

Module 12: Processing Electronically Stored Information

Learning Outcomes

Upon completion of Module 12, students will be able to:

- Interpret the duty to preserve electronic evidence that might be potentially relevant in litigation and when this duty begins.
- Identify the fields of metadata that should be included in any production of electronically stored information (ESI).
- Assess the risks and challenges in the processing/analysis/review stages of e-discovery.
- Compare and contrast the main approaches for processing a collection of electronically stored information (ESI).
- Evaluate some of the software tools that can be used in the processing/analysis/review stages of e-discovery.
- Outline the individual activities that take place during the processing/analysis/review stages in e-discovery, including planning, training the review team, sampling and quality control and first-pass and second-pass review.
- Defend the process for designating some ESI as either privileged or protected by another confidentiality doctrine.
- Summarize the important points in Module 12 and the reading assignment.

Please be sure to take Exam #3 between Friday, November 15 at 5:00 p.m. and Sunday, November 17 at 5:00 p.m. Note that Exam #3 will cover Modules 8-10 and be a mixture of true/false, matching and multiple choice questions. Good luck!

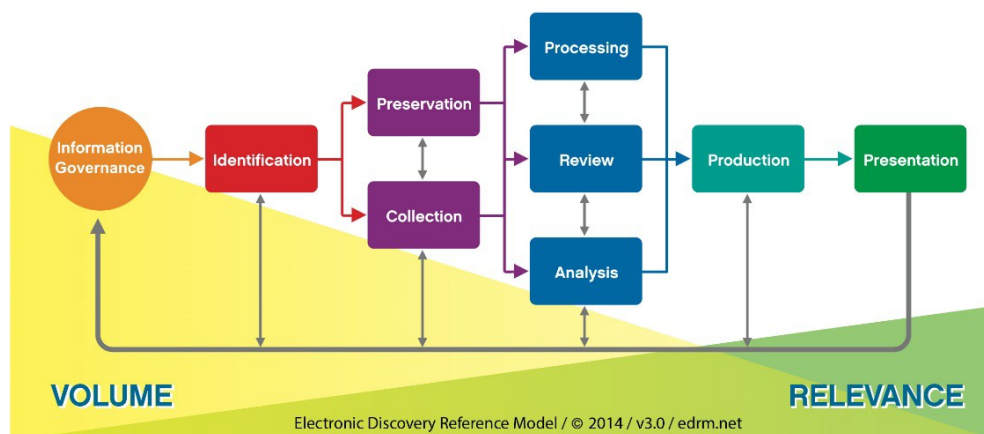
Please read Chapter 5 in *E-Discovery: An Introduction to Digital Evidence*. Please also read the following, included as part of Module 12.

- Chapters Six, Seven and Eight from *Electronic Discovery for Small Cases*, which cover Quick View Plus (<http://avantstar.com/>), dtSearch Desktop (https://www.dtsearch.com/PLF_desktop_2.html) and Digital WarRoom Pro (<https://www.digitalwarroom.com/>) as tools for e-discovery in small cases.
- ESI Review Protocols: No More Shots in the Dark – from *The Indiana Lawyer*
- Common Problems With E-Discovery – and Their Solutions – from *The Federal Lawyer*
- Other suggested listening:
 - TAR 2.0: Predictive Coding Simplified - http://connect.iltanet.org/viewdocument/09-28-16-webinar-recording-tar-2?_ga=1.236684929.1469280009.1476121881&ssopc=1
 - Predictive Coding in the Real World - <http://connect.iltanet.org/viewdocument/08-05-15-webinar-rec>
 - Beyond Predictive Coding: The True Power of Data Analytics - <http://connect.iltanet.org/viewdocument/05-19-15-webinar-rec>

Please read the following mini-lecture.

With Module 12, we move into the processing stage of electronic discovery, which includes review and analysis. It is helpful to visualize this stage using the EDRM, which we were introduced to earlier in the semester. Recall that Module 9 was devoted to information governance/litigation readiness, Module 10 covered collection of electronically stored information and Module 11 examined how to preserve electronically stored information. Note that the processing, review and analysis stages are iterative and not linear and that the volume of electronic evidence that we might have to handle continues to decrease, even as the relevance of the evidence that remains for us to produce has increased.

Electronic Discovery Reference Model



The three chapters from *Electronic Discovery for Small Cases* introduce us to four tools that can be used in the processing/analysis/review stage of an e-discovery process.

Chapter 5 in *E-Discovery: An Introduction to Digital Evidence* describes the litigation process from the perspective of the party that is responding to a request for electronically stored information (ESI). Note that in most cases, both parties will be requesting and exchanging ESI, regardless of whether they are the plaintiff or the defendant. The authors remind us that the preservation stage in an e-discovery process is critical, with the goal of collecting and protecting potentially relevant electronic evidence in a way that is legally defensible and proportional to the discovery request. The authors highlight the importance of cooperation as embodied in FRCP Rule 26(f) and The Sedona Principles.

On page 114, the authors review the duty to preserve, which is embodied in The Sedona Principles and articulated in *Zubulake IV* (and now in the 2015 amendments to the Federal Rules of Civil Procedure). Note that the responding party is responsible for evaluating which procedures and methods are best suited to meet the preservation order and is obligated to preserve and produce relevant electronically stored information in good faith. As the authors advise, to identify all potentially relevant data, the responding party must use a variety of methods and tools for searching and data sampling.

Recall that we covered the collection stage of an e-discovery process in Module 10. Of course, in order to identify the electronically stored information that will need to be collected, FRCP Rule 26(f) mandates a “meet and confer” conference between the parties. Failure to cooperate in planning the e-discovery process can result in sanctions. What is quite interesting about the *National Day Laborer* case reported on page 115 is Judge Schendlin’s observation about metadata, wherein she noted that “metadata is generally considered to be an integral part of an electronic record.” She recommended that nine fields of metadata be included in any production of ESI and suggested several additional fields for email. Finally, she stated that certain metadata fields must accompany images of paper records, such as beginning and ending Bates numbers. As indicated on page 116, courts prefer that requesting parties receive ESI in the same form and with the same content as what the producing party uses, with metadata intact, rather than in a “litigation response format” such as PDF or scanned images. Note that ESI can be collected with software tools such as Discovery Attender and AccessData Forensic Toolkit. As emphasized on page 116, an important factor is ensuring that technically qualified people are used to manage the collection of ESI because they have expertise in knowing where different types of data might be stored.

After relevant ESI is collected, the processing stage of e-discovery takes place that will not only reduce the volume of this ESI, but will also convert it into formats that facilitate review and analysis. One of the benefits of new technology such as predictive coding and technology- assisted review (TAR) is that this technology can reduce the cost of this part of the process, which tends to be quite expensive if it has to be conducted only by humans. Because human beings are prone to error, a smaller set of data that eventually has to be reviewed by humans before production will hopefully mean that ESI that could and should be protected as privileged will not be inadvertently produced to the opposing party. This is where I see the powerful opportunities that technology can provide in e-discovery.

Requests for data can be received via a subpoena or an e-discovery request. As indicated on page 116, using a targeted scope for data collection requires configuring search criteria in the software tools for collecting data, which can include proximity searches, keywords, date ranges, email addresses and file types. One of the most common activities in the processing step of e- discovery is to remove duplicate files. Recall the article from a previous Module that indicated that a law firm might generate an additional 10 copies of the same file even after it had received it as part of e-discovery! As the authors observe, many e-discovery collection and review tools have built-in features to compare files and remove duplicates. At this point, data sets are typically processed in one of two ways, depending on the file format that the parties have agreed upon. One method is to continue to the review stage with the files in native format. The other method is to convert the files into TIF with OCR (optical character recognition) or into searchable PDF. There are advantages and disadvantages to each of these methods.

As indicated on page 117, most tech-savvy industry experts push for native file review so that time and money are not invested in converting files that are already searchable and that have their metadata intact. Figure 5-1 on page 118 shows that even after duplicates have been removed, there are still over 1300 email messages and 199 attachments in a variety of file types. You will need a

review tool that allows searching without altering the metadata. The authors explain how this can happen with such common software as Word, Excel and PowerPoint. Discovery Attender and FTK are good tools for searching which do not alter metadata, but they do not have the features some lawyers would like for organizing their work. For example, we might prefer processing software that automates functions such as Bates stamping. Figure 5-2 shows an exhibit attached to a summary judgment with a Bates number identifying the document and page. Note that we need to be able to redact certain information from documents, such as the example of the proprietary algorithm on page 119. Redacting part of a document requires a tool that can handle this task without altering metadata, which may be a feature of some collection software.

After redacting protected data, lawyers will consolidate all remaining ESI into one location for review and production. This includes paper discovery, which is typically scanned and imported into the litigation database, along with digital files, deposition transcripts, audio and video files, etc. Before production, Bates numbers and redaction marks will be “burned in” and cannot be removed.

Web-based repositories are covered on pages 120-121. As the authors indicate, for all but the smallest cases, web-based repositories are a good option for review and most can handle native file formats as well as paper. By using a web-based repository, the law firm can focus on the legal issues of a case, leaving the problems of technology, infrastructure and database administration to a third-party vendor, who will maintain the chain of custody and coordinate database configuration. Not only does a web-based repository provide security, but multiple lawyers can access the data quickly from anywhere. The authors caution that we need to consider software licensing requirements when choosing between in-house versus web-based review platforms. Fortunately, the market for e-discovery tools has become quite competitive in recent years. The authors focus on iCONNECT as just one example of the possible choices.

Although the American Bar Association has deemed outsourcing to be acceptable (which was one of the issues covered in its Ethics 20/20 Project), remember that the lawyer of record is ultimately responsible under Rule 5.3. The authors (as well as Nelson and Simek) remind us that there may be special concerns with privacy and security if the data is transported and stored offsite, but especially with a vendor outside of the U.S. Thus, security controls must be specified and included in Service Level Agreements (SLA).

The authors describe the use of native file conversions using in-house repositories on pages 124-126. They compare the use of searchable PDF files with TIF and OCR. Note that even litigation databases and trial presentation tools, such as Concordance and Summation, may still work best with TIF even though they can handle PDF and native files. PST and embedded files may present challenges for these tools. On the other hand, the authors advise that when discovery is converted to TIF with OCR, Bates stamping and redacting are available, but metadata has to be captured separately. The authors suggest some other tools on page 124 for conversion to TIF with OCR. The final processing output from these tools is a group of files, usually TIF images, image logs and load or data (.dat) files. These files populate the review platform with metadata, OCR content

(extracted text) and coded information and operate as linking files. Figure 5-7 shows an example of how these files are related. You can see the Enron iCONNECT database on the left with the corresponding TIF image on the right. The .dat file contains database field names as well as OCR. The image log file shows code for finding the correct image in the database to display and lists a page count for each document. The actual TIF images are in a separate folder that is not shown. They are displayed in the viewer when activity occurs in the corresponding database field.

After the processed discovery is loaded into the review platform, additional database fields are coded with information similar to what was once included on index cards (such as issues, privilege, etc.). The authors define coding as the process of examining documents to identify names, dates and relevant keywords or phrases and entering the results into a database, either manually or with software. Note the explanation of objective versus subjective coding. Coding provides information used in fielded searching and sorting. Full-text searching involves all of the contents of documents that have been converted to OCR. A compromise format is called near-native conversion, which helps to deal with Excel, Access and other formats that do not easily lend themselves to TIF conversion. Of course, we still may have to handle paper documents, so the authors describe how to convert them to TIF with OCR. Although the technology has improved considerably, OCR quality can still vary. This is another quality control issue that the review team should monitor, along with adjustments in field coding and search methods (See Figure 5-8). Subjective coding can be added to the database to improve search results and auto-coding tools have been developed. Authority lists and coding manuals can be prepared to improve the consistency in data entry.

Review begins after the forensics consultant, vendor or IT staff have finished identifying, collecting, preserving and preliminary processing of ESI, including deduplication. As the authors note on page 127, by this point, the choice of review platform has been made and discovery has been processed and imported into it. The legal team will then review the preliminary data sets to find relevant material and to identify any material that may be protected by attorney-client privilege or other confidentiality doctrine. As the authors caution, as the volume of discovery increases, there may be difficulties with an automated review process. Some challenges include failing to understand and use the technology correctly, poor communication, loss of quality control and linguistics issues in searching. The authors provide an overview of the review process and describes a few of the difficulties:

- Planning session, including preparing a review manual containing all materials that members of the review team might need. (It is not uncommon for law firms to use contract lawyers for the review).
- Review the core documents (complaint, responses, requests for admissions or interrogatories) and put keywords, phrases and key players in context. The authors suggest that we question the relevance of similar words not initially included in searches as well as acronyms, initials, etc. (Recall in Ms. Zubulake's case how she was referred to as "1" rather than by her name or initials.)
- The lead attorney will explain the preliminary case strategy.

- Training in the review platform's search features, including fielded and full-text searching, keyword, Boolean, fuzzy, stemming, proximity, wildcard, concept and cluster searching. Note that most review platforms can be configured to recognize special symbols.
- All reviewers should be given clear examples of potential search problems, such as poor quality OCR, and workarounds for these problems.

Pages 129-130 describe the process for sorting relevant versus non-relevant data, in what is called a "first pass review." As the authors indicate on page 129, the data set is divided into groups. Document reviewers are assigned portions of these groups to search and tag with suitable categories. As you can imagine, this document review can take considerable time and be expensive, often not always practical when the volume of discovery to review is large. Search protocols, which are established by the lead attorney, are used as a review filter, including when to use wildcards, stemming and other search methods based on OCR quality as well as the frequency of acronyms, misspellings and other word deviations that are identified during periodic sampling for quality control. A common rule of thumb is that an associate or paralegal can review 40 to 50 documents per hour, with a "document" being defined as 4-5 pages. Not only can you see how quickly the expense for this review process would add up, but note the very real risk of mistakes when people have to review at this speed, including that ESI that could be protected as privileged or confidential would still end up in the eventual set of ESI that is produced to the opposing party. Also, the authors observe that discovery requests that are overly broad produce more ESI that is of possibly little value, but that still needs to be reviewed, contributing to additional costs (and such overbroad requests are now prohibited under the 2015 amendments to the Federal Rules of Civil Procedure). The authors recommend that a random data review be conducted to make sure that the collection process has identified data that are relevant to the request.

In most review platforms, case documents are organized using tags and folders. The authors suggest that, at a minimum, each document should be tagged as privileged, responsive, non-responsive or for further review. Figure 5-9 shows folders and tags used in IPRO Eclipse. Note that this document has been tagged as "Attorney Eyes Only" to identify the document as attorney work-product.

As indicated on page 130, next is the second-pass review for material that can be protected under the attorney-client privilege or as attorney work-product. The review team should be given search terms for this review, including the names of attorneys, email addresses and other words such as "privileged" or "litigation." Note that the review team needs to know how to handle parent-child documents (an email is privileged, but the attached document is not). Not only is it important to use consistent review processes and procedures, but quality assurance checks should be done on sample sets of search results. Recall again the risks of inadvertently producing ESI that could and should be protected as privileged or confidential because of a sloppy e-discovery process, but also because of the labor-intensive nature of the review step, which relies so intensively on human beings to conduct the review. Although the parties may negotiate claw-back agreements as part of the "meet and confer" conference, the authors caution that these agreements should not be the

basis for poor review processes. What I always think of is that “the cat is out of the bag” anyway, meaning that it is difficult for a lawyer to erase the content of the inadvertently produced ESI from his/her memory. Also note that courts differ greatly on how they handle waiver of the privilege. As the authors state on page 130-131, many problems can be avoided by documenting reasonable review processes and conducting sampling and quality assurance testing on results. They cite the case of *Holdings LTD v. Microsoft Corporation* as an example of due diligence and a defensible review, even though a software glitch meant that otherwise privileged ESI was inadvertently produced.

