'file' usually refers to a paper folder or container that holds multiple paper documents. If we use 'file' at the base level in an ECM system what term do we use for containers that hold more than one content 'file'? In computer operating systems we use 'folders' and can create multiple levels of folders. In the UK The National Archives use the term 'folder' but in Europe records managers prefer to use the term 'file'.

- **Components**

The standards bodies use the term 'component' to define the basic content object or file. Electronic content is therefore held as a component or file or content object. The authors of the MoReq standard define a component as a 'distinct byte stream' and technically that is what it is. You can have a text component and/or an image component, etc. The content component is therefore level one in the content model (Fig. 3.1).

- **Document**

A document is a container that comprises one or more content components and which stands alone as a unit of information. This is level two in the hierarchy. The information contained within the document may be structured, semi-structured or unstructured. The world of business and government is based around documents. Documents can be updated and reissued on a regular basis and hence many documents need to be subject to version control.

- **Publication**

A subset of documents will be published as publications. Prior to electronic publishing, manuscripts would be sent to a printer who would typeset the content and create metal plates from which the required number of paper copies would be printed. With electronic publishing the content is marked up using a markup language and rendered using a defined style sheet. Using the same marked-up content and different style sheets the electronic publisher can publish the content on paper and via a range of digital channels. With the Internet we are publishing an increasing percentage of our documents so more of us are making use of markup languages and style sheets to ensure we can reuse our content and publish it across a wide range of channels.

- **Record**

Traditionally, a subset of the documents held by organisations would be designated as records and a smaller subset as vital records. The criteria used to determine what was defined as a record included legislation, regulations, operational, historical and others. Increasingly, a subset of all electronic documents are declared and held as electronic records. An excellent definition of a record was provided by The National Archives[1]:

A record is a specific piece of information produced or received in the initiation, conduct or completion of an institutional or individual activity. It comprises sufficient content, context and structure to provide evidence of the activity. It is not ephemeral: that is, it contains information that is worthy of preservation in the short, medium or long term.

- **Folder/File**

As explained above, there is a need to group related documents and/or records together in higher-level containers. Traditionally, filing clerks and records managers used paper files as containers to hold multiple related documents and/or records together. The computer industry uses the term file as the first-level entity in the model to manage content. Increasingly, the records management community now refers to

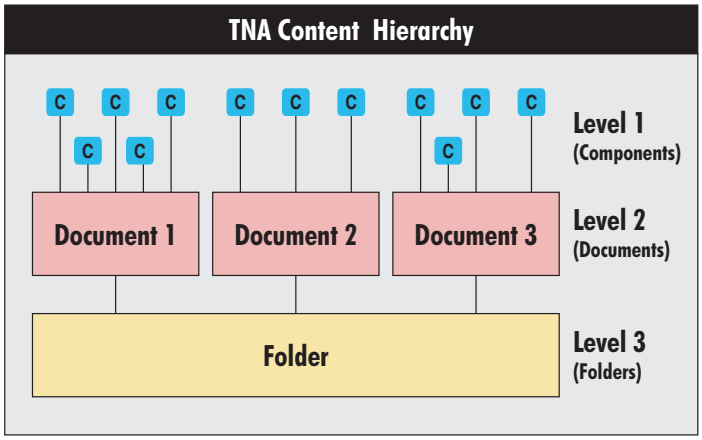


Fig. 3.1
The National Archives content hierarchy

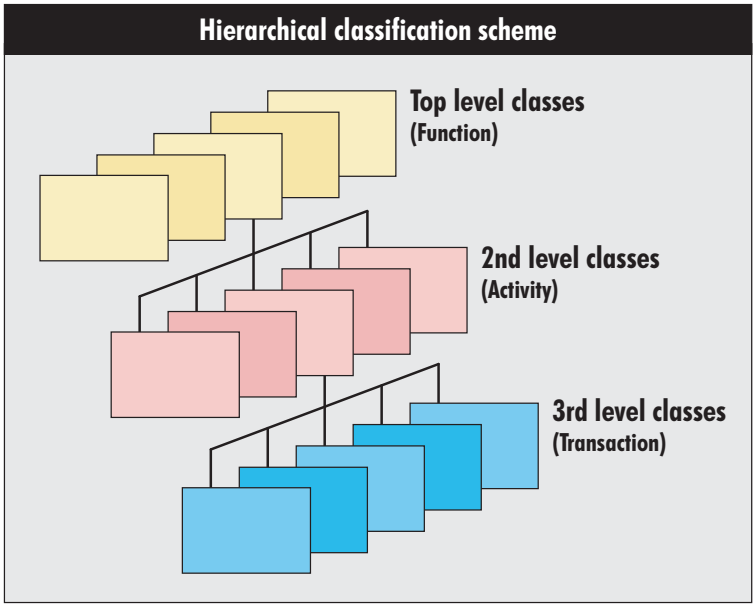


Fig. 3.2 Hierarchical classification scheme

that level as a component. Level two covers documents and/or records. At level three in the hierarchy there are two options available. The National Archives use the term 'folder' which neatly avoids any confusion with the level one component or file, but this is a UK-only standard. The MoReq standard uses 'file' at this level three and uses 'component' at level one. For convenience, we have kept to the TNA use of the term folder in this guide for level three (Fig. 3.1). For electronic content we use the term component for level one. If you use component at level one feel free to use file or folder at level three.

