Content Management

Managing the Lifecycle of Information

AIIM Industry White Paper on Records, Document and Enterprise Content Management for the Public Sector

Content Management

Managing the Lifecycle of Information

AIIM Industry White Paper on Records, Document and Enterprise Content Management for the Public Sector

AIIM International Europe
Chappell House
The Green, Datchet
Berkshire SL3 9EH - UK
Tel: +44 (0)1753 592 769
Fax: +44 (0)1753 592 770
europeinfo@aiim.org

DLM-Forum
Electronic Records
Scientific Committee Secretariat
European Commission SG.B.3
Office JECL 3/36, Rue de la Loi 200, B-1049 Brussels - Belgium
Tel. +32 (0)2 299 59 00 / +32 (0)2 295 67 21 / +32 (0)2 295 50 57
Fax +32 (0)2 296 10 95
A/e: dlm-forum@cec.eu.int

Authors
FileNET Europe North   FileNET Corporation
Royce Murphy (Ingrid O'Brien)   James DeFerrari
1 The Square, Stockley Park,   434 Exeter Street
Uxbridge, UB11 1FN - UK   Sacramento, CA 95864
Tel.: +44(0)20 8867 6341   Tel.:+1(0)916 761 0387
Euro.north@filenet.com   jdeferrari@filenet.com

Executive editors and coordinators
Dr. Ulrich Kampffmeyer
Silvia Kunze-Kirschner
PCI PROJECT CONSULT International Ltd.
Knyvett House, The Causeway
Staines, Middlesex TW18 3BA - UK
Tel.: +44 (0)1784 895 032
info@project-consult.com

Preface

The Information Society impacts in many different ways on the European citizen, the most visible being the provision of access to information services and applications using new digital technologies. Economic competitiveness of Europe's technology companies and the creation of new knowledge-rich job opportunities are key to the emergence of a true European digital economy. Equally, the Information Society must reinforce the core values of Europe's social and cultural heritage – supporting equality of access, social inclusion and cultural diversity. One important element in ensuring a sound balance between these economic and social imperatives is co-operation between the information and communication industries and public institutions and administrations.

Over the past 5 years, the European Commission in co-operation with EU Member States, has worked to create a multi-disciplinary platform for co-operation between technology providers and public institutions and administrations. The Forum aims at to make public administration more transparent, to better inform the citizen and to retain the collective memory of the Information Society. These objectives are at the heart of the eEurope Action Plan adopted by the European Summit in Feira on June 2000. I welcome the way the DLM-Forum has evolved over this period as a platform for identifying and promotion concrete solutions to many of the problems facing our public administrations.

In 1996 the initial focus of the DLM-Forum was on the guidelines for best practices for using electronic information and on dealing with machine-readable data and electronic documentation. More recently, at the last DLM-Forum in Brussels in 1999 a challenge was made to the ICT industries to assist public administrations in the EU Member States by providing proven and practical solutions in the field of electronic document and content management.

The importance of providing public access and long term preservation of electronic information is seen as a crucial requirement to preserve the "Memory of the Information Society" as well as improving business processes for more effective government. Solutions need to be developed that are, on the one hand, capable of adapting to rapid technological advances, while on the other hand guaranteeing both short and long term accessibility and the intelligent retrieval of the knowledge stored in document management and archival systems. Furthermore, training and educational programmes on understanding the technologies and standards used, as well as the identification of best practice examples, need to be addressed. I welcome the positive response from the ICT industries to these challenges and their active involvement in the future of the DLM-Forum, for example in the event proposed in Barcelona in May 2002, to coincide with the EU Spanish Presidency.

The information contained in the following pages is one of a series of six ICT Industry White Papers produced by leading industry suppliers, covering the critical areas that need to be addressed to achieve more effective electronic document, records and content management. I am sure that the reader will find this information both relevant and valuable, both as a professional and as a European citizen.

Erkki Liikanen
Member of the Commission for Enterprise and Information Society

Preface Sponsor

Worldwide, governments and industry continue to connect to their constituents in a way never before possible. Content management is enabling this transformation.

For many, Content management (CM) is seen as the essential strategic IT investment of this millennium, and the public sector is leading the charge. This is clearly evidenced in the DLM Message to the ICT Industry that calls for easily applicable and cost-effective records management and digital archival solutions.

Our information society demands that authentic and pertinent information be easily accessible and securely retained. The challenges ahead are significant, as outlined in the ICT Industry's Answer to the DLM Message. Many can be addressed through content management. The future of government depends upon it.

Lee Roberts
CEO
FileNET Corporation

Table of Content

1.   Introduction

There is a new call to arms mobilising government IT decision makers around the world – namely the priority of e-Government. Leaders in the public sector are promoting the new capabilities that e-Government promises and are committing to become more efficient and more responsive to the constituents they serve.

First, however, there are many questions that need to be answered. How can public agencies manage and process more information electronically via the Web? How can they share content seamlessly with the public, other agencies and their employees? How can back office and front office activities be integrated or, better yet, automated? What are the capabilities needed to realise this transformation in government? The answer to many of these questions lies in content management.

Content management is emerging as a key enabler facilitating this transformation in government. Content management solutions are meeting the immediate needs of today's emerging e-Government initiatives while helping departments and agencies operate with never-before-seen speed and efficiency. The technology will serve as a critical component for better government in the future by providing numerous electronic touchpoints for their citizens.

In this paper, we:

- Define content management and the various technologies it embraces
- Examine the differences between several content management architectures and the different types of solutions being deployed today
- Explain the different functionalities included in content management solutions
- Outline the relevant standardisation bodies, definitions and technologies
- Showcase Best Practice applications, featuring examples from both the private and public sector, and
- Forecast the future of content management and identify possible trends and developments.

It is our hope that this paper will serve as a roadmap for implementing content management and help you begin the next step of your journey towards better government.

John Mancini
AIIM International

2.   From Archival to Enterprise Content Management

Enterprise Content Management (ECM) evolved from previous market terms, including Integrated Document Management (IDM) and Web Content Management (WCM). They evolved as the market demands changed and new capabilities emerged. This section addresses records management, content management, the implications of using the Internet for knowledge management (KM) archival, the challenge of distributed and shared solutions, and the potential costs of implementing – or not implementing – CM, as well as selected legal aspects.

2.1   Getting a Handle on Managing Information

Information is the bottom line in government, as it is in commerce. Content management is the strategic determination of how information is created, used, achieved and, ultimately, disposed of. Managing unstructured content is the major hurdle to moving business and government online. Organisations that understand information management is strategic and mission critical will ensure their competitive advantage. Paper forms still dominate. The first step for e-Government is to push documents and paperwork into the electronic era. Managing content electronically will increase productivity and operational efficiency; after which, organisations will be able to manage electronically the critical processes which depend on the content. Only then will the benefits of e-Government be realised.

2.2   Records Management

Records Management includes the technologies used to create, capture, customise, integrate, deliver and manage records (content that has been transitioned from author-control to its "official" corporate-controlled records phase) to support business processes. An organisation's records can represent either strategic corporate information assets or, if not properly managed, liabilities. Trends that will influence records management and CM include: globalisation, organisational change, digital commerce, and an increased focus on privacy issues. And now, as a result of September 11th 2001, many organisations have begun to examine how they protect their records, especially those vital to their very survival. The cost of protecting or rebuilding these records must be determined in comparison to their value to the organisation.

Tools for maintaining and using records include: file plans, indexes, controlled vocabularies, taxonomies, data dictionaries, and access and security procedures. The main tool used to manage the disposition of records is the retention schedule. More detail is available in a European specification for electronic records management, the MoReq specification, which was produced in response to a DLM Forum initiative (see chapter 5).

2.3   Archival

A crucial part of any records management program is the definition of the time periods for which an organisation's records are archived and maintained, and the procedures for their transfer and disposition. The benefits of archival functionality within records management are many. It reduces space requirements for paper and electronic records, allowing organisations to restructure storage space for other purposes. It also improves operational efficiency by providing access to pertinent records in the best way, while transferring inactive records. By reserving prime space for active records, there are equipment and supply cost savings. In turn, worker effectiveness increases, as well as service levels through increased collaboration between department and agencies, and – in some cases – productivity rises. Customer satisfaction also increases because information can be retrieved faster, and more completely.

2.4   Content Management

Content management brings together several related technologies that enable organisations to link documents, records and fragments with the business processes that govern their use and lifecycles. CM can enable local and national governments to make information available to citizens, across multiple touchpoints, without extensive reformatting, thus meeting increasing requirements across Europe and North America for e-Government. Some examples are given in Chapter 6.

Content managers face the challenge of managing a wide variety of document formats, including: compound documents (linked and embedded), expanding formats (pdf, rtf, wav, video), and XML and XHTML documents from Web sites. They must also determine the relevance to business processes, and conformation with records and information management retention requirements. Managers are concerned about security measures (replication, redaction, public vs. non-public versions) as content moves to the Web. See Figure 1: European CM Market for a prediction of the growth of the European CM market to 2003.

The CM Market in Europe 1999 – 2003 – €m
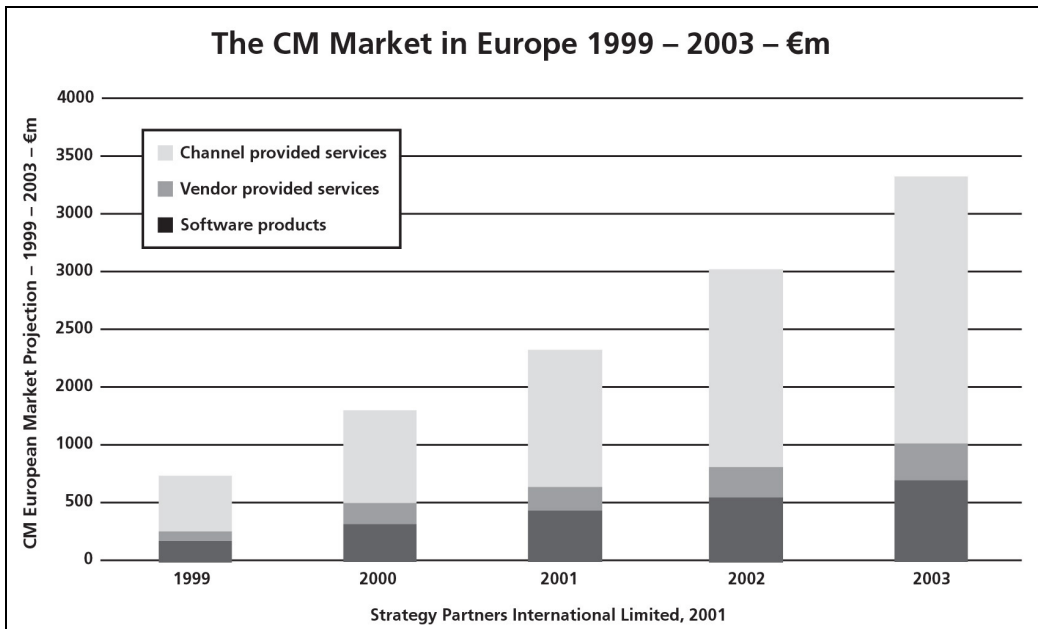
Figure 1: European CM Market

2.5　Knowledge Management

When it comes to knowledge management (KM), there are basically two truths:

1.) There is no universal definition of KM.
2.) No two people seem to be able to agree on just what constitutes knowledge.

One definition states: "knowledge management involves the identification and analysis of available and required knowledge assets and knowledge asset-related processes, and the subsequent planning and control of actions to develop both the assets and the processes so as to fulfil organisational objectives."

In the most general terms, KM is the set of processes that business and government use to engender value from their intellectual and knowledge-based assets. To do so, they must share this information with employees, departments, and other organisations, resulting in a set of best practices. And, while KM can be facilitated by technology, in itself technology is not KM. But technology must come to play in order to break through the paper-based logjam in which most information still resides.

Another thing to remember is that all information does not necessarily have intrinsic value. It is up to the organisation to determine the value of its information and knowledge-based assets. These might include: patents and trademarks, research, business plans, etc. Other, less explicit knowledge is harder to define. The real challenge here is to generate, organise, manage and share the less definable kind of information.

Just as difficult to define, is how KM can benefit an organisation. It's generally agreed that to get the most benefit from intellectual property, it must be shared and used as a collaborative tool.

2.6 Distributed Solutions

No organisation can excel at everything. It is no longer viable to attempt to depend solely on internal skills and knowledge assets. Every day, business solutions are shared and distributed to specialists who can focus on individual segments of an operation and contribute to an overall excellence that the individual organisation cannot.

Some sectors have been slow to adopt e-business practices. This may be partially due to the fact that earlier enterprise resource planning installations took years to complete, while others still struggle to integrate IT infrastructure as global mergers continue. Success will come as business and government learns to manage the knowledge they own and enhance it through a sharing with external knowledge resources. Increased productivity will result from the tools, techniques and services that enable them to accomplish this.

Widely distributed organisations – such as large government departments – have considered the architecture of a content management solution carefully. The benefits and disadvantages of each approach are numerous, and the decision complex.

Key advantages of a distributed solution include local administration – that is, bringing the administration close to the users; reduced reliance on wide area networks to obtain information; and scalability. On the other hand, disadvantages may include security, consistency and the overall management to obtain a single, corporate, view.

2.7 Implications of the Internet

The Internet and intranets have dramatically diminished the barriers of time and space, and increased the ability to share information among a multitude of users. In no other period of history have government organisations been able to share information to the extent they do now.

Government organisations and businesses are increasingly beginning to realise the importance of "knowing what they know," along with the ability to maximise the use of this knowledge. This information can be found on databases, knowledge bases, filing cabinets and in peoples' heads. Thanks to the Internet and intranets, all this information can be easily shared across the organisation and with other entities – save of course information that is in peoples' heads. But, in practice, one department often duplicates work of another department because of the difficulty of keeping track of, and using, the knowledge held by other departments. Content management can be used to reduce this unwelcome duplication of effort.

The Internet provides a solution to the problems of integrating many individual departments and agencies. While the benefits of ease of access and sharing documents are highly desirable, many organisations are still adopting a wait-and-see attitude, waiting until one of the early adopters has successfully implemented its CM functionalities. One very real hurdle to implementing KM on the Internet and intranets is convincing individuals that this move is beneficial to them, as well as the organisation. The major organisational challenge is to convince individuals to agree to share their expertise with others, except as part of their job. Having spent a career validating their

value to the organisation because of what they know, they may be reluctant to "give away" their expertise and, in their view, loose some or all of their value.