Reviewing for protected ESI is often done during the review stage and can include any case material that needs to be protected, such as trade secrets, software code, algorithms and even recipes, or that qualifies under the attorney-client privilege or as attorney work-product. In criminal cases, protected information could be witness identities, phone numbers and addresses. Because case documents filed with the court are normally a matter of public record, materials that contain information that would cause harm to a person or company may be filed “under seal” with the court to keep them from being accessible by the public.

After all review and quality assurance steps have been completed, ESI is produced to the opposing party in the format agreed upon (which we will cover in Module 13). A description of ESI subject to redaction or withholding because of privilege or another confidentiality doctrine should be entered into a privilege log, which is also produced to the opposing party. The authors have provided the text of FRCP Rule 26(b)(5) on pages 131-132. The objective of a privilege log is to disclose information to the opposing party and the judge so that it is clear what has been identified as privileged and why. The authors list the information that should be included for each item on the privilege log on page 132. Note that most review platforms can be configured to generate reports that automatically contain most of this information. After this stage, the producing party continues to organize materials related to case strategy, such as creating folders or notebooks for chronologies, witnesses, experts and issues. The case is then prepared for trial, arbitration, mediation or settlement.

The authors provide a brief explanation of arbitration and mediation on page 132, which are methods of alternative dispute resolution (ADR). One thing that is important to note is that the decision reached through arbitration is generally final; absent a very narrow exception, the party does not then get to appeal the decision in court. Most people would be shocked if they read a lease, mortgage, loan document, etc. because most likely this contract requires that any dispute be settled through binding arbitration. You might have noticed, in locating and reviewing e- discovery cases, how often a special master is used to decide discovery disputes and he/she can also serve as a mediator between the parties about what gets turned over to the opposing party. Chapter 5 of *E-Discovery: An Introduction to Digital Evidence* concludes with a summary of the chapter and a list of key terms.

With Module 13, we will consider production and presentation of electronically stored information, concluding all of the steps outlined in the EDRM.

Please let me know if you have questions about Module 12 or the reading assignment.

THE HUNT

Recovery and Acquisition

5.1 Introduction

In this chapter, we begin looking at the recovery and acquisition procedures you should think about in your organization or even as an individual. These processes and ideas can assist anyone who needs to better understand the “where” and “how” of electronic evidence (Figure 5.1).

We consider the “hunt” and all of the different ways that all important electronic data may elude capture. We also discuss the classification of data as “privileged” or “sensitive” or “unreasonably accessible,” and all the requirements you need to meet to prove those classifications. We give you the means to recognize, organize, and document those classifications in a way that will hold up in court, or for an audit, or simply for your own organization and protection of said data.

We also talk about nonrepudiation versus plausible deniability, the very basis of evidentiary law, and in many cases, the facts on which a court will depend to decide if evidence is admissible.

A big part of that nonrepudiation or integrity of electronic evidence for admissibility is often based on time stamps, so we look at how time stamps work, how they can be corrupted or compromised, and how they can be used for or against your case to prove the efficacy, integrity, and reliability of your data.

Finally, we lay out some suggested procedures for the gathering of the relevant electronic data in a way that the courts will respect and understand as forensically sound. In other words, we offer you tools and processes to allow you to collect, organize, and preserve electronic evidence in a way that you can ensure no alterations were made to the data and that the integrity of the data was maintained in its pristine and original state. Let's get started.



Figure 5.1 The hunt begins.

5.2 Where, Oh Where, Has My Data Gone?

The first thing you need to consider when you begin the hunt for relevant data is the big question of where is it hiding? Like any good hunter, you need to first be familiar with your prey and its habits. In Chapter 3 we spent a lot of time considering all the different types of electronic data and the myriad ways in which it is created and stored. For details on the question of where data could be, I refer you back to that chapter.

But even if you memorized every word of that chapter and completely considered all of the possible places the data you need might be located both physically and virtually, you might still have a problem finding it. How could that be, you ask?

As we noted before, computers and other digital devices that store electronic data can and do lose data sometimes or possibly store it in ways that are inaccessible to the normal user. And in many cases, data are stored in a way that can only be accessed, manipulated, or recovered using a specific front-end application or interface. We look at each of these with examples to make this clearer.

5.2.1 *Applications as a Vital User Interface*

For our first example, we look at the most popular electronic evidence item in most court cases: e-mail (Figure 5.2).



Figure 5.2 E-mail is the most often requested electronic evidence.

Nearly all e-mail applications that reside on a computer have a database of some sort behind them that is responsible for storing and organizing all of the items you receive or input. When you get an e-mail message from your sister in Roxbury, the database stores it in the database in such a way that when you open your e-mail application and go to your Inbox, you will see the new message from your sister.

Usually, that message, being new, will have some formatting attached to it to indicate that it is new and that you have not opened it yet. That might be a closed envelope icon, or something similar. That message also contains a considerable amount of metadata or data about itself, as we discussed in detail in Chapter 3. It will contain information about who the original sender was, what time it was posted, the e-mail servers it started from and went through to get to yours, the types of e-mail applications that were used, what time it arrived, and more. The message may also contain attachments such as pictures and documents that your sister sent to you. And each of those items will also contain metadata that contain information about when they were created or modified, by whom, and so forth.

However, as soon as you open that message, its properties are changed and the back-end database will add new attributes to that data and possibly store it in a different way when you close your e-mail application or shut down your computer. If you open and save the attachments, that will be recorded in the e-mail application database. You might choose to move the message to a folder to keep yourself

organized, such as perhaps a Family Stuff folder or, depending on your relationship with your sister, maybe to the Junk Mail folder.

Those actions will have an effect on the way the database behind your e-mail application treats that message. The message may be stored in a different way to indicate it can be found in the Family Stuff folder. It might have special attributes or rules attached to it as a result of being stored in the Junk Mail folder.

Some e-mail applications will make the assumption that when you store things in Junk Mail you would prefer that any further correspondence from that same e-mail address be stored therein as well. So, it is possible that future e-mails from your sister will then be automatically relegated to the Junk Mail folder.

Of course, in most modern e-mail applications, you are also able to create and store a task list, contacts information, and calendar items. Again, each of those pieces of information is stored in the database in a particular way so that when you use your e-mail application to look them up, they are revealed to you in the place you expect them to be and in the correct format to indicate whether you have read them or completed them yet.

Sometimes in the case of calendar items, they can be formatted to indicate if they are family related, work related, birthdays, holidays, and so on. All of this information is stored and managed by a back-end database.

In many cases, those databases are based on what is known as proprietary programming. This means that the organization that created the e-mail application and its back-end database chose to use a database programming language that can only be read by that application and is specific to that application. In other words, there is no way to locate that database and find the information you stored in it, other than by using the e-mail application for which it was created.

And here is one way that you might know perfectly well where your data live, but be unable to access the data. Any application can fail, go out of date, or be accidentally deleted. So if what is lost is an application that used proprietary programming to store data, it might be impossible to access that information without recovering and reinstalling the original application.

Here is an example of that problem from my own experience.

As the family geek and official tech support guy, I cannot say I was terribly surprised one day when my Mom called to say that their computer was no longer working. When I began my troubleshooting process, I again was not too surprised to hear that my Dad had been doing some “cleanup” on the hard drive because it was getting too full.

This was back in the day when a huge hard drive might have as much as 500 megabytes of storage, and theirs was an old computer with considerably less than that. As they added pictures, documents, and applications with little understanding but a great deal of enthusiasm for their brand-new computer, they began to run out of space.

If they had come to me first, I would have spent some time cleaning things up and could probably have gained back plenty of space by deleting temporary files, and moving some things off to CDs. But Dad thought he had it figured out and simply went through the C: drive directory deleting things that he figured were not important.

When he deleted sol.exe, Mom was a little bummed out that she could not play solitaire anymore, but she let that pass with her typical stoicism. But when he went back and decided that win.exe looked like an unnecessary file, things went south in a hurry. Of course, many of you are probably giggling right now, at least those geeks among you, because you recognize that once upon a time, win.exe was the application that started up Windows.

Microsoft has since become a little more human savvy, and they now hide that and other all-important system files so you cannot mess with them. That, of course, brings up another way that data can be made hard to locate, but we get to that soon.

The end of the story with my Dad and Mom's computer was that I needed to go to their home and reinstall Windows for them before they could use their computer again. And the point of that story, besides entertaining all of you, was to help you understand how important and indeed vital data might become impossible to access simply by virtue of a vital application being deleted, corrupted, or lost in some way.

In my parents' case, they could no longer access anything on their computer even though everything still existed on the hard drive.

If we go back to our example, they could in no way have accessed their e-mail because the application that ran their e-mail depended on Windows to run and to give them the interface they needed to

open, review, store, or manipulate their messages, calendars, tasks, and contacts. All of that important data was as good as lost to them.

Even if they had not deleted win.exe, but had simply deleted outlook.exe, or whatever other e-mail application file they might have been using, the results would have been the same. There is simply no way for a normal computer user to access those back-end databases of information, even though they exist and are perfectly sound on the hard drive, unless they have the front-end application that creates the interface with the database.

Now, all of this e-mail, contact information, task lists, and calendaring items might be the most important things you need or want to produce for a trial or an audit, or simply for your own records or mementos. So, if something happened to the front-end application that gives you access to that data, what can you do?

There are several different possible solutions to that dilemma.

The first and most obvious one would be to acquire and reinstall the front-end application to access your data. However, this can be tricky. You will need to have the exact same version and type of application that you created the database with in the first place if you have any hope of it still being compatible and accessible.

You may also need to do some manipulating of the reinstalled application to allow it to find and open the back-end database that was left over from your original application. This can be a difficult task, requiring some technical expertise to accomplish, depending on the application. In many cases, you may be well advised to ask for help, either from your family geek or from a real expert like that 13-year-old who lives across the street.

Another possibility will be to find a separate tool or application that is made to open and access the back-end database for your particular application. In the case of Microsoft Outlook, for instance, that back-end database file can be stored as what is known as the PST file. It is a hidden file that contains all of the information from your e-mail application, stored in a proprietary programming language that Outlook is designed to access and interface with.

However, there are other programs available that allow a user to access and open up a PST file and search for e-mail, calendar, contacts, or task items. These are specialized tools that are probably most often used by forensics or other investigative organizations. Some law

firms are also likely to have tools to open these and other back-end e-mail databases because it is not uncommon for a discovery request for e-mail data to be produced as the back-end PST file.

Other applications might have a SQL database, or an Oracle or other database system backing them up. This is especially often the case with Internet-based applications.

If that is the case, there are many tools and applications available that will allow you to open and query those databases to find the relevant evidence. Of course, the efficient use of those tools will require expertise and experience that not all of us can claim. Once again you are well advised to consult with an expert or the kid across the street.

5.2.2 *Hidden or Restricted Access Data*

Another place that important information can be stored is in log files or data files that your computer's operating system might hide from view. As noted in my example of my parents' mistaken "cleanup," there can be very good reasons for hiding these files from the normal computer user.

These types of files and folders are called *hidden* or *system* files, and most operating systems have some special storage space just for them that is inaccessible to users without the correct privileges on the device. Yet they may have information that could be important to you in a court case, or just in the general understanding of what might have occurred when on your computer.

There are several different types of files on Windows devices that are hidden from normal view. These include files that contain data for what is known as the registry.

On a Windows system, the registry is yet another database that keeps track of all of the different software and hardware your computer uses, the paths and interdependencies to those applications, and connections between the operating system, applications, and the hardware. It is a very complex and technical database of connections that is the heart and soul of how your computer and its applications interact with the Windows operating system.

There is no reason or need to get into details of the design or even the contents of this vast database in this book. However, it is important to recognize that the registry contains electronic

records of actions taken on your computer that may be relevant in a court case.

Here are some examples of the data that can be found in the registry and that might be of interest to someone trying to understand what has taken place on a computer, for whatever reason:

- A record of every USB device that was ever plugged into your computer—the serial number, model, and date it was first plugged in as well as the date it was last accessed
- Records of all of the last websites you typed into your browser
- Records of all of the most recently accessed files (depending on the version of Windows, these can be records of hundreds of files back into the past)
- Records containing data about all users of the computer, including encoded passwords (that with the right tools can be un-encoded)
- Records of all network and other drive connections that have been made
- Records of what applications have been installed, when and where they are located in the directory system
- Records of when applications were last accessed
- Records of who logged in when as well as what networks or shared directories they were connected to when they logged in
- And much more that can be of great interest to someone with the tools and expertise to winnow that data from the registry in order to understand who has done what on that computer

Other files that are hidden are the files that contain information about your Internet browsing. These are stored in small flat database files with a .dat extension on some Windows operating system computers. For Internet Explorer, these files are most often called index.dat and are stored in various hidden folders depending on their specific purpose.

One folder is for your Internet history, and with Internet Explorer the index.dat file in that folder contains the information you see when you select History from your Internet browser's menu.

Other browsers use different storage mechanisms, but in every case these are not files that you can simply find and open on your own. Some of them are made available through the interface of your

browser. History is one, and another common one would be the cookies file.

Cookies are small text files created by websites. They contain information about your access to the websites with the purpose of speeding up your access when you visit the same site later. They can, however, be less benevolent and store information about your browsing history with the intent to know what you purchase and target advertising toward those favorite items.

So cookies, though generally benign, can be used in more malevolent ways, and there are reasons sometimes for you to check them and possibly remove them. Many modern antivirus programs will look for the malevolent versions of cookies (sometimes called *tracking cookies*) and delete them for you.

However, in order to see and manage cookies, you most often have to access them through the interface of your browser. There are other tools and applications that will also allow you to manage cookies, but in every case they are referencing one of these hidden data files to which we are referring here.

There are other hidden data files like these as well that keep track of every single item that you access either on the Internet or even just on a network. I think it is pretty easy to see how those could contain vast amounts of interesting information about the activities on a computer, especially Internet browsing activities.

Another hidden folder in Windows operating systems contains data on all the files you accessed in a given period of time. It is called the 'Recent' folder, and similarly to the registry items we noted above, it can give a lot of good information to an investigator about what files a computer user accessed recently. Included in that database will be the name of the file, the path where it was located, the creation date, and the last accessed date. All of this might be extremely interesting information.

Other hidden files and folders contain information on all of the applications you installed and any error messages or logs that were created by those applications. Others contain data on security alerts or system issues, such as every time you logged into or out of your computer, or when the time clock was changed. Again, these can be extremely important sets of information to someone trying to establish use patterns or general usage of a device.

There is one hidden system file in Windows computers called \$MFT. MFT stands for master file table, and that is exactly what it is. Again, it is a flat database created the first time your computer starts up, and it keeps track from then on of every single file created on your computer. The MFT of a brand-new computer already has tens of thousands of entries, and that number can grow into hundreds of thousands as you add new applications, create documents, surf the Internet, and so on.

Almost every action you take on your computer will result in files being created. Sometimes it is log files that an application keeps in the background for the purpose of troubleshooting, or simply as a means of evolving and making your experience more efficient.

As we noted before, every trip you take on the Web creates literally hundreds of new files as it opens up the various pictures, text, movies, and so forth, that make up a website.

Every one of those file creations, modifications, last access times, deletions, archiving, or anything else that the operating system needs to track about those files is recorded in the MFT. This can be a wealth of information for someone who needs to know what happened on a computer. In fact, it can be far too much information.

As a computer forensics practitioner, I use the \$MFT file on a regular basis to try to understand when malware infected a computer and what type of malware it is. Yet, even if I have a pretty good idea of when the device was infected, it can be like searching for a needle in a haystack to look through the thousands of files created every day on a computer.

One more type of hidden file that can be of great value in understanding the actions of a computer user is called the “*Restore*” file or “Shadow Copy” in some Windows operating systems. The Restore file contains records that the Windows operating system keeps of the state of your computer at a given time and date. Every time you add a new application, a restore file is created. Windows also creates these whenever it does an update, or just on a regular schedule.

