



*The Enterprise Content Management solution
built on world-class CRM technology*

DECIDING ON DOCUMENT MANAGEMENT

Part I: CONCEPTS FOR BEGINNING ENTERPRISE CONTENT MANAGEMENT

A brief three-part primer on common issues related to implementing file management on a business system.

By Robert L. Gilligan, President, GoldVision Software Corp

Deciding On Document Management - Part I:

CONCEPTS FOR BEGINNING ENTERPRISE CONTENT MANAGEMENT

Overview of ECM

Enterprise Content Management [ECM] is the “umbrella” description for the managing of all forms of unstructured information throughout a business enterprise, including Document Management. Today, the *majority of business information* fits this description of “unstructured”, in that each item is a *composite* of many pieces of information [e.g., correspondence, spreadsheets, billings, sales materials, web pages, financial records, etc.] that are stored in single containers [individual files].

Unlike structured information such as City, State, Zip, Phone Number, Amount, etc., unstructured information is not easily described in a single value, because by its nature it is a “container” of multiple values from often disparate subject categories. [For example, a letter addressed to a person at a company, discussing a sales transaction, for a list of products, expected for a particular delivery date, being shipped by a specific transportation carrier, eligible for a certain discount rate, sent by another person at another company.] Organizing this type of information requires multiple criteria/reference points.

In the information-based world we live in, competition, regulation, as well as litigation, are all motivating businesses to bring this vast amount of valuable amorphous “content” into some form of managed accessibility. Technology has made this possible for decades, but only recently has it become affordable for the average company. And just in time...

Having all your information in an electronic library has numerous important advantages over traditional storage. Some of the most obvious are:

- A small fraction of the cost of traditional storage/duplication [100,000+ pages for \$1]
- Instant access to any piece of information
- Simultaneous access to the same item by multiple people when needed
- Instant transmittal to another party at virtually zero-cost [e-mail]
- Automated duplication of the entire library in virtually transparent/unattended processes [backup]
- Offsite duplication/storage of the entire library at a small fraction of the cost of traditional methods
- Immediate restoration of the entire library if needed, and also at minimal cost

Getting Started

While scanning/imaging is only one element of Document Management [and Document Management is only one element of ECM], scanning is the activity most often referred to interchangeably with Document Management by those starting the task of capturing/storing/managing/accessing their unstructured information. When beginning to manage unstructured information, paper files are an obvious candidate; however, they are only one aspect. While paper does represent the oldest “information technology” still in use in a typical business, most paper used today is a printout of something that already exists in an electronic source.

With the elimination of the typewriter and its replacement by the printer, virtually all information produced on paper today has an electronic source. That source could produce an electronic file instead-of/as-a-copy-of the paper output. Capturing the information first as an electronic file for storage in the electronic library significantly streamlines the entire ECM process. Even externally-generated information can ever-more-frequently be requested in electronic form instead-of/in-addition-to paper [e.g., bank statements, invoices, e-mail-instead –of-paper, etc.] Therefore, managing unstructured information should begin with the intentional and organized storing of the already-electronic files used in a business.

There is another important reason for starting at this point. Ultimately your goal is to have all your information available in an electronic format. Therefore it makes sense to begin with planning how all electronic files will be managed. After that, it is simply a matter of converting the non-electronic files to an electronic [same] basis as all your other files, and managing them all the same way. This avoids the cart-

before-the-horse-trap of designing your electronic file system around non-electronic files. It is a very common mistake, and requires expensive changes as soon as it is apparent.

Implementing Image-Capture – Scanning and Faxing

Most often the papers that companies want to begin scanning are the historical files of their business, where the electronic source is not available. This is usually the next step after setting in place processes to manage your electronic files.

Give some thought to what extent you need access to very old files. If access is unlikely other than under an emergency condition [e.g., unforeseen lawsuit] you may minimize the overall process by sending out your old file drawers to have each one bulk-stored to a CD/DVD. This still provides the advantages of eliminating costly paper copies, and electronic access/transmittal when needed.

Scanned paper, as well as faxes, represent “pictures” [images] of the source document. The scanner and the fax machine both convert the “picture” of the source document into tiny “dots” that can be represented as digital values. Now that the source information is in a digital form, it can be transmitted and/or stored electronically.

Faxing

Faxing has two components: sending and receiving. If your source document is electronic [Word file, digital picture, spreadsheet, etc.] you can both send and receive faxes from your computer with even just an inexpensive intelligent fax modem [\$25]. If you need to send documents that only exist in paper form, then you will also need hardware to scan the paper into the fax process. The best way to handle this is to scan the paper [in a scanner], making a digital file, and then “print” the digital file out to the fax modem. [A fax modem usually appears as another printer on your system. Sending a document to the fax “printer” opens the fax software to configure the recipient information.]

