NJSBA
PRACTICE**HQ**
PREPARED BY
▲ Affinity
CONSULTING

DOCUMENT MANAGEMENT AND PAPERLESS TERMS YOU NEED TO KNOW

IMAGING

A scanned document is essentially a digital photograph of the original paper document. Imaging is the process of using a scanner and computer to capture (scan), store, manipulate, and display an image (an image is a visual representation of something). In document imaging, the emphasis is on capturing, storing, and retrieving information from the images (which are often mainly images of text).

DIGITAL PAPER

This is the product of imaging. After you scan paper documents, you end up with electronic copies of them or "digital paper." They look exactly like the original documents from which they were created and can be printed again if a paper copy is required. However, the digital version is stored on your network and is only printed if necessary. Since digital paper is stored on your network, it's much easier to find.

SCANNER

A scanner is a device connected to a computer which captures images from photographs, posters, magazine pages, and similar sources for computer editing and display. A scanner functions similar to a copier and comes in hand-held, sheet-fed, and flat-bed types and for scanning black-and-white only or color. Very high-resolution scanners are used for scanning for high-resolution printing, but lower resolution scanners are adequate for capturing images for computer display. Scanners usually come with software which lets you re-size and otherwise modify a captured image.

TWAIN – THE COMMON LANGUAGE OF SCANNERS

TWAIN is a widely used program that lets you scan an image (using a scanner) directly into the software program you're using to store or manipulate the image. The TWAIN driver runs between a software program and the scanner hardware. TWAIN usually comes as part of the software package you get when you buy a scanner. TWAIN was developed by a work group from major scanner manufacturers and scanning software developers and is now an industry standard. I've read several articles falsely claiming that TWAIN is an acronym developed from "technology without an important name."

However, the TWAIN Working Group says that, after the name chosen originally turned out to be already trademarked, the group came up with TWAIN, deriving it from the saying "Ne'er the twain shall meet," because the program resides between the hardware and the software.

OCR (OPTICAL CHARACTER RECOGNITION)

OCR is the recognition of printed or written text characters by a computer. Advanced OCR systems can recognize hand printing, but most of them only interpret machine print. When a text document is scanned into the computer, it is turned into a bitmap, which is a picture of the text. OCR software analyzes the light and dark areas of the bitmap in order to identify each alphabetic letter and numeric digit. When it recognizes a character, it converts it into ASCII text. Hand printing is much more difficult to analyze than machine-printed characters. Old, worn and smudged documents are also difficult. OCR is currently in use by libraries to digitize and preserve their holdings; OCR is also used to process checks, credit card slips and sort the mail.

PDF (PORTABLE DOCUMENT FORMAT) FILES

PDF is a file format that captures all elements of a printed document as an electronic image that you can view, navigate, print, or forward to someone else. PDF files are created using Adobe Acrobat, Acrobat Capture, or similar products. To view and use the files, you need the free Acrobat Reader, which you can easily download for free (www.adobe.com). Once you've downloaded the Reader, it will start automatically whenever you want to look at a PDF file. PDF files have become the de-facto standard method for distributing electronic forms on the Internet. This is the format we recommend that you create your digital files in.

FILE

For purposes of this discussion, we're going to use the term "File" to mean any collection of documents and matter-specific information. That information can be stored in paper files or electronically as digital paper or database records.

DOCUMENT MANAGEMENT ("DM")

DM is the process by which we store, classify, search, share, and eventually retrieve our documents. Before going further, this would be a good place to define what we mean by "document." In the context of DM, a document is essentially a file. A file, in this usage, is an electronic, digital container of information. A document may be a word processing file, or it may be a graphic image, or any other discrete, identifiable information unit that can exist within a computer system.

DOCUMENT MANAGEMENT SYSTEMS ("DMS")

Document management systems (DMS) are simply hardware/software systems that automate the DM process. Specifically, a DMS provides an organization with the tools to create, manage, control, and distribute electronic documents.

Traditionally, operating systems such as DOS and Windows have failed to offer the tools and resources necessary for managing documents. The principal case in point is the paltry 8+3 file naming constraints enforced by DOS and Windows 3.x. Not until Windows 95 did the Microsoft+Intel platform offer the possibility of long, "descriptive" file names.

Understandably, the tools used to create and distribute files—word processors, spreadsheets, graphics programs, and the like—have concentrated on their core functionality, leaving document management to the operating system. Because the operating system wasn't very good at this, document management issues were largely left unaddressed. Companies (such as law firms) that create huge numbers of documents, and that have invested their intellectual capital in the content of these documents, often turn to document management software to overcome this problem. As more "corporate memory" is captured in electronic documents, more firms are recognizing the need for a document management system.

Consider that most department managers have a much better idea of the contents of their supply cabinets than they do of the electronic documents generated by their group. We're talking about the critical intellectual assets upon which their business relies. Clearly, there is a problem here.