These are created for the purpose of restoring your computer to a known good state in case something goes bad. Sometimes a new application will cause conflicts with existing applications, or will not work correctly with your device for some reason; perhaps it has bad hardware drivers or some incompatibility with your system.

If that happens and you recognize that the problems started when you installed that software, you can use the system restore points to roll back your computer to the state of things before the problem started. Now Windows keeps these restore points for a long time, going back many years sometimes. Think about the value that might have to understanding what happened on that computer sometime in the past.

If someone was purposely trying to hide his or her actions by deleting applications, Internet browsing activities, documents they created, and so forth, the restore points give an investigator a way to go back into the past to see how things looked before.

Finally, as mentioned earlier, most applications will keep log files that track how they are being used and any problems that come up. Most often these are hidden files that are only reachable through the application interface and sometimes only by knowing advanced options in the application that will allow you to view and manage those files.

There are many more hidden and system files on your computer, and we only touched on some of them and only talked about Windows computers. Other operating systems such as Linux and Macintosh also have system files that are not generally available to a casual user.

Yet in every case, these can contain vital information when you need to understand what has been taking place on a computer. They can hold data that will assist you in locating other relevant information and possibly point you to other sources of information, such as those USB drives that were connected to the computer, or the list of most recently accessed files, or the information in one of the restore points.

If you want to see some of these files, or at least their folders and their file names, on a Windows computer you can select, “show hidden and system files” in your folder settings. However, being able to see them will do you no good whatsoever if you actually hope to gather some information out of them. They are all encoded in such a way as to only be accessible and readable either by an application interface or by the system.

There are special tools that computer forensics experts use to parse and access these files for analysis. So, if access to any of that type of information might be important to a court case or audit or simply

record keeping that you are doing, you will need to either gain that expertise yourself or hire someone with those skills and tools. In the understanding of where data live, do not neglect this rich source of information. In many cases, it contains some of the most important facts and clues to understand what really happened on a computer and when and by whose hand.

5.2.3 *Encrypted Data*

Yet another way that data might be made hard or impossible to access is through encryption (Figure 5.3). There are many ways to encrypt data on a computer, from the relatively simple act of assigning a password to a file, to the incredibly complex process and management of digital certificates that automatically encrypt data in e-mails, or specific folders, or even entire hard drives.

On some newer Windows operating systems, they include a tool called BitLocker. This allows you to create specific encrypted folders such that any data that you want to protect can simply be stored in those folders and a pass phrase is required to access the data.

Other encryption tools work as noted earlier with a certificate system. In those cases, an individual or a device is given an electronic certificate that is basically an encoded file that contains the necessary information to prove the individual's identity. Again, these will use a



Figure 5.3 Encrypted data can present extreme challenges to access or even be impossible at times.

pass phrase of some complexity and length to ensure you are authenticated to open files.

Any in-depth conversation about encryption is far beyond the scope of this book, but it can be a fascinating and mind-challenging study if you lean toward the mathematical geek edge of things. But for those of us normal and mortal humans, we simply need to recognize how this can affect our ability to gather and access data from our computing devices.

If you have a need to access data from a device, drive, folder, e-mail, or file that was encrypted, there is really only one sure-fire solution. That is to find out what the authentication mechanism is and learn the pass phrase or password. Failing that, even if you have a way to locate and try to open the encrypted information, the data in its encrypted form will look like long strings of bizarre characters and will mean absolutely nothing and be completely worthless.

How does one obtain those passwords and phrases? That will often depend on the circumstances.

In an ideal situation that is not confrontational, you may be able to simply ask. Even sometimes in a more combative situation such as a court case, you might be able to convince the person with access to supply his or her passwords if there is a compelling reason to do so. Each case will be different, but this is obviously the first and easiest solution and the one you should always strive and hope for.

Barring that easy solution, there are other possibilities depending on the level of encryption and the technical astuteness of the person or persons who encrypted the data in the first place. With the relatively simple encryption that is done when you password protect a file, such as an Excel spreadsheet, for instance, there are quite a few readily available tools available that will allow you to crack those passwords or sometimes bypass them.

More complex encryption schemes can be extremely difficult or even impossible to break through. Again, there are tools that your friendly neighborhood computer forensics expert will have that will make the cracking of difficult passwords possible. But even they can fail at times, and often it can be a very long and complicated process.

Encrypted data can be a real problem if the data are relevant to a case or necessary for some project you are working on. If you have no way to gain access to that data, you might have to declare it not

reasonably accessible and look for other ways and places to gather the electronic evidence you require.

5.2.4 Deleted or Corrupted Data

Of course, any data can be deleted by accident, or in some cases simply corrupted so that it is no longer accessible by the applications or computer operating system.

Whenever someone deletes a file on a computer, the operating system makes a small change to the data that makes up that file, basically placing a flag consisting of a couple of bits of information at the beginning of the bits and bytes that make up that file. That flag tells the operating system (and any application that interfaces with it) that this storage space is available and that the information stored here can be written over with new data whenever the space it is using on the electronic storage device is needed.

Sometimes, due to a problem with the operating system, a virus, or something wrong in an application, some of those bits and bytes that make up a file or other piece of information can be destroyed, changed, or moved out of sync with the rest of the file. This can make that data unreadable or inaccessible by the application that created it. This is called corruption of the file.

In both of those cases, what is important to understand is that most often a large part of the data is still intact and still available if you have the tools and skills to find it. The bits and bytes that make up the files still exist, all except for the corrupted ones. Or they might all still exist but simply have that “deleted” flag prepended to them. If these have not yet been overwritten or are not too seriously corrupted, it can be relatively simple to recover them, given the right tools and skills.

The best example I can give to illustrate this is one that happens to me fairly often in my role as the community forensics dude.

I will hear someone bemoaning the fact that he or she just uploaded all their pictures from their vacation to their computer, then pressed the wrong buttons and deleted all of them from both the camera and their computer.

Or, someone will tell me that their computer was infected with a virus and he or she can no longer get it to start up and all of their important documents were never backed up—forensics dude to the rescue.

With tools at my disposal, I can often gain access to a corrupted hard drive, or find all those deleted files on a camera memory card or computer. Those tools basically take a look at the hexadecimal code that makes up all those files. Then it goes through and recognizes the pieces of different file types and puts them back together again in a form that I can then export to recover the data.

In computer forensics terms this is called *data carving* because the tools look through the lines of data stored on whatever medium we are searching through and carve out the pieces that make up different file types.

The way it recognizes those files is by standard file headers or bytes of data that tell the operating system the type of file that is stored, its length, date and time information, and possibly information about where on the disk different parts of the file are stored.

All of this information is part of something we discussed in other chapters called metadata or data about the data. The operating system, applications, and my forensics tools read that information and use it to access the data that make up all the different types of files.

Once again, we have another type of data that though it might be incredibly important to your case or your project, might also be difficult to gain access to without the right tools and skills.

What is important to recognize here is that deleted or corrupted or lost data might still be accessible. Depending on the circumstances, such as how important that information is to your case and how available and costly the resources to recover it might be, you will need to make decisions as to whether you consider it reasonably accessible. And be aware that while you might deem it not reasonably accessible due to high costs or difficulty in obtaining the data in question, the courts or other parties may disagree with you.

This makes it important to understand what data might be there and to do your best to preserve it by not continuing to use a device that might result in that data being overwritten or further corrupted.

5.2.5 *Proprietary Data or Data Stored on Obsolete Media*

We already talked a little about proprietary data when we discussed the need for a front-end application in the first section of this chapter. However, another way that proprietary methods of creating and

storing data can result in your being unable to access that data is by virtue of its having been created by an application that no longer exists. It could also have been stored in a format or on hardware that is no longer available. Let's look at some real-life examples.

Some of you, like me, have some miles on you and have been around long enough to remember those halcyon days when computers were a new and exciting invention that only a few geeks in research universities could afford to own.

Over those many years there have been many types of storage devices. We discussed some of the more common and modern ones in Chapter 3 on where data can be stored. But if you happen to have old data kept on some of these older types of media, you will be facing different challenges.

Data have been stored on many different types of media (Figure 5.4). Here is a list of some of them:

- Punch cards and paper tape that used holes punched in the paper to code those ones and zeros we discussed before



Figure 5.4 Storage devices have been many and varied over time.

- Magnetic tape on reels the diameter of an automobile tire, to smaller reels and audio tape-sized cassettes, down to mini tape cassettes you could hide in your fist
- Floppy disks in sizes from 8 inches to 3.5 inches
- Hard drives of various capacities and interfaces
- Various magnetic disk drives such as the Jaz and Zip disks
- Other outliers such as the SuperDisk and the MiniDisk
- Solid-state drives such as SanDisk, and so forth
- Optical drives—CDs and DVDs
- And even one hybrid optical/magnetic drive

If data that you need for a case, or an audit, or just for your own historical records are stored on any of these types of older media, you could face some unique challenges.

Consider first those floppy disks.* Floppy disks were first invented and used in the 1970s, and as noted previously, they came in 8 inch, 5.25 inch, and 3.5 inch sizes. It is rare to find an 8 inch floppy these days, but you might still find some 5.25 inch disks laying around getting dusty in a closet, and the 3.5 inch floppy drives were still pretty popular up until just recently, so you can still purchase them and some computers still have a floppy disk drive for reading them.

A floppy disk is made up of a flexible (floppy) magnetic disk contained within a square or rectangular plastic sleeve that protects it from dust and damage. They could hold relatively small amounts of data by today's standards, but they were still used extensively to store information.

If you have one of these disks and need to gain access to it for any reason, you are faced with some challenges. The first is the very real possibility of damage or corruption of the data. As with any magnetic media, these are vulnerable to magnetic fields. There has also been some research indicating that they may lose some of their magnetic charge over time, thus making them impossible to read.[†] Basically, the ones and zeros are maintained by different levels of positive or

* The chronology of disk storage. http://www-03.ibm.com/ibm/history/exhibits/storage/storage_chrono20.html. Retrieved June 17, 2011.

[†] June Jamrich Parsons and Dan Oja, *New Perspectives on Computer Concepts 2012: Introductory*, Florence, KY: Cengage Learning, p. 78. Retrieved June 17, 2011.

negative charge, and if that faded, there is no code for a drive to read and interpret.

Second, unless you are a true geek and saved every computer you ever owned back into the computer stone age, you probably do not have a running computer with an 8 inch or even a 5.25 inch floppy disk drive installed. If you have the need to acquire data from one of these older types of media, it may require a trip to your friendly neighborhood computer antique store or a search online for someone willing to sell (or loan) that hardware to you.

The same thing applies to any of the other older or rare devices we mentioned above. I imagine that many of the readers of this book never even heard of a Zip or Jaz drive, though for a little while they were a pretty popular choice for storing more data. The Jaz drive could store up to 2 GB of data, which at the time was a huge storage device.

Some of the interfaces of older hard drives (SCSI for instance) are rarely used anymore and may also be hard to find or connect.

If your task is to gather some important information from one of these types of devices, I recommend an online search as your first step. You will find that there are still organizations out there that will sell you the proper hardware to access the data on those media. However, if you know or fear that the data were corrupted or the media was damaged, you will probably have to turn to a reputable and experienced computer forensics organization for assistance.

Another way that data can be made inaccessible is through its having been created in a proprietary format. Again, some of you oldsters will remember word processing applications such as Word Perfect and spreadsheet tools like Lotus 123. Some of these are still in existence and use today, yet they have faded into the background with the ubiquity of Microsoft Word, Excel, and so on.

When applications are used to create and store data, those applications use different types of coding mechanisms and architectures to do so, and in many cases you cannot open or view data created by those applications without using the original applications.

Say your mother spent a lot of time 15 years ago documenting your family history and lineage using Word Perfect version 1.2. She carefully saved all of that important information on her 5.25 inch floppy disks and stored them in a shoebox in the closet.

You went to a family reunion and started talking with her and your aunts about your ancestors, and she remembered that work she did. You decide it would be a great time to gather up that information and update it with recent events like the birth of your children and your nieces and nephews. Well, we already discussed the problems with the floppy disks, but you are in luck. Your mom actually saved her old Commodore computer with a 5.25 inch drive.

You plug it in and fire it up, but it only boots up to a DOS prompt. There are no windows, there is no mouse, and there is no “user interface.” Just a black screen with C: at the top waiting for you to do something. (This is what is known as the DOS prompt, or the computer’s way of telling you in the old digital operating system that it is ready.)

If you are old or geeky enough to remember what to do next, you might manage to plug that floppy into the drive and even navigate over to it to see what is on it. There is a command for that. You can type it in after the DOS prompt (if you are interested, it would be DIR D: assuming the computer was recognizing your floppy drive as D:).

Let’s assume you got this far either because you are old and geeky like me, or because you found some help somewhere. You can actually see file names that end with the extension .wp for Word Perfect. Excellent. Now you just double click on them. Oh wait, there is no mouse and no cursor.

OK, so you do some more research or call that old guy back and you figure out the command that you need to type to open those files in Word Perfect. You try that out, but it does not work. Unfortunately, the way those old systems worked, you had to have the floppy disk with the Word Perfect application on it and load it up first. Once you loaded it on the computer, you could then open those files. Do you have that old application floppy disk? No? Oops, game over.

Again, this is all meant as an illustration of one of the many ways important data might be difficult or in fact impossible to gain access to when needed. In the example just given or in a similar situation, your choices would be to find someone or an organization with the correct software to open and recover those files in a meaningful way.

Now, a computer forensics expert, even without the correct software, might be able to gather enough of the raw text in a text-based

document to at least give you something to work with. However, if it is something with formulas, or graphics, or pictures that were saved in some proprietary format, it is unlikely that even the best experts will be able to decode and restore that to anything usable.

The bottom line is that data stored on obsolete or rare media, or stored by a n a p p l i c a t i o n i n a p r o p r i e t a r y c o d e, m a y b e c o n s i d e r e d unreasonably accessible or may be impossible to recover.

If you have to deal with any such information, keep this in mind and consider the resources and time that you will need to be prepared to expend if you choose to attempt the acquisition, restoration, or production of such data.

5.3 Privileged, Sensitive, and Inaccessible Data Management

In order to manage and understand the implications of privileged, sensitive, or inaccessible data, we have to begin by defining and documenting what we mean when we use those terms. In this section we consider each of these carefully and give you some things to consider as you strive to define and document your understanding, procedures, and management of these types of data.

In the courts, it has pretty much always been the case that if you wanted to assert that evidence could not or would not be produced for a case, you needed to prove that it was privileged.

We begin by looking at what the courts consider privileged data and how that can be established, documented, and managed. If you are in the legal profession, this should be old news to you, but for the rest of us mere mortals, this is important to understand.

Anyone, at any time, can be embroiled in legal issues such that they must understand what parts of the information they have access to will be produced. All of us therefore will benefit from understanding that some information that is exchanged or stored in certain prescribed ways might be exempt from being produced in court because it is considered “privileged.”

Let’s begin with the legal definition of privileged information:

n. statements and conversations made under circumstances of assured confidentiality which must not be disclosed in court. These include communications between husband and wife, attorney and client, physician

or therapist and patient, and minister or priest with anyone seeing them in their religious status. In some states the privilege is extended to reporters and informants. Thus, such people cannot be forced to testify or reveal the conversations to law enforcement or courts, even under threat of contempt of court, and if one should break the confidentiality, he/she can be sued by the person who had confidence in him/her. The reason for the privilege is to allow people to speak with candor to a spouse or professional counselor, even though it may hinder a criminal prosecution. The extreme case is when a priest hears an admission of murder or other serious crime in the confessional and can do nothing about it. The privilege may be lost if the one who made the admission waives the privilege, or, in the case of an attorney, if the client sues the attorney claiming negligence in conduct of the case.*

This definition does a good job of summing up the legal understanding of privileged communications. This can also be called attorney-client privilege if it is specifically addressing your communications with your lawyers about a case.