Use of the Internet in this way also brings technical challenges, not least of which is the difficulty of tracking what information is available through the Internet at what time. This can be especially difficult in the case of highly dynamic databases published on the Web. Government organisations, which have a basic requirement to keep records of their actions, need to find ways to achieve this.

2.8   Content Management Today

Philip Hunter (Information Officer at UKOLN) has expressed the view that content management systems are equivalent to the tools developed for the publishing revolution of the 15th century, in that they are tools that revolutionise both the assembly and delivery of information. This perspective neatly summarises the potential importance of CM for our future.

CM predecessors were proficient in managing such things as graphics and, more recently, Web page layouts. They were not, however, able to manage different forms of content, particularly unstructured content (images, documents, etc.). Departments and agencies were not able to efficiently integrate the vast amounts of documents, reports and forms that existed in a hard-copy format.

The CM market has evolved from the need to address today's economic objectives, the necessity for managing unstructured content, and the need to electronically manage complex processes. These capabilities are dependent on one another and one cannot happen without the other when addressing e-Government requirements.

As government departments and agencies adopt e-Government activities as a means to lower costs and improve services to constituents, CM will become an absolute necessity. CM will help many different departments and agencies meet their immediate objectives and redefine the way they conduct business and serve the public.

CM can have an immediate impact, and offers benefits that can be sustained over the long term, including: reducing storage costs by better management of paper-based information; streamlining processes central to operations; improve service levels; provide self-service options to other departments and faster access to the public; improve internal collaboration and speed up operations; and overcome complex issues through a transparent, easy-to-use government.

2.9   Enterprise-Wide Compatibility

The need for CM is not isolated to any one group within an organisation. CM must, therefore, be an open-enterprise platform that is capable of being integrated into any environment or application that the organisation determines is appropriate. For example, CRM and ERP are popular software applications that enhance customer service and optimise asset management. Both have their own content and process management tools, but typically neither is compatible with the other, let alone other applications that might be purchased or written in-house. Organisations that accept proprietary CM systems are forced to create silos of valuable content that cannot be readily shared for the benefit of everyone throughout an organisation.

Enterprise Content Management (ECM) is a new concept that emphasises a single open CM system built on industry standards, and enhanced with two-way connectors to other industry standard platforms and internal applications such as CRM, ERP, collaborative working applications, e-mail, portals etc. This is an organisation's best insurance policy for the investment that they are making in CM, and the only way to ensure that critical information and processes are available to be shared throughout the organisation, its customers, and suppliers.

2.10 Costs

It should be apparent from the variety of approaches already touched on that CM can be implemented in many ways, with more or less functionality, with greater or lesser scope, and so on. Further, the nature of a CM project will depend as much on the organisation's starting point as on its objectives. It is therefore not possible to provide a useful, single, cost metric such as "software cost per user". In any case, the total cost of a CM project may well include other costs that dwarf the software angle – costs such as changing ways of working, customising interfaces, and converting legacy data can be material.

While the costs can be substantial in some cases, it is worth considering also the costs of not going ahead with CM. Once implemented, CM solutions can lower costs, for example when dealing with commercial partners; and can increase effectiveness, for example when dealing with citizens. These points mean that the cost of CM should not be evaluated in isolation; it must be offset against the costs of not proceeding.

2.11 Legal Aspects

The legal ramifications of electronic content management are, in the general case, numerous and complex. The laws applicable are the same as those which apply to any other type of IT system; however, in content management applications it tends to be the case that many requirements apply simultaneously in a complex environment – and in extreme cases, across the boundaries of different jurisdictions. To simplify somewhat, the following areas need to be considered:

- Data protection: any application holding "personal data" must be designed and operated so as to conform to data protection legislation. The implications here are chiefly on the human procedures used to maintain and dispose of personal data; but there are also impacts on the technology used to secure the information.
- Freedom of information: the application may have to comply with legislation intended to allow citizens to make arbitrary enquiries. This can be a particular challenge in large distributed systems.

- Evidential weight: in many public sector contexts, there will be a need to ensure that information in a CM environment can confidently be submitted as evidence in litigation – that is, without an unacceptable risk that its completeness, accuracy or integrity can be successfully challenged. Where this is the case, appropriate steps need to be taken during design and operation to ensure sufficient controls are in place and that their correct operation can be demonstrated if needed.
- Industry regulation: some applications, for example those interfacing to regulated areas such as the nuclear industry, the defense sector or the financial sector, may also be subject to sectoral regulation or other controls.
- Security: some jurisdictions will impose further controls to ensure the security of data, and to reduce the risk of unauthorised access or "hacking".

The precise implications have to be worked out case by case, depending on the nature of the information, the nature of the application, the user base, and the legal environment.

3.   Content Management Architectures

Today, the volume of information in any large organisation is daunting. To be successful, an organisation must find ways to share information internally and with the outside world. This information takes many different forms: letters, memos, faxes, printed reports, e-mail, spreadsheets, drawings, blueprints, microfilm, Web content, and electronic forms, just to name a few.

With state-of-the-art technology and industry-standard design, content management can provide a secure, robust foundation for your document management solutions. CM is a major step forward in addressing an organisation's goals for universal access to information repositories and uniform administration across workgroups and departments throughout the organisation.

In addition to the seemingly limitless variety and change of industry standards and expectations (See Chapter 5), document management systems must be open and flexible enough to accommodate emerging technologies. Document management system architecture must not only bridge disparate data repositories, but also bridge file-centric PC operating systems still in use today to the newer object-oriented operating systems now becoming available.

CM systems face a number of challenges in their design. Content managers need to focus on key aspects of these systems, which typically include:

- Scalability of the overall system and software architecture
- Modular architecture
- Open-systems design
- Easy deployment and implementation
- Integration with automated business process capabilities (workflow software)
- Sophisticated security
- Maximisation of data/document/record security
- Complete lifecycle management
- Dual thick and thin client capabilities

CM systems should provide common interface services to all types of documents, shield users from the complexities of LANs, and centralise and simplify network administration. All desktop information becomes a secured and quantifiable asset—safe, yet available on demand to authorised users.

3.1   The Anatomy of Content Management

CM implies storage, management and business process integration with many different types of information and media. In order to do this CM must integrate imaging, electronic document management, and workflow technologies. Each of these components uses a database, and each may be managed independently in order to balance system distribution or performance requirements.

3.1.1   Imaging

Many organisations still function almost entirely in a paper-driven environment. Mission critical information is scattered throughout mountains of paper that is stored all over the organisation. Information is difficult or expensive to find and retrieve, or it is duplicated many times so that it is more readily accessible. Important information stored on paper is often forgotten, overlooked, or even lost. Scanning paper documents and storing the images in a CM system improves access to documents and substantially reduces the cost of storage and retrieval. Electronic storage can also enhance the privacy and security of information that has traditionally been stored on paper.

The four basic elements of an imaging system are:
* Capture: a scanning operation which converts hard copy documents into a digital (i.e. electronic) format.
* Indexing: allows users to identify scanned documents so that they can be retrieved at a later date. Indexing can be done in a variety of ways, including key entry, bar code, OCR and ICR.
* Storage: performed by one or more servers. These servers place the indexing information in a database, and keep the scanned images in an area commonly called a library or repository.
* Retrieval: accomplished by invoking the index information to locate the desired images. There is virtually no limit on the number of people who can view an image simultaneously.
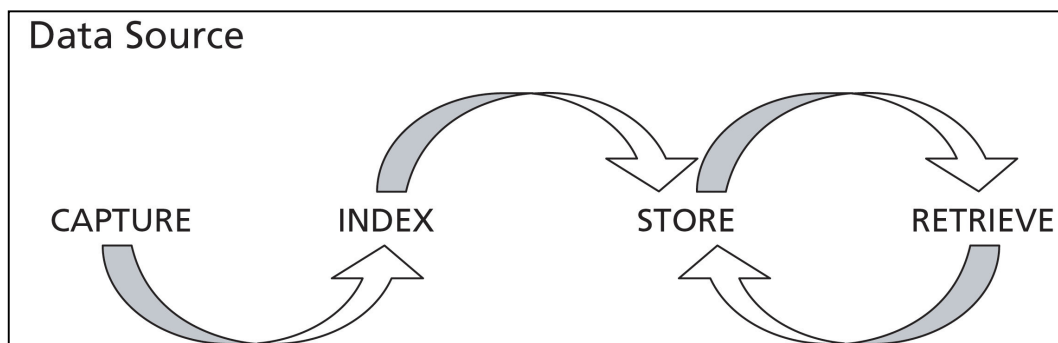
Figure 2: Imaging

3.1.2   Electronic Document Management

Ever since the PC was introduced, electronic documents have proliferated at a dizzying pace. Spreadsheets, word processing documents, electronic forms, and e-mails threaten to outnumber paper documents; they are becoming a records management nightmare. Storing electronic documents in a content management system improves ease of access by making documents easier to find and by reducing the need for proprietary software (Microsoft Word®, WordPerfect®, etc.) on a viewer's desktop. Proprietary software is not necessary because the system handles the display of hundreds of the most common

document types. CM also improves the utility of electronic documents by enforcing version control, automating Web content management, and enabling electronic publishing through renditions to other formats, such as PDF, and the creation of compound or related documents.

Web-based document management systems enable applets, or thin-clients, to check information in/out of the repository, allowing users to make changes and check the modified information back into the repository. The browser client becomes a combination reader and editor for all types of information. This has been achieved through the use of applets that can be downloaded when required or requested, or can be pre-loaded on the user workstations. These applets can be further managed from a centralised point, enabling system administrators to control not only which functions are available, but also which users have access to those functions. With these advances in technology, information can be tightly tied to business processes and collaborative information creation and management is greatly facilitated in a secure and controlled environment. The products that specifically exploit the dynamic nature of Web content and transform documents into entities capable of customised one-to-one communication provide the highest level of Internet-related functionality.

3.1.3   Workflow

Workflow is probably the most important of all of the elements of CM because automating business process offers the greatest return on investment. According to Jack Welch, former CEO of General Electric: "Technology is simply a vehicle for carrying out processes. The power of your organisation is contained in the processes themselves. The most valuable opportunities for establishing competitive differentiation are in how a product or service is created, sold, delivered and supported."

Workflow provides for the automation of business processes and enables users to control the process logic in all the various environments throughout their Inter/intranet. This ability to control the various business processes, the document management system's control over content and integrity, enables mission-critical, document-centric business applications to operate in an environment otherwise cumbersome to implement and manage. This has resulted in most document management vendors offering an integrated workflow engine or integrating the workflow engine with various workflow products readily available throughout the industry. The primary difference between these two approaches is whether the product consists of only those components developed by the primary product supplier or whether the primary product supplier has tightly integrated specialised technologies that other suppliers have developed.

The maturity of workflow technology and the associated trends are based on the separation of the processing rules from the processing scripts or work routing. In more sophisticated Web-based environments, workflow scripts could be tightly integrated to specific documents, making the routing, editing, approval, and submissions of documents manageable at the user level. Interaction with the various thin-clients would trigger sub-processes, as defined in the workflow script, resulting in the appropriate applet being downloaded and/or launched.

Workflow computing is the automation of work processes performed daily throughout any business. A workflow application automates the sequence of actions, activities or

tasks used to run the process. This includes tracking the status of each occurrence of the process and providing tools to manage the process. Taking a simple view, there are four basic components to a workflow system:

- Processes: An automated workflow application is made up of the different tasks or activities that must be completed to achieve a business goal. The workflow engine manages these processes. The workflow application works in conjunction with the engine to manage the work process.
- Participants: Usually individuals, but sometimes application systems, or automated agents taking the role of individuals, which perform processes based on business conditions or rules.
- Tools: There are various tools, excluding word processors, terminals, etc., the user accesses. These tools are used to access existing host applications and perform office related activities.
- Objects: This is another term for data that the workflow system uses. The term became more prevalent after the computing technology became sufficiently sophisticated to support video, audio, and other forms of information into the workflow system. These objects become the work item to be processed during the normal course of business.

Figure 3 shows a typical structure, or topology, for an ECM system.
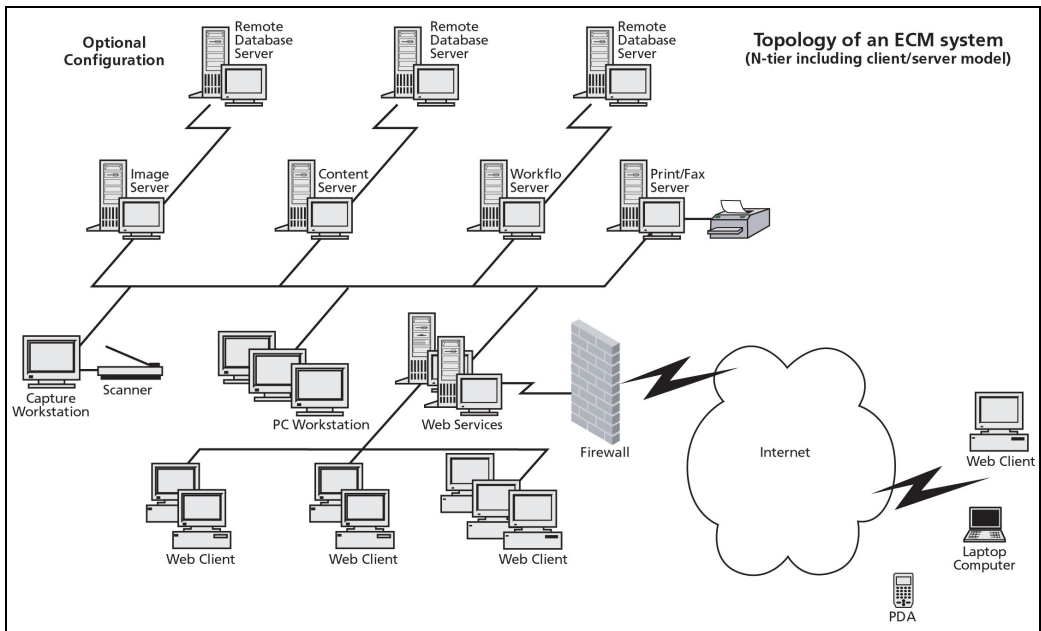
Figure 3: ECM Topology

3.2   Scalability: Expanding a System

One of the most critical capabilities of an enterprise system is to accommodate growth – both in terms of the numbers of users of the system and the number of transactions the

system handles. A content management system's architecture allows the system to grow with an organisation. It is the design of this architecture that makes CM scalable. Depending upon the applications, volumes, locations, and functionality, systems should be able to be broken down into multiple services and distributed to different servers (in different locations if necessary) to optimise performance and throughput.

