

Document Management

DOCUMENT MANAGEMENT

A document management system (DMS) is a computer system (or set of computer programs) used to track and store electronic documents and/or images of paper documents. The term has some overlap with the concepts of Content Management Systems and is often viewed as a component of Enterprise Content Management Systems and related to Digital Asset Management, Document imaging, Workflow systems and Records Management systems. Contract Management and Contract Lifecycle Management (CLM) can be viewed as either components or implementations of ECM.

Overview

In the broadest sense, document management systems can range from "the shoebox" to ECM. There are several common issues that are involved in managing documents, whether the system is an informal, ad-hoc, paper-based method for one person or if it is a formal, structured, computer enhanced system for many people across multiple offices. Most methods for managing documents address the following areas:

Location

Where will documents be stored? Where will people need to go to access documents? Physical journeys to filing cabinets and file rooms are analogous to the onscreen navigation required to use a document management system.

Filing

How will documents be filed? What methods will be used to organize or index the documents to assist in later retrieval? Document management systems will typically use a database to store filing information.

Retrieval

How will documents be found? Typically, retrieval encompasses both browsing through documents and searching for specific information.

Security

How will documents be kept secure? How will unauthorized personnel be prevented from reading, modifying or destroying documents?

Disaster Recovery

How can documents are recovered in case of destruction from fires, floods or natural disasters?

Retention period

How long should documents be kept, i.e. retained? As organizations grow and regulations increase, informal guidelines for keeping various types of documents give way to more formal Records Management practices.

Archiving

How can documents be preserved for future readability?

Distribution

How can documents be available to the people that need them?

Workflow

If documents need to pass from one person to another, what are the rules for how their work should flow?

Document Management

Creation

How are documents created? This question becomes important when multiple people need to collaborate, and the logistics of version control and authoring arise.

Authentication

Is there a way to vouch for the authenticity of a document?

History

Beginning in the 1980s, a number of vendors began developing systems to manage paper-based documents. These systems managed paper documents, which included not only printed and published documents, but also photos, prints, etc.

Later, a second system was developed, to manage electronic documents, i.e., all those documents, or files, created on computers, and often stored on local user file systems. The earliest electronic document management (EDM) systems were either developed to manage proprietary file types, or a limited number of file formats. Many of these systems were later referred to as document imaging systems, because the main capabilities were capture, storage, indexing and retrieval of image file formats. These systems enabled an organization to capture faxes and forms, save copies of the documents as images, and store the image files in the repository for security and quick retrieval (retrieval was possible because the system handled the extraction of the text from the document as it was captured, and the text indexer provided text retrieval capabilities).

EDM systems evolved to where the system was able to manage any type of file format that could be stored on the network. The applications grew to encompass electronic documents, collaboration tools, security, and auditing capabilities...

Document Management and Communication

Electronic document management is in particular worked out by Carzaniga and Wolf (2001) in their paper "Content-based networking: a new communication infrastructure". The authors introduce content-based networking as a communication infrastructure where information is driven by the content throughout the network. The users express their interests, and the senders simply input the message into the network. From that point the network delivers all the information to the right people. Sprague (1995) delivers a more elaborate work in which he introduces document management through using IT. He calls it electronic document management: EDM. He defines managing of documents as the "creation, storage, organization, transmission, retrieval, manipulation, update, and eventual disposition of documents to fulfill an organizational purpose" (pp.32), and he further states that EDM improves communication among people and groups of people (pp 42-43).

There are several other examples from the literature for the link between EDM and communication. Hansen and Haas (2001) elaborate on the role of the suppliers and users of information in electronic documents. Another research with a very clear link between EDMS and communication is that of Thorpe and Mead (2000). They showed that an EDM system changes the communication patterns. Of the three case projects they researched, EDM acquired a central role in two of them, (the third project was abandoned after three months). A research of Howard and Pettersen (2001) about the way of communicating in a construction project had as result that EDM (Howard and Pettersen call it project web) was number three communication tool just after telephone and a meeting, leaving e-mail, paper-post and fax behind. Rene Brohm (2005) introduced in his dissertation the theater model. The theater model illustrates metaphorically how document management systems correspond with a stage in a theater. His argumentation is that the interaction in a play on the stage is similar with the functioning of a document system.

Document Management

If all the data and information would be put in a central database/intranet, which can be used by everyone in the organization, there would be a clear link between IT and dissemination of information according to Marin & Poulter (2004). They argue that because of the easy access to the information, it would flow through the organization. The authors confirm this in their paper (2004) by stating that distribution of intelligence can be aided by technology.

