ALM
LAW JOURNAL NEWSLETTERS

LJN'S LEGAL TECH Newsletter®

Volume 26, Number 1 • April 2008

Declaring Digital Documents As Records

By Eric Mosca

Many law firms have joined the quest to go "paperless." This holy grail of efficiency and order in the digital age has been touted for years as being just over the horizon. The answer to the question "how do we get there?" has been just as elusive. Sure, some of your colleagues may have achieved this ideal within the confines of their personal office, but frequently at the expense of significant additional burden and paper for those who support or collaborate with them. In a world where so many attorneys and support staff still rely on hard copy versions of documents, the only way to quickly go paperless has involved massive amounts of document imaging, even when many of the documents being scanned originated in an electronic format.

What is the attraction of a "paperless office"? This phrase conjures images of a pristine working environment, free of clutter, where any piece of information is available at a moment's notice. There is no doubt that electronic information offers the promise of easy searching and reuse coupled with a very slim physical footprint. Paper, on the other hand, turns very quickly into a disorganized mess and tends to clutter our sight for years because this hardcopy information sits in piles on desks and floors or in boxes in basements and warehouses.

Hard-copy documents and information also present a real challenge for backup plans and disaster recovery. Duplicating a store of information in a remote site in case of disaster can be done with a few clicks if the information is stored electronically. As a result, we tend to backup massive amounts of information whether or not it will ever be useful again. The standard practice for vital physical documents includes laborious microfilming, though this process has gotten to look and feel much like running paper through a scanner.

There is no doubt that the world is inevitably moving toward electronic creation, transmission and retrieval of documents, but the digestion of this information is still frequently handled in a hard-copy format. When reading a lengthy document or making edits, many users still turn to a printed paper copy. Processes such as this may lead to the creation of a brand new, and frequently more important, record. Conflict reports with attorney signoffs and edits to draft client bills are still done in paper at many of the largest law firms. Then there are the documents that only become "official" once they have a signature seal or stamp on the hard-copy. These records also contain more value than the originally created electronic version.

What this means is that there will always be a place for imaging of physical documents to obtain more easily referenced and shared versions. This does not mean we should image documents when a suitable electronic version is available. Electronic filing of documents is now commonplace everywhere — from courts in jurisdictions all over the country to the United States Patent and Trademark Office to the IRS. State bar associations are authorizing maintenance of electronic files in place of hard-copy documents.

In many cases, the electronic version may actually contain more information. Assuming that more detail is a good thing and useful to your organization, you stand to have access to considerable metadata when dealing with electronically stored information. This is not, however, always desirable if you are facing litigation or working your way through a massive amount of discovery material. This may include information about the author, time spent editing, tracked changes and access history, among other things.

Imaging hard-copy documents is still an expensive process. While the equipment and software has come down in price, the biggest costs are usually associated with coding this information. Finding the skilled operators to review information and tag it with metadata can be very difficult, but a massive collection of images is not useful unless you can find what you are looking for. Depending on the collection, you may be able to pull information automatically from the documents. Forms, for instance, where the information appears consistently in the same locations can be processed according to definite rules.

Eric Mosca is the Director of Operations for InOutsource. He provides project management expertise to assist clients in every aspect of records management. Eric is a Certified Records Manager and a member of ARMA International. He frequently lectures and contributes articles to industry publications on records management topics.

Large, complex collections, like a defendant's e-mails over the course of a few years, can be much more difficult to code. What information is relevant enough to be captured? Are attachments included? While this is not a small task to accomplish electronically, you can imagine how much easier it is to categorize and sort standard fields and search for important terms in such a scenario. Vendors have touted OCR (optical character recognition) processes to make images searchable and the technology has improved greatly. Depending on the collection, the effectiveness of making documents searchable may vary widely. Handwriting and small fonts cannot be captured easily. There is a certain accepted error rate that still may necessitate manual review, though this can be an attractive option for making images searchable when forced to scan from hardcopy documents.

When dealing with electronically stored information, issues of document format become very important. Documents in native format exist as files associated with the system or application that created them. There are many benefits to working with and sharing documents in their native format. We have already discussed the important metadata that may be included. This metadata is frequently more detailed for files in their native format. Information in its native format is frequently more easily editable. When reusing a quality brief, it is beneficial to be able to make slight edits or cut and paste large passages, while this may be impossible with an imaged version of the document that only gives you an effective "picture" of the page.

There are also drawbacks to documents stored in their native format. There have been embarrassing occurrences of re-purposed documents going to clients who look into the document properties to find that it was originally drafted by someone else for a different client. PDF documents have become a standard because they allow documents to appear with native formatting intact, but are not easily edited and may be smaller in size than the native version containing formatting information and formulas.

Even the amended Federal Rules of Civil Procedure recognize that issues of document format are important details that have to be discussed at the start of a piece of litigation. Rule 34B states: *"(i) a party who produces documents for inspection shall produce them as they are kept in the usual course of business or shall organize and label them to correspond with the categories in the request; (ii) if a request does not specify the form or forms for producing electronically stored information, a responding party must produce the information in a form or forms in which it is ordinarily maintained or in a form or forms that are reasonably usable"*

IT professionals also have to worry about maintenance and migration issues associated with electronically stored information. Documents created with older versions of an application may become difficult or impossible to access. While a standard such as PDF may reduce some of these concerns, there is no guarantee that even this format will be around in 10 or 20 years. Older image formats such as *.tif (Tagged Image File Format) have been around for longer periods of time and are good for archival imaging, even though the format is not as feature rich as PDF for text selection, comments and notations.

The quality and format of electronic collections has much to do with who has created them. Filing from the system that created the content typically means that documents will be stored in their native format. When you save to a document management system from Word, your document is stored with a .doc file extension. Frequently, a copy of a final document will be printed for the official physical file. If at some point in the future the firm embarks on an imaging operation for disaster recovery or "paperless" purposes, an imaged version of this document may be created from this official file even though an electronic version already exists.

Individuals have been more successful in relying primarily on electronic documents as opposed to larger organizations because the individual avoids the larger issues surrounding collaborating electronically and reducing duplication. When an organization looks to share documents and information, it can be very difficult to standardize document naming and profiling standards. The "paperless office" may be an unrealistic model, but capturing content as it is created electronically and avoiding the printing of paper is a step in the right direction.