- **Part/Volume**

Paper folders can be divided up into parts based on chronology or number of documents, etc., so electronic folders can be divided up in a similar way. The TNA uses the term 'part' and allows folders to be divided into parts. MoReq uses the term 'volume'. In MoReq2 if sub-files are being used then each sub-file can be divided into volumes—if not then files can be divided into volumes.

- **Class**

Traditionally, librarians and records managers have used a classification scheme which divides the world of knowledge into classes on a hierarchical basis with as many levels as are needed. In libraries, books were catalogued or classified by being assigned to the right class. Classification is used today in electronic content management and records management. Folders belong to a class, and there can be a variable number of class levels defined in a classification scheme (Fig.3.2).

- **Metadata**

Metadata is the term used in the computer industry to describe data which is held and associated with components, documents and/or records, files/folders, classes and classification schemes to uniquely identify each entity and to provide information on its content, context, structure and use. The metadata is what is used to identify and retrieve the content, to organise it, control access to it and assure its integrity.

3.2 A Brief History of Enterprise Content Management

3.2.1 The all-paper environment

One hundred years ago all the information which an organisation held—the corporate memory—would have been held on paper or in the brains of its employees. Data was recorded in tables in ledger books (Fig. 3.3) and documents comprised handwritten text on a range of paper sizes. When mainframe computers were first introduced some of the ledgers were replaced by temporary digital storage, but for safekeeping most computer data was also output to paper in the form of bulky computer printouts.

3.2.2 Paper and microfilm

A photographic medium called microfilm had been in existence for many years in niche markets such as banks where cheques were regularly microfilmed. In the 1970s, alongside the rise of the computer, the use of microfilm increased as a way of reducing the space occupied by inactive and archival paper documents and/or records. Paper folders were used while documents were active but, when they became less active, the paper was microfilmed to save space and for preservation. Large volumes of valuable reference documents were also microfilmed and duplicates made and sold as a low-cost form of publishing. There were roll microfilm formats including 16mm and 35 mm (Fig.3.4) and flat or unitised formats including microfiche, microfilm jackets and aperture cards.

Government departments, financial institutions and manufacturing companies all made use of microfilm. Several suppliers also developed computer output to microfilm (COM) recorders, which recorded computer data directly onto microfilm. The use of microfilm peaked in the 1980s and billions of documents were recorded onto this medium. Microfilm made a secure archive format and huge collections

Fig. 3.3 Ledger books

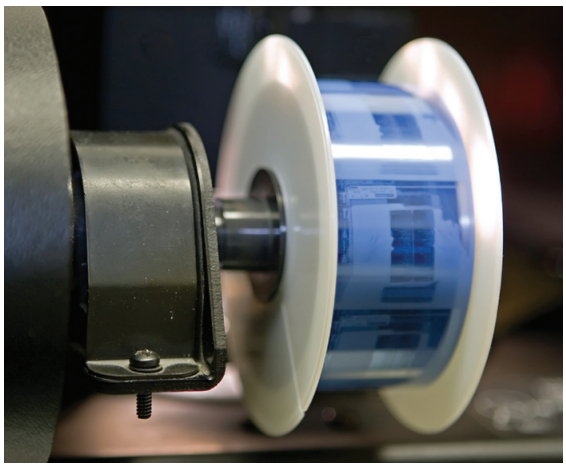


Fig. 3.4
35mm roll microfilm in use on a reader

remain on microfilm today. However, it was never a very easy format to use and was not practical for the management of active documents. It survives today as a proven archive medium.

3.2.3 Computers in the office

In the 1960s and 70s medium-to-large organisations had their own mini and mainframe computers but, with some exceptions, they were largely used for data processing rather than document creation. Tools for managing structured data developed to include relational database management systems, simple flatfile databases and free-text systems. Government departments and large commercial organisations started to use computer systems to manage paper records. A computer database was used to hold the metadata needed to uniquely identify each folder and folder part, to track the location of folders in the registry or out on loan to users, and the status of folders as open or closed.

In the 1980s the entry-level cost for computing came down considerably and we saw the introduction of personal computers in the office. Word processing became widespread with staff creating their own documents in electronic format. This was followed by spreadsheets and many other personal productivity applications. As Microsoft achieved dominance in the personal computer world, with its Windows operating software and Office applications, a level of de facto standardisation was achieved and users started to develop ad hoc, hybrid, records management solutions where internally-generated electronic documents were held in electronic folders on shared drives while incoming documents were held in paper folders in shared locations.

3.2.4 Document image processing (DIP)

In the 1980s suppliers began to introduce a range of scanners that could scan paper sizes from A4 to A0 and create digital images from them which could be stored, retrieved and viewed on a new range of high-resolution PC displays. Small-scale systems were, in effect, electronic filing cabinets (Fig. 3.5). Active, scalable systems were supplied with workflow management (WFM) software so organisations could scan the incoming post, route the images to designated staff to process, and keep track of the status of each work item on the system. Systems were often provided with a forms-processing capability so they could capture data from forms, validate it and load it into business administration databases[2].

