

TechnoFeature™

Practice management and technology articles written by experts.

eDiscovery: Everything You Need to Know About Collection and Processing

By Bruce A. Olson

INTRODUCTION

In the days before electronic discovery, the process of collecting and processing a client's case-related information was relatively straightforward. You identified potentially relevant documents given the issues in the case. You identified the appropriate custodian of those records. You worked with your client to locate the paper files, and then you photocopied the materials you needed for the case and took them back to the office in banker's boxes.

Generally speaking a photocopy was sufficient for discovery purposes, and you certainly didn't need to worry about producing the documents in electronic form or providing them in native format with the associated metadata included. Spoliation was rarely a concern because the likelihood of inadvertent spoliation was slim. It would take an intentional act on the part of the client to destroy records, and if that happened and was discovered, your case was already dead in the water.

All that has changed. The process of collection and processing, not to mention the review, analysis, and ultimate production of the materials, has become complex, extremely expensive, and fraught with the risk of missteps that could result in spoliation sanctions. It is therefore necessary to thoroughly plan your electronic discovery efforts and understand the universe of electronic information you are dealing with right from the moment of case intake. You no longer have the luxury to wait until you receive your first discovery demands to address these types of issues.

PLAN AHEAD ...

The process of collecting and processing electronically stored information should begin with a thorough planning process. This is true in terms of developing your own plans for dealing with your client's infor-

mation, and in planning for the discovery you will seek from the opposing party. The importance of initial planning is recognized in the new eDiscovery amendments to the Federal Rules of Civil Procedure Rule. Rules 26(f) and 16(b)(5) and their associated comments make clear that early and comprehensive planning is required so your initial meet and confer conference can address the myriad issues that accompany electronic discovery, and your initial disclosures meet the current requirements of electronic production.

The process of collecting and processing electronically stored information should begin with a thorough planning process.

The planning process should include participation by everyone who will participate in the electronic discovery process. This includes the attorneys, paralegals, IT, and litigation support personnel from the law firm. It also includes knowledgeable persons employed by the client, including their IT staff, and in-house legal staff if available. It helps to include decision makers, even if they were not involved in the events giving rise to the litigation, so they understand what is involved in eDiscovery and what can happen if it is not handled properly.

Finally, early use of an outside eDiscovery expert is advisable since many of the things that now need to be done are better off outsourced.

An important part of the planning process includes ensuring appropriate security for the data collections involved. Consideration must be given at the outset to protecting work product and privileged information

(Continued on next page)

so that it does not become part of the information produced to the other side. You must address chain of custody concerns. You must also take steps to ensure that employees do not access the information in the course of the collection phase as that could result in the alteration of the metadata.

It seems that in every case I've dealt with, by the time I've been retained, employees, managers, and even in-house counsel, have gone rummaging through the saved email and other electronic files. At a minimum, they have changed the last accessed date in the process, but it is possible they altered other metadata that could come back and pose a problem. Thus, one of the discussion items in your very first telephone call should include a warning to your client to advise everyone to stay out of the electronic information until you have had an opportunity to put an appropriate plan in place.

An important part of the planning process includes ensuring appropriate security for the data collections involved.

COLLECTION ...

Your collection efforts should include the identification and collection of live data from the client's email and file servers, databases, and any other applications that contain live information. You must also consider archived records or those organized in some type of records management system. It's important to determine whether there is off-line data stored on local PCs or laptops, CDs, DVDs, flash drives, external hard drives or other personal storage devices.

Finally, you must identify all available backup media. Keep in mind that in appropriate circumstances, where employees have remote access, their home systems may need to be included. Once this information has been identified you need to take appropriate steps to protect it from inadvertent modification prior to the actual process of searching the data sources for responsive materials, again to avoid inadvertent spoliation.

Data Culling ...

One of the big problems with electronic discovery is that there is often so much electronically stored information that needs to be dealt with that the process is extraordinarily time-consuming and expensive. You'll often find multiple duplicates of documents and email in an enterprise environment, and the production of every single duplicate item does nothing to advance discovery. Likewise, there may be technically discoverable electronically stored information, but it is of such marginal value to the case that it makes little sense to pursue it. Thus, a good plan to filter and cull the data down to manageable amounts is imperative.

There are several things to consider in this process. You may want to filter by custodian, limiting a search to key custodians and files associated with them. You may also want to filter by date and time to limit the search to a finite period. You should also consider identifying custodians whose information is privileged and take steps to avoid production of privileged information at this point in the process. Keyword searching can be used to limit the universe of materials to produce. Boolean searches may be appropriate. Stemming searches and fuzzy searches may also be employed. More recent technology also enables one to conduct contextual searching or concept searching as an additional means to scale back the size of the collection.

