

The Difference Between Paper and Electronic Files

Toby Brown

Paper is wonderful. It is comforting to the touch. It is portable. It is easy to read and browse. You can read through pages and back again. But for lawyers perhaps the most useful things about paper are; 1 – When you discard it, (for the most part) it's gone, and 2 – It's stupid.¹ For a lawyer these two qualities have greatly limited the damage paper-based information can cause. As well, these qualities have kept the management of legal information relatively simple.

Then along comes electronic files. These types of files are really messing things up. And lawyers are just beginning to realize the depth of the challenges these file types present. In order to better understand the emerging landscape of electronic-based information, we will take a close look at the nature of electronic information. We'll start with the 'discarding' topic.

When a piece of paper is thrown away by a lawyer, generally one of two things occurs; it is thrown away or shredded. If discarded, it is likely carted off to the local landfill where it becomes the proverbial needle in the haystack, albeit a very odorous haystack. The likelihood that one might come into possession of a discarded paper document is extremely low. If a document is shredded, again, the resulting information is relatively inaccessible.

Too bad the electronic file paradigm couldn't follow along with this thoughtful approach.

When an electronic file is deleted, it is actually still there. The ones and zeros that make up a document still exists. All that happened was the name of the document was removed from the filing index. Consider a library where a card from the card catalog was removed.² The corresponding book is still on the shelf and accessible by those who know how to find it. This presents a problem for a lawyer from a couple of angles. You may well want the document gone. I mean really gone. And if the information is not really gone, you end up with new and interesting discovery problems.

How do you find this stuff? Who knows how to find books when the index card is missing? And if that isn't bad enough, even if you find a way to 'remove' that book from the shelf, in this library we have a backup copy. Or more precisely we have backup *copies*. Our book exists within a number of other locations in the library and hopefully one is stored safely in another building (our offsite storage).

Since the legal profession owns its share of masochists, we'll throw in one more twist. When the book in our library was checked out, six copies were made of it and passed

¹ "Stupid" in a good way. I'll explain this later. Also, while I have your attention, another fine aspect of paper is the simple ability to sign it. But that's an issue best left out of this article.

² If you do not know what a card catalog is, you can skip to the end of the article ... the very end.

along to other people. Of course nice trails were left (in the e-mail system) of where those copies were sent.

What's a lawyer to do? The ability to maintain control over the existence of electronic information is considerably lower than good ole paper. The defense bar now has the task of finding ways to bring this situation into focus so that clients' interests are protected.

We will start down that path now. But as we do, you should be thinking more about policy than technology. Information, whether it is in paper or electronic form, is managed through policy. The more you understand the technology, the better equipped you are to make sound policy decisions for yourself and for your clients.

When we talk about finding deleted electronic information, we are referring to computer forensics. This is the process of finding and retrieving deleted electronic files in a trusted fashion.³ As you might expect, there are a number of vendors emerging in the market that specialize in this service. Most of these vendors are focused on serving the emerging e-discovery market⁴, since discovery is the legal event that usually triggers the search for deleted information.

To further illustrate the issues surrounding deleted information, we will explore e-discovery as a process. Finding the deleted information on its own is not good enough. A defined process is required to maintain the integrity of the information and put it in a producible form for discovery requests, depositions and court proceedings. This process should include; 1 – acquiring the data, 2 – indexing & organizing the data for review, 3 – image creation (for further review, coding and presentation), and 4 – loading the data and images into in a database for ongoing search and retrieval.

Acquisition

The geek term of getting a copy of the information from a device (usually a hard drive⁵) is called acquisition. This step is very important. Acquiring the information from a hard drive is not as simple as making a copy of its files.⁶ A bit-by-bit image of the hard drive needs to be obtained. And it needs to be obtained in such a manner that the bits are not altered during the process. The mere turning on of a computer can change the bit-level information. Therefore you need an expert at acquiring this 'bit mapped copy' so that no evidence is spoiled.

With this bit-by-bit reproduction of a hard drive, a forensics expert can begin reconstructing deleted information to prepare for the next step.

³ An important part of computer forensics is trusting that the retrieved data did in fact originally exist and was not created during the forensics process.

⁴ E-Discovery has become the industry term for the discovery of information in electronic formats. Initially the services have focused on producing e-mail files, but are quickly expanding to include a great variety of other files.

⁵ Other devices can include PDAs (ala Palm) and even the newer digital copiers.

⁶ Side note: Your first step should be serving some sort of "preservation of evidence" notice to ensure that discoverable electronic data is not compromised.

Indexing

The forensics expert will now create an index of all of the information retrieved. At this point we need to return to a paper analogy. The equivalent set of information retrieved from a hard drive would be a file room, including all recently discarded files. Opposing counsel would never be allowed to fish through your client's file room. So we first need to have attorney review of this relatively raw data to see if it matches the discovery requests and whether there might be privileged information. An initial review can occur with just the index. However, some of the review will need to include the documents themselves. For this level of review, deleted information needs to be recombined into readable files. Again, you should have this done by a trained or certified forensics expert so that the files are trusted. This is especially a problem for reconstructed e-mail. Unlike a Word file, e-mail is made up of separate data files that are "put together" with your e-mail software (e.g. Outlook). When these separate data points are recombined, care is required so that the result is actually a valid file.