3.2.5 Electronic document management (EDM)

In the 1990s a new generation of systems was introduced to manage active, changing digital documents created on Microsoft Office applications and other popular application packages. These electronic document management (EDM) systems managed all the documents in a repository or library and assigned index/metadata to each document to manage the relationships between them and to provide access control. They were provided with check-out and check-in facilities and version control and were widely used in head-office applications and by legal firms and pharmaceutical companies. Over time, DIP and EDM systems merged to form integrated systems. The document and data capture subsystems were mainly provided by specialist software providers who designed their capture subsystems to interface with the leading EDM systems.

3.2.6 Text retrieval and knowledge management

Ever since books were first printed, publishers have employed specialist indexers to extract key words from the text of the book and arrange them in alphabetical order in indexes. Shortly after organisations started to process text on computer systems, a number of suppliers developed full-text software engines that could take a large library of electronic documents and create a full-text index listing in alphabetical order all of the words contained within the documents and every instance where they were held. A whole industry was born with publishers creating huge online databases of documents on specific subjects that information professionals could search online to create a shortlist of relevant material. The EDM suppliers integrated the leading text engines into their solutions. Users could search for the specific folder or document they wanted or carry out a text search and find all the documents in the library that contained specific words or phrases. The Internet is powered by powerful search engines from Google, Yahoo, etc.

Fig. 3.5
An early document image processing (DIP) system

which allows us to identify relevant content across millions of web sites

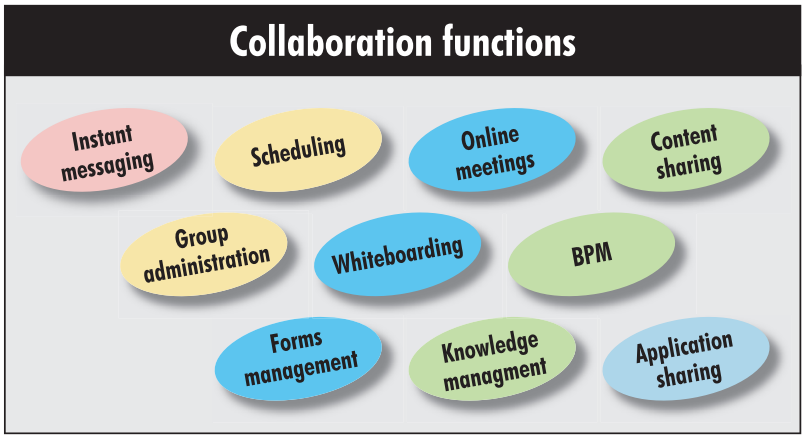


Fig. 3.6 *Functions provided by collaboration software*

3.2.7 Collaboration software

In addition to needing tools to manage and search for structured data in databases and documentary information in EDM systems, organisations began to identify a need in the mid 1990s for electronic tools to support co-operative working. Users did not simply need to access information in electronic filing cabinets; they needed tools to help them collaborate to achieve their goals (Fig. 3.6).

A range of tools began to emerge to support collaborative working. Different tools were developed to support unstructured, ad hoc processes on the one hand, and high-volume, rules-based structured processes on the other. The simplest and most flexible tool to aid collaboration was email. On top of email we saw the development of groupware software which was epitomised by Lotus Notes in the 1990s and which spawned many followers. Over time, groupware has given way to collaboration suites which take full advantage of developments in the ICT infrastructure (Fig. 3.7). The most widely used collaborative software platform is now Microsoft's SharePoint. This was launched in 2001 and has progressively developed and expanded over the years to provide most, if not all, of the elements required in an ECM suite.

The late 1980s and early 1990s saw the launch of a range of workflow management (WFM) software that allowed large organisations to redesign and automate their business processes, to manage change and ensure that their processes better met business objectives. When linked with DIP and EDM software, WFM software led to the complete redesign of functions such as claims processing and loan processing.

3.2.8 Electronic publishing and content management

Almost as soon as word processors were installed in offices, the suppliers went on to develop and market electronic publishing software to both commercial and corporate publishers. The initial benefit of electronic publishing to print publishers was that publishers could take electronic content, mark it up and then render it to create masters for conventional print publishing on the one hand, and electronic publishing on the other.

The key to the widespread use of electronic publishing was standards and the availability of new, cost-effective, digital publishing channels. Initially, all the major suppliers had their own markup languages. In 1969 IBM developed their Generalised Markup Language (GML) and then the International Standard (ISO 8879) called Standard Generalised Mark up Language (SGML) was developed which added many features and options[3]. SGML was a powerful and intricate standard and became the de facto standard for the interchange of large, complex documents such as airline maintenance manuals.

The advent of the Internet and the World Wide Web saw the arrival of new standards including Hypertext Markup Language (HTML) that was based on a simple example document type in the SGML standard. The simplicity of HTML allowed programmers worldwide to quickly build systems and tools to work with the Web. Most of HTML's element types are generalised and descriptive but it only uses a fixed set of element types. HTML is not extensible and hence cannot be tailored for particular document types.

According to Charles Goldfarb[3], as the Web grew in popularity developers found HTML's fixed document type a limitation. Browser vendors started to make incompatible extensions to HTML and the great strength of the Web—interoperability and standardisation—was under threat. The World Wide Web Consortium (W3C) invented a simple HTML oriented style sheet language called Cascading Style Sheets (CSS) and decided to develop a subset of SGML that would retain the power and flexibility of SGML but also remain as simple as possible in line with the general philosophy of the Web. The new standard was named Extensible Mark up Language (XML) and they also developed standards for advanced hyperlinking and style sheets—Extensible Linking Language (Xlink) and the Extensible Style Language. XML was specifically designed to facilitate reliable interchange on the Web. This included the interchange of data and documents but also paved

Fig. 3.7 *Collaboration software helps people to work together as a team*

the way for electronic business with computers communicating with each other across the Web via XML.