How would this affect you as someone engaged in a legal issue? For one, you should understand that if you e-mail or text or otherwise communicate anything to your attorney, or your wife, physician, therapist, or a religious counselor/minister of some kind with some assurance of confidentiality, those communications are out of bounds for that issue and should not be allowed as evidence.

However, what is important to recognize is that, as with all things in the legal arena, the idea of privilege can and will be challenged. There are many ways that these types of communications might in fact end up as part of the case, and most of them come down to good and careful practices by you and your legal or personal counselors.

In order to assert privilege and avoid confidential communications from becoming public or part of a court case, an attorney is required to document all of these types of communications and keep a privilege log. Further, the attorneys are required to note in all such correspondence that the information contained therein is privileged. Other counselors are similarly charged with the duty to manage confidential

* law.com, Legal terms and definitions, "privileged communication". <http://dictionary.law.com/Default.aspx?selected=1615>. Retrieved June 17, 2011.

communications in such a way as to be able to prove to a court or to the opposing party that information is privileged and not available for discovery or production.

To you as a n individual, this means that if you think you are communicating to one of the entities noted in the definition above on a confidential basis, you should look for and require that they are maintaining the proper management of those conversations.

For instance, in an e-mail correspondence with your attorney, you should see something in the subject line saying "Attorney-Client Privileged." If you are ever concerned about a professional's management of your confidential information, you should ask and require a sufficient answer and explanation of how they are managing privileged information.

In most all court cases where there is any question of privilege, an attorney is required to keep and produce a privilege log that documents privileged communications and proves that those communications should be privileged and kept out of evidence. That privilege can be lost, however, if data are inadvertently produced through the error or oversight of an attorney or their staff. This has happened in some cases, and it can result in bad outcomes for the interested parties.

On the other hand, I read a case where very competent paralegals who had to go through literal mountains of data inadvertently produced a small number of documents that should have been privileged. When those documents were referred to by the opposing counsel in deposition, the attorneys for the privileged party objected and asked the court to remove them from evidence based on their privileged status. In that case, the court agreed based on the huge volume of data that the paralegals had been working with, their general competence, and the fact that the information should in fact have been considered privileged.

As usual, this can go both ways. In the end, it is always a matter of careful practice and oversight. You can assist with this by being very clear with your attorneys or other professionals when you think something should be confidential and ensuring the proper handling and documentation are being maintained.

Another type of communication that a court will consider unavailable as evidence is what is known as work product. Again, if we go back to the law dictionary, we find the definition is as follows:

n. the writings, notes, memoranda, reports on conversations with the client or witness, research and confidential materials which an attorney has developed while representing a client, particularly in preparation for trial. A “work product” may not be demanded or subpoenaed by the opposing party, as are documents, letters by and from third parties and other evidence, since the work product reflects the confidential strategy, tactics and theories to be employed by the attorney.*

In legal terms, you will hear this referred to as the work product doctrine. This doctrine is basically the practice of ensuring that any of the work done by an attorney or his consultants or agents[†] is protected from discovery in a court case.

This can be hugely important for an attorney or anyone working for an attorney to understand because, again, this protection can be lost by mismanagement of the information.

For instance, if you are consulting as a forensics examiner with an attorney, all of the reports and findings, even including the “mental impressions, conclusions, opinions, or legal theories of a party’s attorney or other representative concerning the litigation”[‡] (as specifically stated in the Federal Rules of Civil Procedure), are considered work product. Therefore, they are not producible as evidence in court. That makes it extremely important that you as an examiner or agent for the attorney take the necessary precautions to protect that information and not inadvertently deliver it to the court or opposing counsel.

A big caveat for anyone working in this capacity is to understand that unless you are working as the attorney’s agent, this work product doctrine will not apply. In other words, if you are hired by the actual plaintiff or defendant, your work is not protected and could be discoverable for a court case.

* law.com, Legal terms and definitions, “work product”. <http://dictionary.law.com/Default.aspx?selected=2261>. Retrieved June 18, 2011.

† Forensic focus: Articles/papers, barristerharri, is your client an attorney? Be aware of possible constraints on your investigation (Part 1 of a multi-part series), September 24, 2011. <http://articles.forensicfocus.com/2011/09/24/is-your-client-an-attorney-be-aware-of-possible-constraints-on-your-investigation-part-1-of-a-multi-part-series/>. Retrieved June 18, 2011.

‡ Fed. R. Civ. P. 26 (b)(3)(B). Retrieved June 18, 2011.

It is also quite often the case that experts or consultants are hired by the courts as third-party neutrals or special masters. In that case, he or she is working for the courts, and all evidence, including reports, findings, and opinions, can be considered relevant evidence to be admitted for the case. This can be advantageous to the finding of fact by virtue of the objectivity of the expert who was not hired by either side in a dispute.

Next we need to consider the issues around information that might be confidential or sensitive.

Whether information is considered confidential, sensitive, or public only matters if someone is maintaining some type of classification of that data. Unless and until data are classified and that classification is documented, there is no way to give it the proper protections. Confidential data might include information such as that we discussed with privileged communications. This type of data might also be personally identifiable information (you might hear this called PII) which if revealed could result in harm to a person or his or her financial well-being. This would include data such as

- Credit or debit card or social security card numbers
- Driver's license or other license information, especially if combined with other personal information such as full name and date and place of birth
- Information about your health such as medications or diagnoses (this is sometimes referred to as PHI, or personal health information)

These types of information are considered confidential and private in the United States. In Europe, privacy rules are even more proscribed and include information about where you live or work, what you purchased, and so forth.

Confidential information can be difficult to define precisely. But again, it is in your interest to understand and carefully manage your own exposure of confidential information, or the exposure of your organization or family's data. Find out what the privacy laws in your jurisdiction cover and protect that information carefully.

As an owner or manager of an organization, you may be responsible for a nother large portion of private or confidential data, such as your employees' personal information, health and insurance data, banking information, and so forth.

Your company may also have its own banking information that requires special protection and management. If you sell or purchase anything, you may also be holding confidential information for your clients or suppliers. And if you accept transactions using credit cards, all of that credit card information must be protected as confidential. All of these data can be and most often are stored electronically and thus require a level of diligence and care to maintain security. This is not a book about information security so we will not go far down that path.

However, it is important to understand that as an owner or manager of an organization, a legal professional, or even just a citizen, we control confidential data that must be protected both for legal reasons and just for the sake of privacy and the security of our assets.

Everyone should have management practices in place to ensure that this type of data is held in a safe and secure method and with good security practices. Some of the most basic of these practices would include

- Strong passwords and authentication policies
- Encryption
- Good monitoring and incident response
- Security-minded development of new applications

Now there are other data that might be considered sensitive and, depending on the importance of the data, might even be treated as confidential. For example, information about critical infrastructure such as the physical or logical location of control systems for electric or other power grids, or water/sewer pumping stations.

By logical I am referring to their location on a network or on the Internet, if they are available over those medium. More and more of these types of systems are moving to Internet access because of the convenience and efficiency that type of access brings with it.

Yet that type of information in the wrong hands could be very dangerous. That data would be considered at least sensitive and possibly should be guarded and secured at least as carefully as other confidential information.

If such information was requested or required for a court case or an audit or other reasons, it would be the responsibility of the producing party to justify its refusal to produce that information by proving

the data's classification is based on real and justifiable concerns. The courts or auditors might still decide the data are of enough importance or relevance to require their production. If that is the case, the responsibility for the security and integrity of that data then passes to the court or the opposing side or other parties who will receive that data.

Finally, there will always be data that are simply inaccessible. In this chapter we showed various reasons that might be the case.

In a legal case, it is often true that attorneys will assert inaccessibility as a reason for failing to produce certain discoverable and relevant information. The legal acronym you will hear is NRA, which has nothing to do with the National Rifle Association, though they may have some NRA data. NRA in this case stands for not reasonably accessible. There is a ample and various case law that sets precedent for what can and cannot be considered not reasonably accessible. And as with all case law, it often depends on the circumstances of the case at hand.

In some cases, the fact that important data would result in extremely high costs both in finances and resources required, was enough to convince the courts that it should be legitimately considered out of bounds and not entered as evidence. However, in other cases, the courts felt that the information was so important that the costs were immaterial or justified and the data must be produced.

Sometimes the courts would do what is known as cost shifting and require the party with either the most resources or the most at stake to pay for the acquisition of data that were deemed NRA.

In at least one case, the court decided that the information was important and that the party asserting its lack of accessibility had contributed to that lack by virtue of extremely poor records management practices or outright spoliation (a legal term meaning destruction or corruption of evidence). In that case, the courts basically slapped the offending party by requiring that they spend the money and time to acquire and produce that evidence no matter how long and expensive it was, because it was through their own fault that the data had become inaccessible.

In many cases now, the courts are requiring an accessibility log, just as they are requiring a privilege log. This means that they will want clear documentation that justifies the assertion of inaccessibility.

They will want to know about your records management policies and even more important your actual practices around management of your data.

This is a great lesson and a good way to end this section. In the end, it all comes down to good practices and careful management of your data. Whether you are the responsible party in an organization, a legal professional, or just a citizen in our digital world, you are the owner of digital data that may be needed in the future. You have a responsibility to understand and document that data in such a way that it is properly classified, indexed, and protected against corruption or loss. If it is confidential data, it should be classified as such and treated with the correct practices that ensure it is safe from exposure.

Whatever type or classification of data is within your realm of responsibility, you are well advised to manage it carefully and understand where it exists and how to access it if needed. A large part of that management is knowing what parts of that data need to be treated differently due to their classification as privileged, confidential, or sensitive, or for some reason inaccessible.

5.4 Proving Ownership and Integrity

One of the biggest issues that can get in the way of electronic evidence being accepted in a court of law or any other venue is the ability (or lack thereof) to prove who actually created that piece of evidence.

This can create a situation that is referred to as plausible deniability, meaning that it is possible for a party to deny having created a piece of electronic evidence based on the inability to prove that he or she ever had access to it.

Creating and managing data in a way that maintains its integrity, documents its access, and monitors any changes allows a party to ensure nonrepudiation. Nonrepudiation is simply the ability to prove, based on carefully controlled records management practices, who actually had access and made changes to or saved a piece of electronic evidence.

Similarly, by the very nature of electronic data, it can be relatively easy, whether by mistake or intentionally, to corrupt the integrity of a document. By this I am referring to the idea that it can be difficult to prove that the electronic record in question is in fact the same record

you started with. You need to somehow be able to document that the record has no changes of any kind from the one that is required for the case at hand.

In this section, we consider all of the steps you as an individual or an organization can take to ensure the integrity of data such that the possibility of plausible deniability is minimized and nonrepudiation is maximized. We use some examples to illustrate these concepts.

As we discussed earlier, a digital record, whether a document, a picture, a spreadsheet, or anything else, is created using some kind of user interface application. Even if you are using the most basic text editor, there is coding going on in the background both by that application and by the computer operating system. These metadata (or data about the data) are extremely important to our conversation about integrity and ownership.

When one is attempting to establish the ownership of an electronic record, understanding when and where it was created is going to be your best clue. The data are contained in the metadata of every electronic record, or in some cases the data are collected by the operating system and stored in some type of log.

On Windows computers there is a hidden system file called the MFT that we talked about earlier. This is a flat database containing information about every file and folder ever created on the device since it was built and fired up for the first time. Other operating systems use similar methods of tracking files and their creation, modification, and access times, as well as their locations within the folder hierarchy.

How does this work in the real world? Imagine a case where the plaintiff alleged that the defendant stole patented information. These kinds of cases go on all the time.

In our imaginary case, the facts as we know them are that the plaintiff worked for the defendant for many years and was a model employee. She was an engineer who developed many different products that the defendant company developed, patented, and eventually sold. Everyone was happy. However, some of her ideas either were not that great, or at least the company found they did not have the potential of others, so they told her they were not interested. Please go on to your next great idea. Toward the end of her employment with them, one of these rejected ideas seemed to her to be an extremely excellent and wonderful invention that she was very excited about.

Now, as an employee with this company, she agreed that anything she invented belonged to them and was theirs to patent if they wished. When they turned down this idea, she was disappointed and sent an e-mail to her boss asking if she could consider this rejected idea exempt from their agreement and take it up on her own. She received several e-mails from management that eventually agreed to release this invention from their agreement and allow her to pursue it on her own if she wished. Shortly after this, she tendered her resignation from this company and went out on her own to develop her great idea.

As it turned out, she was right to be so excited about her invention. It became a huge success and she and her new product began to become extremely valuable. Everything was looking pretty rosy until suddenly her old company came out with an extremely familiar looking product and started competing with her. Because they were a much larger and more established company, they had a considerable advantage in marketing and brand recognition, and her company began to quickly lose market share for her wonderful invention.

At this point she sued her old company, the defendants, alleging that they had stolen the patent for which they had given up their rights while she was in their employ. She had electronic documentation to prove her assertions including the previously mentioned e-mails and an actual document that was attached to one of the e-mails stating that they released their interest.

It seemed like a cut-and-dried case until in the first days of discovery the defendants produced e-mails and electronic documents from an earlier date than the ones she had, showing that they had not in fact released their interest and that they asserted their ownership of the patent rights. They threatened a countersuit if she did not drop her case.

Now, as was noted, these folks had an established company, and along with that came powerful and experienced legal counsel who felt pretty assured that they could bully our friend the plaintiff into giving up her lucrative patent and bowing to their superior might. However, this is a good illustration of one part of the electronic evidence revolution that we have not really dwelt on too much.

In some cases, at least, it can enhance the justice system by creating a more level playing field. In this case, because they were dealing with electronic evidence and the need to prove or disprove its integrity,

both sides had an equally good chance of prevailing if they had the capacity and expertise available to prove their assertions.

Now enter the forensics experts.

The truth of this matter rests squarely on proving the integrity and ownership of the two sets of documents being presented with opposing conclusions. If our plaintiff can prove that her electronic evidence was in fact created when and where and by whom she is asserting, she is one step closer to prevailing in this case. However, she also needs to prove the contrary—that the defendant's evidence is false and was not created when, where, and by whom they are asserting.

In our imaginary case (which I admit has its basis in several different cases I have read), the forensics experts for the plaintiff ask for and receive complete sets of records from the defendants' computers and begin creating a timeline of events to try to understand the discrepancy between the assertions of their client the plaintiff and those of the defendants.

They are looking through the manager's computer, and they are noting regular and expected activities on a daily basis. They can see when he logged in for the day, when he went online and checked his e-mail and his stocks, when he did some Google searches, when he worked on some documents, when he created and sent e-mail, and so on.

The timeline goes along pretty normally, but for some reason on the day that the alleged e-mails and the contrary document were supposed to have been created and sent, they are not finding any evidence of that activity. So, they go into the MFT looking for the creation time and date of the document in question, and there it is with approximately the right time and date. Now this is a puzzle and an anomaly, and experts do not like those.

They next take a look at some event logs for the day in question, and again there does not seem to be any evidence of the application that was supposed to have created the document having been opened at that time and date. Being the careful sleuths that they are, they look further and deeper and suddenly they find something of extreme interest.

On a day several weeks after the supposed day of creation of the document in question, the event logs show an odd thing. The time clock on the computer is set back to a day 2 weeks earlier for a short

time and, lo and behold, the application that created the document is fired up during that window as is the e-mail program. After some time goes by those applications are closed again and the time goes back to normal and proceeds in a regular way.

The next day in court the plaintiff's attorney asks the court to allow him to depose (deposing means to question in a special meeting called a deposition), the defendant's IT manager. The court allows this and the deposition is scheduled and occurs some days later. In deposition, the IT manager for the defendant admits that he was asked by management to set the computer time clock back on the manager's computer and did so on the date noted by our heroes, the forensic sleuths.