There are different ways of improving this communication tool. Hansen and Haas (2001) see the electronic document management as a market, with competition. According to them, suppliers should have a strategy about how to share information and how to persuade their clients (employees) to use the system

One way to do this is introduced by Yan & Garcia-Molina (1999 pp.2) who use EDM to: "make long term profile consisting of a number of standing queries to represent his information needs". Through this they state that dissemination of information is improved. Users receive information in their field of interest because of a profile that was submitted. Therefore search costs and search time for employees are decreased.

Components

Document management systems commonly provide storage, versioning, metadata, security, as well as indexing and retrieval capabilities. Here is a description of these components:

Metadata

Metadata is typically stored for each document. Metadata may, for example, include the date the document was stored and the identity of the user storing it. The DMS may also extract metadata from the document automatically or prompt the user to add metadata. Some systems also use optical character recognition on scanned images, or perform text extraction on electronic documents. The resulting extracted text can be used to assist users in locating documents by identifying probable keywords or providing for full text search capability, or can be used on its own. Extracted text can also be stored as a component of metadata, stored with the image, or separately as a source for searching document collections.

Integration

Many document management systems attempt to integrate document management directly into other applications, so that users may retrieve existing documents directly from the document management system repository, make changes, and save the changed document back to the repository as a new version, all without leaving the application. Such integration is commonly available for office suites and e-mail or collaboration/groupware software. Integration often uses open standards such as ODMA, LDAP, WebDAV and SOAP to allow integration with other software and compliance with internal controls. [Citation needed]

Capture

Images of paper documents using scanners or multifunction printers are known as capture. Optical Character Recognition (OCR) software is often used, whether integrated into the hardware or as stand-alone software, in order to convert digital images into machine readable text.

Indexing

Track electronic documents. Indexing may be as simple as keeping track of unique document identifiers; but often it takes a more complex form, providing classification through the documents' metadata or even through word indexes extracted from the documents' contents. Indexing exists mainly to support retrieval. One area of critical importance for rapid retrieval is the creation of an index topology.

Storage

Document Management

Storage can be summarized by Store electronic documents. Storage of the documents often includes management of those same documents; where they are stored, for how long, migration of the documents from one storage media to another (Hierarchical storage management) and eventual document destruction.

Retrieval

Retrieve the electronic documents from the storage. Although the notion of retrieving a particular document is simple, retrieval in the electronic context can be quite complex and powerful. Simple retrieval of individual documents can be supported by allowing the user to specify the unique document identifier, and having the system use the basic index (or a non-indexed query on its data store) to retrieve the document. More flexible retrieval allows the user to specify partial search terms involving the document identifier and/or parts of the expected metadata. This would typically return a list of documents which match the user's search terms. Some systems provide the capability to specify a Boolean expression containing multiple keywords or example phrases expected to exist within the documents' contents. The retrieval for this kind of query may be supported by previously-built indexes, or may perform more time-consuming searches through the documents' contents to return a list of the potentially relevant documents. See also Document retrieval.

Distribution

Security

Document security is vital in many document management applications. Compliance requirements for certain documents can be quite complex depending on the type of documents. For instance the Health Insurance Portability and Accountability Act (HIPAA) requirements dictate that medical documents have certain security requirements. Some document management systems have a rights management module that allows an administrator to give access to documents based on type to only certain people or groups of people.

Workflow

Workflow is a complex problem and some document management systems have a built in workflow module. There are different types of workflow. Usage depends on the environment the EDMS is applied to. Manual workflow requires a user to view the document and decide who to send it to. Rules-based workflow allows an administrator to create a rule that dictates the flow of the document through an organization: for instance, an invoice passes through an approval process and then is routed to the accounts payable department. Dynamic rules allow for branches to be created in a workflow process. A simple example would be to enter an invoice amount and if the amount is lower than a certain set amount, it follows different routes through the organization.

Collaboration

Collaboration should be inherent in an EDMS. Documents should be capable of being retrieved by an authorized user and worked on. Access should be blocked to other users while work is being performed on the document.

Versioning

Versioning is a process by which documents are checked in or out of the document management system, allowing users to retrieve previous versions and to continue work from a selected point. Versioning is useful for documents that change over time and require updating, but it may be necessary to go back to a previous copy.