Alternatively, you might use a traditional paper-feed fax machine to send the outgoing document. In any event, receiving [incoming] faxes should always be handled by a fax modem attached to your computer, so that the digital file it receives is automatically stored on your system, without requiring additional conversion. If a paper copy of the fax is needed in your workflow, it can be easily printed from this electronic file. Once again, the source file is already electronic; manage it that way.

Scanning

Every scanner comes with interface software [referred to as a “driver”] to output an electronic image file onto your system, representing the scanned item. The most common types of files produced are TIF and PDF. Many scanner manufacturers include Nuance’s “PaperPort” as this free interface software. As with Adobe Acrobat, PaperPort can also be used to view a PDF file. However, similar to Microsoft Office, these programs are for viewing and editing files, they are not ECM solutions. Once the image file has been produced, it can be processed for long-term storage/access just like every other electronic file on the system.

Storing/Indexing

Managed Storing and Indexing are the most important processes in establishing an ECM system. Managed Storing assures the physical file is available; Indexing assures it can be found. Each is useless without the other; so both are usually accommodated in the same process.

Managed Storage

The managed process of storing files should address not only the individual physical file, but the subsequent need for files to be added in the future. The system should not only accommodate, it should also initiate and manage, placing each file in the appropriate physical location *without user intervention*.

The storage location must:

- be accessible to every user that needs the files [network location]
- have an efficiently-scalable structure to accommodate the ever-increasing number of files in a business
- be organized to accommodate eventual archive/removal of files when no longer needed
- facilitate scheduled backups/archives

- be restricted to authorized access
- NOT be a Binary Large Object [BLOB] structure
- NOT be a proprietary location/method limited to access by one vendor's product
- NOT be an elaborate descriptive hierarchy structure, a pseudo-database

Indexing

The process of indexing a file is where the access to it is organized. It is where the content is categorized/cross referenced with terms to make it instantly accessible in the future. Therefore, when planning your storage/indexing process, it is important to anticipate how do you want to be able to find a file when you need it? The expected retrieval criteria determines the required indexing criteria.

The most important association[s] made when indexing a file are the entity/event[s] that correspond to the content. Often these entity/events are not described within the content itself. This is why key-word indexing based solely on file-content is an inadequate method of indexing content for business.

For example, while storing correspondence related to an order, in addition to *file-content* indexing, the file should also be *entity/event*-indexed by items such as the customer name, the order number, and the sales rep. Not all these items might appear in each file of related correspondence, but all are useful criteria for retrieving the file in the future. They also allow all related files to be retrieved by those common items/criteria. Other entity/event items may be process-related criteria such as: who stored the file, what department do they work in, when was it stored, when was it last accessed, is it needed frequently, how was it found? The automatic capture of these attributes allow for later access based on these criteria.

In addition, *generalized* indexing improves searching when more specific criteria are unknown, or larger numbers of files are needed. Including general classifications, e.g. categorizing as "General Correspondence", or "Notes To File", "Financials", "Specifications", etc. effectively "groups" files in familiar categories. These are particularly useful when used in conjunction with other search criteria, such as Company or Contact name.

Indexing must be consistent, employing naming conventions that are both specific and familiar to the users of the information. Businesses generally establish their own terminology for processes and items; that terminology is often evident in the paper-based filing systems that contain their historical records. Adopting this terminology and adapting it to the electronic environment continues familiar classifications while introducing new capabilities to the users of the system.

Accessing

The second most important process is retrieving information when it is needed. Files should be accessible through multiple access points, as by their nature they represent multiple data elements. The methods for retrieval should also anticipate the user not knowing what they need. When it comes to information systems, if you have to already know the information you need, you don't have a system. The purpose of a system is to provide you with what you need even when you don't know where/what it is.

An important aspect of information design is the ability to coordinate data of different types so that it is inter-related. The most inter-related design is building everything in the same database. In this way, maintenance is minimized, and related data is not out of date. Separate databases have minimal relationship, and so searching in one for information in the other is very limited. As discussed below, the extensiveness of this inter-relationship is extremely valuable in an information system. Keep in mind that the purpose of these systems is information; the more the information is inter-related, the more value it has in your business.

More than just historical information, another valuable use of content is in identifying associated or inter-related information. Analysis of inter-related information is one of the advanced methods of data-mining. This uses a content search as a proxy for the event surrounding a document. For example, an invoice represents a sale. Searching for invoices within a specific date range identifies customers likely to need to re-order a product. Likewise, searching for resumes that mention specific skills identifies candidates for a recruitment request. Or searching for service complaints might identify a zip code that needs service improvement/repair.