The arrival of electronic publishing created a specialised demand for content management systems rather than electronic document management systems. Content management (CM) was the term used to differentiate systems that managed at the content or component level from systems that managed at the document level. CM systems were designed to help organisations exploit, reuse and publish their content components. A CM solution was designed to keep the content separate from the delivery mechanism or the presentation format.

With a DIP or simple EDM system you have no choice about how you view the document: you see either an image of a page in an image viewer or you see the native file presented by the application used to create it. With a CM system the content can be marked up using SGML or XML and then the user defines, using a style sheet, how they want that content to be rendered on one or many delivery media or channels (Fig. 3.8).

Hence a CM system treats content more like data in a database. The content can be held in a neutral format and marked up so it can be reused as an asset in a number of different ways. A CM system manages semi-structured information to create a reusable resource that enables an organisation to fully exploit that content, just like a database enables them to fully exploit their structured data.

One of the key features supported by CM systems is assembly: the ability to assemble a document from all its component content objects or components and render it on screen or on paper or however it is required to be rendered in future. For many years CM was a relatively specialised marketplace supporting large commercial or corporate publishers. The advent of the Internet and the World Wide Web widened the market for CM, or at least for an important subset of CM called Web content management.

3.2.9 Web content management

In the 1990s interest in the Internet and the World Wide Web grew rapidly and the first generation of Web applications (Web 1.0) took us all another huge step further towards enterprise content management. The application with the greatest impact for ECM was the World Wide Web, a document-based hypertext system which operates as an overlay to the Internet. It protects users from the need to know where information is physically stored and allows authors and users to set up, manage and navigate information bases or websites that span different locations, resources and suppliers.

The World Wide Web originally comprised a range of software tools[4] including (i) web server software to delivers web pages to users, (ii) HTML (Hypertext Markup Language) to describe the structure of documents (pages) to be delivered via the Web and to define URLs which are used to point to other resources on the Web, (iii) the browser which is the client program used to view information on the Web and was made available for almost any environment and (iv) HTML editors to make the process of creating HTML pages as easy as desktop publishing.

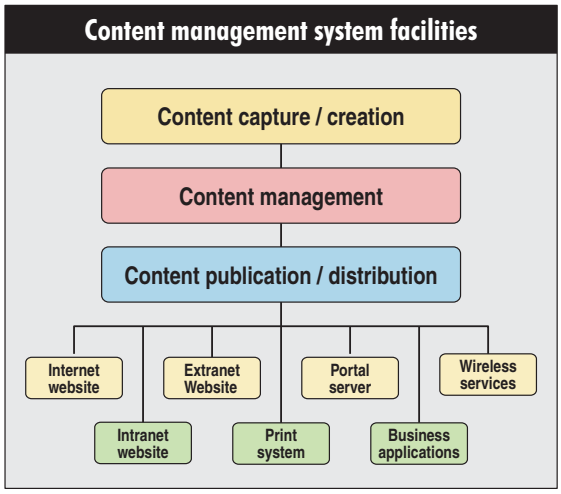


Fig. 3.8
Content management system facilities

Internet websites grew at a rapid rate and we saw the introduction of corporate intranets and extranets. The power of the Web was obvious, but while it was easy to set up websites and load documents onto them it was less easy to keep the material up to date and maintain all the links as documents were changed or deleted. It was difficult to monitor the quality or consistency of the material published on a website and users got bored with disorganised websites.

This created a demand which was initially met by a specialised type of content management system designed to manage the content to be published on websites and, not surprisingly, this software came to be known as Web content management (WCM) software.

To support WCM you needed to deal at the content and at the document level and to capture existing content, transform existing content and support the creation of new content specifically for the Web.

A WCM system had to support the development and management of websites. Organisations needed facilities for the development of content-based web applications, supporting both code and content elements. They also needed to create web pages that were reusable, and index and manage them in the repository alongside the content. One of the key functions of a WCM system was to check all the links between pages and content and hence guarantee the integrity of all the links within a specific website.

WCM systems supported workflow and processes designed to pass content from authors to editors where it is reformatted or transformed ready for publishing and where it is reviewed and approved for publishing on the Web. A core function supported by all WCM systems is web publishing or delivering approved content to the correct live web server and ensuring that the correct content appears on the right site in the correct format. WCM software also monitored the usage made of content and incorporated archiving software for pulling inactive content off a website.

The World Wide Web on the Internet was based around a standard browser that effectively provided a gateway into the world of knowledge stored on all the websites on the Internet. Over time, the raw search engines have developed into public portals— literally software designed to act as a gateway to the world of knowledge held on the Web. Companies such as Google and Yahoo provided personalisation features that enable users to configure a personal page to contain selected links to sites, information and applications.

WCM was ideal for managing large editorial websites which involved organisations and commercial publishers creating, organising and publishing their content to a defined audience in an organisation via the intranet or worldwide via the Internet.