Industrial strength content management offers a three-dimensional growth path:
- Multiple servers: library services can coexist on a single server or be split among several servers.
- Multiple services: Each library can have multiple services per server.
- Multiple libraries: each content management installation (database engine) can have multiple libraries.

For example, as Figure 4 illustrates, the three major services of a typical document management system can be split across multiple servers as transaction rates and throughput increase. Starting with a single server running all services, an organisation can add servers to the configuration until, ultimately, there is a server for each service, each in a different library system.
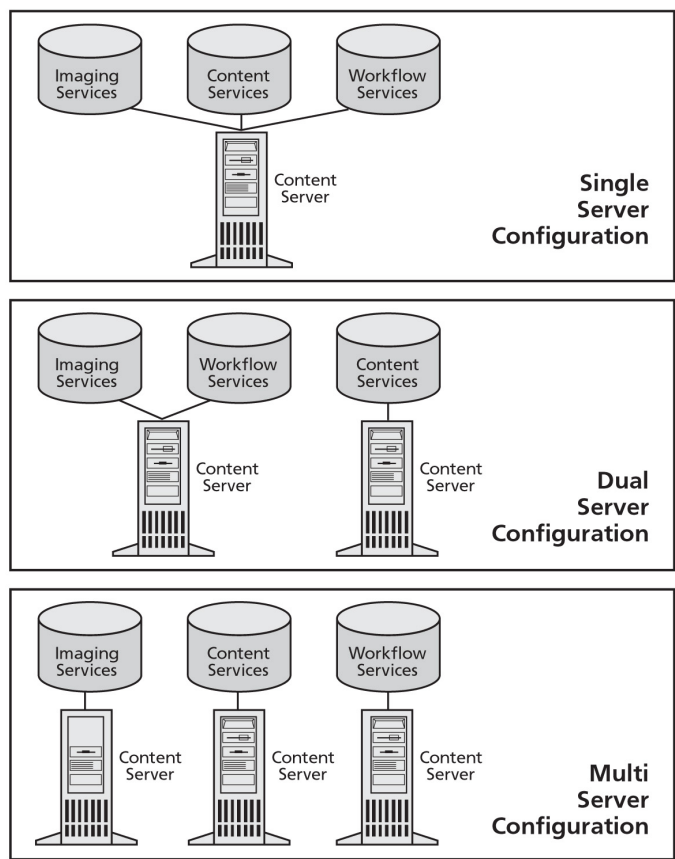
Figure 4: System Flexibility

By configuring multiple servers to run multiple instances of the same service, system administrators can incrementally expand the system to accommodate the growth of a library. It allows load distribution to help relieve network and hardware bottlenecks, such as traffic bandwidth limitations, hub and router delays, and thrashing read/write heads, which overwhelm server capacity.

A Storage Manager transmitting documents from Frankfurt to Paris may use too much of a WAN's bandwidth. To get around this problem, additional storage services can be added to the Paris installation; the Paris server can then transmit documents to client desktops it serves, freeing up the WAN.

3.3   Performance: Multi-Threading and More

CM system services modules are designed to be distributed and coordinated as peers. All user requests are processed through multi-threading (multi-processing in UNIX®). This allows simultaneous requests to be processed in parallel. If a server has reached its limit for processing requests, the next request can be automatically sent to another server.

Contention management services prevent the performance degradation that occurs in systems that rely on network file and record locking to handle user requests.

Other technical approaches which can be used to achieve good performance include the use of inter-process communications – which can be much more efficient than alternatives – and stored procedure calls to minimise network traffic and speed up the response to queries.

3.4   High Availability

The Web is an integral part of doing business in the new millennium. Mission-critical commerce and information exchange among business partners, consumers, suppliers, and employees are now fully dependent on the availability of critical Web sites. When a critical Web site is not available, commerce stops and revenue can be lost. Information doesn't flow. An organisation's image can suffer, leading, indirectly, to lost credibility or, directly, to customer complaints to legislators.

What is high availability? High availability for a Web site is defined in terms of the percentage of time the site is available to users during scheduled time over the period of a year. High availability is generally discussed in terms of 99%, plus some number of nines to the right of the decimal point. The so-called "five nines" availability, 99.999%, which allows as few as five minutes of downtime per year, is at the top end of the range of high availability. Contrast that to two nines availability, 99%, which sounds high, but still allows 87 hours of downtime a year.

META Group uses the analogy of insurance to describe how to choose a high availability solution: "Selecting a (high availability) solution is like choosing insurance. The premium is weighed against the possible loss of life or property. Best-practice organisations both ensure the appropriate level of availability and against the possibility of unplanned downtime. Selecting a cost-effective high availability solution is the product of weighing the cost of the solution's hardware/software against the cost of the loss of the application for some period."

3.5   Storage Management: Cost effective data storage

Image documents are relatively large – typically 50kB per page, but sometimes much more. Because of their size, they have traditionally been stored on optical discs. Optical discs are excellent for the storage of images because of their relatively low cost, large capacity and write-once-read-many times (WORM) characteristics.   Unfortunately, compared to magnetic storage, optical drives are slow, for two reasons. Firstly the data transfer rate from a magnetic platter is less than the rate available from magnetic discs. Secondly, optical discs are usually stored in a jukebox; consequently they have to be retrieved by a robotic arm before they can be read, which is much slower than accessing data on a magnetic disc. In the past few years the price of magnetic drives has dropped. They are now a reasonable alternative to optical devices in situations where WORM technology is not an absolute requirement.

CM can manage the migration of objects from magnetic to low-cost storage devices, such as optical disc jukeboxes or tape. It keeps track of graphics, documents, images, and other objects that have been migrated to near-line storage devices, and provides immediate access to them when they are needed.

The system administrator can define the archive rules and, on a daily basis, scan the library for documents that are candidates for being archived. This allows the migration of data from high-cost to low-cost devices in a controlled manner.

3.6   LDAP Support

LDAP (Lightweight Directory Access Protocol) allows administrators to maintain user information in a single, centralised directory for a set of applications, including network and computer workstation access. By modifying a single set of records, administrators can easily grant or revoke user access privileges for many applications in a coordinated and unified manner. User names, group names and user attributes employed by CM applications can be stored in, and retrieved from, the directory, allowing a shared, non-redundant usage. The central directory is accessible via LDAP.

There are three approaches to using the information held within an LDAP-compliant directory:

- Extract this information, transfer it, manipulate it and use it in the security system of each repository, or
- Enable an application to link directly to this information live via the LDAP protocol and over the network.
- The application can link to the data it currently understands and manages itself, or it can be enhanced to support and utilise more of the concepts and features of the directory service (i.e. common name plus e-mail address).

The first method can be achieved by scripting, and it is easily managed; as a result it can meet the vast majority of requirements where instantaneous synchronisation of the information is not essential. Thus, the ability and benefit to perform user and group

maintenance in the central directory is achieved, while special needs can still be handled through existing application-specific security systems.

3.7 Making the Connection: How Content Services Communicate

In a client/server environment, communication between the clients and the servers starts with a local area network (LAN) . It is important to support multiple LAN connections simultaneously to provide maximum scalability and redundancy. Note that with today's hardware technology, the speed of the LAN can be key to system throughput, and is often an overlooked bottleneck.

CM communicates not only service-to-service but also service-to-database via Open Database Connectivity (ODBC). ODBC is an industry standard protocol database that is widely supported by vendors.

3.7.1 ODBC

Most relational databases today use the Structured Query Language (SQL) as an access method. ODBC provides a common interface for accessing heterogeneous SQL databases. It allows a single application to transparently access different SQL databases through a common driver, without tying the application to a vendor-specific application program interface. This allows a single desktop to access different versions of content services, as well as different databases such as SQL Server® and Oracle®.

There are four components to the ODBC architecture, illustrated in Figure 5:
- Application – Performs processing and calls ODBC functions to submit SQL statements and retrieve results.
- Driver Manager – Loads/unloads ODBC drivers on behalf of an application. Processes ODBC function calls or passes them to a driver.
- Driver – Processes ODBC function calls, submits SQL requests to a specific data source, and returns results to the application.
- Data Source – Provides the data to be accessed via an associated operating system, DBMS, and the network infrastructure used to connect to the source database.
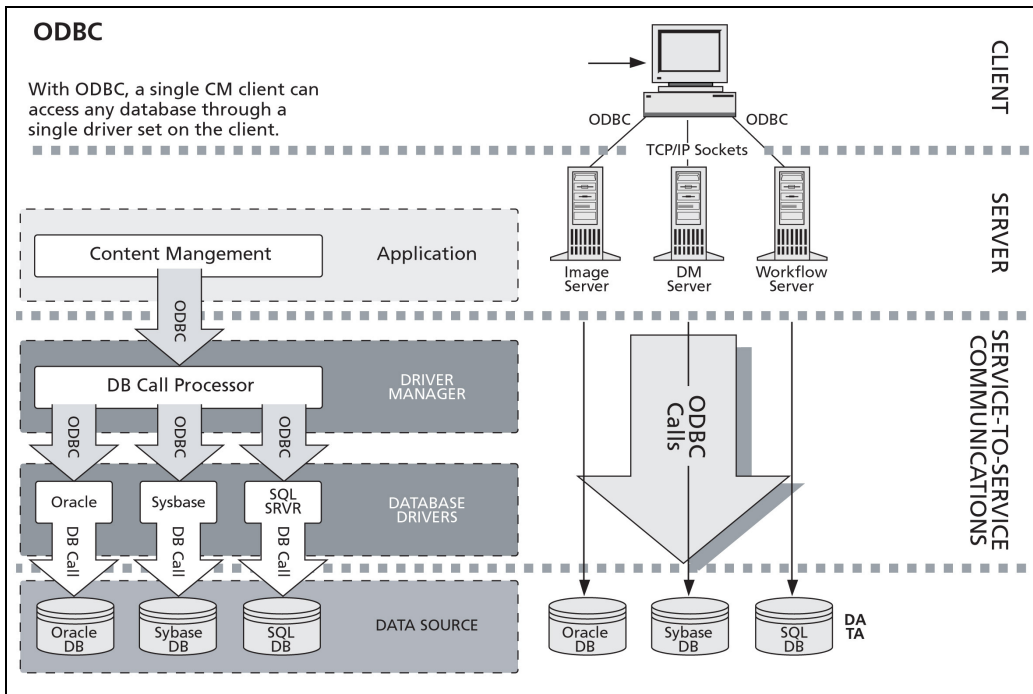
Figure 5: ODBC Content Services

The powerful combination of content management and ODBC makes a true Enterprise Solution. From a single client desktop, users can access different DBMS platforms and a range of CM Services versions through a single interface.

3.8 WebDAV

WebDAV (Web-based Distributed Authoring and Versioning) defines a communication standard, which uses HTTP between a WebDAV Client and a WebDAV Provider. WebDAV is not specific to any particular object technology (e.g. Java or COM) or any particular Web server (e.g. IIS or Apache)

One of the strengths of this standard is that no vendor-specific code is required on the client; meaning that the only thing that needs to be done to deploy WebDAV (after the provider has been set up) is to give the WebDAV clients the provider's URL. One of the results of this is that the user interface and functionality of WebDAV clients depends solely on the WebDAV client implementation. Thus from the user's perspective, the WebDAV client appears fully integrated into whatever application they are using.

Windows Explorer, Internet Explorer, Microsoft Office, Adobe Acrobat, Mac OS X, Dreamweaver, Adobe GoLive and Arbor Text Epic, are among the products that have built-in WebDAV clients.

3.9 Digital Signatures

Information security and integrity have always been key requirements of any user of IT systems. Today, however, there is a more critical need due to the Internet revolution, with government organisations moving towards online interaction with other agencies, citizens and businesses. Digital signature is an emerging tool that can assure full security and high reliability for electronic document interchange over the Internet. This will be a major innovation, impacting both the public sector and the business environment. It will allow government agencies to operate via the Web with legally valid acts.

Digital signatures are based on two "keys". Each key is a long string of digits. One, referred to as a "private" key, is used by the sender to sign a document. The second, called a "public" key, is used to verify the signed document, in combination with a certificate supplied by an external certification authority. Assurance of the integrity of a digitally-signed document requires the correct combination of public and private keys.

4.   Functionality of Content Management Solutions

Chapter three discussed content management architectures. This chapter moves on to cover CM functionality, including version management, workflow, rendition support, Web publishing, replication services, compound documents, record retention and document lifecycle management, facsimile services, interoperability with industry standards, security, system administration, internationalism and localisation.

CM provides extensive search capabilities throughout vast amounts of information. Through use of the Web or a customised application, users have access to simple searches that allow users to search for a document based on indexed information associated with the document. An advanced search allows users to form a complex search based on any document property, including document type, user, or any other index information stored in the relational database. A full text search allows a simple or advanced search of the complete text stored in a document. Stored searches let a user specify a standard set of search criteria that can be reused. For example, a user can create a stored search that searches for documents of a particular document type; say a certain court case, created on or after a particular date by a specific author. Then a user can run the search monthly to retrieve the same type of information on a regular basis. A universal resource locator or a URL may store saved searches. This allows quick and easy access to stored searches.

4.1   Version Control

Content management systems protect the integrity of the information they store. A document, once checked in, is "protected" by the library and is never changed. It also protects the version history of documents by sending only a copy of the document version each time it is checked out.

If an important section of a document is inadvertently deleted while open, the checkout can be cancelled. Any edits made to the document during that session, including deletes, are discarded and the document is retained in its pre-checkout status. This safety feature ensures that the original document for each version of a file is safely stored on the network and can never be lost or corrupted.

Every time a user checks in a document (after having checked it out), a new version of the document is created. A new version object uniquely identifies any changes, and then stores the updated version in the system. Each document may have several version objects.

