



Digital Workflow

The basics of scanning

The right tool for the job

Many people have tried scanning, found it frustrating, and given up on the idea of using a scanner. On the other hand some people have had no trouble getting started with scanning. The difference between success and failure is usually tied to the scanner.

Multi-function devices (e.g. scanner, copier, printer all-in-one) have become prevalent and many lawyers believe that they can use these devices as part of their scanning workflow. In almost every case this is a recipe for failure, or at least a path to more problems. If you are serious about scanning then you need to get a machine that is dedicated to on purpose only: scanning. And this advice applies even if you have an expensive workstation printer/scanner.

The reason why you want a dedicated scanner is this: you need the machine to be set to default settings that are optimal for scanning. If the machine serves other purposes then the odds that the default scan settings will be accidentally changed are greater.

A short 'war story'

The electronic filing system that is used in federal courts requires PDF files to be uploaded. There is no page limit on filings, but there is a file size limit. I know of at least one attorney who had problems filing a 4 page document because he had scanned the document in at exceptionally high settings. He was shocked that his scanner could be the source of the problem. What kind of scanner did he have? A very expensive high-end workstation printer/scanner.

A scanner recommendation

If you don't know what kind of scanner to get (or if you have a multi-function scanner and have realized that it is not optimal), then just buy a Fujitsu ScanSnap. As of this writing the current model is the S510. The Mac version is called the S510M. The ScanSnap retails for about \$400 on Amazon.com, but you can often get it for less. It comes bundled with Adobe Acrobat, which is worth at least \$200.

The ScanSnap is dead simple to set up and use. It has a 50 sheet document feeder and one large button. You put the documents into the feeder and push the button. That's it. If you get a mis-feed just stop the scan and re-feed. You can mix and match letter size and legal size and it figures out the difference on the fly. It reads the front and back of the paper at once and tosses out blank pages. It will automatically rotate documents that have been put in upside down, or have the wrong orientation.

In short the ScanSnap was built to work easily and reliably in almost any work environment. People who start with this scanner never have serious problems.

Configuring the scanner

After you have gotten the right scanner the only important thing left to do is to configure it properly. If you aren't generally going to scan in color (and you really shouldn't scan to color unless you have a good reason) then you need to set the default to 'black & white.' Sometimes you have the option to set to 'greyscale,' but avoid this. Black & white scans create the smallest file size, and that's what you usually want. (Especially if you are going to upload to federal court as part of an e-filed document).

Many people set scan quality to 200 dpi, but if you want to OCR (see below) your documents then you might consider setting the quality to 300 dpi. Just remember the higher the scan setting, the larger file size and the longer it will take to scan. On the ScanSnap scanner the settings are 'normal,' 'better,' 'best,' and 'excellent.' I have my default set to 'better,' which I figure is about 300 dpi. I set the scanner's color mode to 'black and white,' and the scanning side to 'duplex scan' (i.e. double-sided).

Scan output

The next thing to configure (if you have the option) is where you want your scan job output to, and what file name you want to give it. I set my ScanSnap to output to a folder called '2008 Scan jobs.' Next year I will create a new folder and called it '2009 Scan jobs.' The scanner automatically puts a completed scan into the target folder and gives it a file name based on the date/time of the scan. The first thing I do after the scan (unless I'm interrupted in which case I come back to the document later) is rename the scan and put it in the proper folder. For example, I'd rename a scan of a letter I received in the Hodgkins case to, say, '2008_03_11 Marcus ltr to EES' and then file that in the 'Correspondence folder in the main folder called 'Hodgkins Case.'

The nice thing about this system is that I have two copies of everything I scan. One is in the '2008 Scan jobs' folder where it is sorted by date of scan, and the other is placed in the electronic case file with a name that is understandable. Every month or so I go into the '2008 Scan jobs' folder and delete documents that are several months old. The value of this folder is that it allows me to find documents that I just recently scanned quickly. And, if it turns out that I misfiled the document I can find it quickly here. Also, some documents aren't really important enough to store. For example, if I just need to email someone a document, I usually don't save it to a special folder. But I do have it in the 'Scan jobs' folder for a few months in case I need it. And, of course, my email system will have a copy for as long as it keeps old emails.

Optical Character Recognition (OCR)

If you want to be able to text search the documents that you scan then you need to OCR the documents. Whether you decided to OCR all of your documents is something that

you need to decide upfront. You can always OCR specific documents later on, so there really isn't a big need to OCR up front. The benefits of OCRing every document is that it gives you a powerful way to find information across all of your documents. The downside is that it takes longer (at least 5 times longer) to scan a page and then have it OCR'd. Also, OCRing a document makes the file size slightly larger (generally about 25%).

I generally don't OCR my documents, with the exception of documents that are produced or received in my litigation cases as part of the discovery process. In that situation, my workflow system is to 1) scan, 2) OCR, and then 3) bates-stamp. It's important to note that if you plan to bates-stamp a document you usually must OCR it before you bates-stamp, otherwise you won't be allowed to OCR (for technical reasons that aren't very interesting so I'll spare the details).

Scanning challenges

If you have a good scanner, set to the right settings, and a simple and consistent workflow, then you'll have no trouble with scanning. In the beginning you'll have to learn to handle small problems like making sure that the pages are all scanned (if you get a double feed and miss a page then you'll never know about the problem until it's too late). Sometimes the scanner will jam or mis-feed, so you shouldn't just leave it to scan unattended.

If you have really large scan jobs then consider hiring a company to do the scanning for you. The cost of scanning at a commercial service is now the same as (or possibly slightly cheaper than) printing. Commercial bureaus have excellent quality control (or at least better than you're likely to have) so the problem of mis-feeds and so forth will disappear. And they'll usually be able to OCR the documents and a very small additional cost.

When you first get started scanning you'll have a tendency to keep the paper around (just in case). This actually will create a problem that isn't obvious until you experience it. If you are trying to become completely paperless you need to know which paper has been scanned already and which paper needs to be scanned. If you throw the paper away immediately after you scan it (as I do) then you don't need to keep track of what's been scanned and what hasn't been scanned.

So what do you do if you aren't willing to throw out the paper that has been scanned? Get two stamps: one that says "SCANNED" (which you will use to stamp any paper that has been scanned already), and another that says "NOT SCANNED" (which you will use to mark paper that you don't want to scan, so you know that it doesn't need to be scanned). With this system you can assume that any paper that hasn't been stamped needs to be scanned.