The main model for these Web 1.0 applications was the publishing model where organisations could publish a subset of their content on the Web and consumers, other organisations and customers could access this vast amount of content via their standard browser on the desktop. While e-commerce applications were supported the bulk of consumers were cast as passive consumers of content. This is something that has now changed with the advent of Web 2.0 applications (see below).

3.2.10 Electronic records management

In the 1980s many large organisations managed their paper records using records management software and they had a print-to-paper policy whereby all documents deemed to be records had to be printed out and stored in paper folders. From the early 1990s a growing percentage of documents were created electronically and the paper-based RM regimes began to fail. Users were reluctant to print out a copy of all their electronic documents and, increasingly, incoming paper documents were held in paper folders and outgoing digital documents were held in directories on magnetic disk drives.

Records managers looked at the new DIP and EDM systems but there were very few facilities provided for corporate classification, declaration and disposition. A number of the suppliers of RM software for controlling paper records enhanced their systems to manage electronic records as well. In addition, some new suppliers focused on the electronic records management (ERM) market. But there were no standards for best practice in ERM.

In the UK, The National Archives (TNA) initiated a project to identify functional requirements for ERM in the UK government context and to evaluate available systems against those requirements. The first version of the requirements[5] was published in 1999 and identified three core functional areas for ERM systems:

- **Declaration**—the ability to capture the document(s) that make up the record and freeze its content so that it cannot be changed thereafter.
- **Classification and organisation**—support for the structuring and categorisation of records so that all relevant records are brought together at the same point in a corporate fileplan or taxonomy for consistent access and lifecycle management. A fileplan defines a hierarchy of classes, folders, folder parts, documents/records and components (content).
- **Disposition**—facilities to support the maintenance of sustainable records over time, to retain only those that should be kept and to manage the controlled disposal of those no longer needed.

Those organisations that did not want active document management facilities could purchase and implement a standalone ERM system. Where companies had already implemented an active EDM system, a growing number of EDM suppliers integrated their products with ERM software to provide integrated document and records management solutions which were referred to as electronic document and records management (EDRM) solutions.

Following the work by TNA, the European Union commissioned and published a more-detailed set of guidelines and requirements called MoReq[6] to assist users specifying their requirements for an ERM system. The arrival of these documents, together with a drive by UK government to persuade public-sector bodies to manage their new records in electronic format, led to a rapid growth in the market for ERM and integrated EDRM systems. In 2002 The National Archives issued a new version of its requirements document[7] with more-detailed metadata standards. More recently, the European Union supported the development of an updated version of the MoReq standard—MoReq2 in 2008 and now MoReq 2010 which is described in more detail in section (2.4.6) above.

3.2.11 Enterprise content management

After nearly thirty years of evolution we arrived at a situation in 2003 when organisations identified the need for a suite or framework of software designed to meet their content, document and records management and collaboration requirements. The suppliers responded with a series of acquisitions and mergers and new software development that has allowed many of them to develop integrated solutions. Today, there are broadly three types of suppliers serving this market: (i) suppliers of point solutions, (ii) suppliers of EDRM suites that interface with collaborative software, and (iii) suppliers of ECM frameworks (Fig. 3.9).

- **Point Solutions** Suppliers in the first category offer one tool and are targeting small to medium-sized customers who need a solution to a specific problem. They will typically offer an archive solution, an electronic document management solution or a web content management (WCM) solution.
- **EDRM Suites** Suppliers in the second category comprise those who have combined document cap-

ture, document management, records management and often business process management (BPM) functions to create an electronic document and records management (EDRM) suite. Such a suite meets all the document and records management requirements of a large organisation that needs to set up a tightly-controlled records management regime for regulatory compliance. They can usually also be integrated with one or more collaboration suites and web content management (WCM) software. Most will integrate with Microsoft SharePoint 2010.

- **ECM Framework** Suppliers in the third category comprise those who have combined an even wider range of software to create what they refer to as an enterprise content management (ECM) framework. A full ECM framework includes document, content, records, web content and knowledge management and the full range of collaboration tools including BPM tools.

Fig. 3.9
The types of suppliers

3.2.12 Social networking and content management on the Cloud

Today, some fifteen to twenty years on from the start of the Web 1.0 era of Web content management, a new wave of Internet applications is growing rapidly that support a second, fundamentally-different model of content management on the Web; a collaborative or social networking model where most, or at least a significant percentage, of the content to be managed is user-generated rather than editorial, and where one of the key objectives is to support collaboration and reuse of the content. This new generation of web software came under the banner of Web 2.0. Tim O'Reilly provided one of the clearest explanations of what Web 2.0 was[8] using the following headings:

- The Web as platform
- Harnessing collective intelligence
- Data is the next Intel Inside
- End of the software release cycle
- Lightweight programming models
- Software above the level of a single device

For those of us in the document and records management world, Steve Bailey[9] produced a useful and thought-provoking guide to the Web 2.0 world and the challenges it poses for corporate information and records management. He referred to a useful attempt by van Harmelen[10] to classify Web 2.0 applications into seven distinct types:

- blogs
- wikis
- social bookmarking
- media sharing services
- social networking systems
- collaborative editing tools
- syndication and notification technologies

Many of these applications currently are aimed at individual consumers and provide them with the opportunity to share content and use services that will manage their content for them. In addition to the passive publishing of content on websites the new applications support the reuse of the content and provide facilities for the content to be reviewed, tagged and linked to other relevant content.

