| **Component** | **Description** |
| --- | --- |
| **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. |
| **Capture** | Capture primarily involves accepting and processing images of paper documents from scanners or multifunction printers. 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. Optical mark recognition (OMR) software is sometimes used to extract values of check-boxes or bubbles. Capture may also involve accepting electronic documents and other computer-based files. |
| **Validation** | Visual validation registration system and important data. E.g. document failures, lack of bells,[[clarification needed](http://en.wikipedia.org/wiki/Wikipedia:Please_clarify)] missing signatures, misspelled names, this can be printed on paper documents or images on paper. |
| **Indexing** | Indexing tracks 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** | 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** | A published document for distribution has to be in a format that cannot be easily altered. As a common practice in law regulated industries, an original master copy of the document is usually never used for distribution other than archiving. If a document is to be distributed electronically in a regulatory environment, then the equipment tasking the job has to be quality endorsed AND validated. Similarly quality endorsed electronic distribution carriers have to be used. This approach applies to both of the systems by which the document is to be inter-exchanged, if the integrity of the document is highly in demand. |
| **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, in the United States, 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. Document marking at the time of printing or PDF-creation is an essential element to preclude alteration or unintended use. |
| **Workflow** | Workflow is a complex process and some document management systems have a built-in workflow module. There are different types of workflow. Usage depends on the environment to which the electronic document management system (EDMS) is applied. Manual workflow requires a user to view the document and decide whom 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. Advanced workflow mechanisms can manipulate content or signal external processes while these rules are in effect. |
| **Collaboration** | Collaboration should be inherent in an EDMS. In its basic form, a collaborative EDMS should allow documents to be retrieved and worked on by an authorized user. Access should be blocked to other users while work is being performed on the document. Other advanced forms of collaboration allow multiple users to view and modify (or mark-up) a document at the same time in a collaboration session. The resulting document should be viewable in its final shape, while also storing the mark-ups done by each individual user during the collaboration session. |
| **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 or reference a previous copy. |
| **Searching** | Searching finds documents and folders using template attributes or full text search. Documents can be searched using various attributes and document content. |
| Publishing | Publishing a document involves the procedures of proofreading, peer or public reviewing, authorizing, printing and approving etc. Those steps ensure prudence and logical thinking. Any careless handling may result in the inaccuracy of the document and therefore mislead or upset its users and readers. In law regulated industries, some of the procedures have to be completed as evidenced by their corresponding signatures and the date(s) on which the document was signed. Refer to the [ISO](http://en.wikipedia.org/wiki/International_Organization_for_Standardization) divisions of ICS 01.140.40 and 35.240.30 for further information.  The published document should be in a format that is not easily altered without a specific knowledge or tools, and yet it is read-only or portable. |
| **Reproduction** | Document/image reproduction is key when thinking about implementing a system. It's great to be able to put things in, but how are you going to get them out? An example of this is building plans. How will plans be scanned and scale be retained when printed? |