



E-Discovery for the Rest of Us: What You Need to Know Now!

Written by Dominic Jaar

Presenters:
Dominic Jaar
Sharon Nelson

April 2-4, 2009
www.techshow.com

Session Agenda

In the first part of the presentation, we will look at the process and tools that could be used to reduce the burden and costs of e-discovery in smaller cases. We will also discuss the best approaches to focus on relevant material.

Small case e-discovery: Re-discovering e-discovery

Introduction

“Civil litigation has priced itself out of the market.”; “The rules governing electronic discovery need to be re-thought. They cannot be fairly applied to all cases across the board.”; “One set of rules cannot accommodate all cases.”; “E-discovery is crushing the system.” are just a few quotes from Appendix B of the Interim Report & 2008 Litigation Survey of the Fellows of the American College of Trial Lawyers which consists of “selected comments from survey respondents” [available at <http://www.du.edu/legalinstitute/form-ACTL-survey.html>]. And you know what? They are right!

Don't get me wrong! I love reading judges Facciola and Grimm's dissertations on e-discovery in sophisticated cases. Their opinions are really helpful to large corporations, their counsels and e-discovery consultants like me. I am so much into these judgements that I even found myself waiting for them at the secret joint where they are said to meet to discuss their next opus! However, when it comes to your average small and medium cases, you can't rely on them to help you find a cheap but legal way to handle your electronic documents.

Contrary to what the Fellows mentioned, I don't think the rules are problematic. Their application is!

I think the notion of proportionality is not given as much attention as it should. In the paper world, the lawyer could rely on a witness to provide the potentially relevant documents without having to personally go through all of the corporation's documents “manually” in search of the needle in the haystack. Can anyone tell me which rule changed this principle?

To make this point clear, I'll quote Ken Withers, director at The Sedona Conference, reacting to the abovementioned interim report:

“Looking at the American College report, I had the impression I was reading a report from the Buggy Whip industry, complaining about the dangers and high cost of transportation caused by the advent of automobile. These were the Buggy Whippers' findings, I'm sure: paved roads would be a waste of taxpayer resources; since only the rich can afford automobiles, poor and middle class people will be deprived of any means of transport; and rules should be adopted requiring that all these newfangled horseless carriages be led by a person on foot waiving a warning flag. For the Buggy Whippers, learning to drive, or even getting a bicycle, would not be considered viable options.”

So, let all lawyers learn to ride PSTs, dupes, near-dupes, load files and the likes! In this article, I will provide you with my personal opinion and experiences handling smaller, and sometimes

larger, cases with no fancy software, no vendor, no costly discovery, just plain good sense and reasonable solutions to the new litigation issues raised in the world we live in.

***N.B.** Before we start, please be aware that playing tech-savvy can be dangerous and that most of the tips you will find in this article are not 100% accurate due to many factors and variables such as the operating system (OS), the software used and their settings, the custodian's information hygiene, etc. and the numerous nuances that cannot be discussed in this article. Accordingly, you should not start doing your own e-discovery unless you are fully computer literate and have tested the following software in the context of your particular case to ensure that the output conforms to your goals and obligations. Furthermore, playing the technician for a lawyer can be tricky because of potential personal and professional liability but also because of privilege issues. Finally, note that I have not used all of the tools mentioned in actual cases and that I don't have all the answers to all the questions and issues that could arise from using these tools in a particular case. This article objective is to start a discussion to find better ways to handle e-discovery in small cases that do not require forensic investigation. I invite you to contribute your own experience with small case e-discovery.*

We will look at some phases of the e-discovery process, following the Electronic Discovery Reference Model [available at <http://edrm.net/>], namely: information management, identification, preservation, collection, processing and review.

Information management

Allow me to let other articles elaborate on information management. Suffice it to say that this is probably the part where individuals and corporations could save the most, not only during litigation but also on a daily basis by improving efficiency and productivity. In the paper era, who would have invested in a corporation that did not provide filing cabinets, bankers' boxes, folders, Post-It, pen, etc. to its employees and let them pile up stacks of papers underneath and on top of their desks, in their cars, at home, etc.? Well, if you still have some money left from the current economy, you won't find many corporations to invest your money in as most of them are letting their employees manage their information on an ad hoc basis. Therefore, the first cheap e-discovery tip is to make sure your clients get a good document management system and a clear and implementable document retention policy.