For example, consider a memo added to a system. After it is in the system, the author realises last-minute updates were left out. The first version of the document is checked out and the necessary modifications are made. When it is ready to be checked back in, there is no need to name the memo to indicate that this is the second version. The system takes care of version tracking. When the check-in is complete, the system stores the modified document as "version 2" of the memo, without corrupting the original version. In this way, the original document is protected from being overwritten or corrupted.

Further, the CM provides a sequential audit trail of all changes made to the document during its lifecycle. If an older version of a document is required, its version is readily available in the system.

4.2   Process Management and Workflow

Much of the content that government organisations need to provide to their customers, employees and business partners, is produced during a business process. Therefore managing business processes is an integral component of CM. Business processes can be as simple as the approval of documents prior to posting them on an intranet or Web site, or as complex as processing a corporate tax return. Process can also relate to the internal aspects of managing a document throughout its lifecycle, or the steps a customer must take to open and track information on a current transaction. Business processes govern relationships that define the rules, interfaces and sourcing that delineate how government works.

Process management is based upon workflow technologies that include various types of routing, including ad-hoc routing, administrative routing, and production routing:

- Ad-hoc routing enables the user to specify a specific process, for one document only, which a document is to follow.
- Administrative routing enables users to define specific routing for a specific type of work that is always followed, regardless of the data within the work being routed.
- Production routing enables the users to define rules and work methods that are based on the document type and data contained within the work item. As the data changes, the production routing system processes the document accordingly, including the ability to support work timeouts, escalation, and work reassignment.

Workflow allows the re-use of process definitions; and routing can be based on contextual information, allowing the deployment of a single process to be used in a number of different scenarios. In addition, workflows provide, by default, an auditing tool that tracks every event associated with processing a piece of work, as well as tools for monitoring and managing workloads.

4.3   Rendition Support for Effective Publication

Creating a document rendition is the process of converting a document from its native format to a different format. Typically, the rendition is a near exact copy of a document. The primary use for rendered documents is to allow electronic publication to the widest possible audience. Content management can automate and manage the rendition process. (See Figure 6)

The Rendition process starts with the ability to flag a document from one format to another, such as from Microsoft Word to Adobe's Portable Document Format (PDF). The process then stores the rendered document in a particular content library. For example, suppose a technical manual is created using an uncommon word-processing application. To read the manual, all of the recipients of the published document would have to have that particular word-processing software installed on their computers. This

is not practical and limits the usability of the technical manual, defeating the purpose of publishing it. If, however, the manual is rendered into PDF format before being published, any user with widely available, free, PDF viewer software (Adobe Acrobat Reader®) can read it.

PDF can maintain the exact look and feel of virtually any source document, and has widespread acceptance as a legally admissible original in a court of law. Other benefits of PDF include its compact size, platform and application independence, as well as its ability to support content search. PDF is the most widely accepted rendition standard in the world. Its popularity is due in part to its acceptance as a format that maintains the original look of the document, and that end users cannot manipulate or change it.

The Adobe Acrobat Server executes the appropriate application plug-in (a copy of each native application must be loaded on the Adobe Acrobat Server), for the format of the requested document. In turn, the executing application generates a Postscript representation of the input document. The Adobe Distiller reads the Postscript output, and then generates a PDF representation of the document. The PDF Administration Tool then places the PDF document in the document repository and notifies the client that the process has been completed.
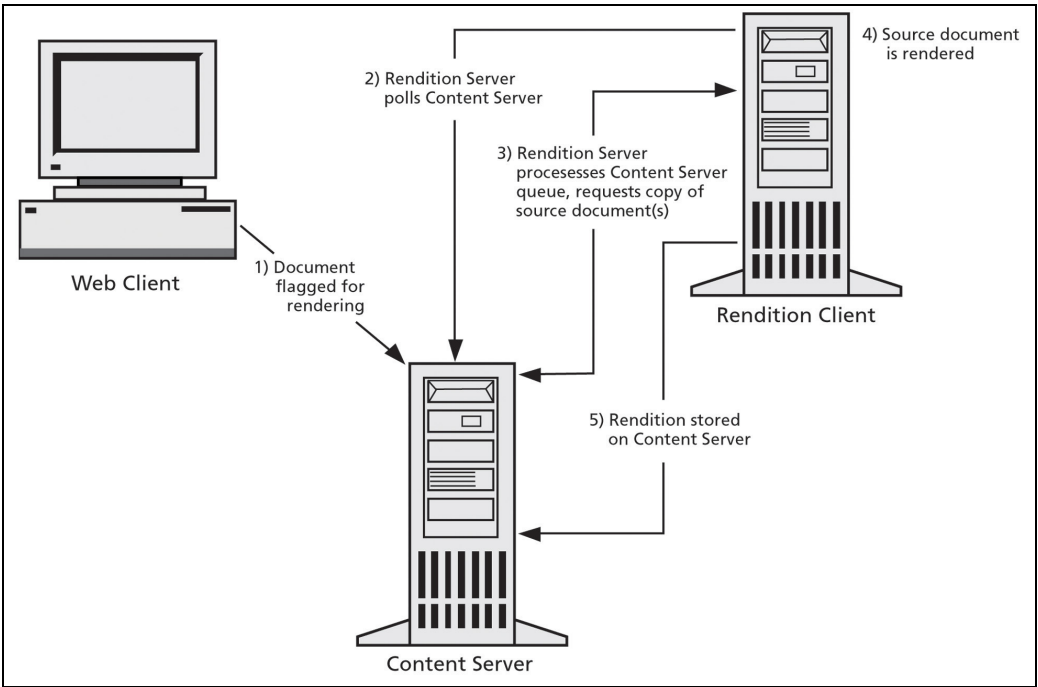

Figure 6: Document Rendition

4.4   Publishing and Web Publishing

Publishing is the ability to deliver a document copy (rendered or original format) to a predetermined and unique repository location. The unique repository location enables the published document to have a viewing audience different than that of the source

document. This also means the original document can be published as a new document on an as-needed basis, independent of any source document version updates.

Content management systems have the ability to create and publish documents and images to the Web. Frequently, the documents need to be rendered into publishable format, as described above. Automatic table of contents and user navigation tools can also be included as part of the publishing process. Items can be published ad-hoc or via a scheduler.

All published items are managed by the content management system; all of the security and versioning functions, therefore, are handled automatically. In other words, unreliable and time-consuming Webmaster functions are replaced with predictable processes that automatically cycle content through publishing processes. When a user checks out a document that has been posted on the Web, makes a change and checks it in, the document is automatically routed through an approval process and then posted on the Web. This linking and automation of version control to publishing represents, in some cases, a major benefit of content management. It brings the additional benefit of freeing Webmasters to focus on tasks more important than updating Web sites.

Advanced integration will support Web content management functionality, such as personalisation, subscriptions and advanced searches across heterogeneous repositories.

4.5 Replication Services

Replication services maintain one or more copies of documents in separate libraries, in a manner that is completely transparent to end-users. In effect, this gives users the impression that a document is stored locally, in a readily accessed library, even if it is not. This makes documents more available across the organisation, automatically and without compromising integrity. Rather than attempting to provide all users access to a document residing in a single library, the document can be automatically copied – replicated – to other library systems.

When a document is replicated, both the content and properties of the document are copied into the library participating in replication services. The benefit is faster document search and access times for users, less network traffic, and higher overall resilience. All instances of replicated documents or folders (replicas) in all participating libraries are automatically synchronised with any modifications to the source folder or document, even if the document is added/checked in to a library other than the source library.

4.6 Compound Documents - Complexity with Control

The ability to manage compound documents is a useful aspect of any content management system. The foundation of this capability is in forming relationships between documents, and dynamically tracking the links and relationships among the various components of a compound document.

Creation of a compound document happens when a document is checked into a content management system and users are given the choice of associating other documents with it. In addition to link creation, the type of link, whether parent, child, dynamic or static, is also established. These relationships are stored and associated with the specific

version of a checked-in document. As new versions of parent and child documents of a compound document are created, the relationships between these entities are associated with the new versions.

Compound document support provides several benefits. First, it allows documents to be associated with and accessed from a single reference point, simplifying accessibility. It also allows users to view, edit and print different documents of different formats through interaction with a single parent document. This is an important timesaving feature. Each linked document concurrently exists as both a part of the compound document and as a separate stand-alone document. This allows different authors to work on different parts of the same compound document concurrently, improving individual user productivity.

Compound documents can be divided into two types.

- "Complex" documents, which are single documents containing other documents that are linked, usually with Microsoft's OLE technology. Complex documents represent a simple compound document structure, and are usually produced by a single individual or, at most, a small group. An example of a complex document would be a sales proposal, developed in Microsoft Word, that includes appendices, such as a pricing spreadsheet developed in Excel and a PowerPoint presentation.
- "Relational" documents are made up of a set of independent documents arranged for the purpose of review and ultimate publication. Relational documents do not have a physical component; rather, they exist as a series of interrelationships and links that make up the larger parent document. They are typically hierarchical in nature and their content is displayed as one, even though the single complex document itself does not exist within the library.

4.7 Record Retention and Document Lifecycle Management

Virtually every large organisation establishes rules that govern proper classification and retention of corporate records. Unfortunately, without CM, records management, or automated processes, the organisation has to rely solely on other resources, such as Records Manager personnel, to enforce the rules and policies. As a result they are rarely enforced consistently. The Records Manager typically designs document classification systems; and is then responsible, among other things, for:

- Ensuring established retention rules are followed.
- Archiving or destroying inactive records.

Electronic records management accomplishes these retention functions more effectively and efficiently while eliminating the element of inconsistency. It allows the Records Manager to create detailed document disposition rules and multiple levels of archive to ensure proper review of documents before they are destroyed and automatically ensures these rules are followed. It also generates "certificates of destruction" if necessary, and maintains the essential audit trails in a manner acceptable to regulatory authorities.

Retention control is one part of the records management continuum. Records management focuses on the full lifecycle of record creation through to final disposition.

Structured classification and retention provides the rules to ensure that "official" records are properly managed throughout their lifecycles.

4.7.1   Fax Services

CM is not limited to managing electronic information. CM can also manage facsimiles (faxes). Using fax services in CM enables users to send and receive faxed documents at their workstation. When considering these services, departments should evaluate the following requirements to support their specific business needs:

- Outgoing fax without document viewing: provides the ability for users to fax documents directly from their computer without viewing each document first. The user should have the ability to select a range of documents and have them routed to the fax "server" for transmission.
- Outgoing fax after document viewing: This provides the ability for a user to fax a document during viewing. The user should have the ability to attach other documents to the outgoing fax, as appropriate.
- Incoming fax processing: As incoming documents are received, the system should support the ability to receive incoming documents and automatically route the document based on configurable rules (via a system administration interface), either by incoming telephone number or by the use of forms processing.
- Fax status reporting: The system should provide a fax reporting capability, enabling users to view status and historical information related to faxes sent by users.

4.8   Openness: Interoperability with Industry Standards

An open system is one that adheres to published industry standards, so that using organisations choose hardware and operating system platforms, network media and protocols, databases, access methods and peripherals according to their needs. Here are some of the standard platforms that are available in the industry:

- Operating System Platforms: HP-UX®, Microsoft Windows NT®, Microsoft Windows 2000®, Microsoft Windows XP®.
- Database Management Systems: Oracle® and Microsoft SQL Server®.
- Networks: Popular LAN/WAN technologies including Ethernet, LAN Manager, Token Ring, FDDI, ISDN, T1, PPP and Frame Relay.
- Network Protocols/IPCs: TCP/IP Sockets.
- Database Connectivity: ODBC.
- Document Formats: Documents are stored as objects; any type of document can be added to a Content Services library. Examples are XML, HTML, TIFF, JPEG, GIF, AVI, WAV, word processing and hundreds of others.

4.9　　Security: Multiple Levels

Content management systems should provide multiple levels of security for documents. Each time a document is added to the system, the author determines what users and groups will be able to access that item and its associated versions. This information is kept in an access list – a list of users and groups who can access the object – of the Item object. Default access lists are provided for all items.

The access list for an item can control access on a user-by-user basis. To make access control easy to implement and maintain, administrators can assign access rights to groups. Groups contain sets of users who have common access requirements and privileges

Different CM implementations identify different access models.　A typical model includes five levels of access rights that can be applied to users and groups specified in an access list, as shown in the following table..

Access Right	Privileges Granted
None	None. The user will not see that the item is in the system.
Viewer	The user can view the item object and can make copies of the associated versions.
Author	Viewer access rights, plus the user can check out and check-in associated versions and modify property values in the version and selected property values.
Owner	Author access rights, plus the user can modify most property values and delete the item.
Admin	Owner access rights, plus the user can modify all property values, except the audit trail.

Figure 7: Access Rights and Privileges

4.10　　Additional Security　Features

There are six additional security features worth mentioning, as illustrated in Figure 8. A system of secure checks and balances guarantees robust integrity of sensitive corporate information.
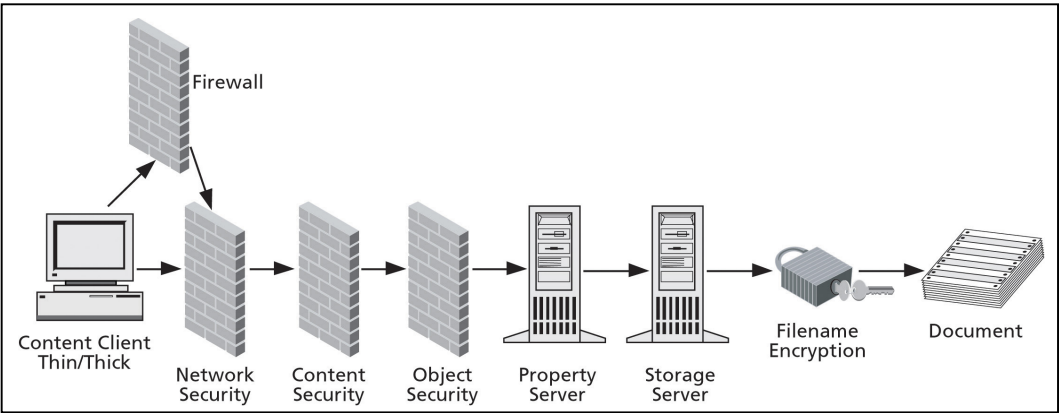
Figure 8: CM Security

4.10.1   Firewall Security.

There are three basic kinds of threats to a site: people trying to steal data either externally or from within the company; defacing information; and viruses. To protect an organisation's information, it is important to use firewalls, which restricts unauthorised access between the Internet and internal networks and to keep inside connections from reaching the Internet without any authorisation.
Firewalls, of course, will not protect information assets alone. They are only one part of a layered security that also uses anti-virus software and vulnerability assessment tools. Together, they are stronger than any individual part.