The take up of Web 2.0 applications has been dramatic with millions of blogs, widespread use of wikis and high levels of usage of applications such as Flickr—for the storage of digital photographs—YouTube for the storage of video clips and Facebook. For these users the Web is their IT infrastructure and all they need to access it is a desktop running a browser.

This concept of the Web as a global content management system has always been true but, with Web 1.0, it was a global publishing system and most organisations still created content using office applications and stored the content on their inhouse systems before opting to publish a subset of that content to their websites. Web 2.0 applications have the potential to manage content throughout its life on the Web including collaborative creation, review, publication, review, revision, archiving, etc.

For individuals this trend is unstoppable and the benefits clearly outweigh any issues. For organisations the opportunities and challenges posed by some of these Web 2.0 applications have yet to be fully understood. We will look at blogs, wikis and collaborative editing tools in chapter 4. In this historical section all we can do is look at two issues that Web 2.0 applications are already raising for corporate information and records managers.

3.2.12.1 Personal versus corporate information

The first issue is that Web 2.0 applications will further blur the boundaries between when an individual is working for their organisation and when they are simply being an individual. Staff use social networking sites to access content and share their content. If that is their personal content then that is their decision, but what if the content was created in work time and hence could be considered to be owned by their employer? Can staff just give valuable corporate information away in this way? What happens if they store the content on a third-party website and the third party deletes it? Clearly, as well as offering great

potential this trend also creates risks and records management and intellectual property rights issues. Organisations will need to develop policies and procedures that cover this.

3.2.12.2 Information and records management on the Cloud

The second issue is more fundamental for this guide. Traditionally, all our organisations have used commercial email packages and office suites and our IT departments have provided storage facilities where all the emails created and received are held, and where all the office documents created are stored. When we implemented an EDRM or ECM solution we were typically just adding a layer of software that allowed us to hold more index data or metadata about the content and to classify it and define who can access it, etc. But we were still keeping the creation applications, the metadata and the content on servers running on an in-house network.

This model was first challenged by the growing trend to outsource IT functions. Many organisations found it difficult and costly to provide their own IT infrastructure and to respond quickly and flexibly to changing user IT requirements. The answer for some was to outsource the provision of IT services. In some cases this simply involved a third-party supplier taking over the IT staff and managing the organisation's IT infrastructure on site. Increasingly, however, the trend has been to then move the organisation's databases and applications onto centralised data centres run off-site by the third-party provider. So while the user still has a desktop connected to an in-house network the organisation's network is connected via a wide area network (WAN) to the third-party data centre and all the main servers are held there. This reduced the organisation's hardware and software budget but raised issues of intellectual property rights and service levels, etc. Negotiating a contract to outsource IT services like this can be complex but, provided all the core issues are considered and resolved up front, the model has been shown to work for many organisations in the private and public sector. The recession and cut backs in the public sector have accelerated this trend in both the public and private sector.

The increased use of Web services is now ushering in the era of the 'Cloud' where organisations can access computing power and applications on demand via their browsers. To the organisation it can be very similar to the outsourced option. Users have their desktops on the in-house network which is then connected via the Internet to servers which hold their applications and data including content and records. Google and Amazon were the first suppliers to promote the Cloud. Users of Google Applications can access office software and email and store all their content on the Cloud. In response to this Microsoft is offering users email and office applications and a version of SharePoint on the Cloud.

Such services in their basic form can be free to individuals and paid for on-demand by organisations. In a time of recession the potential attractions of an era of on-demand computing via the Cloud and the widespread use of shared services are obvious. In addition, the data and content can be accessed by any authorised user whether they are working on the organisation's premises or working remotely from home, a hotel room, etc. Lower costs and improved access are the key selling points.

However, when it comes to holding your organisations valuable information and records on the Cloud there are also some serious issues that need to be addressed, including performance service levels, reliability and intellectual property rights. What level of availability does the service provider guarantee? What level of security is provided to avoid the risks of other users gaining access to your content? What happens if the service provider goes out of business or decides to delete your content after a set period of time? None of these issues is insurmountable but, clearly, the more service levels and security requirements are taken into account so the cost of the service is likely to increase.

The early offerings on the Cloud tend to be low-cost or free, standard offerings. If you sign up to them you are using a standard service that is proven and low cost but there is little scope for customising or tailoring the service to meet your specific requirements.

Popular applications include shared storage, office applications and email. On top of these mass market applications we are also seeing the development of a wider range of services on the Cloud that can provide large companies or public-sector consortia with shared spaces that can be made more secure and offer scope for much greater customisation. The potential here is endless and certainly one of the applications of interest would be shared services including information and records management. Welcome to the decade of information and records management on the Cloud!

3.3 Reviewing the Solution Options for Information and Records Management

3.3.1 Background

In chapter 2 we have looked at the case for improving your corporate information and records management; what is involved in developing a policy and the best practice guidance available to help you review where you are and where you could make improvements. In this chapter so far we have given some definitions and charted the historical development of solutions that enable organisations to manage their information and records more effectively.

In chapter 4 we review the functions that together make up an enterprise content management solution. In chapter 5 we look at a methodology for managing an information and records management project, and in chapter 6 we look at how you can make the business case for investment in an information and records management project.

Before you start planning a project and making a business case you need to take account of where your organisation is with information and records management and consider which solution options are likely to be most appropriate for your organisation.

