

# **Why We All Hate Our Scanners: The Paper LESS Office™ is the Answer**

**By**

**Ross L. Kodner, Esq.**

©2004 Ross L. Kodner, All Rights Reserved

Scanning has been a sore spot in law firms for many years. Why? Lawyers have viewed scanning as being synonymous with optical character recognition (OCR). The problem? Even with the best OCR products, results often fall short. Many documents are not good "candidates" for recognition. Without a clean laser-printed source document you'll end up with gobbledygook. Your staff will tell you it would have been faster to type the document than to OCR it and have to clean up the resulting mess.

Instead, view scanning as a way to turn physical paper into digital paper. This is like photocopying the documents onto one's computer screen. When scanning as images, the process can be 20 times faster than the processing-intensive OCR approach. Further, imaged documents on screen look **PRECISELY** like the originals: handwriting, pre-printed lines / boxes: all scan perfectly. This is a core part of the concept that I call the Paper LESS Office(tm) (see [www.microlaw.com/cle/plessindex.html](http://www.microlaw.com/cle/plessindex.html), also LawCommerce.com's Online CLE section).

Lawyers and their staff universally have one thing in common . . . they are buried in an unending sea of paper. Pleadings. Correspondence. Briefs. Exhibits. Memos. Pink phone message slips. Sticky notes. You name it, paper is everywhere, choking and clogging the flow of work in both private and public law practices. Sometimes getting client work out is more an issue of managing mounds of paper than applying legal brilliance. So is there any hope at the end of the paper-lined tunnel? Maybe, just maybe . . .

For years, lawyers have been on a holy quest for the mythical and fabled "paperless office." This endlessly elusive concept is likely the "Greatest Lie of the Technology Age." We're never going to become "paperless," at least in the foreseeable future. We just need to accept the fact that even if we reduce the amount of paper we generate, others will continue to send us paper.

Microfiche was supposed to be the answer, at least at one point. But microfiche just really isn't used very often in law firms because of the general inability to access the material from the PCs we use to do our work. Scanning was the next great answer. But let's be realistic. How many of you have had "bad scanning experiences"? Yes, we see all of you raising your hands out there.

And why has scanning been so generally unsatisfactory? Because since the dawn of document scanning, the term "scanning" has been synonymous with "OCR" (Optical Character Recognition). In other words, most people equate scanning with trying to use software to identify the characters on a page and turn it into an editable word processing document. Good idea conceptually, but in practice, even with the latest greatest technology available, this process is far from perfect. Even with the cream of the modern OCR software crop hitting text recognition accuracy levels as high as 97%, it's just not good enough. There are four problems here and any OCR veterans/victims will immediately identify with all of them:

- 1) 97% accuracy in text recognition is akin to a package of bologna proudly trumpeting its 97% Fat-Free status. Since about 90% of the calories are from fat, the thing is a veritable artery-clogging, love-handle expanding nightmare. With OCR, think of 97% accuracy this way: that's three screwed up characters out of every 100 and with a single-spaced page of text containing about 2200 words on average, that's 66 errors per page on average. And what if one of those errors is a nearly-impossible-to-detect-but-a-bet-the-case-on-it number?? Not good. Not at all.
- 2) OCR software tends to have a significant number of problems retaining the formatting and layout of the original scanned document. For example, you get a local state court pleading and give it to your secretary to scan. Seems like a pretty simple request, doesn't it? It's a "clean" document that has all appearances of being a solid candidate for being OCR'd: a mainstream typestyle and an original aser printed document (not some smudged, skewed third generation photocopy of a FAX of a photocopy). Should be NO problem, right? Wrong. What you likely get back could very well be a nightmare of reformatting, with a caption that defies clean-up, line spacing and odd tab stops that are equally baffling. OCR software tries its best to figure out what codes or styles to apply in the target word processing format--but it's really guessing at best and often it guesses wrong.
- 3) OCR is not terribly speedy. Even if you have a new high-end 3.2 ghz Pentium 4-equipped PC the OCR process can be pretty slow, and it seems with every increase in accuracy we have a geometric leap in the processing requirements. So even if you can afford one of the megabuck Intelligent Character Recognition systems like those from Kofax (<http://www.kofax.com>) that use costly processing boards you need heavy-duty PC horsepower for adequate text recognition. Forget about those 733 MHz Pentium IIIs with 128 meg of RAM. But either way, with OCR there's waiting involved.

- 4) And finally there's the always wide expectation gap between what we think is OCRable and those documents that actually can be OCR'd. We can't tell you how many times we've talked to some of you who have said, "When I try and scan this thing, all I get is garbage. How come??" and what they show you is a pre-printed state-specific divorce financial disclosure form replete with boxes and lines galore. You figure, it should be able to at least read the text right? Wrong. What we technologists have to realize is that the average lawyer user who expects this to work has a more legitimate claim to reality than those of us who make excuses for present technology by saying, "well of course it won't be right--look at all those lines and boxes--nothing can recognize those."

So the bottom line is that equating scanning with OCR is a fallacy that no longer needs to be the case. This is the case where my "Paper LESS Office™" concept comes in. First put forth in an article of the same name in Law Office Computing back in late 1995, presentation audiences all over the world have favorably received the concept. Here's the concept in a nutshell:

Using low-cost, high-simplicity image scanning, physical paper is turned into "digital paper." Image scanning is the process of using a scanner and using it to effectively photocopy your documents onto your computer system. This creates "digital paper," ideally stored in the universally readable PDF format.

Digital paper takes up no physical space and is manipulated easily by software on your PC systems. And the beauty of digital paper is that it is perfect--it is a picture of the original document, exact in every way without any of the vagaries of the OCR process. Of course, you don't have editable text at this point--you merely have a picture of the document. But most of the time, that's all we need.