At that point the defendants for some odd reason agree to settle the case, cease and desist from marketing the product in question, and pay all court costs as well as a substantial award to the plaintiff. Case closed.

This example illustrates many of the ways that the integrity and ownership of electronic evidence can be manipulated and used to either obfuscate or prove the integrity of that data.

We look a little closer at the issue of date and time stamps in the next section, but in the end, the ability to prove who did what and when is often the entire crux of any assertion in court or otherwise.

To establish the "who" is often difficult, especially in situations where more than one person uses a computer or other electronic device. There have also been cases where defendants have used the "virus" defense and insisted that they were not guilty of whatever crime was shown to have taken place from their computer because they had a virus and the "virus made them do it."

This defense might have some merit, but there will need to be some evidence to show this is the case, such as actual virus evidence on the device.

In establishing ownership, one can sometimes find evidence in the metadata. Some applications will record the user name either from something a user typed into the application itself or from the person logged in at the time. However, that can be relatively easy to dispute as without an actual camera taking pictures of who is typing on the computer at the time the document is created, there is no sure way to tell if someone else had control of the computer. The logged-in person

may have left it running and went to lunch, or that person might not have specific and controlled login credentials for everyone, so anyone in the family or the office can log onto anyone else's computer.

Given that those metadata may be difficult to prove or at least easy to refute, what other methods can be applied to prove who did what and when? One method often employed is the one we mentioned in our example. Experts will create a timeline of events on a computer that establishes what was taking place when. For instance, they can show that at 8:37 A.M. on Thursday, December 2, the computer was logged into by someone with the user name of Vjohnson. At 8:42 A.M., an Internet browser was started up and the Facebook account for Violet Johnson was logged into. A session of the game Farmville was played for the next hour or so under the user ID of VJ. At 9:56 A.M., a Google search occurred that looked up information on how to poison your husband. Then at 10:23 A.M., another search was typed in with the search term strychnine. Finally, at 10:54 A.M., a Google mail session was opened up under the user name Johnsv1@gmail.com and that session shows evidence of lasting for approximately an hour. Traces of e-mails to and from that address are found with date and time stamps that correspond to that time frame, and all of the ones from that e-mail address have ample evidence of having been authored by the same Johnson.

Now while the above does not establish perfect evidence that Johnson was planning to kill her husband with strychnine poison, if he should show up with that condition in the future, this could certainly be important electronic evidence of her culpability. Even without using metadata that might be difficult to prove, the forensic experts can show, using this methodology, what in this case would be pretty compelling evidence of who was actually on the computer when the actions that establish culpability for a crime took place.

The other part of proving integrity is somehow establishing that a piece of electronic evidence was not tampered with in any way. Because it is electronic, this type of data is susceptible to many different types of changes, often simply by turning on a computer or starting up an application.

In order to use evidence in a trial or for other uses and be sure it was not changed in any way, processes, tools, and protocols have to be carefully followed that ensure against these types of changes.

One often-used tool is what is known as hashing. A hash is a mathematical process run on all of those ones and zeros that make up a file. Because all electronic data are represented by the bits and bytes of numbers that we discussed in other chapters, these numbers can be mathematically manipulated through a known hash algorithm that will result in a very long set of numbers and letters in a hexadecimal format. This long hexadecimal number is called the hash, and it can be made of a file or folder, a group of files or folders, or even an entire hard drive or other storage medium (Figure 5.5). That large and unique number is like a fingerprint of what is contained in the data at that moment. If someone changes even the period at the end of a sentence and you run that same hash algorithm against the file again, the resultant number will have changed.

To guard and maintain the integrity of electronic evidence and to be able to prove that what you are producing for a court or audit is the same evidence as was requested or was created originally, experts will often begin by creating a hash value of that evidence before it is moved or copied. They use special tools that do not manipulate or write any data to the device they are copying from, and they carefully document the chain of custody or any changes that might have occurred and what and why they occurred.

Then, at the other end of things, when they are producing that evidence, they can use the hash values that were recorded, along with their chain of custody documentation to prove that the evidence in



Figure 5.5 A hash of a file, folder, drive, or any piece of digital data is like a fingerprint.

question is the same as when it was created. Or if that is not the case, they can show why it is not and establish whether it is still legitimate and relevant evidence that should be admitted.

That last part probably needs an example to make more sense. Assume you were asked to bring to court a set of e-mails that were created on a certain computer. Those e-mails were copied into a Word document by someone with less expertise in these issues and then deleted from the original e-mail application. Those e-mails saved into a Word document, though they lost the integrity of the original documents, might still be admissible as evidence if

- You carefully documented when and by whom the e-mails were copied.
- You can demonstrate how that Word document was protected from any changes (perhaps by creating a hash of the document).
- You did everything in your power to attempt to recover whatever you could from the e-mail application.

If the courts feel they are important enough and there is good documentation establishing that the core text was not changed, the e-mails might be admitted.

In an interesting example of the importance of maintaining the integrity of electronic documents, the U.S. Congress in 2011 was considering a new law called the Uniform Electronic Legal Material Act. This would encourage all of the states to adopt methods of authentication for any of their official government records. In an article from the Library of Congress,* it is noted that this process can be complicated and expensive, but that states should be considering the best ways to accomplish this for obvious reasons.

Depending on the sensitivity of the documents in question, they might have a simple mechanism that compares a downloaded document to an original, or they might have something more complex that compares hash values.

* Library of Congress, Butch Lazorchak, UELMA: A law we all can love. http://blogs.loc.gov/digitalpreservation/2011/11/uelma-a-law-we-all-can-love/?doing_wp_cron. Retrieved 2011-07-09.

All of this will require states and other government entities that come under the purview of this law to create methods of authentication up front when they create and store documents. In other words, each new official document will probably need to have a hash value or some other type of authenticating metadata attached to it that can be referenced in the future.

Establishing and maintaining the ownership and integrity of electronic evidence can be paramount in any type of case. It is always going to be an important role that an organization should create policies and procedures around in order to ensure they have the means to prove that integrity. These procedures should include the following:

- Careful records management and documentation
- Log management
- Good practices around the handling of potential electronic evidence

Those good practices will require an organization to have policy and established processes for litigation holds that ensure all affected people in the organization understand their responsibilities for preserving the integrity of any potentially relevant electronic evidence should there be any indication of a possible court action. We discuss litigation holds in more detail in Chapter 6.

5.5 Marking Time: How Time Is Recorded and Ensuring Integrity

As noted in the last section, the time and date information included in or about an important piece of electronic data can often be the crux of a matter when trying to establish what happened when and to whom. In this section, we spend a little more time considering time (Figure 5.6).

Modern computing devices operate in cycles of time that are tiny compared to what we are normally aware of. They operate in and record those operations in milliseconds (one thousandth of a second). This was not always the case. Some of the first computers did not bother tracking or recording the time of day, but they still operated in timed cycles that they did track.



Figure 5.6 Time is of the essence: The question of when something happened can be one of the most important facts to establish.

Tracking the time is done in different ways by different operating systems. Microsoft operating systems track time on the basis of how many 100 nanosecond periods have passed since January 1, 1601, while UNIX systems track seconds since January 1, 1970.

These are arbitrary choices made by geeks with not enough to keep them challenged in my humble opinion, but in any case that is how these operating systems keep track of the time and record it internally. However, as with everything in our electronic world, those geeks knew better than to expect us normal humans to be able to decode their schemes. All modern electronic devices decode and represent these system time data to users as a time and date we recognize, such as October 15, 1978, at 12:01:06:13, which actually does take us down to the millisecond level if we care to know it (in this case 06 is the hundredths of seconds and 13 is the milliseconds). But the important thing to understand is where the computer gets that time from in the first place. Does it come out of midair? I guess in some cases you might say that it does.

The bottom line is that the time a computer records and stores is entirely dependent on input it gets from outside. There is no magic time finder hiding inside of each computer.

Computers and all of the devices that use computing systems (phones, cameras, toasters, VCRs, televisions, etc.) will contain a chip that keeps track of milliseconds and in many cases will have a backup battery that keeps it running so it does not lose track when you turn off the power. Some of these devices, as noted above, will get

their information out of midair by picking up radio signals from the atomic clock at the U.S. Naval Observatory or other sources. These are accepted standard keepers for the “real” time worldwide.

Other devices get their time reference from a network time server that they are connected to via a wired or wireless network. That server has in turn gotten its time from some other source, perhaps the atomic clocks noted above.

Because this is so important to the smooth running of our society, standards have been established and are adhered to pretty much worldwide. The base time standard originally came from the Royal Greenwich Observatory in Greenwich, England, where they carefully monitored the rotation of the planet and calculated the exact agreed-upon “mean” time or what is known as Greenwich Mean Time.*

More recently, this was replaced by something called Universal Time that is calculated by comparing many different atomic clocks around the world and adjusting for anomalies to come up with a time standard that is as accurate as possible. This is the base time upon which other time zones around the world are based. Greenwich is still considered the “prime” or base meridian of time calculations, and Universal Time, or what you will hear referred to as UTC (Universal Coordinated Time), is still the time you will find on all the clocks in England.

Depending on the distance from Greenwich and sometimes some arbitrary and even politically motivated lines, the time in a certain zone will differ from Greenwich time, usually (but not always) in hourly increments somewhat corresponding to the movement of the planet around the Sun and its approximate 24-hour spin.

Now you might be wondering how all this is related to understanding the time of an activity on a computer—good question. The reason we are delving into this fascinating discussion on time standards and time zones is that most all computers will record creation, access, and modification times of the activities that occur on them in UTC.

If you are producing records and hope to prove or disprove a point with those records based on the time they happened, you may need to

* History of legal time in Britain. <http://www.polyomino.org.uk/british-time/>. Retrieved June 17, 2011.

understand and translate between UTC and the time zone in which the activity took place.

That last part of the sentence above is very important to note. To establish what time an activity took place, you need to consider where it took place and the date it took place as well.

As an example, think about a case where you are tasked with establishing exactly when a document was created and last saved. According to the evidence you have, the document was saved on a computer in Florida on December 16. You are examining the computer 6 months later in Colorado. The time recorded for that document's creation and last access are recorded as noted above in UTC. Your job will be to consider what the time zone difference would be for Florida during December. That would be Eastern Standard time (note that between November and March we have to adjust for daylight savings time) which is -5 hours' difference from UTC. Your time stamps have to be adjusted to reflect the actual local time where the activity took place.

Now, think about how this might be complicated if the activity took place in one of the oddball places like Indiana, where they do not observe daylight savings time, or other time zones like some of the Caribbean Islands that choose to adjust their time by only a half hour instead of an hour. And of course, all of this assumes that the time on the computer in question was accurately set to begin with and was not tampered with or simply lost time due to malfunction of the time chip or power loss in the battery.

Because this can be so important, it is imperative that some type of auditing or other means of verification of the accuracy and integrity of the time be established. If you are in the position of needing to prove the integrity of the time, you may need to gather corroborating evidence such as other devices that show the same time for the activity in question, or perhaps some type of documentation of the time having been audited and monitored for accuracy on a regular basis.

As an organization that might have to prove the integrity of time information recorded on its electronic data, it is important to maintain that type of auditable documentation to show that you have some type of time server or standard that is monitored regularly and carefully maintained.

5.6 Legal and Forensically Sound Acquisition

This chapter has focused on the idea of how to locate and recognize data and the challenges and issues that can arise. Once you have accomplished that task and worked through all of those challenges, you are ready to acquire that information in order to organize it and produce it, either for a court proceeding, some administrative or audit review, or just as a part of your own archival records management.

No matter what the reason for gathering that data, doing so in a forensically sound manner is a good practice as it will ensure that the data you produce can be proven to be the same data as you started with and have not been manipulated or changed in any way from the original.

We talked a little earlier about some of the tools that professionals use to accomplish this maintenance of integrity, but in this final section we will delve into those more deeply and give you some practical policy rules that you can incorporate into your organization's practices.

As noted before, the number one priority is to maintain information in its original state and be able to show that you have done so. To accomplish that, you should begin with good documentation around your records management policies. Much more important than the policies, however, are the actual practices to which you or your organization adhere. In every case, a court is going to want to know and understand what you actually do, not so much what you suggest you or your staff should be doing in policy documents.

So, let's begin with records management policy basics. Of course this could be a book all by itself, so we just hit some main points that you can put into practice, and I encourage you to learn more and apply that knowledge to your specific needs and requirements.

Records management policy should first include classification of data. You and your colleagues and staff must have a good understanding of the types of data you are producing and the legal and archival ramifications and requirements for each of those types.

For instance, there will most always be some records retention rules that attach to certain types of information such as contracts, invoices, and human resource records. These are usually determined as regulations either for your industry or for the particular jurisdiction of which you or your organization are a part.

Spend some time learning about and understanding which pieces of information your organization creates that are subject to these regulations. Ensure that everyone who creates, accesses, or in any way handles that data is aware of the regulations that apply. Create written policies that are available to all employees that explain those regulatory obligations. If you are smart, you will also insist that anyone who has access to this type of data has read, understood, and signed off on these policies. This will go a long way in a court case toward showing that your practices follow your policies.

Another good way to prove that those practices are adhered to is to do a regular audit of your employees and of the regulated information and their record retention practices to ensure they are being handled according to policy.

Once you deal with the regulated information, you still may have other data that are important to your organization and should be managed carefully. This might include e-mail in general or intellectual property, or perhaps log information from your network or a specific application. Again, your task is to have a thorough understanding of the importance of that information as well as considering its possible importance either to your organization or yourself, or as possibly relevant information in a court case.

If you are doing this for yourself as an individual, it is mostly a matter of considering all of the types of information that you store electronically and its possible value in the future. Make good choices about how the data are preserved, organized, and maintained so the data will be available when needed.

If records management practices and policies need to be developed for an organization, I would recommend gathering a group of stakeholders including top-level management, auditors, legal counsel, and IT managers and staff if those positions exist. If any of those positions are outsourced or contracted for, it would be useful to have them in the meeting as well.

Start, as noted above, by understanding the records retention regulations that apply, and then carefully consider all of the other data that you produce. Use the information you learned in this book as a guide to ensure you are considering all of the different places data can be stored and all of the issues and challenges around being able to access that data.

Come up with a set of policies and guidelines that takes all this information into account. Your auditors and legal counsel can assist you with the regulatory requirements, and your IT folks can help you to understand where the data live, who owns and controls the data, and how you will be able to access and manage the data. Management will be able to help you understand all of the different types of information created and preserved within your organization.

Top-level management or executives need to be included in this work as well, because it is extremely important that they understand and buy off on the policies. They will ensure these policies are adhered to throughout the organization if they believe in the importance of these practices.

Document those understandings in policy and guidelines, and when you feel they are as complete as possible (with the understanding that such policies need to be dynamic and flexible as your organization expands or contracts and as technology changes), arrange for regular and ongoing training on those policies for everyone in your organization who will have access to those types of records.

Good records management practices are the most important foundation for ensuring the integrity and availability of records. With those in place and well documented, you have taken the important first steps.

When it is time to gather that data for whatever reason, you now have the information readily available to know who owns it and how to gain access to it. The next important step is to acquire that data in a forensically sound manner. To do that, start with some type of tool that allows you to make one of those fingerprints we spoke of earlier called a hash. Run that application and create a hash value for every piece of information that you will be preserving.

Many computer forensics applications have this capacity built into them. If you or your computer forensics specialist are using those tools to create an exact copy (sometimes called an image or a snapshot) of a set of data, those applications will automatically verify the integrity of the data by creating a hash value before and after the data are copied. In many cases, they will use two or more types of hash algorithms to add even more surety to the integrity of the copied data. They will sometimes store those hash values in a text file that you can save with

the copied data, or sometimes these applications actually attach those hash values as metadata to the copied data.