Typically the reconstructed files are converted to an image-based file format such as .tiff or .pdf. This approach serves two purposes. First, the software to view these files is ubiquitous. This aspect will come in useful for future coding⁷ and subjective review. Second, since the files are images, you have less concern about end-users altering files when they are viewed.⁸

Production and Metadata

Now that our forensics experts have given us useful and trusted information, what next? To explain "useful" we mean a collection of image files along with the corresponding metadata available from the file systems. Metadata is data about data. Many of our electronic files include metadata within the files themselves. This is a double-edge sword for lawyers. On the good side of things, it means you can more easily index the discovery data. If we know who was listed in the "To:" and "CC:" fields for a collection of e-mails, we can easily create an electronic index of addressees and copied recipients.

The Bad Side of Metadata – ETHICS Warning

You may recall my original comment about paper being "stupid." This is where we get to that issue. Paper only contains the data you see. It is not smart enough to know where all of its information came from, which means it is unable to reveal its secrets. In contrast, we'll refer to electronic files as 'smarter,' since the metadata they contain is knowable.

⁷ Coding is the manual process of retrieving the important data from a document, such as author, date, subject and recipients.

⁸ Ensuring the integrity of the information in electronic files is a growing concern. Image-based file formats reduce the likelihood of alteration, but do not eliminate it. It is likely courts will create rules for managing document integrity as things evolve.

For example, MS Word includes metadata for every Word file. If you go to the 'File' pull down menu in an open document and select 'Properties' you can see some of this metadata. Depending on your Word configuration, Properties can include differing levels of data. Some of this data might reveal client confidences, especially if a file was created by using an existing document, then modifying it. The metadata can reflect the information from the original file as well as the current document.

The Even Worse News

In MS Word there is a feature called "fast saves." This feature can be very useful in the event of hardware failure, as it reduces the chance of lost changes to a document. However, with the fast saves option enabled, deleted information can remain hidden within a document. Imagine a contract Word document is pulled up from a past deal. Then changes are made to it and a copy is e-mailed to the opposing side. Then the opposing side opens the document with another program, such as Notepad. All of the deleted information from the original deal is now viewable to them.

Has this scenario actually happened? Yes it has. If it makes you feel any better, the one example I heard of from its source, involved a contract a very large software company ("VLSC") sent out. The VLSC told this person that it did not give certain concessions in the type of deal proposed. Yet when the deleted data was viewed it became clear that the VLSC had given those concessions to someone else. Oops.

In addition to fast saves, there are other metadata traps. If you are using the Reviewing or Versioning features to track changes, then deleted information can be retained. Or if you have used the Comment feature, you may have planted some confidential information somewhere inside the document.

Lawyers obviously want to be very careful about the metadata contained within their own files and the files of clients. The ethics dialogue on this issue has been quite interesting. The main issue is pretty clear-cut. Lawyers who reveal client confidences get disciplined. The interesting part has to do with the lawyers on the receiving end of the metadata. Is it ethical to look for metadata within electronic files? And if you find it, what then? On one of my e-mail lists I even read a good argument that it should be unethical for lawyers NOT to look for metadata. The point was made that lawyers should have a responsibility to search for and review this information to best protect their own clients' interest.

In any event, you do not want to be the test case for discipline and malpractice for this issue. You should find and eliminate any metadata that could comprise your clients' interests before you share electronic files via e-mail or in an e-discovery production.

But How?

The next obvious question is how can you get rid of metadata? First consider that metadata exists in other file types beyond word-processing, such as spreadsheets. There

are software programs that can help you deal with this broader problem. One example is from Payne Consulting (www.payneconsulting.com) which provides a product called MetaData Assistant that works with Word and Excel. Another option, the poor mans' approach for removing metadata, is to save the documents in .pdf or .rtf file formats. The .pdf approach will do a better job of preserving formatting, but will produce a larger file size. Whereas the .rtf approach will be much smaller, but may well loose some formatting.

Back To Production

We've strayed a bit from our e-discovery process, but with good reason. The metadata issue is valuable in that it further demonstrates that paper and electronic documents are definitely different. The result is that lawyers should treat the two differently as it relates to protecting client information.

Bringing It All Together – Policy!

Now that you have a better understanding about the technology, this should help you make better policy decisions. These policies will apply to both you and your clients. For your own systems, think about policies that will best protect the client data you hold. For instance, what happens to your old computers? Now that you now how much client data could exist on the hard drive, you might want to implement a policy for wiping⁹ or destroying the hard drives of these computers before they leave your office.¹⁰

For your clients, you may want to help them implement sound e-mail policies. Better information management policies will further reduce the discoverability of your client's e-mail in the future.

Fear as a Motivator

If after reading this article you are a bit concerned or even afraid, that's a good thing. You should be more concerned about the security of the client data you hold and transmit. Changes in the way people share information have been incremental, but the rate of change continues to accelerate. Those incremental changes are adding up to qualitative changes, which mean it is time to address your policies for how you manage electronic information.

A great resource for ideas in managing electronic information is the Association of Records Managers and Administrators (www.arma.org). At their web site you will find good articles on information management and also on evolving standards.

⁹ Wiping is a term for completely erasing data from a device. Examples of wiping programs include Cyber Scrub, Eraser and DataGone.

¹⁰ Recently some MIT students purchased a number of used hard drives via eBay as a class project. They were able to retrieve significant data from most of the drives.

Hopefully this article has given you a nice foundation of knowledge about the nature of electronic information and motivated you to take a closer look at how you and your clients manage information in an electronic world. Taking proactive steps now to bring these issues into focus will pay off well into the future.