Another option to consider is the use of sampling. A more limited set of data can be analyzed using the above referenced techniques. If significant numbers of hits are generated, the dataset can be expanded. If the opposite is true these datasets can be set aside. Where large volumes of data are involved rolling production can also be used to obtain the priority items first. Oftentimes it is appropriate to work with third party vendors to address these types of issues at the outset so you can take advantage of the available technologies they have to assist in the culling process.

Chain Of Custody ...

To avoid charges of spoliation, and frankly to make internal management of the information easier, you need to develop procedures for logging chain of custody information, ensuring that the information

(Continued on next page)

collected has been properly copied without additions or changes to the data or metadata, and that the collection process has remained secure. If you process in-house you need to develop your own written guidelines and procedures. If you use third-party vendors you need to ensure that they have comparable procedures in place so they can provide you with auditing information. You should use audit history logs that identify the person who performed the work, the search methodologies employed, the collection method utilized, and any other pertinent information.

I have found through experience on large cases that it is helpful to assign a particular paralegal the task of logging collection and production information, including the receipt of incoming information from the client or from opposing counsel. In large cases you may have many people working with the information, and it is next to impossible to manage the information effectively if you don't have someone who knows how everything you're working with ended up in your office.

Metadata ...

One of the key things to keep in mind when beginning the collection process is that you must avoid alteration of the data and/or metadata. Failure to do so could result in spoliation sanctions. Simply copying selected files or folders using drag-and-drop methods may present risks. Ghosting an entire hard drive may present risks as well. The only way to ensure absolute accuracy when copying data is to create a forensically valid mirror image copy of the drive. However, depending on the circumstances this may well be overkill.

Oftentimes you don't really care about the information contained in the metadata, you are really just looking for the electronic equivalent of the paper file. Thus, in your meet and confer one of the first things you should consider is whether you are concerned about metadata, or system files, or other information that is hidden. If both sides understand the risks and are only interested in seeing the specific content of the document or email you can agree to replication methods that are far less expensive and easier to perform than a full forensic image.

PROCESSING ...

When dealing with the processing phase you need to keep specific issues in mind relating to specific types of data.

Restoration and Backup ...

One of these is restoration of backup tapes. This is perhaps the most costly and time-consuming area to deal with because backup tapes are ordinarily intended for disaster recovery, not as substantive archives of information. Often the client maintains an inadequate record of the backup tapes, including a failure to adequately label the tapes themselves so you know what's on them or when they were created. You may find that backup tapes exist, but that the hardware and software originally used to create them was long ago abandoned, and the client no longer has the capability of accessing the information on the backup tapes.

One of the key things to keep in mind when beginning the collection process is that you must avoid alteration of the data and/or metadata.

Furthermore, if you do need to restore information from backup tapes, even if you still have the necessary software and hardware, this requires a process of restoration that could conflict with current live information on your client's servers. Thus, when it comes to tape restoration you really should not expect your client to handle the restoration. It is extremely important that you use a third-party vendor equipped to deal with the restoration and subsequent search of the restored files.

Again, discussion of the feasibility of restoration and the accessibility of such backup information should be addressed at the meet and confer. Simply because it is backup does not necessarily mean it is inaccessible as that term is used under the Federal Rules of Civil Procedure. However, it may raise issues

(Continued on next page)

of cost shifting that should be considered. Furthermore, it may be prudent to use a method of sampling of selected tapes to determine if there is any relevant information on them. You could easily waste a tremendous amount of money restoring every backup tape only to find that the information contained on them is redundant or not particularly relevant.

Finally, the issue of backup tapes needs to be addressed early on to avoid the unnecessary cost of buying additional tapes when a litigation hold has been put in place. As a general proposition, in order to avoid any charges of spoliation, when a litigation hold is put in place this should include a hold on recycling or destruction of backup tapes. However, I've already experienced a number of cases where the purchase of additional backup tapes to meet the litigation hold requirements ultimately exceeded the settlement value of the case. Thus, it is important to arrive at an early agreement on which backup tapes need to be preserved and which can be recycled. It is also important to negotiate an agreement that enables a client to return to its normal backup procedures, lifting the litigation hold, as soon as possible.

When it comes to tape restoration you really should not expect your client to handle the restoration.

Parent-Child Relationships ...

In the old days we always wanted to view documents in the order in which they were maintained in the normal course of business. Seeing that a particular item was stapled to another, or finding a Post-it note attached to a particular document, could provide some useful information on what transpired historically. It is important to keep this concept in mind when dealing with the collection and production of electronic files. This is particularly true with respect to email and attachments. You want to ensure that any method used to extract email files and/or attachments retains the link information or embedded data and maintains any parent-child relationships. As recent case law has demonstrated, not all third party vendors are capable

of preserving those links, and the choice of a vendor that destroys the links in the process of extraction can be very expensive.