4.10.2   Network Security

All networked systems have some form of access control. In all but the most secure, this usually takes the form of network login IDs and passwords. CM solutions vary in the way that these can be managed and, in some cases, integrated to the operating system password management.

4.10.3   Content Management Systems

These typically have their own rigorous security that requires users to have unique identifiers and passwords to gain access to a library system.
In some cases, they can be configured to rely on secure operating system user identities.

4.10.4   Property and Storage Servers.

Once users have navigated the network and the content management system, and have successfully logged into a library, they still cannot directly access stored documents unless they are members of a group that has access to those documents.
In some cases, they may have to meet other application-level security requirements. Typically, even if users are able to gain unauthorised access to a library, they have no way of knowing where specific documents exist or how to directly access a desired document without going through application-level security controls as a member of a designated group.

4.10.5   Recovery.

If nothing else, what happened to hundreds of businesses in New York on September 11th 2001 proves once again that it is important to pay attention to the fundamentals of backup storage in order to recover data in case of a catastrophic event.
Basic rules include:

- Backup all critical data daily.
- Backup not only desktops and servers, but laptops as well.
- Maintain all backups off-site.
- Perform systems maintenance regularly.
- Ensure that the system can realistically be restored from the backups.

All CM systems must be capable of being backed up and restored; the sophistication of the backup regime should be determined by the criticality of the application.

4.10.6   Filename Encryption

If an unauthorised user gains access to the directory containing the documents, a content management system's filename encryption scheme offers yet another level of security.

4.11   Audit Trail

Audit trails should be able to track almost every action users take. In secure environments, an audit trail of unauthorised access attempts can be logged. Each audit trail entry should capture who initiated the action, when the action took place, what object was being accessed, what the outcome of the action was (success or failure) and which client workstation was used.

In some systems, auditing can be turned on and off. In most cases, because audit trails are an essential part of the framework that provides confidence in the integrity of electronic content, it is important to ensure that it cannot be turned off.

System administrators should be able to view the audit trail with a reporting interface. At a minimum, the system events shown in the following table should be audited.

Object Related	User Related	Version Related	Other
Add Object Get Object Delete Object Modify Object Check Object Existence Get Access Info Rename Object	Modify Password Login Logout Change Group Suspend Session Resume Session	Check-in Version Checkout Version Cancel Checkout Archive Version Index Version	Perform Search Attempt unauthorised access Create, modify or delete workflow Change audit trail settings

Figure 9: Reporting events for audit trails

In practice, a system performance can be a significant consideration, as some audit trail approaches are especially resource-intensive. It is important to ascertain from the early stages of system design that the desired performance level is consistent with the desired level of auditing.

4.12   Internationalisation and Localisation

In a European context, it is often necessary to implement solutions such as CM in several languages. This means that the software used must support localisation.

Where a Microsoft environment is required, CM products should conform to Microsoft globalisation and localisation standards. These standards require that language-specific user interface components, including text-searching capabilities, are self-contained modular components. They also require that application translation does not involve modifying program logic. This reduces the possibility of introducing new bugs in the conversion process.

Translation from one language to another language requires simply selecting the language and then translating the text in the interface components; this makes it relatively simple and fast to deploy applications internationally. This approach is far preferable to trying to maintain multiple localised code sets. Changes and new functions

can be added more quickly and with less risk of new bugs within a single code base. This is a key competitive advantage as it helps organisations stay ahead of evolving business requirements. Adherence to these, or comparable, globalisation standards speeds the rollout of production applications across national boundaries and reduces training and support costs. It also improves product reliability for all users regardless of their locale.

4.13   System Administration

System specifications sometimes concentrate on user functionality, at the expense of administrative functions; however, it is important that the latter are not overlooked.

Using administration tools, administrators can add and maintain user and group data, as with any other application. There should also be a capability to perform additions, modifications and deletions in bulk. In one step, for example, an administrator should be able to deny access to multiple users, or to change properties or securities for multiple documents. This is an important timesaving feature.

Administrators should be able to perform administrative functions while users are logged on and using the system; and ideally it should be possible to add storage and services without interrupting users' work.

5.   Content Management Standards

This chapter introduces the importance of content management standards. It covers relevant technologies and the challenges of system design, as well as the benefits of XML to managers, interfaces, relevant standardisation bodies and initiatives. Included under this heading are formal standards, de facto industry standards, and definitive guidance such as technical reports issued by standards organisations.

As explained in chapter 3, content management incorporates many technology components that enable organisations to link documents, records and fragments with the business processes that govern their use and lifecycle. One view is that, at a high, management-oriented level, there is a shortage of standards to unify and pull these components together. Another view is that at a lower level – the technology component level – there is an overabundance of standards. It is as though there are no standards governing the building as a whole, but a multitude of standards and guidelines defining individual parts such as the adhesives, screws and other fixings which hold the building together.

The component-level standards have for the most part been developed independently of each other – by different bodies, at different times, to different levels of detail, and so on – and the standards are not necessarily consistent. For example, guidance on image compression may be fine for office environments but inappropriate for Web applications. However, the situation is not as bleak as these plain statements may suggest; in practice, several key standards and components are widely understood and accepted as de facto standards. For example, SQL is pretty much a default assumption for databases; this makes it practical to integrate the database-using components which make up CM applications.

The technologies that are most often incorporated under the umbrella of CM include:
- Digital or document imaging software
- Electronic document management software (EDMS)
- Web-based content management software
- Workflow or business-process management software

As organisations continue to seek ways to manage information consistently, content management solutions have become a key component in efforts to manage the "unstructured data" presented by the paper-based and microfilm-based systems of the past. For example, one department may create documents that can be shared by others, thus saving time and the expense of having to recreate the information, also reducing the possibility of introducing errors.

Governments in particular are keenly interested in developing new methods of electronic content management because they often must retain and retrieve records from archives for decades. To accomplish this, government agencies are normally the first entities to set overall content management standards to provide a framework for the definition of agreements about data exchange – agreements that, if sufficiently standardised, could serve as the basis for laws and commercial codes.

5.1 XML as an Interoperability Standard

XML stands for eXtensible Markup Language. It is defined by the World Wide Web Consortium (see http://www.w3.org/XML). XML is a "meta-language," which means it can be used to develop other languages, each with its specialised terminology. It allows system designers users to build the meaning of information into new computer systems. In other words, designers can build applications that can send information to and from unrelated external systems, using a common, non-proprietary, method of describing the content and nature of the information; the commonality of this method ensures that the external systems can do the same.

Although XML dates back only to the late 1990s, it is already accepted widely; the industry is confident it will have a rosy future. All content managers – in government and elsewhere – will benefit tremendously from XML, which will enable departments to reuse information across divisional boundaries. As an example, some national governments are already working to define consistent XML definitions for information resources held across government.

Because XML is a meta-language, government organisations and commercial companies can work to agree on a standard XML vocabulary. There could be difficulties, however, if major industry sectors develop their own set of agreed-upon vocabularies without regard to other sectors. To guarantee e-commerce can successfully use XML, there must be a consensus on such things as tags used to describe data.

XML is also a powerful complement to the Java programming language, and will provide a level of interoperability never before possible. With improved interoperability, legacy systems' lifecycles will be extended dramatically. This will be particularly attractive to governments that have to archive records for decades. As governments play a leading role in the spreading and acceptance of XML, they also help private industry follow suit.

There is a need to develop technologies that comply with the law and best meet the needs of government organisations. XML, as part of the new advances being made to assist organisations to become more efficient, effective and profitable, could be hampered because of the inconsistencies in laws and the slow pace lawmakers follow to recognise new technologies.

While governments, businesses and even programmers wait, the XML experiment continues to march forward confidently. This is the future of the Internet and, indeed, of data organisation. With each new device that uses the Internet – from handheld devices to information appliances – XML is finding its niche in the data-sharing arena.

5.2 Document Management and Records Management Standards

- Model Requirements for the Management of Electronic Records (MoReq Specification).
  Published by the European Commission, 2001. Available in various formats from several Web sites including http://www.cornwell.co.uk/MoReq. This provides detailed guidance on how to specify an electronic records management system, with guidance on document management and other aspects.

- US DoD 5015.2-STD
  Design Criteria Standard for Electronic Records Management Software Applications - Defines the basic requirements based on operational, legislative and legal needs that must be met by records management application (RMA) products acquired by the Department of Defense (DoD) and its Components.
- UK Public Records Office (PRO)
  Functional Requirements for Electronic Management Systems - Published standard set of functional requirements and requirements conformance testing on electronic records management systems for UK government.
- Records Management,
  ISO 15489:2001 - This international standard briefly defines the essentials of records management, without concentrating on electronic issues. The accompanying technical report provides further detail.
- DMWare
  The open-source distribution and development clearinghouse operated with the institutional sponsorship of AIIM International.
  Document Management Alliance Specifications - DMA defines software component interfaces that enable uniform search and access to documents stored in multi-vendor document management systems.
- Open Document Management API
  ODMA specifies a set of interfaces that applications can use to initiate actions within a document management system.

5.3   Workflow Industry Standards

- Workflow Management Coalition (WfMC) Application Programming Interface (Interface 2 & 3)
  This specifies standard workflow management Application Programming Interfaces (APIs) which can be supported by Workflow Management (WFM) products.
- WfMC Audit Data Specification
  The purpose of this document is to specify what information should be captured and recorded about the events occurring during a workflow enactment.
- WfMC - Interoperability, Internet, e-mail MIME Binding
  This document maps to the WfMC Interoperability Abstract Specification, which defines the functionality necessary to achieve a defined level of interoperability between workflow engines. It defines a binding that gives concrete type definitions and message formats for the realisations of the abstract specification.

5.4   Document Imaging Industry standards

- ANSI/AIIM MS44 - 1993
  Recommended Practice for Quality Control of Image Scanners – This recommended practice provides procedures for the ongoing control of quality within a digital document image management system.

- ANSI/AIIM MS52-1991
  Recommended Practice for the Requirements and Characteristics of Original Documents Intended for Optical Scanning – This standard describes the physical characteristics of paper documents that facilitate black-and-white optical scanning and the characteristics that make scanning difficult or impossible. It provides general recommendations for the design of documents in order to make them easier to scan.
- ANSI/AIIM MS53-1993
  Recommended Practice; File Format for Storage and Exchange of Image; Bi-Level Image File Format: Part 1 – The purpose of this standard is to standardise a self-contained file format for the transfer of bi-level image documents in environments other than facsimile telecommunications.

Figure 10: CM Standards

5.5   Other Content Management Related Standards and Guidance

The following standards and other guidance documents contain information which may also be relevant.

- BSI DISC PD0008, 1999
  Code of Practice for the Legal Admissibility and Evidential Weight of Information Stored Electronically.
  This code of practice contains guidelines intended to assure the integrity of electronic content. Intended for the English legislative framework, its use in other countries must be tempered by the applicability of relevant local legislation, though it contains ideas which will be valuable in any legal framework.
- BSI DISC PD0009, 1999
  Compliance Workbook for use with PD0008

A workbook in the style of a comprehensive audit checklist, designed to allow system users and suppliers to verify the extent of compliance with PD0008.

- Dublin Core Metadata Initiative
  This initiative, though not a formal standard, has developed a set of metadata elements – the Dublin Core" – which is widely accepted as a starting point for resource discovery metadata in governmental content management and similar applications. See http://www.dublincore.org.

- ISO 12651, 1999
  Electronic Imaging Vocabulary – This international standard contains ISO definitions of key imaging terms.

- ANSI/ARMA-10-1999
  Glossary of Records and Information Management Terms.

5.6   ANSI/AIIM Guidance

In the USA, The American National Standards Institute (ANSI) and AIIM International have published several guidance documents which may be relevant in a European context.

- ANSI/AIIM TR2-1998
  Technical Report for Information and Image Management Glossary of Document Technologies.

- ANSI/AIIM TR25-1995
  The Use of Optical Disks for Public Records.

- ANSI/AIIM TR15-1997
  Planning Considerations, Addressing Preparation of Documents for Image Capture.

- ANSI/AIIM TR27-1996
  Electronic Imaging Request for Proposal (RFP) Guidelines.

- ANSI/AIIM TR31
  Performance Guideline for the Legal Acceptance of Records Produced by Information Technology Systems, Parts 1 to 4, 1992 to 1994, reaffirmed 1998.

- ANSI/AIIM TR32-1994
  Paper Forms Design Optimisation for Electronic Image Management (EIM) .

- ANSI/AIIM TR33-1998
  Selecting an Appropriate Image Compression Method to Match User Requirements.

- ANSI/AIIM TR34-1996
  Sampling Procedures for Inspection by Attributes of Images in Electronic Image Management (EIM) and Micrographics Systems.

- ANSI/AIIM TR35-1995
  The Human and Organisational Issues Guidelines for Successful EIM System Implementation.

- ANSI/AIIM TR40-1995
  Suggested Index Fields for Documents in Electronic Image (EIM) Environments.

- AIIM ARP1-2001
  Implementation Guidelines and Standards for Web-Based Document Management Systems.

6.   Best Practice Applications

Content management solutions can be implemented in the public sector in a number of ways. For government agencies, in particular, the possibilities for cost savings, streamlined operations, and improved service levels run far and wide.

This chapter presents case studies of real implementations, followed by descriptions of other possible applications.

6.1   Online Services via a Web Portal
Project of INAIL

The Italian National Institute for Insurance Against Industrial Injuries (INAIL) is a central public agency, with a widely distributed network of offices and more than 11,000 employees throughout the country. INAIL's mission-critical processes are to collect and manage insurance funds from more than three million businesses and assist injured workers.

6.1.1   The Challenge

INAIL needed a comprehensive, integrated content management infrastructure to automate and optimise its working processes in order to increase efficiency and responsiveness to the "customers" it serves, citizens and companies. At the same time, the Institute had to comply with the new set of rules and laws for the modernisation of the Italian public administration, which calls for the mandatory introduction of "Electronic Protocol" (electronic filing) in all public agencies by January 1, 2004. In this framework, the electronic protocol is not simply seen as the computerisation of manual processes, but as the first step of a more general approach to electronic document management and workflow control. This will allow the public sector to automate work processes, make them transparent to the public, interoperate with government bodies and deliver services online.