You will not be able to make a final decision about which option best meets your needs until after you have carried out your information gathering exercise and defined your core objectives—in other words when you have your project set up. However, it is useful to have an overview of the options available to you before you start as it can help you scope your project. If, like many organisations currently, you have significant budget constraints, or your organisation is already committed to a particular solution, e.g. SharePoint, or has already outsourced the IT function, etc., then you can rule out a number of options and focus on the one or two remaining options for you organisation.

The methodology outlined in chapter 5 defines ten stages for an information and records management project:

1. positioning ECM and EDRM systems
2. defining and managing your project
3. information gathering and analysis
4. feasibility study and options review
5. making the business case for the preferred approach
6. defining the statement of requirements (SOR)
7. procuring the solution
8. managing the implementation
9. measuring the results
10. project closure and solution support

In all cases we recommend that you follow stages 1–4 irrespective of which options you have ruled in or out of scope. Even if you do not have a budget for a new system, and hence are going to use existing tools, you should follow Stage 1 and review what functions such systems provide and then consider whether you need them and, if so, how you can meet the requirement with existing tools.

However, depending on which options you have ruled in or out you only need to consider the remaining options in Stage 4.

If you opt to use existing tools, or a solution that has already been procured, then you may not need to go through a formal procurement process but you should still define your requirements in Stage 6 and define the in-house or third-party resources needed to implement the solution so that you can make the business case and ensure your project is adequately resourced.

There are many ways in which the range of solution options available to you today can be divided up. At a high level we have used three options:

- use existing tools
- implement an in-house ECM solution
- outsource ECM

The first option means that procuring a new solution is out of scope of your project. Your project must focus on improving information and records management policy and procedures using existing tools. This option is reviewed briefly in section 3.3.2 below and is the subject of a one-day Cimtech course.

The second option means that procuring a new solution and implementing it in-house is within the scope of the project. Within this option we make a distinction between Microsoft SharePoint and other ECM solutions. This option is reviewed briefly in section 3.3.3 below.

The third option means that using a new solution is within the scope of the project but, either the organisation has outsourced all its IT applications already, in which case an ECM solution will also be outsourced, or the organisation wishes to outsource ECM specifically. This option is reviewed briefly in section 3.3.4 below.

3.3.2 Information and Records Management Using Existing Tools

This option is adopted by users who either do not have a budget to procure a new system or who are not currently convinced of the need for a new system, i.e. they think they can achieve the improvements they are looking for without investing in a new solution. It can also be adopted as an interim strategy by users who want to develop information and records management tools first and improve the organisation of existing information and records, and then move forward to procure a solution as a separate project.

Existing tools usually includes one or a number of the following:

- shared network drives and operating software
- email systems
- business administration systems that support content and document management
- intranets
- paper storage

For many users it involves looking closely at how information is currently stored. Digital information may be stored on personal or shared drives, in databases, on websites and in email inboxes. Analogue information may be stored on paper in folders held locally or in registries, or in a range of other formats including microfilm, slides, photographs, video tape, etc.

Most clients Cimtech has worked with will focus on improving how electronic documents are stored on shared network drives, ensuring that no valuable information is held unprotected on personal drives that are not backed up, and imposing a structure on shared drives. Key factors include file naming conventions, preventing records from being overwritten, implementing a classification scheme and retention schedule, handling the migration of existing content, controlling the creation of new folders and controlling deletion.

Many users will also focus on existing paper records centres, cross referencing folders to the classification scheme and integrating them with the retention schedule.

Overall, organisations can achieve considerable improvements using existing tools, but if full regulatory compliance is called for then using existing tools can be a very labour-intensive solution. Cimtech recommends this approach to organisations who wish to improve their information and records management and start to migrate their content into a classification scheme in preparation for the implementation of an ECM solution in the future. Cimtech runs a one-day, introductory course specifically devoted to *Information and Records Management Using Existing Tools*[11].

3.3.3 In-house enterprise content management solutions

3.3.3.1 ECM solutions

If organisations decide that they need a new in-house solution to manage their content and documents as part of an information and records management project then they need to define their requirements, make a business case, go through a procurement process and decide how much support they will need from the software provider and/or a third party. This is the second option to be considered.

As outlined in section 3.2, above, the market for document and content management solutions has evolved over the last twenty years to the current position where there are, broadly speaking, three types of system on the market—point solutions, electronic document and records management (EDRM) solutions, and enterprise content management (ECM) solutions.

The way to decide between these three categories is to define your requirements in advance. Broadly speaking, if you want to use the solution just in one application or department you may select a point solution. If your prime concern is web content management then you could select a solution dedicated to WCM only. If you are looking to manage your documents across the whole organisation and to define a subset as records and manage them then you could look at an EDRM solution. If you want to manage documents and records and all your web content with one solution, and you want to support collaborative working as well, then you will need a full ECM solution.

The full range of functions offered by these systems are reviewed in chapter 4. The process you need to go through to gather information and define your requirements and select a solution is reviewed in chapter 5.

Details of the full range of solutions and their suppliers are contained in the relevant sections of *The Cimtech Directory* which is freely available online.

3.3.3.2 Microsoft SharePoint

Microsoft SharePoint is the best-know ECM solution on the market. We deal with it separately here for two reasons.