However, you do not have to have these specialized tools. Some operating systems will do what is known as a verified copy. This is usually done from the command line with special switches in the command that you type.

An example of this in a Windows operating system would be to open up the command window by selecting Start, Run, and typing `cmd` to get the command window (a window will pop up with a black screen and usually a DOS prompt, `C:`). Then you can type in “Copy [source—path and name of the file you want to copy] [destination—where you want to copy it] and/V” (without the quotes). An example of this type of command line might look like “`C:Copy c :myfilessample.doc d :evidencesample.doc/V`” (again without the quotes).

The/V option at the end tells Windows to verify that the copied file is the same and was copied successfully. To do this, Windows hashes the file in the background using a relatively simple hash algorithm called a checksum before and after and checks to ensure the hashes are the same.

There is another command in Windows that you can use. (These are all actually old DOS commands as this is the basic structure that Windows was built on, and it is still backwards compatible with these commands.) You would use this instead if you wanted to copy both folders and files. The syntax of the command is the same, but you use `XCOPY` instead of `COPY`. Again, use the/V for verify, to ensure everything you copied is the same on both ends.

On a Linux operating system you can use a command line that compares the copied files to ensure their integrity: `cp “$file” “$dest” && cmp “$file” “$dest” || echo “failed.”` In this case, you would substitute the name of the file for “\$file” and the path to that file for “\$dest.” You would insert the two files and paths for the files you want to compare on either side of the `&& cmp` in the command line. The echo at the end of the command will respond on your screen with the word “failed” if the two files do not match.

Unix and Linux operating systems have several ways that you can copy different quantities of data and ensure they are the same on both ends, but this example will work for smaller groups of files. To copy directories, you add the `-r` switch into the command line right after

the c.p. If you choose to use this more basic approach, you simply need to document carefully that this is what you did so you can show that you have done your best to maintain the integrity of the copied information.

If you need to preserve larger volumes of data, such as entire hard drives, you can use forensic imaging tools to do so. Some of these are actual hardware and some use software to accomplish this. In either case, you connect the hard drive or the computer containing the hard drive to software or hardware that includes what is called a write blocker.

As noted, every time a computing device accesses information it can potentially change that information subtly by changing the access time, or making other changes to the metadata. A write blocker, whether it is hardware or software based, does not allow any data to be written to the files or data storage to which it is connected.

Once your device is connected to a write blocker, you next make a connection to a storage device. Then you use either an application or possibly command lines to initiate a byte-by-byte copy of the hard drive or other large portions of data.

This is sometimes also called a mirror image, though I do not like that term because in a mirror everything is backwards, and in a successful forensic image everything is the same as in the original. These images become a file with an extension of .dd, or .e01, or other types of recognized image files.

Of course, if this is an image of a hard disk, they can be very large files. In some cases, the forensic imaging applications that create them will automatically divide them up into pieces the right size to fit onto CDs or DVDs so that you can transport them on those media. The forensic imaging software will also sometimes allow you to compress the resultant image files by keeping track of but not actually copying empty sections of the hard drive. When it sees empty parts of the drive, it tracks and records where those are so that if you need to reproduce the exact hard drive it can put those empty sections back where they were on the original.

If you prefer using command lines to create your image, the dd command will work in both Windows and UNIX operating systems. This command syntax is a little different from most UNIX commands and can be used for many different tasks. The basic UNIX command

line for copying one drive to another might look like: “dd if=/dev/ad0 of=/dev/ad1” (without the quotes) where /dev/ad0 refers to one drive partition and /dev/ad1 another. The “if” and “of” in this command refer to the input file (if) and output file (of).

The Windows version of that command would be: “dd.exe if=.PhysicalDrive0 of=d:\images\PhysicalDrive0.img” (again without the quotes) to which you could add some hash commands to verify that it copied correctly, such as: “--md5sum --verifymd5 --md5out=d:\images\PhysicalDrive0.img.md5” (again without the quotes).

Now this is not meant to be a technical how-to book, and there are many different tools and applications that assist in accomplishing the creation of a forensically sound image of a digital storage device. So we will not get any more detailed about how to make all the connections and transfer all of the data. It is only important for you to understand that there are procedures and tasks that must be accomplished in order to ensure the integrity of data on either end of an acquisition process.

If you are going to do this work yourself, you need to take the time to become familiar with these methods and comfortable with these tools. You will also need to purchase and license the tools you want to use. All of that will take time and money, but it is important to both you and your organization that it be done and done well.

If you do not feel the urge to become a forensics investigator or expert, your job will be to ensure that you have access to someone with those skills and tools.

No matter what method you choose to create and preserve the forensically sound copies of data, the important thing to understand is the necessity to ensure the integrity of the data so that you can prove the data are the same after being copied as they were before.

As noted, there are many good tools out there that will allow you to do this using an application interface if you are not comfortable with command lines. One of these tools is free from AccessData and is called FTK Imager. Of course all of the command lines are free to use as long as you have that type of operating system and the comfort level to work with them.

We have reached the end of this chapter, so remember the most important points when you consider the where and how of electronically stored information. Those are as follows:

- ESI can be made difficult to acquire in many ways, whether due to corruption, loss, deletion, encryption, or loss of the application or device that created it in the first place.
- The most important thing you and your organization can do to ensure your ESI is accessible is to have good records management policies and, most important of all, good practices that follow those policies.
- Maintaining and documenting the integrity of data will be paramount in the data being legally admissible as evidence, or usable for audits or simply valid records.

If you are able to use this chapter, along with the information from other chapters to begin to address these important issues, you will be well on your way toward better management of the ESI that surrounds us all.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Common Problems With E-Discovery—and Their Solutions

ANDRES HERNANDEZ

Discovery is all about being able to track down information that is pertinent to your case. A few decades ago, this meant getting your hands on paper documents, but of course, that's not true anymore.

As our lives and business dealings became increasingly digital, so did discovery. It moved from filing cabinets to computer servers, but unfortunately, the change was not simple or easy.

Even as e-discovery became more and more crucial to building legal cases, it increased in complexity, and a number of common issues have started to make themselves known. Here are the most common e-discovery problems that you are likely to face as well as solutions for how to make the process run more smoothly.

There's Just Too Much Data

You have probably read reports about how much digital data is out there and how fast it is growing. Most, though, have a tough time putting the numbers into perspective. What, really, does 4.4 zettabytes look like versus 44 zettabytes?

Cyclone Interactive tried to put it into perspective in a 2014 infographic,¹ using the memory in tablets from the year 2013 (which, of course, has only increased in the time since). They explain that if you put all of the digital information that existed in 2013 (4.4 zettabytes) into tablets and stacked them up, they would reach two-thirds of the way to the moon. By 2020, that data is estimated to grow to 44 zettabytes, which would result in 6.6 such stacks.

Obviously, you will not have to go through even close to that amount of information on a single case. Doing so would be impossible. But the tiniest slice of a zettabyte is still a stunningly high amount of data.

And sifting through massive amounts of data is no longer a problem solely for mass tort and class action cases. More and more often, relatively small cases—say, for example, a breach of contract case worth a few hundred thousand dollars—are eliciting e-discovery requests that might end up yielding tens of thousands of non-duplicative items. Getting through so much data takes an enormous amount of time and manpower, not to mention cost.

All of us—companies and regular people—have a lot more digital data these days, and it's only going to get worse. When attorneys have to delve into that data to litigate a case, everything takes longer, becomes more complicated, and costs more.

Solution: If you know exactly what document you are looking for, but the main problem is the sheer amount of data provided, you actually have an advantage. Traditional Boolean searches, like you use on Google or Lexis-Nexis, can turn up what you need. This is particularly true if you know the exact file name, but it's also the case if you know specific keywords in the target document. The more you know, the better.

Data is Everywhere

The most common source of e-discovery data is email, but it is far from the only one out there. Today, e-discovery might include data from a variety of digital sources, such as: email systems, social media accounts, collaboration systems, real-time communication systems,

work computers (desktops and laptops), home computers (desktops and laptops), work-issued mobile devices (including tablets and smartphones), employee-owned mobile devices (including tablets and smartphones), corporate blogs and wikis, file servers, USB storage devices, and other repositories of both structured and unstructured data.

There are issues involved at every step along the way. First, you have to track down the data from all of these disparate sources. Then you have to gain access to it. And once you reach that point, you still have to sift through it—potentially mountains and mountains of data—to determine what, if anything, is worthwhile. But you also do not want to ignore a potential data source because it could end up netting you a piece of information crucial to your case.

Solution: There are many ways to collect data, so you should consider the pros and cons of each strategy to determine what's best for your particular case.

Data Collection

"Self-collection" means allowing the data custodians themselves to copy the files on a storage device. This is generally the least reliable option, since individuals may not be technically savvy, could make mistakes, or might overlook documents. Courts have also questioned whether this collection method constitutes a defensible e-discovery response. But it can be cost-effective if it is a small matter, involves

"Remote collection" means integrating data sources with a centralized internal collection system, allowing data to be collected remotely. IT professionals are a part of the process, but don't directly interact with the data sources. This means the process can be completed quickly and efficiently. This strategy involves an upfront investment and deployment process, however, so it is best suited to organizations with regular collection demands.

Not All Data is Created Equal

Beyond the fact that digital data is continually growing and lives pretty much everywhere, there's another issue: some types of data are inherently more difficult to look through.

What we're talking about is structured versus unstructured data. The simplest possible way of explaining these two types of data is that structured data is easy to search using simple operations, while unstructured data isn't. In other words, some data sources are easy to search through for specific terms and information, while others require you to pretty much look at each individual digital document to ensure you're not missing anything important.

Want some examples? Spreadsheets make for good "structured" data because the information is in a relational database system that computer programs can quickly and easily scan. But emails, social media posts, texts, and other phone data are unstructured.

"IT collection" means that members of the IT department collect the data. Unlike individual data custodians or general employees, they have a better understanding of the data landscape, as well as the technical skills necessary to ensure the data is collected properly.

a low volume of data, involves only highly conventional data, and the opposing party and judge have agreed to the plan in advance.

"IT collection" means that members of the IT department collect the data. Unlike individual data custodians or general employees, they have a better understanding of the data landscape, as well as the technical skills necessary to ensure the data is collected properly.

But at smaller organizations with limited internal IT resources, the data collection process can be burdensome for maintaining regular business functions. Also, IT professionals tend to provide forensic imaging versus logical copies when gathering data. This is important to note because forensic copies include additional information, such as previously deleted data.

This can result in too much data being collected, driving up e-discovery costs. And unless there is a suspicion of data tampering, a logical copy will usually meet the court's expectation.

"Third-party collection" means allowing an external professional to collect the data. This is beneficial because it has a smaller impact on the day-to-day functions of the organization. And the third party will have expertise in the procedures, tools, and skills to collect data in a way that will stand up to judicial scrutiny. The downside is that it comes with a larger price tag.

For example, if email was truly "structured," it would be arranged by content and exact subject, in addition to being organized by date, time, and who sent it. But that's not how people communicate. Even if an email conversation is relatively focused, it is rare that we only speak about one subject per email or that we don't use synonyms or slang terms. All of these things complicate a search and make it more difficult to unequivocally say that certain files are useful—or not—without looking at them one by one.

Solution: There are a number of different ways that you can sift through unstructured data depending on what you know coming into the search such as:

• If You Know Exactly What You're Looking For...

Use metadata analysis. You can group documents together based on topics that you know you want to look for and "drill down" into those classifications to get more and more specific. Using email as an example, you could categorize things by: To, From, CC, originating domain, when the email was sent, and so on.

This type of search can be particularly useful if you are searching for information where a topic is being discussed using a variety of different words that may include euphemisms, code names, and other

forms of nontraditional language. In other words, the way people often talk or write.

Use textual analytics. Another way to search is to feed a particular document that includes the type of information you're looking for into the system and have the program search for conceptually similar documents and data. One of the best things about this type of search is that, if you don't have a "control" document with the right kind of information, it's possible to simply create your own—even using multiple documents with "good" data to do it—and tell the system to look for documents like that.

- **If You Are Trying to Fill in Knowledge Gaps...**

Use keywords intelligently. A cornerstone of e-discovery review is coming up with search terms to define the documents that you believe you need. Many lawyers don't go any farther than that—but they should.

Keywords can be used not to just provide you with a "data dump," as it were, but also to prioritize what you should review. One way to help you in this prioritization is to look at the keyword hit count report. This will tell you how many times specific keywords show up in the documents that you have requested, which can help you determine which files are likely to be the most relevant to your search. Keep in mind that just because a keyword appears a lot does not mean it will actually be useful, but this approach can help you with that issue as well.

How so? If you see a keyword turning up a lot in documents that end up not being helpful, you can eliminate that keyword and try different searches to fine-tune your approach.

And finally, it can be helpful to turn on keyword highlighting in the documents, so you can more readily scan and focus on those areas where keywords appear. This can really speed up the review process without completely eliminating files, and it's especially helpful when looking at documents in groups or families.

- **If You Are Still Trying to Understand Your Case...**

Use concept-clustering. This analytics tool searches for words and concepts that appear together frequently and creates a report telling you what they are. For example, let's say you are litigating a class-action personal injury case. If you receive a set of documents about the injuries, it is possible that some of them will be talking about the nature of injuries, while others might be arguments over whether a particular injury occurred, and still others will discuss who is at fault.

What a concept-clustering tool does is separate the documents by the concept or content rather than group them all together because they share a keyword. It's a great way to keep documents organized and have a better idea of what you're looking at, as well as providing you with a way to eliminate groups of files if the concept doesn't apply to your case.

Use word frequency hit count. A keyword hit count report tells you how many times the keywords you've chosen appear in individual documents. A word frequency hit count tells you how many times all words appear.

Why is this useful? Because when you begin the e-discovery process, chances are good that you won't necessarily know all the terms that are going to be important. For example, what if you are litigating a case about a particular project and that project is often referred to by a code word or slang term? Running a word frequency hit count

can help you to find those documents where a project may appear under a different name than the one you knew originally.

Use TAR. Technology-assisted review, or TAR, can cut the amount of time you spend on review by up to 75 percent.² To put that in real dollars, let's say your e-discovery costs for a case are \$100,000. According to estimates, \$70,000 of that would be due to review. Three-quarters of \$70,000 is \$52,500. So using TAR could potentially save you over \$50,000—or more than half of the total e-discovery price tag in this instance.

Seems worth it, no? So what exactly is TAR? Think of it as a combination of textual analytics and concept-clustering. Here's how TAR works in a nutshell: You select documents that are valuable and relevant to your case. TAR looks at the content of these documents and performs a search to find similar documents. The more documents you select as relevant, the better TAR becomes at finding the types of data you need.

Some attorneys worry that TAR will eliminate the need for lawyers altogether or prove indefensible in court because they are not manually looking at documents. These fears, however, are largely unfounded. TAR will not eliminate lawyers, but free them up to advocate for their clients.

Moreover, assisted review still requires attorneys to establish the process and evaluate the outcome. You'll still be involved, but you'll no longer need to wade through file after file.

And as for TAR being indefensible, the old "eyes on every document" process has already been determined to be unduly burdensome. Translation: using TAR should be just fine.

It's Just Too Expensive

This has been touched on numerous times under other points herein, but it's worth discussing in-depth. Because large-scale e-discovery used to be the purview of mass tort and class-action cases—in other words, cases with big budgets and even bigger payouts—the pricing model for handling e-discovery needs was largely developed for these types of situations. In other words, it's expensive. Ridiculously expensive. Prohibitively expensive.

Let's say that the collected data in your case equals about one computer hard drive. Not even a big one—500 GB. Of that, there's about 100 GB of reviewable material. To house that data over a year, a typical e-discovery company will charge over \$160,000. Factor in that housing data tends to be about 30 percent to 40 percent of the total cost and 100 GB of e-discovery data will cost you about a half million dollars.³ For small and medium-sized firms handling small and medium-sized cases, that's just not possible.