Identification

As was the case before the amendments to the FRCP, any lawsuit starts with the identification of potentially relevant information in order to preserve and eventually review it for production. The challenge in the electronic world is to deal with the volume, numerous locations and formats of the information. However, we went overboard trying to find everything. We should instead be focussing on what is most relevant and allow ourselves, as in the old days, to rely on the potential witnesses to provide us with the necessary documents. Down the road, they should be the ones explaining what and how they did their identification, preservation, etc. True, it generally means educating witnesses on how to approach the task, but this time invested in educating these people normally generates great returns in the day to day business in terms of improved information management. In order to help their clients, lawyers should have templates explaining the best ways to identify relevant information that they could customize and provide

to their clients. Accordingly, in smaller cases, custodians could identify and preserve the relevant information and, as we will see, collect it by themselves.

To identify the relevant information, custodians could use free or cheap search engines such as:

- a) Copernic
- b) dtSearch
- c) Google Desktop
- d) X1

With the new Windows Explorer preview pane under Vista and the power of Windows Search 4, if it weren't for the latter's incapacity to search email attachments, I would go as far as saying that, in some cases, we do not need anything but Windows default software to identify desk/laptop data.

For identification purposes, it is important to be competent and knowledgeable about keyword searches to create Boolean searches that will find the relevant information you are after. It is also necessary to understand that when you run a search, images, audio and video recordings will not come up in the search results. Accordingly, if the custodian has scanned documents that have not been processed with OCR, i.e. optical character recognition, you should isolate them for independent treatment. You should also think about the languages of the documents found on the computer and the ability of the search engine, and custodian, to find them.

Such an approach during the identification phase implies that you understand how computers and software work and behave, which is a knowledge that still needs to be acquired by most lawyers.

Preservation

Onsite

With respect to preservation, the information can be kept onsite, i.e. on the hard drive of the computer. However, to prevent any data alteration or loss, all of the data on the computer should be backed-up and the back-up media set aside. This back-up could be created using basic tools such as:

- a) Windows Backup [Start>Programs>Accessories>System Tools]
- b) Reflect
- c) SyncBack
- d) Second Backup
- e) Truesafe

The most important benefit of onsite preservation is that it prevents over-collection and the ensuing costs and delays.

Manual collection-preservation

Another alternative is to do a collection-preservation similar to what professionals would do. However, instead of acquiring particular hardware and software to complete a forensic or quasi-

forensic copy of the relevant data, a focussed “manual” collection of the relevant data could be performed for preservation and later, review purposes. In laymen’s terms, you could basically ask the custodian or someone else to find the data himself on the computer and copy it onto a USB drive, CD, DVD, etc. or simply email them to you.

Many people say that the integrity of a document will be affected if it is not “harvested” by a professional. Even if, in certain circumstances and for certain types of data, this is true, this reality is often exaggerated. In fact, what will likely be modified is the metadata of the documents. However, in many instances, metadata are not material to the case. Granted, they can be helpful to manage the information particularly if you use litigation support software. However, the time and skills needed to collect the electronic documents with their metadata can be deterrents. Furthermore, alternative means of preservation and collection often maintain the metadata or most of them, as we will see later. Therefore, in circumstances where metadata are of low marginal utility, parties could agree that they won’t guarantee the perfect preservation of all of the metadata. Should the opposing party disagree, it would probably be reasonable to move for a cost shifting order.

This second approach is particularly fruitful when the parties are interested in the information on the documents for its face-value, i.e. what is written on them as opposed to their history and context which is generally derived from the metadata.

Collection

As mentioned above, one approach is to have the custodian collect “manually”. Here are a couple of tips to do it to least affect the metadata.

Email

Normally, most of the documents that are of interest are emails and therefore, looking at the market share, we will focus on Outlook. For an in-depth look into different email clients and webmail applications, I urge you to read Craig Ball’s excellent technology primer entitled [Meeting the Challenge: E-Mail in Civil Discovery \[PDF\]](#).

The easiest way not to mess with metadata, which is used by professionals, is to run a search for “.pst” and “.ost” files on a computer. Make sure you select “include non-indexed, hidden and system files” before running your search. However, when you use this approach you will likely have a couple of MB of data that include emails, calendar events, notes, tasks, etc. particularly if you are dealing with a packrat custodian who has folders entitled “2002” which are full of sub, sub-sub and sub-[...]-sub folders!