So you have your Digital Paper/PDF of the document you scanned; a letter from opposing counsel, a set of interrogatory responses, the C.V. of a prospective expert witness, a stack of hospital records. What does it accomplish having all these pictures, these pieces of non-editable text? Glad you asked! What you accomplish is saving the time it takes to track down the physical file, or rummage through a roomful of banker's boxes to find the documents. All the document searching time costs money--economic dollars wasted whether it is lawyer time or staff time. In other words, we can use a scanner for its best purpose -- creating images -- rather than always using it for its less-than-perfect ability -- OCR.

And it gets better. One of the core problems in working on client files is that they are always split into two locations. The documents we create are located internally on our PC systems. The client

documents we receive from outside sources are stored in our paper filing systems. So, if you want to view all the correspondence on a client's file, you have to look in two separate places--on-screen for your own documents and then you need to track down the paper file and rifle through it to view the externally generated letters. That is, of course, if no one happened to have that particular file in their briefcase at home.

Next, whether you are using a great document manager like Worldox ([www.worldox.com](http://www.worldox.com)) or the document management capabilities inside case managers like TimeMatters, Amicus Attorney, PracticeMaster, ProLaw, etc. you go to that client file's folder/directory on your system and look in the "folder" where you store the correspondence for the client, you will see document names that begin with "Letter to . . ."-- which are word processed documents you created -- and document names that begin with "Letter from . . .", which are the scanned images of externally generated documents. Now internally-created and externally-received documents are all in the same convenient place! Then just double click and up pops that perfect picture of your document in the Adobe Acrobat Reader software.

The bottom line is that your client files become electronic and totally contiguous--they're all in one place. You just can't help but save all sorts of otherwise non-billable wasted time you would otherwise spend just LOOKING for things. Not to mention the ease at bringing a few client files home for the weekend, or take them on the road to a depo or a trial--without lugging back-breaking boxes of paper (and subjecting the potentially irreplaceable originals to coffee spills, misplacement and other forms of folding, spindling and mutilation).

And when you close the file, it's already "digital paper"-- you can store it in a convenient byte-sized package (sorry, pun intended). This is a far better alternative for closed file storage than the costly space-hungry storage requirements for physical paper files--which usually end up commandeering an area the size of small starter home.

What kind of scanner should a firm deploy? What software should be used to scan, organize and then search through the content of "digital paper?" Factors to consider: (1) intended volume of documents to be scanned, (2) number of pages scanned per job, (3) budget for internal scanning v. cost-effectiveness of outsourced scanning. As to volume, read the specifications for duty cycles. Buying a \$100 scanner rated for 2000 pages monthly when your firm needs to scan 10,000 per month will surely smoke that "bargain" scanner. The scanning marketing stratifies this way, roughly:

1) Entry-level flatbed - usually flatbed scanners without automatic document feeders - \$50-\$300. Unsuitable for law firm use because of cumbersome paper handling.

1.5) Portable scanners - Visioneer's Strobe XP100 is under a pound. This smaller-than-an-egg-carton-sized scanner can pull 5 imaged pages per minute into your computer system for under \$200 and the best news is that it has . . . drum roll please . . . NO POWER BRICK! It's actually powered by your PC or laptop's USB port.

2) Entry level document-fed scanners - \$250 - \$600 flatbed scanners with automatic document feeders. Suitable for lower volume scanning situations up to 15,000 pages monthly. Look at Visioneer's 9450PDF, models in Hewlett-Packard's Scanjet series and several from Microtek. Scanning speeds: 4-8 ppm.

3) Lower Mid-range document-fed scanners - \$600 to \$1300. Can feed 25-50 pages at speeds from 12-25 ppm Can handle up to 50,000 pages a month. Leading products range from Visioneer's 9650 at 12 ppm, Visioneer's Strobe XP450 at 20 ppm and several Fujitsu models' starting at \$600 and 25 ppm and heading up from that point.

4) Mid-range document-fed scanners - The Fujitsu 93GX has been replaced by the 3093 series - reliable workhorses. With 27 ppm capacity the 3093 series has a rugged 50 page feeder and fast SCSI interface. \$1800 to \$2400 depending on the configuration.

Above this level, the sky is the limit. Spend enough money and you'll end up with a riding model with 12 hp engine and a pull start! Well almost . . . Fujitsu, Panasonic, Bell + Howell, Canon, Ricoh, Kodak produce scanners that push the 100 ppm mark with massive paper handling ability.

You now have these images in your computer system; what's next? Organizing and searching them. Document management and work product retrieval systems are the best answer. These software systems can gently impose a file cabinet-like consistency on the way any law practice organizes both its internally-created documents as well as its externally received and scanned documents. Worldox is the undisputed leader in the small firm marketplace and has been digging into the larger firm segment for several years with great success. For larger firms, iManage and Hummingbird Docs are popular. Most legal case management systems also incorporate document management functions that can adroitly handle scanned "electronic paper."

All document managers let you organize and search scanned image files. This presumes, of course, the images are stored in a format that actually permits content searching of what would otherwise be only a picture. Documents scanned with Adobe Acrobat 6 or Adobe's Capture systems are stored in the universally viewable PDF format. PDF documents can now be "Captured image over text" documents. This means that if the software can recognize the underlying text, it may be searchable by a document manager that has PDF-search capabilities. Worldox excels at

such a role as part of its overall complement of document organization, management and retrieval functions, but isn't the only tool that can accomplish this.

Quick tip: a common misconception is that if one scans at a higher resolution, the text recognition results will improve. In fact, often the opposite is true - lower scanner resolution settings can yield better recognition. At higher resolutions, modern scanners can actually be "confused" by the fibers of the paper. Set the resolution to 150 - 200 dpi for better text recognition results.