Solution: Luckily, low-cost options have started to appear, and the overall pricing structure is changing and adapting to meet the needs of both larger and smaller firms. Some examples of cost-effective solutions include:

- Acrobat Legal Edition: www.adobe.com
- CasePoint: www.legaldiscoveryllc.com
- Cicayda: www.cicayda.com
- CloudNine: www.cloudninediscovery.com
- CS Disco: www.csdisco.com
- Digital WarRoom Pro: www.digitalwarroom.com
- dtSearch Desktop: www.dtsearch.com
- Harvester and SafeCopy: www.pinpointlabs.com
- Intella: www.vound-software.com

- Lexbe Online: www.lexbe.com
- Logikcull: www.logikcull.com
- Nextpoint: www.nextpoint.com
- Quick View Plus: www.avantstar.com/products
- Z-Discovery: www.zapproved.com
- Wind: www.windlegal.com

You can also keep costs under control by narrowing down the data to review first. Focus on what you know about what you are seeking. Think about specifics such as the date range and key players. This will allow you to narrow the data down to a smaller set of documents through metadata analysis. This doesn't mean you should completely discard other data, but it can help you to save time and money if you locate what you need early on.

Falling Into the 'Scope Creep Trap'

The flip side of needing to make sure you don't miss any important information or data sources and looking individually at unstructured files is that it is all too easy for the scope of your search to become too broad. You believe there may be a single text that can help your case, so you look through six months of text communication. When you come up empty, you decide to expand your search, looking back a year. Then two. Or you decide that you aren't looking for a text, but an email.

Many lawyers—even those who have been practicing for many years—do not focus enough on the e-discovery process. Perhaps they believe there are more important issues to deal with on a case. Or that the data they need to go through will be relatively small. Or that the process itself has become so automated that it will be a breeze.

"Project parameters" for e-discovery are inherently volatile as new facts are uncovered in a case and potentially useful data sources are discovered. Unfortunately, what this means is that it is all too easy for a seemingly simple e-discovery process to become bloated both in time and money spent.

Solution: Project changes are to be expected, but even small changes can add up over time. Monitor the situation regularly, and have key stakeholders sign off on proposed changes. Set a budget at the start of e-discovery, and make sure all scope changes take this factor into consideration.

What is the likelihood of finding the information you need? Why is the scope changing—simply because you didn't find what you need? Or is it because new information has come to light that leads you to believe you were looking in the wrong place? Can those discovery dollars be spent better elsewhere?

Not Starting Early Enough

Many lawyers—even those who have been practicing for many years—do not focus enough on the e-discovery process. Perhaps they believe

there are more important issues to deal with on a case. Or that the data they need to go through will be relatively small. Or that the process itself has become so automated that it will be a breeze.

Whatever the reason, they wait to begin the e-discovery process. This puts them behind the eight ball when they need to complete e-discovery quickly. They may end up with incomplete data, information that is incorrect, or a court violation if they are unable to comply with a data request.

Solution: The best way to ensure that e-discovery goes as smoothly as possible is to start the process immediately. Do your best to anticipate requests and determine where the more valuable information is likely to exist.

E-Discovery is Done by Project

While it is true that every case and every matter is different, that doesn't mean your e-discovery practices need to be reinvented each time you take on something new. Sadly, though, this is exactly what happens in many instances.

An ad hoc, project-based approach to e-discovery is taken because legal professionals often have to find unique ways to track down information they believe will be pertinent and helpful. But constantly having to figure out how to conduct e-discovery in a particular case leads to a number of problems.

First, it adds time and money because you have to come up with the process on the fly rather than diving in and actually beginning e-discovery immediately. Second, it can make clients feel like you are milking them since you don't have a specific process in place before you begin—why should they pay for you to figure it out?

Solution: To fix this problem, law firms need to work with businesses to enact better information management practices that use comprehensive technology solutions to increase data organization. This will not only speed up e-discovery, but set down a clear plan that can be followed for these types of cases in the future.

Your Analytics Just Aren't Good Enough

Of course, most law firms don't want to handle e-discovery on a case-by-case basis. You want there to be a clear process, because it will make you more efficient and clients will be happier.

But how do you come up with that process? By utilizing more useful metrics.

Solution: You need to use software that provides you with spending totals per matter, spending totals per file type, spending

totals per location, volume analytics across cases, production statistics (including how many times a document has been produced), collection and preservation information that shows past discovery activity, and search metrics.

These things can help you to estimate the necessary time and budget, come up with an overall e-discovery process, and generally streamline things for your clients—and your firm!

A Lack of Convergence

Chances are good that you have digital e-discovery tools. In fact, you may have several of them. Maybe you invested in a review tool as well as an Early Data Analyzer. And then something else to handle file processing. And so on.

The point is that you opted for these tools to streamline the process to make it faster and more affordable, but what you've actually done is lost a lot of time and money. Every time you move from one software "solution" to the next, all of your electronically stored information has to be exported and processed by the tool down the line. And that's assuming that the new and old softwares don't have any communication issues that can slow things down or even cause information to be lost entirely.

Solution: In the last few years, technology in this area has evolved significantly. Law firms started to recognize that a single tool that was able to handle e-discovery from beginning to end would be much more efficient, and legal service providers listened. While there is certainly plenty of room for improvement, there are a number of tools and services that handle multiple aspects of e-discovery without having to continually move data around.

Some of the best out there include Logikcull, Nextpoint, CloudNine, and Wind.

Unwilling to Work With the Opposition

Your job as an attorney is to win cases, which means that the other side has to lose, right? They are the enemy, and you should oppose them at all costs. But where e-discovery is concerned, is that really the way to go?

One of the biggest issues that drives up costs for e-discovery is one side sticking to unreasonable exchange protocols. This can literally price smaller firms out of e-discovery by demanding they use services that far exceed their budget.

Solution: If you run into this kind of situation, you have two choices—battle it out in front of the judge or play nice with the other side in an effort to get them to agree to use lower-cost e-discovery solutions.

Difficulty Recovering E-Discovery Costs

The laws in place for recovering e-discovery costs are slowly evolving. However, they are still largely, for lack of a better term, "analog." For example, they refer to the price of making copies or the cost of getting to data that is difficult to access.

But unfortunately, access isn't the problem that legal professionals face in our increasingly digital world. Getting data is easy, generally speaking. What is currently driving up the cost of e-discovery is the fact that there is so much of it out there that it takes an inordinate amount of time and money to review. There have been anecdotal accounts of \$4 million lawsuits where \$3 million was spent on e-discovery—\$3 million that is not recoverable.⁴

Solution: Sadly, this isn't something that individual lawyers or

firms can fix, but rather a systemic problem. Courts are starting to get more creative in reining in disproportionate e-discovery, but it's still a big issue.

Laws Are Complex and Ever-Changing

E-discovery laws need to evolve as new technologies emerge and big data becomes even bigger data. But when legal cases involve international elements, trying to navigate the disparate legal systems across the globe and keep up with the hundreds of laws added each year can be incredibly daunting. What was true six months ago may no longer apply.

Much in the same way that e-discovery used to largely impact big mass torts and civil-action lawsuits, but now affects smaller cases as well, international law is coming into play in more and more cases. Data can now be easily housed anywhere in the world, which sometimes brings up questions of who owns it and what laws apply.

Solution: There's not much to do here other than strive to keep up with laws related to data ownership and e-discovery as much as possible. Perhaps one day there will be unified international data laws in place, but until then, it is up to legal professionals to navigate this tricky space to the best of their ability.

Comparing E-Discovery Providers Is Next-To-Impossible.

Even while more cost-effective e-discovery tools have become available and providers of e-discovery services have lowered their prices, it is still incredibly difficult to price compare. Simply put, the tools, services, and price points differ from solution to solution—but they do so in a manner that makes it almost impossible to tell how much you might end up paying for e-discovery services using one provider or tool over another.

Solution: More standardization is needed to alleviate this problem, and possibly more regulation and oversight. But until that becomes a reality, your best bet is to work with a legal IT company that can advise you on what software solution best fits your case or firm. They'll be able to draw on the experience of all of their clients to advise you for your specific situation.

Technological Incompetence

A lot of lawyers probably won't want to hear this, but a huge part of the problem is that, as a group, they're just not as tech savvy as they need to be. This causes all kinds of problems with e-discovery, including overpreservation of data, overpayment for services, methods that are outmoded, shortsighted source identification, using outdated tools, failing to identify and address new issues, and discovery that is poorly executed and unfocused.

In particular, experts point to a lack of understanding about the nature and pervasiveness of smartphones, BYOD (bring your own device), texting, and instant messaging versus email, as well as social media in general. It's not that lawyers do not know about these things, but that their understanding does not go deep enough to allow them to adequately navigate and explore them as it pertains to e-discovery.

Solution: Don't go it alone. Hire the right project manager or project management company. Beyond experience with e-discovery, you want to seek out someone who can share their expertise effectively. The goal shouldn't be just for them to manage the project, but also to guide and educate your legal professionals through-

out the process. You'll have better, more cost-effective results if everyone is informed and working as a team.

Data is Sorely Mismanaged

Rare is the company these days that truly has a good handle on its data and knows exactly what it needs to hang on to—and where that data lives. Mismanaged data is bad for e-discovery because it is more difficult for you to track down and—since there is likely to be a lot of extraneous information—it takes longer to sift through to find facts and details pertinent to the case.

Solution: Unfortunately, this is an issue that starts with businesses and individuals themselves. If you are a lawyer taking on a case, you have to step in and deal with the situation as it is. In-house attorneys for corporations, however, should pay careful attention to data management and suggest comprehensive solutions regarding how to deal with it, including creating an information governance policy that can allow companies to defensibly delete data.

Conclusion

If you wanted to sum up the problems with e-discovery, it would be this: The process is completely and utterly erratic. Costs vary. Laws vary. Processes are underdeveloped. And data keeps growing and spreading to more places. As a society, we all need to get better at managing all of the information and data in our lives.

But since attorneys often have little say in how data is managed before they take on a case, it is incredibly important to pay attention to emerging trends in communication technology and business and utilize innovative solutions to increase efficiency. ☉



Andres Hernandez is the co-founder of Wingman Legal Tech, which is a technology consulting firm that specializes in law firm technology. Their technology solutions are designed for simplicity and efficiency.

Endnotes

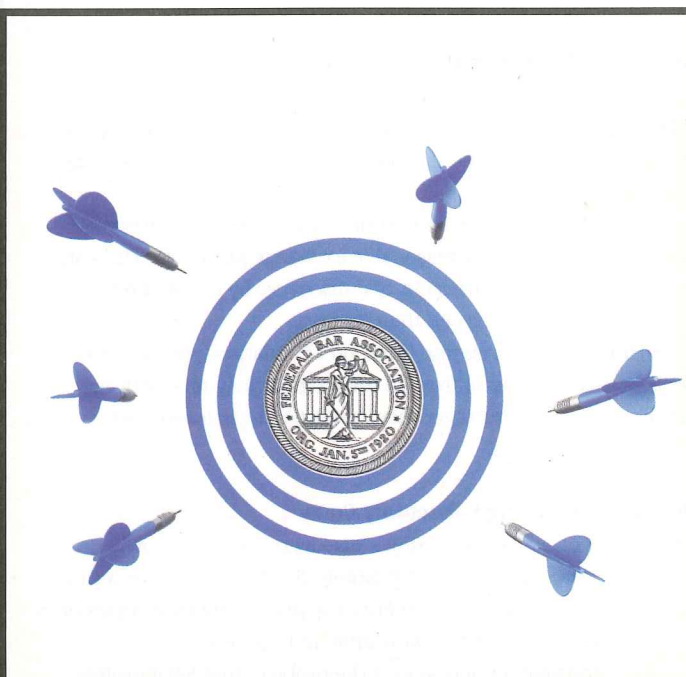
¹*The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things*, Infographic, Cylcone Interactive (2014), <http://www.emc.com/infographics/digital-universe-2014.htm> (last visited June 14, 2016).

²Case Study: Predict Proves Effective for Small Collection, Catalyst Repository Systems (2015), available at https://www.catalystsecure.com/pdfs/case_studies/Catalyst_CaseStudy_Predict_Proves_Effective_for_Small_Collection.pdf.

³Tom O'Connor, *Cost-Effective E-Discovery for Small Cases*, A.B.A. (2013), http://www.americanbar.org/publications/gp_solo/2013/january_february/cost_effective_ediscovery_small_cases.html (last visited June 14, 2016).

⁴Steven Baicker-McKee, *The Award of E-Discovery Costs to the Prevailing Party: An Analog Solution in a Digital World*, 63 CLEV. ST. L. REV. 397 (2015), available at <http://engagedscholarship.csuohio.edu/cgi/viewcontent.cgi?article=3846&context=clevstlrev>.

Missing Your Target Market? Hit the Bullseye with Our Reach



Partner with the Federal Bar Association today—contact Heather Gaskins, Director of Development at (571) 481-9106 or hgaskins@fedbar.org to hit your target!

Cohen and Mattingly: ESI review protocols: No more shots in the dark

August 10, 2016



Cohen



Mattingly

By Hamish Cohen and Jonathan Mattingly

Much has changed from the good old days when associates reviewed documents in huge warehouses. Back then one could stand and survey his or her domain of Bankers Boxes, many of which were gently dusted with what was probably asbestos and had been invariably damaged in some long-forgotten flood or fire. The possibility that your client wrote something damaging (or helpful) in an offhand email was non-existent; the likelihood of someone finding useful documents in some moldy box about the same. We were all, in many ways, shooting in the dark.

Technology has changed things. Big clients no longer open their warehouses to opposing parties and tell them with passive-aggressive glee to enjoy a winter half-frozen with carbon-paper stained fingers looking for goodness-knows-what.

Today, a producing party gathers, processes, reviews and produces information that the receiving party can analyze and search immediately using sophisticated diagnostic software and complex Boolean search protocols. Document productions, if done incorrectly, are often overly and underly broad; unnecessarily expensive and inefficient; and potentially damaging. These days if you, knowingly or unknowingly, produce a needle in a stack of hay, it will be (or should be) found.

Attorneys today must closely monitor client “needles” while still producing responsive documents, removing non-responsive documents and identifying privileged documents. How? You need good document review procedures and protocols. What are good procedures and protocols? The answer to this question, as any experienced lawyer knows, is “it depends.” Fear not; principles exist to guide us.

1. Know your case. Use your case knowledge to develop a written review protocol that introduces the case, explains the relevant legal concepts, reiterates the reviewer’s legal and ethical obligations (their duty of loyalty and confidentiality, for example) and describes on a category-by-category basis how any document should be coded for responsiveness, privilege or in relation to the issues. Provide, as part of the protocol, key documents or document excerpts such the complaint, discovery responses or specific sample documents. Then train the reviewers. Train them thoroughly. Train them in the database you are using. Train them using the protocol and real-world document-based examples. Then train them again.

2. Keep it simple. Complex review protocols take longer to implement, cost more, result in slower

review times and introduce more opportunities for error, so keep coding focused and on-point. This includes limiting the number of issue codes for the reviewers to consider. Rather than develop a code for every eventuality, come up with a few overarching categories and perform searches for subsets of documents.

3. Evolve in real time. Accept that your initial protocol will not be perfect (or even close to it). Review protocols must evolve and change as information is learned during the document review. Establish a chain of command through which questions, concerns or documents that require further review (because they are problematic or key) are passed up the chain and information, answers and protocol changes are passed back down to the review team in real time. Don't just train the review team once. Train them as needed. Communicate these changes to your team through emails and/or at your team meetings. Additionally, keep a log of the questions and the responses provided so consistent feedback is provided.