Metadata ...

As previously discussed, it is important to take steps to secure the metadata. Typically you should begin the process with the expectation that electronic information will be provided in native file format so that you have the metadata available to be produced. That means when extracting the information you want to ensure that the metadata is properly collected and archived as part of the processing of the source data. If this is not done, the integrity of the collection process can be challenged.

However, it may well be in a particular case that you don't really need the metadata, or you only need certain metadata fields. For example, when it comes to email, you may want the to, from, blind carbon, and date created information, but you're not really concerned with any of the other metadata that typically accompanies an email file.

Again, these are issues that can be addressed at the meet and confer, and the eDiscovery vendor who will be processing your data can advise you on appropriate methods to limit the production of metadata and hopefully thereby limit some of the costs of production.

Deduplication ...

Deduplication is a processing technique that is intended to identify electronic files that are duplicates of one another so that every single version of a particular document need not be produced. Deduplication is one of the most important cost-saving mechanisms available. The need for deduplication may exist for a given individual, or for an entire IT infrastructure including all users and all data.

There are several approaches to deduplication that may be employed. Vertical deduplication is used to deal with the records and data of a single custodian. Horizontal deduplication applies across all custodians and all data sets. Each method has its pluses and minuses and depending upon your needs you should discuss which approach to use with your eDiscovery vendor.

(Continued on next page)

Once again, the deduplication process should be discussed at the meet and confer since any deduplication process may result in the loss of some information, and the agreed upon recognition of that risk at the outset will avoid unnecessary motions later in the day.

Format ...

Once the electronic information has been secured the next step in the process is to determine the format in which the information will be reviewed. Electronic information can be produced in quasi-paper formats, e.g., TIFF or PDF, it can be produced in native format, or, of course, it can be printed to paper. Obviously, producing the information in paper form eliminates the advantages that electronic review can offer. Even quasi-paper formats are of limited value.

The best practice in today's world is to produce the information in native format, at least for purposes of internal review prior to production to opposing counsel. The native files may be reviewed with a litigation support program that has native file review functionality, or your eDiscovery vendor may recommend that they be uploaded to an online repository where review can be conducted using the proprietary review tools that individual vendors have developed.

Schedule of Production ...

Finally, depending on the amount of information that needs to be collected, processed, analyzed, reviewed, and produced, the time involved in completing all of these tasks may be significantly longer than a similar paper production. Again, at the meet and confer consideration should be given to an achievable schedule of production, the identification of a list of priority items to produce, and an agreement to engage in a method of rolling production. Consideration of these issues early on will enable you to manage the process of dealing with your client's electronically stored information in the most organized and cost-effective way possible.

CONCLUSION

With proper planning and a thorough understanding of the eDiscovery process, you can collect and process a great deal of your client's electronic information without the need to review and analyze non-relevant or redundant data, or risk the spoliation of data that you will ultimately need to produce to the other side.

Not knowing what you are doing or waiting until your uninformed client has done something wrong can lead to very unfortunate consequences. Reference to Socha-Gelbmann's [Electronic Discovery Reference Model](#) and selection of an appropriate eDiscovery expert will help you deal with this new and expanding area of the law. You might as well jump on board because the eDiscovery train left the station a long time ago.

Copyright 2007 Bruce A. Olson. All rights reserved.

ABOUT THE AUTHOR

Bruce A. Olson, is a shareholder in the Milwaukee based law firm of [Davis & Kuelthau, S.C.](#) A trial attorney and nationally recognized legal technologist focusing primarily on the areas of electronic discovery and litigation technology, Olson is AV rated and Board Certified by the National Board of Trial Advocacy. He is co-author of "The Electronic Evidence and Discovery Handbook: Forms, Checklists and Guidelines," published by the American Bar Association. He received the prestigious "TechnoLawyer of the Year 2002" @ Award from the TechnoLawyer Community, and was Chair of ABA TECHSHOW 2004, Vice Chair of ABA TECHSHOW 2003, and served on the TECHSHOW Board of Directors from 2000-2004.

Contact Bruce:
E: bolson@dkattorneys.com
T: (920) 431-2230

About TechnoFeature

Published on Tuesdays, *TechnoFeature* is a weekly newsletter containing in-depth articles written by leading legal technology and practice management experts, many of whom have become "household names" in the legal profession. Most of these articles are TechnoLawyer exclusives, but we also scour regional legal publications for superb articles that you probably missed the first time around.