A leaner alternative is to export only the relevant information to a new .pst [File>Import and Export>Export data to a file>Personal Folder (.pst)]. Once you have created the new PST, you should make a backup and a working copy. You should then be able to use a working copy to load onto your favourite litigation support software. In theory, you could load the PST in your own Outlook client [Tools>Account...>Data tab>Add] for filtering and review. Then you could sort the material you gathered via the different column fields provided by Outlook [right-click the column header>field selection]. Alternatively, if there are not many relevant documents, you could ask the custodian to simply forward them to you. However, they should not select the

email or calendar event and forward it to you as it will affect its integrity. Instead, they should attach [“attach element”] the original elements to an email addressed to you or make sure the right forward mode is activated [Tools>Options>Preferences>Messaging options>change “When forwarding a message” to “attach the original message”]. However, the unresolved challenge with the last two suggestions is the fact that you will not be able to produce from that batch because most of the metadata will have been modified, that is unless you have agreed otherwise.

Documents

It is impossible to discuss all document types that exist so we will once again focus on your classic Microsoft Office, i.e. Word, Powerpoint and Excel, which is the bulk of the unstructured relevant data we find in most cases.

These document metadata are generally not affected if they are not opened or copied. However, the tough part for custodians is to find the relevant information without opening the documents... Thanks to the abovementioned search engines and now to Windows Vista, it is possible to see a document in preview mode without affecting most of the metadata. Therefore, after a quick preview, you can drag and drop these files to any media for transfer purposes. Note that we refer here to drag and drop, not copy! When you copy a file, the OS changes the access and created date timestamps, which explains why a file’s created date can be more recent than the last modified date... However, be aware that once you drag and drop a file, it is “removed” from its original location and accordingly the “location”, another metadata, is modified.

Nevertheless, at this stage, it remains safer to err on the side of caution and drag and drop everything that seems relevant without previewing it. Once again, upon reception of the relevant documents, a backup copy should be created to have an independent working copy to upload onto your lit support software.

Two important issues that arise from such harvesting techniques are the difficulty of following the chain of custody and the relationship between the “original” document and the produced copy. We will look into these in the next sections.

Chain of custody

If the custodian collects the information by himself, he would be the one testifying about the chain of custody, should anyone second-guess the process. On the other hand, when many people are involved it is generally more effective to have a chain of custody form filled by all of the custodians or collectors involved. Many templates can be found on the web by searching "chain of custody form". Once again, the content of these forms and the way they should be filled could be the object of a full article.

Document signature

If the collection is not done by a professional, how do you prove that what is being produced is what was collected and identified in the first place? Do like the pros!

Each document has its own signature, aka hash, which is derived from an algorithm, aka hash function. Before collection, you should get that “signature” for all of the documents that you intend to collect. Here are a couple of free utilities that will do just that if you know how to use

them and have the right case. However, only the first enables batch calculation of many files at once.

- a) [VisualHash](#)
- b) [Pinpoint Hash](#)
- c) [Pinpoint MetaViewer](#)
- d) [Fastream MD5/SHA1 Hash Extractor](#)
- e) [Easy Hash](#)
- f) [Iside](#)

While we are at it, I'll mention that hash could also be used as an alternative to bates stamping [hat tip to Ralph Losey's [HASH: The New Bates Stamp](#)].

More tech savvy people could use the following tools to copy documents without affecting the metadata and more particularly the aforementioned timestamps.

- a) [XXCopy](#)
- b) [Robocopy](#)

Although Robocopy is a command-line tool, [Microsoft Technet](#) has provided a front-end graphical user interface (GUI) [direct download]. You could also use [Pinpoint SafeCopy](#) on top of Robocopy. For those of you interested in technologies like Robocopy, I invite you to read Craig Ball's cautionary tale [When Out-of-Box Means Out-of-Luck](#).

Paper and Images

Even if we are talking about e-discovery, it is important to remember to look for the relevant information found on paper. However, we are not interested in reviewing paper and electronic documents. Therefore, once identified, the paper documents should be scanned, using a documented methodology, and OCR'd, i.e. run through optical character recognition software which will recognize words in images like PDFs or TIFFs. Here is some free OCR software:

- a) [FreeOCR](#)
- b) [GOOCR](#)
- c) [SimpleOCR](#)
- d) [Tesseract](#)

Once you are done, ask your assistant or paralegal to code the scanned documents, i.e. create the metadata we find in electronic documents. You should focus on the most important fields such as "from", "to", "cc", "date" fields, "subject", etc.

The paper documents scanned, OCR'd and coded, it is possible to process them with electronic documents to "near-dedup" and filter them, as we will see below. Working with scanned documents also enables you to review and centralize all of your documents into one platform.