The first reason is that many organisations will have access to at least some of the SharePoint software through their corporate Microsoft licenses. Hence, in one respect, it can be seen to be an existing software tool and so organisations may not need to go through a competitive procurement process to acquire additional SharePoint software when needed. Depending on how the organisation choose to use SharePoint this makes it a slightly different option. For many organisations SharePoint will be a given.

The point that Cimtech would make here, however, is that if a medium-to-large organisation is looking to use SharePoint as their corporate ECM solution, and they wish to implement a tight set of records management procedures consistently across the organisation, then, in addition to the software, they will need support to configure, develop and integrate the solution over the long term. This would be true with all ECM platforms on the market. Hence, although they may not need to go to a competitive procurement to acquire the software, they should consider going out to procurement to a SharePoint developer to support them in developing, configuring and implementing the solution. They will also need to define their requirements and make the business case for the project. A list of SharePoint services providers is included in *The Cimtech Directory*.

The second reason is that SharePoint is more than just an ECM solution. The latest version, SharePoint 2010, is subtitled, 'the business collaboration platform for the enterprise and the Web.' Microsoft splits SharePoint functionality into six areas:

- sites—web content management
- communities—social networking
- content—ECM
- search—text retrieval, knowledge management
- insights—business intelligence
- composites—rapid application development, mashups, etc.

Many organisations defined SharePoint as their collaborative environment before they started to look in detail at ECM. Significantly, almost all of the other leading suppliers of ECM solutions have either adopted SharePoint as their preferred collaborative tool or have at least integrated SharePoint with their ECM platform.

SharePoint has gone through a number of versions over the past nine years culminating in SharePoint 2010. Early versions of SharePoint did not support a full set of records management capabilities and there were some scalability issues. SharePoint 2010 has addressed many of these issues.

Today, most small-to-medium size organisations and many large organisations will have selected SharePoint as their ECM platform. Many others will have selected a different ECM platform but will have integrated SharePoint with that ECM solution. We look at the options in more detail in chapters 4 and 5.

If you are in the public sector and are considering the use of SharePoint 2010 for information and records management then The National Archives has pub-

lished a useful guide entitled *Records management in SharePoint 2010: implications and issues*[12].

For all organisations that have implemented, or are planning to implement, SharePoint 2007 or 2010 and plan to use it for document and records management then Cimtech offers a one-day course *Information Architecture for SharePoint Document and Records Management*[13].

3.3.4 Outsourced enterprise content management

The third, high-level option is that your organisation opts for an ECM solution (including SharePoint) but opts not to implement an in-house solution but, rather, to go for the outsourced approach.

As described in section 3.2, above, the outsourced option covers a number of sub options.

On the one hand, your organisation may have already outsourced the management of its IT infrastructure, and/or its corporate IT applications, to a third-party supplier. If this is the case then your organisation will simply need to define its business requirements for an ECM solution and the third-party supplier will specify a solution and add it to the range of managed applications. This will simplify the procurement process compared to the standard competitive procurement approach, but be aware that it might limit significantly the choice of solution to those which the third-party supplier has experience of supporting.

On the other hand, your organisation could be looking to trial the outsourced approach and decides to procure ECM capabilities as an outsourced or shared service.

There are a number of elements of an ECM solution that are good candidates for outsourcing even if you decide to implement the software in-house. These include scanning the post in the post room, scanning any back files that you wish to convert into digital images, and indexing and managing legacy paper files which you do not wish to scan. Service bureaux offer scanning and indexing services and there are a number of companies that offer secure, off-site storage of paper records and a range of retrieval options. See The Cimtech Directory for listings of both.

References

1. Management, appraisal and preservation of electronic records. Public Record Office (The National Archives), 2nd edition, 1999.
2. Hendley, A. M. Document image processing—a technical introduction. Cimtech/NCC/The National Centre for Information Technology, 1999.
3. Goldfarb, Charles F. and Prescod, Paul. XML Handbook. 4th edition, 2002, Prentice-Hall. ISBN 0-13-065198-2. http://bit.ly/AnNH2Y
4. Carroll, Ted. Document Management and the Internet. Document Management Guide and Directory, 8th Edition, 1997, Cimtech. ISBN 0-00458-77-1.
5. ERM functional requirements document. Public Record Office, 1999.
6. Model requirements for the management of electronic records (MoReq). ISBN 92-894-1290-9, DLM Forum. http://bit.ly/wXb9DA.
7. Requirements for electronic records management systems. Part 1: functional requirements. (2002 revision). The National Archives. http://bit.ly/xCOGD
8. O'Reilly, Tim. What is Web 2.0. Design patterns and business models for the next generation of software. http://oreil.ly/aa1sP.
9. Bailey, Steve. Managing the Crowd: Rethinking records management for the Web 2.0 world. Facet Publishing, 2008. ISBN 978-1-85604-641-1. http://bit.ly/ABKnPz
10. van Harmelen, M. Briefing Paper on Web 2.0 Technologies for Content Sharing: Web 2.0 An Introduction. http://bit.ly/ADo5gr.
11. Information and Records Management Using Existing Tools. One-day course, Cimtech. http://bit.ly/kELACE.
12. Records management in SharePoint 2010:-Implications and Issues. http://bit.ly/yb5ot5.
13. Information Architecture for SharePoint Document and Records Management. One-day course, Cimtech. http://bit.ly/yvbmBN.