6.1.2   The FileNET Solution

In cooperation with Getronics (a FileNET value-added reseller), INAIL defined an extensive project with the objectives of implementing the Electronic Protocol, integrating applications and procedures within the content management infrastructure and expediting the e-Government initiatives, thus addressing the strategic requirements of the Public Administration reform. FileNET software was selected and about 200 systems have been installed in the INAIL organisation network of 20 regional offices, 148 local offices and 70 territory operation centres.

The first phase of the project, Electronic Protocol, has been successfully implemented at INAIL. In the different sites, incoming paper and electronic documents are captured, stored and classified in optical libraries and made locally available online. They range from business payments to reports on worker' injuries and wealth certificates, from e-mail messages to Word texts or Excel spreadsheets. A total of 6,000,000 documents

(one or more pages each) are required annually, with an average of 30,000 documents retrieved and accessed daily. These are scanned and converted to text by an ICR server; then images and data are transferred into the system. Electronic Protocol automatically routes the registered documents to the various offices for their processing.

In the second phase, now being implemented, INAIL is integrating Electronic Protocol and document management with the application procedures, while introducing work-process automation in the primary activity areas. A centralised content repository, to be connected to the local archives, is being created to allow authorised users to view and manage files from any site. Digital signature solutions (for archiving documents on optical discs, to eliminate paper filing) and electronic document interchange with other agencies will be introduced.

The project's final phase will be totally Web-centric. The agency Web site will become a portal to deliver services to citizens and businesses; internal and external users will be able to track online a file processing state; and interoperability with other government bodies through the Internet will be complete.

See Figure 11 for an illustration of the INAIL solution.
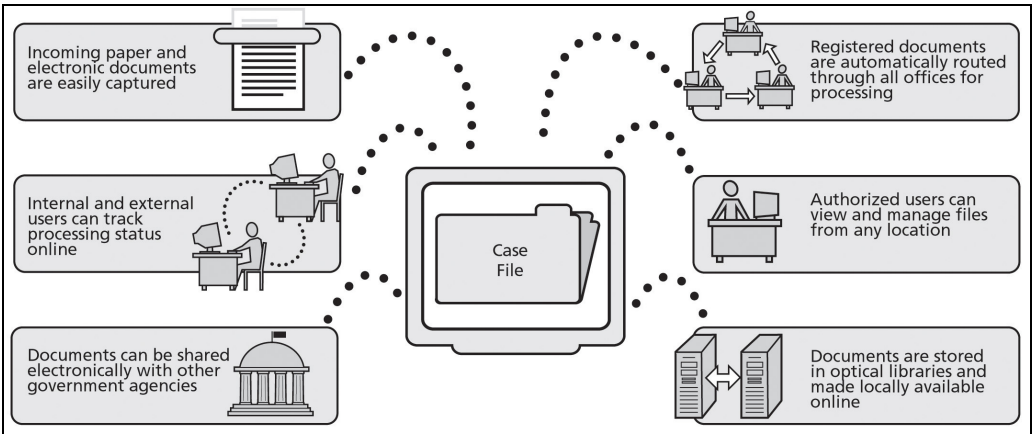
Figure 11: INAIL solution

The integrated content management system will be extended to other application areas (worker case histories, legal and personnel files); will be able to register and classify documents coming via e-mail or a Web portal; and will provide the Contact Centre with timely and complete information. A planned joint portal with the Italian Social Security Agency (INPS) and the central IT agency of the Italian Chambers of Commerce (InfoCamere) will give a single access point to businesses for many applications.

6.1.3   The Bottom Line

FileNET software for managing business processes and their related contents, inside the organisation and over the Web, is the basic infrastructure that supports INAIL's critical processes, and will make possible the exchange of information with other public agencies and the online interaction with citizens and businesses.

This Electronic protocol solution can easily be evolved into a more advanced and complex integrated content-management system. Critical processes, electronically managed, are streamlined and extended to the Web, which improves the organisation's efficiency, work quality and transparency. Using a FileNET platform, INAIL has built a solid, scalable base to achieve the strategic objectives defined in the public administration reform plan.

By enabling the effective migration to e-Government, INAIL will better serve its customers, workers, companies and trade associations.

6.2 Language Services and Payroll Management Project of the German Federal Foreign Office

"We have to become faster, more flexible, and above all more innovative," said Joschka Fischer, German Federal Minister for Foreign Affairs, in characterising the Federal Foreign Office reforms he aims to implement after 50 years of successful foreign policy on the part of the Federal Republic of Germany. Political goals are one thing, and creating the requisite organisational structure is another. Reform has always been an integral part of the process. So the Foreign Office has drawn on its experience in the administration and organisation of language services to convert its microfilm-based archives into a modern archival and content management solution for its translation section and its vast payroll management system, all in the space of just a few months.

Although Berlin is now the centre of German foreign policy, little could be accomplished without the organisational units that remain in Bonn. For example, payslips for some 8,000 individuals are handled in Bonn, and without a salary, not even the minister himself would be particularly flexible. Of course, the process involves the use of appropriate data processing solutions. An IT team in Bonn with around 150 staff ensures that systems are stable and run efficiently. The team has developed a broad array of its own applications, ranging from a software-supported worldwide visa issuance system to the set-up and administration of the extensive language services department.

6.2.1 Initial situation

Until late 1999, all pay records and language-services documents were archived on microfiche (and previously on microfilm). To provide faster, more flexible access to documents for payroll processing and workflow in the translation section, an electronic archive solution had to be implemented with the option of expanding into a comprehensive document management system (DMS), and with subsequent workflow integration. A small, limited DMS solution had been used for that purpose within the translation section since 1992; starting in 1996, the solution was based on FileNET technology. The aim was to use that solution as a starting point. Until then, the existing DMS solution had been used to manage translation texts (creation, saving, processing, archiving, retrieval, etc.). In practice, a staff member creates a document (the text of a speech, for example), saves it, and sends it to a translator. The translator translates the document, saves it, and then forwards it to a reviewer. Once the review and any necessary corrections are completed, the reviewer saves the document in the system.

When the document is needed, a secretary opens it, formats the text according to Federal Foreign Office regulations, prints out the required number of copies, and saves the formatted text as a new document. It remains absolutely essential that this final version of the text be stored in the archives, because that version is the only evidence that can settle any subsequent differences of opinion around the world that might arise concerning what was said, or other readings. In order to grant subsequent access to the documents or to make any changes, authorised work stations had to be equipped with the necessary microfilm printing equipment.

6.2.2 Selecting a System

Under Foreign Office policy, all payroll documentation was also filmed and housed in the archives in Bonn. Finding a single-provider payroll solution was vital for making the leap into the digital future. That automatically gave FileNET a shot at becoming the systems provider. The Panagon solution from FileNET was guaranteed to be available as the basis for the DMS solution and for the document archive. Moreover, the translation section had already been working successfully with them for years.

6.2.3 The Solution

Revision-proof archiving was a key requirement for the electronic archive system, particularly for payroll documents. Of course, additional benefits were also expected, since the existing archives already provided revision-proof storage. These further benefits included:

- faster access from any location,
- simultaneous access by multiple authorised individuals,
- greatly simplified searching (easy-to-learn retrieval system),
- mapping of filing structures that had developed within the Federal Foreign Office over many years,
- mapping of existing classifications,
- indexing based on readily adaptable index fields and keywords,
- possibility of multiple storage of documents (filing a document under several "folders" in the filing structure),
- working with documents in the archive system,
- cross-referencing to documents in the archive through additional archive nomenclature management (including numerical systems),
- long-term archiving, for example for 30-year contracts,
- custom-tailored user management, including various levels of access rights,
and
- user interface similar to Windows (Outlook or Explorer) to streamline familiarisation and promote acceptance.

FileNET's Panagon software was able to meet these requirements easily, with reasonable outlay for customisation. At the same time, the decision was made not to enter any of the documents from the existing microfiche archive into the new system.

One disadvantage of this approach was that microfilm printing equipment would temporarily continue to be necessary, but there would be no document conversion costs and the new system could start to be used as soon as installation, testing, and acceptance were complete. The old files "expire" automatically at the end of the statutory retention periods. Consequently, the majority of the files on microfiche can be destroyed over the course of the coming decade.

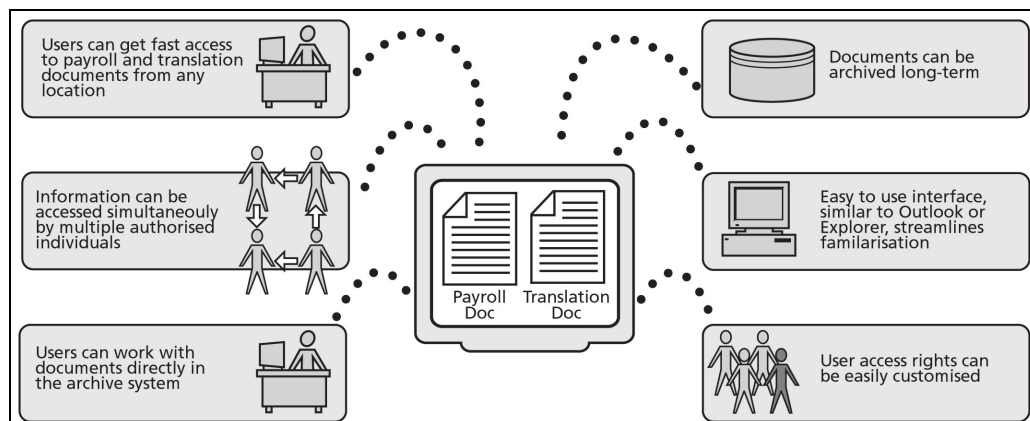See Figure 12 for an illustration of the German Federal Foreign Office solution.

Figure 12: German Federal Foreign Office solution

6.2.4   Current procedures

Today Berlin and Bonn work together closely, and individuals who access the system are unaware of where the infrastructure is physically located. The translation section is housed in Berlin, and its DMS server is located there, too. Through the Integrated Information Network, Berlin-based staff can access the archive housed in Bonn, which also contains all payroll documents. This solution meant that the system software had to be installed only once.

Work orders are now generated electronically in the translation section – the translator receives a work order and text by e-mail, makes the translation, and returns it to the reviewer by email, etc.

The procedures of the various payroll printouts are now almost entirely paperless. On the host computer, two main calculation runs are carried out each month, one for the salaried employees and one for the civil servants. The spools involved are up to 12 MB, and require a run time of only one-and-a-half to two minutes. The digitally generated documents are archived directly via the integrated COLD (Computer Output on Laser Disc) solution. This solution already complies with the new statutory requirements for archiving electronically generated documents effective as of January 1, 2002. Of course digitally generated documents are still printed out in practice, so that they can be signed by hand, for example. These printouts are then scanned and archived.

The decision was made to scan incoming documents later in the handling process. After preliminary sorting and distribution to the appropriate staff members, documents are indexed, scanned, and archived after processing. Generally speaking, the processing of

payroll information is a very sensitive topic. For that reason, and for reasons of special data protection, each staff member in the payroll department receives only a limited number of payroll cases to process. In other words, each payroll document "belongs" to a particular staff member, and only that staff member has the authorisation to access it. This authorisation, of course, is reflected in the indexing so that only specific individuals may access the archival documents.

The integrated archive solution for the payroll department and the translation section is stable and successful. User acceptance among the staff members is very high. Four work places have been set up for scanning and indexing. In-house staff are assigned to operate them. A registrar oversees the work and is specifically responsible for indexing. Outside support is brought in as needed for system maintenance.

In the translation section, an internationally unique solution was also implemented through an application developed in-house at the Federal Foreign Office. Documents in all languages that use the Roman alphabet are generated using the German keyboard layout, i.e. no country-specific keyboards are used as is otherwise common practice within the industry. This offers translators considerable advantages in terms of productivity and quality. They all deal with a number of languages and without this application specific to the Foreign Office, they would have to keep switching back and forth among the various keyboards.

6.2.5   Outlook

The comprehensive archive solution and the expandable DMS solution are, of course, not the end of the road. The next issue to be addressed is the integration of a workflow solution, moving toward earlier document scanning, with an inbox solution and simultaneous processing.

It is noteworthy that the Federal Foreign Office does not have any central filing facility, so no single integrated archiving or even DMS solution will be valid for all departments and areas. Filing systems are still almost exclusively paper-based and department-specific, which is why there are nearly 100 of these scattered archives. Consequently, the solution adopted in the payroll department and translation section could act as a guide for the Foreign Office as a whole, provided that the system is equipped to handle all the options that will be needed in the future, such as workflow integration.

The fact that these two areas of the Federal Foreign Office have such a modern archival and content management solution is significant, since both service providers within the Foreign Office have a strong international orientation.

6.3   Digital Processing for Tax Documents
Project of New Jersey Division of Revenue, USA

The State of New Jersey, USA, has an annual budget of $19.2 billion. The Division of Revenue is responsible for processing fees, remittances and tax forms for several state departments.

6.3.1   The Challenge

The New Jersey Division of Revenue needed a more efficient way to receive, process, store, and manage its large volume of incoming tax remittance documents including personal and business income tax forms.