Containers

This part should be in the next phase if we had followed the EDRM since it is in the processing stage that the tricky type of files known as containers are extracted. Outlook .pst and .ost that you read about earlier are good examples. However, there are many others such as .nsf (Lotus Notes),

.zip or .rar files. These unique files are in fact containers which hold numerous documents together. Therefore, these documents need to be extracted from the container to be read or sometimes searched. Accordingly, all of these files should be isolated for independent review.

Early-case assessment

An extra step that will help you when you go to a meet and confer session is to assess your case as early as possible in terms of e-docs. This section is found in this part of the article because the free tools we suggest are most helpful once you already have collected your data. However, early-case assessment should be completed at the identification stage and updated thereafter. Here are two good tools to help you in your assessment:

- a) EarlyCASE
- b) Onsite3 eDiscovery Estimator

They will provide you with information such as file counts by type and the number of duplicates. They will also help you assess the costs of your project and provide graphs and stats.

Processing

It is difficult to get rid of vendors in this phase, in part due to the numerous software and expertise required to process the collected data. Nevertheless, we will provide you with a few tools that might help you reduce and organise your information for review and production purposes in smaller cases.

One SAAS, i.e. software as a service, that offers many of the tools you need to process your data is DiscoveryForce.com. It provides you with a number of tools that we will discuss thereafter: conversion, dynamic index, de-duping, near-de-duping, metadata culling, image creation, organization, foldering, privilege files sorting and exporting. It also covers parts of the next phases, namely review, redaction and production.

Document Conversion

This is when you convert electronic documents to metadata, text and image. The conversion formats mostly used are TIFFs and PDFs. As aforesaid, text can be extracted from images by OCR software. To convert other types of electronic documents to TIFF or PDF, here are 2 free or cheap software:

- a) Omniformat
- b) VeryPDF

Should you face documents in other languages than English, you should make sure the software you use is Unicode compliant to handle accents and symbols.

Indexation

In order to filter, sort and organize your data, you should create an index of all your documents. If you converted to PDF, Acrobat Pro is probably the best solution. Otherwise, I have yet to find good free indexation software and invite you to send me an email with any you know of.

De-duping

Deduping is the task that enables one to identify duplicates to prevent review of the same document many times. Electronic documents are duplicated so easily by forwarding, saving in

different places, etc., that it is primordial to get rid of these duplicates. Pinpoint FileMatch is a utility that scans for duplicate files. However, it only scans one file at a time... Better free alternatives, which compare numerous files and folders at once, are:

- a) Dedup
- b) Clone Cleaner Lite

Near duplicates

This is the crucial part of processing your data where you de-duplicate files that are similar but not identical. For instance, a Word document converted to PDF or a string of emails you compare with a part of the string that have been forwarded, CCed or BCCed to other people. In order not to review all these documents you want to use near-depuplication software which will compare part of the documents as opposed to the whole documents and their hash value. Unfortunately, to this day, there is no free or cheap solution. The leader in the market is Equivio. However, DiscoveryForce.com does a decent near-duplicate analysis at no extra charge while processing your data. If you know which documents are near-duplicates, you could use Word's compare tool, or free or cheap software like:

- a) File Compare
- b) Active File Compare
- c) Visual Comparer

Metadata culling and organization

Metadata culling is the part when you sort, filter and organise your documents based on their metadata, e.g. "from", "to", dates, etc. to help with the review. In more sophisticated cases, you could group your documents by issues or concept. The idea is to enable the reviewers to follow a learning curve and become expert in one area of your case to be able to review more effectively.

Review

In small cases, you generally don't need a sophisticated tool to review the potentially relevant documents. Many lawyers successfully use Acrobat Pro to perform their review and production. It enables them to have an index, run searches and redact. I prefer to use litigation support tools like Summation which offers more functionalities, flexibility and search capacities.

One last tip that should be used when processing and reviewing document is to never delete irrelevant documents, duplicates or near-duplicates, but only suppress them from the body of documents. You might need to go back to them later on.

Conclusion

Always remember not all cases are created equal. Therefore, when doing an early assessment of your next case, you should scope e-discovery as proportionally as possible and agree with the opposing counsel as soon as possible on ways to alleviate the burden of handling electronic documents. When lawyers will learn how to cooperate and find common grounds and interests, they will realise how e-discovery is one of the numerous ways to improve access to justice. Until then, happy DIY edd!

P.S I wish to thank Craig Ball and Ken Withers for their invaluable comments on the first draft of this article. I must add that the result does not necessarily represent their opinion and in certain cases, is the complete opposite: you sometimes need to agree you disagree and let a judge decide!

E-Discovery for the Rest of Us:
What You Need to Know Now!
April 3, 2009