6.3.2   The FileNET Solution

The Division of Revenue processes more than 3.5 million personal tax returns and handles five million coupon-sized remittance items each year. Physical storage for these documents was expensive and there was only enough on-site space for six months of returns. As documents aged, they were moved to less costly facilities, but document retrieval was lengthy, resulting in delays to taxpayer inquiries.
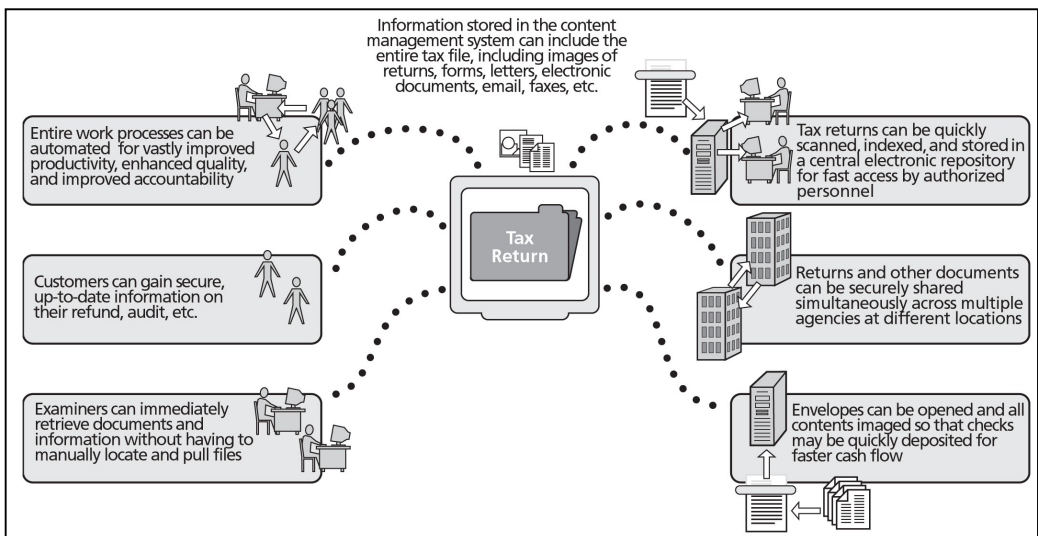
Figure 13: New Jersey Division of Revenue solution

In 1997, Governor Christie Whitman called for a more efficient, more business-friendly government. The Division of Revenue's mission was expanded to include document and fee processing for the Division of Motor Vehicles, the Department of Environmental Protection, the Division of Consumer Affairs, and the Department of Labor.

To manage this doubled workload, the Division looked to FileNET for a customer relationship management solution. FileNET deployed a Panagon™ system to assist the Division in managing its increased paperwork, ensuring fast customer response, and eliminating lost and misfiled documents.

See Figure 13 for an illustration of the NJ Division of Revenue solution.

6.3.3   The Bottom Line

Today, the Division of Revenue's image-enabled processing systems handle in excess of 633,000 pages daily during peak periods, while more than 250 FileNET users in seven

State agencies retrieve over 23,000 documents each day. The Division of Revenue has seen several improvements result from its FileNET implementation. Taxpayers now receive timely responses to their inquiries and are sent their refunds one week earlier, on average. Seasonal staffing and tax processing costs have both been reduced by 50 percent and storage costs have been cut in half. Due to these successes, the FileNET system has been expanded for use in all of New Jersey's revenue collection activities.

"Our long and successful partnership has made FileNET the leading document management vendor for the State of New Jersey." Nicholas Manocchio, Chief of Technology Services, Department of theTreasury, Division of Revenue.

6.4   Other Government Applications

6.4.1   Case Management for Revenues and Benefits, Social Services and Employment Services

Social Services departments promote self-sufficiency, attempt to put people to work and manage payments of social security and housing benefits. Paper-based folders, the repositories for critical case information, typically:

- restrict mobility and delay processes, limiting clients a caseworker helps each day;
- impede information sharing used to coordinate services to clients receiving benefits from multiple departments and agencies;
- adds expense for the distribution of paper and the storage space for large files.

CM can address challenges, and puts people back to work faster by:
- Improving caseworker productivity: electronic case folders help caseworkers see more people, process more applications and complete more tasks.
- Saving time: electronic folders reduce processing time and how long a client waits for service.
- Saving costs: electronic folders eliminate costs of processing, distributing and storing paper. It also reduces fraud and overpayment.
- Delivering demonstrable compliance: an e-Process framework, used around casework-based CM, helps case officers comply with complex regulatory environments, at the same time providing a thorough audit trail.

6.4.2   Engineering Change Management

Public Works Agencies, which are responsible for highways, bridges, waterways, water lines and sewers, manage countless drawings, maps, contracts and other documents. Multiple parties at different locations often need to modify and approve these documents before engineering changes can be executed. This leads to:

- low productivity of engineering resources due to slow access to critical documents;
- high cost of storing large-format maps and drawings;

- security issues involved with storing unique, highly-confidential documents in filing cabinets or on microfilm;
- version control and managing updates between multiple organisations with varying revision standards.

CM solutions help improve engineers' productivity and department efficiency by:
- Improving procurement processes: closes gaps between parties involved in planning process and executing procurements and changes of large capital projects.
- Enhancing document security: limits when and where approved parties can view and alter classified documents.
- Improving resource productivity: uses a common desktop interface to access a library, reduces time and expense of assembling drawings and documents from numerous locations.
- Reducing costs: automates the change management process, reducing costly, time-consuming, errors.

6.4.3   Tax Processing (Local and Central/Federal Government)

Processing millions of returns and requests for information from taxpayers, government departments and agencies, requires proficient document processing and customer service. Challenges include:

- Entering large amounts of data from extensive document backlogs and checking for accuracy and/or missing items
- Processing erroneous returns through the exception handing process requires access to source documents and multiple approvals.
- Providing customer relations personnel with information needed to answer requests for timely information.
- Depositing tax payment cheques in banks on time.

CM solutions help tax agencies address challenges and provide better service by:
- Improving Productivity and Accuracy: Electronic document entry, management and storage for tax processing purposes allow employees to work efficiently.
- Managing the Appeals Processes: Caseworkers can apply appropriate processes to improve productivity and demonstrable compliance.
- Improving Customer Service: Electronic document management speeds up processing and response times, freeing employees to serve customers.
- Reducing Costs: Allows employees access to information from any location, any time, eliminating need to securely store paper documents.
- Balancing Workload and Reducing Time: Multiple users have concurrent access, allowing non-linear processing of returns.
- Accelerating Deposits: Quick processing permits checks to be deposited quickly and begin earning interest.

6.4.4 Court Records

The need to serve justice leads to a tremendous importance on addressing cases in a timely manner. Ever-increasing caseloads and large volumes of paper-based documents place many challenges on the courts:

- Legal counsels (advocates and other lawyers/paralegals) create complex court briefs that often include hundreds of pages citing legal precedents, which increases storage demands.
- Scheduling difficulties and increased continuances when counsels are not given adequate notice before scheduled hearings, so they can evaluate deficiencies in their filings and/or prepare for hearings. For each continuance, case files are retrieved from storage, delivered to court and returned to storage, an expensive process that often results in misfiled documents.

CM solutions are helping the courts address these challenges by:
- Reducing costs: eliminating paper-based files reduces strain on storage facilities and cost of employees who must file, retrieve and re-file documents.
- Streamlining operations: automating processes associated with case documents, expedite a case's handling and allows interested parties to monitor the progress of any action associated with a case.
- Providing immediate access to case documents: electronic case management gives counsels and employees the power to access folders anywhere, any time, eliminating the need to search for and retrieve paper files.

6.4.5 Criminal Records

Historically, police work is a paper-intensive process. Field interviews and various crime reports are accumulated, from which the judicial system draws crucial data. Challenges abound:

- Increased labour by manually generating and processing reports on paper.
- Cost of storing and accessing documents, incident reports and offender information.
- Increased time and expense of gathering records and information from scattered sources, as well as accommodating information requests from various entities.

CM solutions help law enforcement agencies with challenges and keep communities safer by:
- Streamlining procedures: generating and submitting reports electronically to a common repository, making critical information available via desktop computer or mobile devices, allows investigations to proceed smoothly and timely.
- Increasing resource productivity: eliminating management of paper-based documents, such as manually generating reports, locating scattered files and re-filling, allows personnel to focus on tasks that impact community safety.

- Reducing costs: an electronic repository for managing reports and related documents reduces overall expenditures.

6.4.6   Immigration

The complexity of processing immigration policy, the requirement to execute law and policy accurately in a fast-changing environment, the political sensitivities of immigration issues, and the sheer volume of applications received all place massive demands on a secretariat, and on its case file processing and maintenance functions. The many challenges they face today include:

- handling the volume of applications and increasingly complex documents;
- converting paper files into electronic format, and the decision-making processes to determine which ones to convert, and when;
- allowing executive judgement within an overall framework or semi-structured workflow;
- providing decision support tools and guidance, along with access to knowledge bases, to the executives making the judgements;
- self-service capabilities so claimants can apply, check on the progress of applications and/or provide additional information electronically, as part of an e-Government initiative;

CM solutions help governments with these challenges and streamline sensitive immigration processes by:
- Streamlining procedures: generating and submitting reports electronically, while making critical information immediately available via PCs or mobile devices, allows application processing and investigations to proceed smoothly and swiftly.
- Improving caseworker productivity: electronic document management speeds up processing and response times, freeing employees to serve customers.
- Saving costs: electronic case folders eliminate processing, distribution and storage costs.
- Delivering demonstrable compliance: an e-process framework, around casework-based CM, empowers immigration case officers – while being monitored and audited – to comply with complex and ever-changing regulatory environments.

CM is revolutionising automatic document processing, tracking and archiving of constituent correspondences safely and in an efficient manner, allowing citizens to get secure access to their vital records through the Web. As more government departments and agencies manage and process information electronically and extend their processes with the Web, new applications for CM will come online.

7. Outlook

7.1 e-Government…

The move to e-Government is underway with full force. To take only three examples,

- In the UK, the government has a goal to offer 25% of its transactions (such as tax forms, benefits claims, etc) electronically by 2003, with the ultimate target of making all government services available electronically by 2005.
- In Germany, the government is aiming for 50% of services to be online by 2005.
- Ministers of the EU Member States recently agreed (e-Government Ministerial Conference, 29 November 2001) that government services must be delivered online as well as in traditional ways.

7.2 …and Beyond e-Government

CM-enabled government will include departments and agencies that operate more efficiently and effectively. As more and more sectors of government adopt CM technologies, the boundaries that have always existed among departments, between jurisdictions and across borders will begin to fade away. The fluid yet secure exchange of information and the linking of critical processes, will be the catalyst for a better government and, consequently, a better society as a whole.

This integration is a necessary step towards the next step beyond e-Government, Enterprise Electronic Government (E2 Government). A consortium of associations encouraging e-Government initiatives recently issued a blueprint promoting this concept. In it, P.K. Agarwal, CIO of NIC said "The public sector's promise should go beyond offering customer service and begin facilitating economic development – helping businesses and the public succeed."

Over 250 years ago, James Madison, U.S. President, wrote "A government deriving its energy from the will of the society ... is the government for which philosophy has been searching and humanity fighting from the most remote ages". Over time, and through pressures not of their own making, governments have grown to become distant from the publics they serve, and often difficult to access. It is through technologies like Content Management, with the direct and interactive interfaces they allow, that the original purpose of a government for the people may finally be realised.

7.3 e-Merging Technologies

New or improved applications advance CM value each day. While experiencing slow initial growth, collaborative CM systems are likely to be the most successful in coming months as organisations try to provide Yahoo-like functionality. CM systems can provide on-line selling and cross-selling of complex financial products, enabling Web sites to provide 24x7 customer service. ERP and supply chain software vendors are adding or connecting with CM products to provide and control unstructured content, as well as data to connect links in manufacturing supply chains. There is a developing

market for CM systems to support associations, not-for-profit groups that share common interests. Major IT training providers and academic organisations have well-developed systems to control the content and structure of e-Learning systems. And e-Catalogue is an emerging market for CM components purpose-built for supply chain applications, supporting online marketplaces and retail B2C sites.

Only visionaries have so far figured out what new applications Content Management can add to vertical markets. Most vendors can be expected to produce vertical market solutions over the short term, as they seek to increase the value of their generic products.

7.4   The Future of Content Management

There will be significant benefits to those governments that realise the vision of e-Government. They will reduce their costs, they will realise greater productivity from government employees, and they will be able to quickly respond to the increasing demands of the public they serve. They will, moreover, be able to fundamentally change the way government works.

A higher standard for productivity and operational efficiency begins by managing all content electronically. Only when this is achieved can government organisations, and the public they serve, begin to realise the true benefits of e-Government:

- Faster access to information, with more complete and more reliable retrieval, leading to faster and better decision making at all levels.
- Better management of information, thus reducing storage costs for millions of documents.
- Streamlining processes that are central to their operations and enhancing worker productivity at all levels.
- Improving service levels both to the public and to government employees.
- Providing self-service options to other departments and agencies and constituents, in order to facilitate faster access to information and a higher level of customer satisfaction.
- Collaborating better with other departments and agencies and jurisdictions to speed operations and provide better service.
- Overcoming the complexity and difficulty often associated with accessing services to reveal a truly transparent government.

CM has emerged a key enabler facilitating this transformation. It is the next evolution in document and content management, as well as process management technologies. Content Management solutions are meeting the immediate needs of today's emerging e-Government initiatives, and will serve as a critical component for better government in the future.

Glossary

ADL (Advanced Distributed Learning)	ADL is an initiative by the U.S. Department of Defence to achieve interoperability across computer and Internet-based learning courseware through the development of a common technical framework, which contains content in the form of reusable learning objects.
Associative Access	Knowledge retrieval based on pattern matching between an unstructured query (text paragraph) and a document content store.
Authoring tools	Tools/SW to create and adapt content to the web for use in an online course. They assist in creating e-learning solutions and provide a "do-it-yourself" option for placing content and materials online.
Categorization / Category	Assigning documents to different groups by performing content-related analysis - so called categories. Categorization schemes are typically built upon business processes and business rules or rely on knowledge domains within an organisation.
CD-ROM assessment	An assessment or survey that can be accessed and completed by using a CD-ROM launched through a company's intranet. CD-ROM based assessments also can be used on a desktop stand-alone computer if the assessment is a self-assessment for the benefit of the trainee only. Alternatively, a CD-ROM-based survey can be printed (if the CD-ROM has a print capability) and used as a paper-based survey.
Computer-based training	A term used to describe any computer-delivered training, including CD-ROM, the Internet and Intranets. Sometimes referred to as Computer-assisted instruction (CAI), CBT is asynchronous learning.
Classification / Class	Collection of methods applied to categorize documents by analysing their content. In many cases, categories and classes are identical. Categories incorporate the semantics of the application, whereas classes may also be of formal nature.
Classify	Classification is a method of assigning retention/disposition rules to records. Similar to the Declare function, this can be a completely manual process or process-driven, depending on the particular implementation. As a minimum, the user can be presented with a list of allowable file codes from a drop-down list (manual classification). Ideally, the desktop process/application can automate classification by triggering a file code selection from a property or characteristic of the process/application.
Content Search	Information retrieval based on pattern matching between a query (text paragraph) and a document repository.
Distance learning/ Interactive Distance Learning (IDL)	Traditionally refers to a broadcast of a lecture to distant locations, usually through video presentations. IDL is a real-time learning session where people in different locations can communicate with each other. Videoconferencing, audio conferencing or any live computer conferencing (e.g., chat rooms) are all examples of IDL.
Document	A document (any form or format), an email message or attachment, a document created within a desktop application such as MS Word, regardless of format. There are two forms of document: Electronic Document: Body (text) of the document is stored in electronic format and can be read. If declared as a record, an electronic document becomes a managed record (i.e. a document may or may not be a (declared) record) Non-Electronic Document (Ndoc): A physical document of any form (maps, paper, VHS video tapes, etc.). Body is not recorded in electronic form, but descriptive metadata is stored and tracked within CM (profile). If declared as a record, an Ndoc becomes a managed record (i.e. an Ndoc may or may not be a (declared) record).
Document Life Cycle Management	The records life cycle is the life span of a record from its creation or receipt to its final disposition. It is usually described in three stages: creation, maintenance and use, and final disposition. e-Records applies management to all three stages. With e-Records, the records manager can create and maintain the official rules that will dictate when to destroy (or permanently keep) electronic records, as well as record

	and enforce any conditions that apply to destruction (e.g. destroy 2 years following contract completion). Finally, the records manager can carry out the physical destruction of electronic records, maintaining a legal audit file.
Document Security Control	Access control to documents (non-declared records) Note: Document security control is different from Records Security Control.
Electronic Recordkeeping	The practice of applying formal corporate recordkeeping practices and methods to electronic documents (records).
Electronic Signature	A signature is a bit string that indicates whether or not certain terms occur in a document.
Enterprise Content Management	Manage all content (i.e. unstructured information) relevant to the organisation. It embraces three historically separate technologies: web content management, document management, and digital media asset management. While outwardly dissimilar, all of these forms of enterprise content share similar needs for mass storage, search and access, personalisation, integration with legacy applications, access and version control, and rapid delivery over the internet.
EPSS (electronic program support system)	An electronic system that provides integrated, on-demand access to information, advice, learning experiences and tools. In essence, the computer is providing coaching support (i.e. the principal of technology based knowledge management).
File	A disk "file", something stored on electronic media, of any file. Does not necessarily denote a record. For example, "image files are stored an a server" simply refers to the electronic images, and implies nothing about the records status. Will be used in the context of describing the storage of documents and related information to electronic media.
File Plan Administration	Design and administration of the corporate file plan. The records manager can design file plan components. With Tarian's file plan designer, the records manager can design classes of file plan objects (files, records, folders, etc), then define the attributes of these classes. Relationships between classes are then defined (i.e. files can contain files, records and folders). Various views of the file plan may be defined. For instance, a warehouse view might present a view of the physical folders in the organisation, whereas a numeric view might present the sorted numeric structure for maintenance purposes. The records manager can create pick-lists enforcing consistency within the file plan, component profiles that define the characteristics of the file plan, and default values to simplify daily file creation tasks. Policies, Permissions, and Suspensions may be assigned to file plan objects.
Information mining	Linguistic services to find hidden information in text documents on content servers
Information Retrieval	An information retrieval (IR) system informs on the existence (or non-existence) and origins of documents relating to the user's query. It does not inform (i.e. change the knowledge of) the user on the subject of his inquiry. This specifically excludes Question.
Keyword Search	Information retrieval method based on literal match of words.
Learning Resource interchange (LRN)	LRN is the Microsoft implementation of the IMS Content Packaging Specification. It consists of an XML-based schema and an LRN toolkit. It enables a standard method of description of content, making it easier to create, reuse and customize content objects with an XML editor, whether initially developed from scratch or bought under license from vendors.
Neural Networks	In information technology, a neural network is a system of programs and data structures that approximates the operation of the human brain. Typically, a neural network is initially "trained" or fed large amounts of data. A program can then tell the network how to behave in response to an external stimulus (for example, to classify a document based on its content).
Pattern Matching/ Recognition	Matching/Recognition of objects based on features. Pattern Matching with regard to text documents means to identify and match words and phrases from different documents under the assumption that the more features match, the more similar the contents are.
Personalisation	The ability to provide the user with the right content both from the user's and Web

	site owner's perspective. A personalisation algorithm determines whether content is presented to the user, and if so, in what order of priority.
Portal	A single integrated point of comprehensive, ubiquitous, and useful access to information (data), applications, and people.
Record	Any form of recorded information that is under records management control. Records are either Physical or Electronic. Records may take any of the following four forms: Document: A document (see above) that has been declared as a record. Once declared as a record, the document is under records management control Folder: A folder of (paper) documents. Individual documents within the folder may or may not be treated as records (declared Ndocs). The physical handling of the folder is managed by Tarian's Physical Records Module Box: A box of (typically) paper documents. Usually contains folders (see above), which are individually managed as records, but may alternatively contain records other than folders such as loose documents of a given subject. The physical handling of the box is managed by Tarian's Physical Records Module Ndoc: A declared Ndoc (See above for definition of Ndoc) Important: A document (electronic or Ndoc) will not be considered to be a record until has been declared.
Record, Electronic	Electronic Records (e-Records). Any information (document) recorded in electronic form, on any digital media, that has been Declared to be a record. Characteristics of an e-Record: Document is in electronic form Metadata is associated with the document Document has been classified against a file plan Only the authorized Records manager has the means by which to apply retention/disposition to the document.
Record, Physical	Folders, Boxes, Ndocs to which records management control has been applied. A document (electronic or Ndoc) becomes an e-Record only once it has been declared.
Records Administration	The administrative infrastructure represents the tasks that the records manager carries out on the entire organisation's collection of declared records. Conducted within Tarian's Records Administration Client, a browser-based web application. End users never see this process. Consists of the following four broad activities; File Plan Administration, Records Security Control, LifeCycle Management, and Reporting.
Records Manager	Conducts one or more records administrative functions.
Records Security Control	Access control to declared records. Users and Groups of users may be created, and assigned roles and policies that will interact to determine the records users are able to access. Note: Records security control is different from Document Security Control.
Reporting	The process of generating reports from data managed by eRecords solution. It is a tow-step process. Reports are first designed, and the design is saved for later reuse. Second, reports are generated by running the report design against the data.
Repository	Physical storage are for documents and/or electronic records.
Retention Rules	(Retention Schedule). The set of rules which specify how long to keep (retention) records, and what to do with them at the end of their lifecycle (disposition).
Syntactical Analysis	Syntactical analysis derives the syntactic category of words or phrases based on (language dependent) dictionaries and grammars. Example: house – noun.
Thesaurus	A book that lists words in groups of synonyms and related concepts.
Volume	Folder. A Volume will be referred to as a folder (common US terminology).
Virtual Reality (VR)	Virtual Reality simulations (usually involving wearing headgear and electronic gloves) that immerse users in a simulated reality that gives the sensation of being in a three-dimensional world.

Abbreviations

ASP	Application Service Provider
AVI	Audio Video Interleaving
BCR	Bar Coding
BPM	Business Process Management
CBT	Computer Based Training
CCD	Charge Couple Devices
CM	Content Management
COLD	Computer Output to Laser Disk
COM	Component Object Model
COOL	Computer Output On Line
DBMS	Database Management System
DMS	Document Management System
DRT	Document Related Technologies
ECM	Enterprise Content Management
E-Learning	Education, training and structured information delivered electronically
ERM	Enterprise Report Management
ERP	Enterprise Resource Planning
E-Term	European programme for Training in Electronic Records Management
FDDI	Fibre Distributed Data Interface
GIF	Graphic Interchange Format
HTML	Hypertext Mark-up Language
ICR	Intelligent Character Recognition
ICT	Information and Communication Technology
IDM	Integrated Document Management
ISDN	Integrated Services Digital Network
ISO	International Standards Organisation
JPEG	Joint Photographic Experts Group
KM	Knowledge Management
LAN	Local Area Network
LDAP	Lightweight Directory Access Protocol
MoReq	Model Requirements for the management of electronic records
MPEG	Moving Pictures Expert Group
NAS	Network Attached Storage
OCR	Optical Character Recognition
ODCB	Open Database Connectivity
OLE	Object Linking & Embedding
OMR	Optical Mark Recognition
PDF	Portable Document Format
PPP	Point-to-Point Protocol
RMS	Records Management System
RTF	Rich Text Format
SAN	Storage Area Networks
SQL	Structured Query Language
TCP/IP	Transmission Control Protocol/Internet Protocol
TIFF	Tag Image File Format
WAN	Wide Area Network
WAV	Audio Format File
WCM	Web Content Management
WebDAV	Web-based Distributed Authoring & Versioning
WORM	Right once read many times
XML	eXtensible Mark-up Language

References

AIIMAssociation for Information and Image Management International,	Implementation Guidelines and Standards Associated with Web-Based Document Management Technologies, AIIM Recommended Practice, 2001
Federal Computer Week	December 13, 2001
Mazza, Edward	X Marks the Spot, Government Technology, October 2000
Strategy Partners	Content Management – Europe 2001: Content Management for eBusiness, 2001
Vernon, Mark	The Executive Guide to Content Management, AIIM and IBM, 2001

Authoring Company

FileNET

FileNET provides Enterprise Content Management (ECM) solutions that improve operational efficiency and leverage information resources. Many recognise FileNET as a long-term leader in the ECM market place because of its strong legacy in content management, its ongoing innovation in eProcess, and its proven and substantial technology offerings.

FileNET markets and supports its innovative ECM solutions in more than 90 countries through its own global sales, professional services and support organisations, as well as through its ValueNET® Partner network of resellers, system integrators, and application developers.

Today, FileNET's ECM solutions are helping government departments and agencies around the world to improve productivity, reduce costs, and better serve the public.

FileNET Iberia
Avda. Del Partenón, 10 – 3 ^{a} planta, Campo de las Naciones, 28042 Madrid (SPAIN)
Phone: +34(0)91 7214980

FileNET GmbH
Dietrich-Bonhoeffer-Strasse 4, D-61350 Bad Homburg v.d.H., Germany
Phone: +49 (0)6172/963-0

FileNET France
1 Rue de La Cristallerie, 92316 Sevres Cedex.France
Phone: +33(0)1.46.23.76.76

See http://www.filenet.com for full details of FileNET's global & European offices.

Contact Authoring Company

FileNET Europe North
Ingrid O'Brien
1 The Square, Stockley Park, Uxbridge. UB11 1FN. - UK
Tel.: +44(0)20 8867 6341
E-Mail: Euro.north@filenet.com

Industry White Paper Series on Records, Document and Enterprise Content Management for the Public Sector

Capture, Indexing & Auto-Categorization
Intelligent methods for the acquisition and retrieval of information stored in digital archives
ISBN 3-936534-01-2

Hewlett-Packard GmbH
Conversion & Document Formats
Backfile conversion and format issues for information stored in digital archives
ISBN 3-936534-02-0

FileNET Corporation
Content Management
Managing the Lifecycle of Information
ISBN 3-936534-03-9

IBM
Access & Protection
Managing Open Access & Information Protection
ISBN 3-936534-04-7

Eastman Kodak
Availability & Preservation
Long-term Availability & Preservation of digital information
ISBN 3-936534-05-5

TRW Systems Europe / UCL - University College London / comunicando spa
Education, Training & Operation
From the Traditional Archivist to the Information Manager
ISBN 3-936534-07-1

Publishing Information

The series of six Industry White Papers is published to address the needs of public administration and archives at the European, national, federal and local level and to educate the public sector throughout Europe about available solutions for archival problems on relevant topics about acquisition, management, long term storage, multilingual access, indexing and training issues.

DLM-Forum
The current DLM acronym stands for *Données Lisibles par Machine* (Machine Readable Data). It is proposed that after the DLM-Forum 2002 in Barcelona this definition be broadened to embrace the complete "**D**ocument **L**ifecycle **M**anagement". The DLM-Forum is based on the conclusions of the Council of the European Union, concerning greater co-operation in the field of archives (17 June 1994). The DLM-Forum 2002 in Barcelona will be the third multidisciplinary European DLM-Forum on electronic records to be organised. It will build on the challenge that the second DLM-Forum in 1999 issued to the ICT (Information, Communications & Technology) industry to identify and provide practical solutions for electronic document and records management The task of safeguarding and ensuring the continued accessibility of the European archival heritage in the context of the Information Society is the primary concern of the DLM-Forum on Electronic Records. The DLM-Forum asks industry to actively participate in the multidisciplinary effort aimed at safeguarding and rendering accessible archives as the memory of the Information Society and to improve and develop products to this end in collaboration with the users.

European Commission SG.B.3
Office JECL 3/36, Rue de la Loi 200, B-1049 Brussels, Belgium
A/e: dlm-forum@cec.eu.int

AIIM International - The Enterprise Content Management Association
AIIM International is the leading global industry association that connects the communities of users and suppliers of Enterprise Content Management. A neutral and unbiased source of information, AIIM International produces educational, solution-oriented events and conferences, provides up-to-the-minute industry information through publications and its industry web portal, and is an ANSI/ISO-accredited standards developer.
AIIM Europe is member of the DLM-Monitoring Committee and co-ordinates the activities of the DLM/ICT-Working Group.

AIIM International, Europe
Chappell House, The Green, Datchet, Berkshire SL3 9EH, UK
http://www.aiim.org

Industry White Paper Series on Records, Document and Enterprise Content Management for the Public Sector

The Industry White Papers are published by the DLM-Forum of the European Commission and AIIM International Europe to address the needs of public administration and archives at the European, national, federal and local level and to educate the public sector throughout Europe about available solutions for archival problems on relevant topics about acquisition, management, long term storage, multilingual access, indexing and training issues. The leading suppliers of Enterprise Content Management technologies participate in this series and focus on electronic archival, document management and records management for the public sector in the European Community.

Content Management – Managing the Lifecycle Of Information
This paper defines content management and the various technologies it embraces. It examines the differences between several content management architectures and the different types of solutions being deployed today.
The paper explains the different functionalities included in content management solutions and outlines the relevant standardisation bodies, definitions and technologies.
The mentioned best practice applications feature examples from both the private and public sector. It forecasts the future of content management and identifies possible trends and developments.