4. Technology is our friend. We know what you're thinking. You're secretly scared of technology — #Idontgetit; #whatisahashtagusedfor. Don't be afraid of technology. Embrace it. A good review platform permits a more efficient, less-expensive and more accurate document review by permitting data analysis, grouping similar (but not necessarily identical) documents together, and providing easy access to email or document families. Reviewing and coding similar documents together (rather than having them reviewed independently by different reviewers in different custodial mailboxes for example) is faster, cheaper and considerably more accurate because like documents are coded and grouped with other like documents. Technology also can be used to identify documents that are more or less likely to be responsive or more or less likely to be privileged via the use of key terms, custodian-related information or computer-assisted review. Vendors can help you use these tools efficiently.

5. Develop a clear workflow. Any good document review should establish a thorough and efficient workflow diagram for the first pass and second pass review of documents. The workflow diagram should, *inter alia*, identify what occurs with potentially privileged documents, documents that require redaction, documents that will be marked confidential pursuant to a protective order or agreement, and "hot" documents that address key points or issues in the litigation.

6. Quality control matters. Implement substantive and technical quality-control procedures. As to the former, have someone who really understands the case review a statistically relevant sample of reviewed and coded documents right at the outset. Doing so provides a critical opportunity to provide the review team with immediate real-world feedback so as to improve the review quality from the outset. As to the latter, technology can help ensure that the review and any production is complete and consistent (or prove that the other side's production was neither).

These protocols and procedures require up-front work, but in the long run they save a tremendous amount of time and money while improving review results. Implement them and you'll no longer be shooting in the dark. •

? Hamish Cohen and Jonathan Mattingly are founding partners and directors of the law firm of Mattingly Burke Cohen & Biederman LLP and ESI consultant Proteus Discovery Group LLC. Through both companies, Hamish and Jon bring their experience with ESI to work with lawyers and clients in order to manage data protocols and discovery. The opinions expressed are those of the authors.

CHAPTER SIX

Quickly Viewing a Mixed Bag of ESI

IT IS NOT UNUSUAL in small cases to receive a CD or DVD with a relatively small number of computer files. Typically, this is a mixed bag of electronic files in multiple formats. You might receive ESI that includes document files; spreadsheet files; PDF, TIFF, and ZIP files; and extracted e-mail files. If you want, you can view each one by opening it in its native application. The problem with this approach is that you must open each file one by one, taking the time to open the specific application and then load the file for each item you wish to view. Not only is the process time consuming, but you may not happen to have all the native applications installed on your computer to view the documents. You can go out and buy and install the software, but is it really necessary to own every program in order to view the files typically found in a CD or DVD of files produced in response to a request for production of documents? The simple answer is "no."

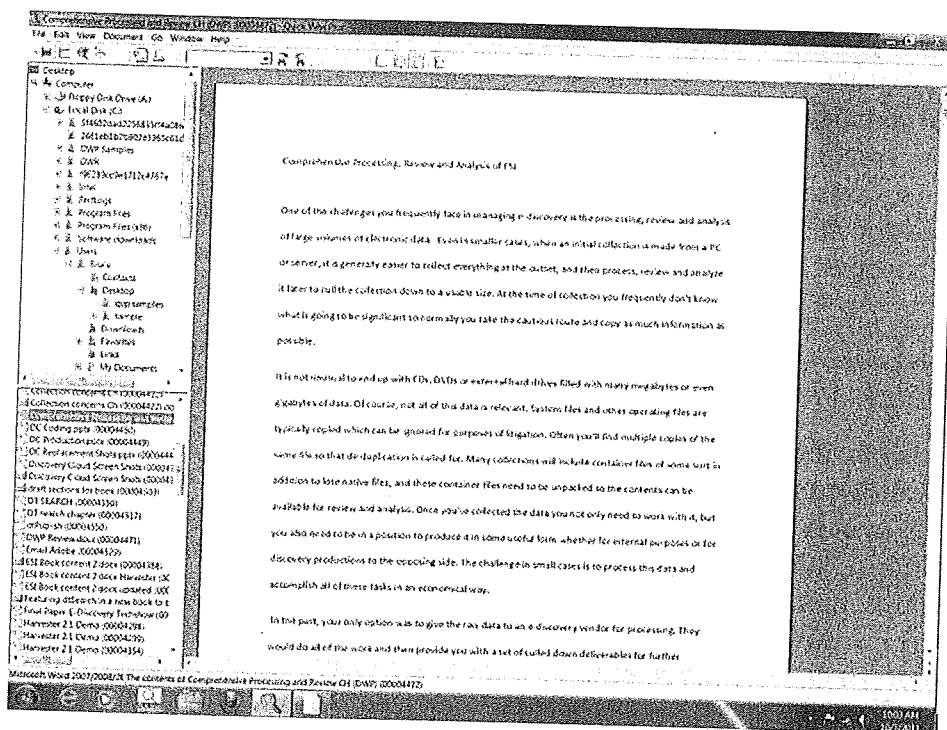
Quick View Plus

Using an inexpensive viewer like Quick View Plus Standard Edition (QVP) (available as a \$49 download from www.Avantstar.com), you can view files in over 300 formats without the need to own the actual native applications. This includes word processing, spreadsheet, database, presentation, compression, graphic, and other formats. Using QVP you can rapidly navigate through the list of files on your CD or DVD, opening one after the other in quick succession without the need to invoke the native application. Everything you need to view a collection of documents in one program can be done within QVP.

In Figure 6.1, you can see that the QVP desktop is divided into three panes. The viewer is on the right, and on the left you will find the navigation panes that include a tree view of your computer folders on top, and below that a specific list of files contained in the folder you have highlighted up top. Click on any specific file to open in the QVP viewer. On the button bar, you can choose to close the navigation panes to have a full desktop view of the document. You can move from document to document using the arrows on the button bar. You can also open multiple windows to simultaneously view multiple documents. For documents containing text, you can search for a specific word and move from highlighted hit to hit. You can also launch the document for editing in its native application by clicking on the icon in the button bar. Finally, you can print the document by clicking on the "Print" button.

QVP is an inexpensive solution for quickly viewing a collection of miscellaneous files. Avantstar also offers a more advanced version of QVP in its Professional Edition. The Professional Edition costs \$570 (and includes the first year of required support and maintenance). Five seats are

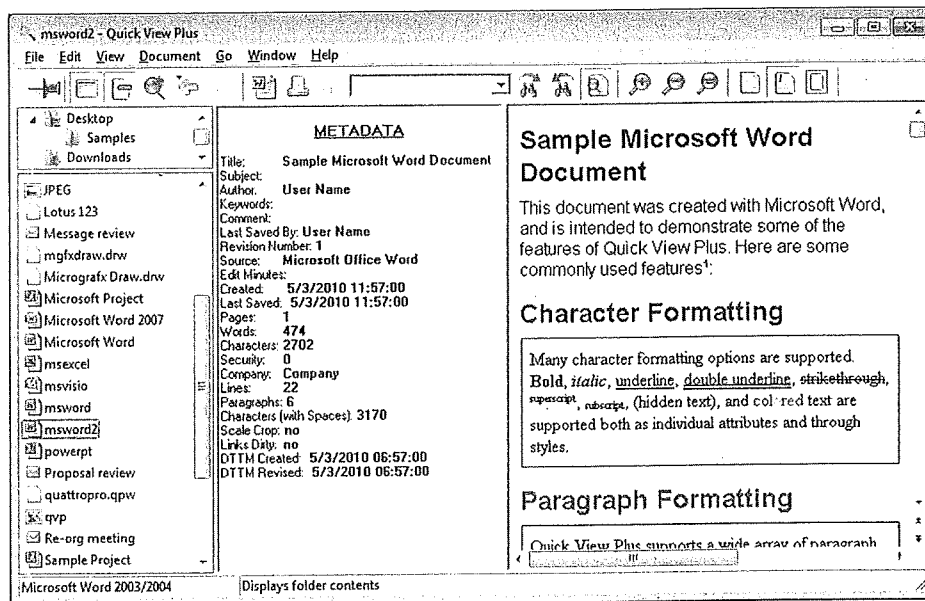
Figure 6.1



included with the license. The Professional Edition is specifically targeted to the legal market. It has all the features and functionality of the Standard Edition plus additional features.

Perhaps the most significant feature for e-discovery purposes is the fact that you can see everything, including the metadata, in a document you have selected to view. An example showing the metadata and rendered view of a Microsoft Word document is set forth in Figure 6.2.

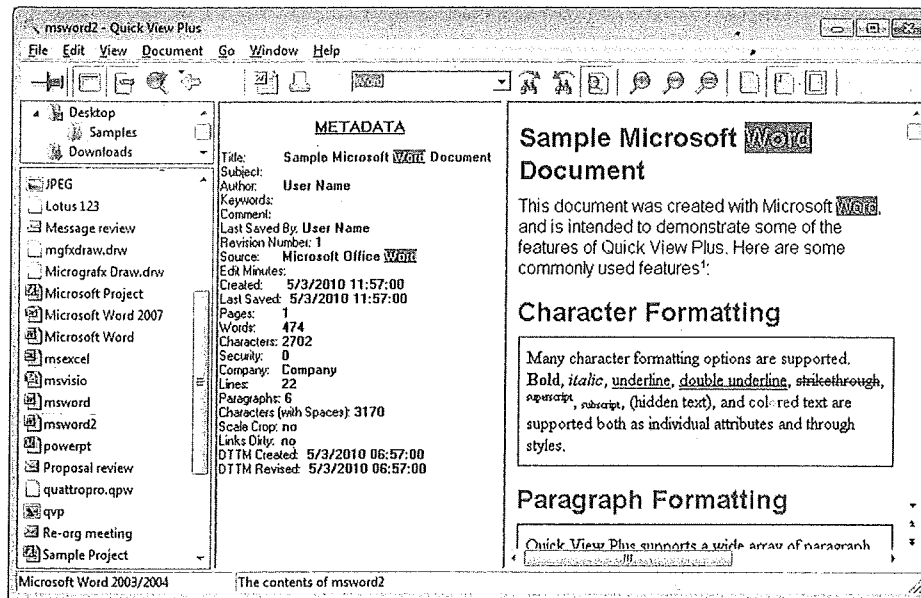
Figure 6.2

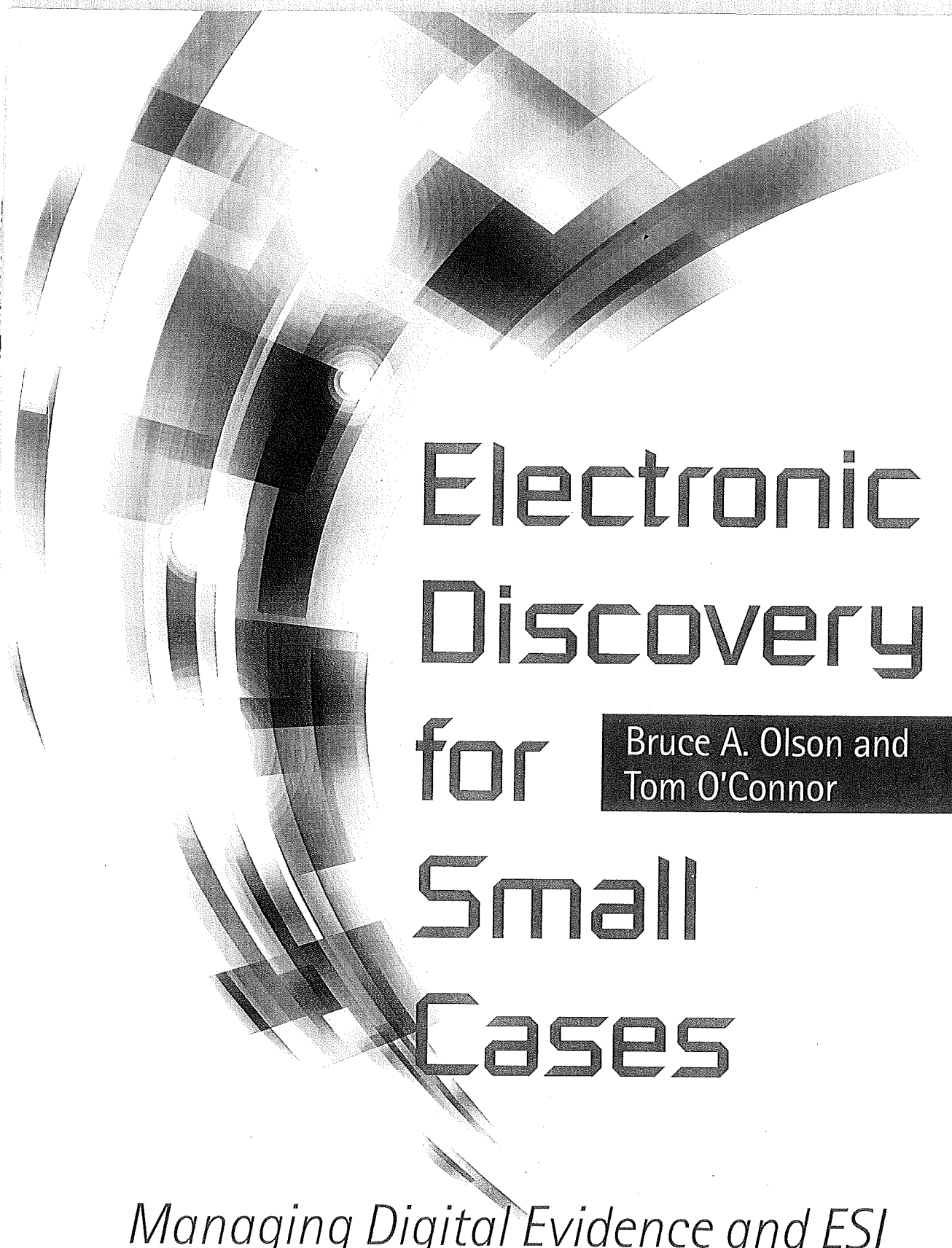


Unlike the Standard Edition, where you can search for key words one document at a time, the Professional Edition has an Auto Search and Highlight function that allows you to navigate multiple documents with a single search. Furthermore, the hits are highlighted in both the text of the document and in the metadata. See Figure 6.3.

You can also copy and paste any portion of a viewed file into another application. Finally, you can bulk print documents without the need to view each one prior to printing. This is a useful feature when you have done a quick sampling of collected documents and know they are the ones you want to print to PDF or paper for production purposes. You can even create a ZIP file within the program to create a production set.

Figure 6.3





Electronic Discovery for Small Cases

Bruce A. Olson and
Tom O'Connor

Managing Digital Evidence and ESI

ABA **LAW PRACTICE MANAGEMENT SECTION**
MARKETING • MANAGEMENT • TECHNOLOGY • FINANCE

CHAPTER SEVEN

Small Budget Searching Solutions

IN SMALL CASES WHERE preservation of metadata concerns do not exist, it is not unusual for parties to a litigation to simply bulk copy multiple computer files that may have some bearing on the case. These files typically will be presented to you on a CD or DVD in the form of raw files in mixed file formats. The files will not be searchable as presented, and you must open each file in its native application in order to view it. A small collection may be limited to several hundred files and, in such cases, you do not really need to employ a full-blown litigation support application. Nevertheless, you need a way to make sense of what you have received with the ability to both search and view the documents in the collection.

dtSearch Desktop

One of the most economical solutions to meet this need is dtSearch Desktop. At a cost of only \$199 per license, dtSearch Desktop offers a robust tool for processing, searching, and viewing a mixed collection of raw files and e-mail files. At this price it is ideal for use with small collections. However, do not make the mistake of thinking that because of the modest price dtSearch Desktop cannot be used with much larger collections. It works equally well with collections that are terabytes in size. In fact, dtSearch Desktop is the search engine that is used in many of the popular and more expensive litigation support applications on the market. Those software manufacturers have licensed the dtSearch Desktop technology and, as licensees, do not advertise that they are actually using a third-

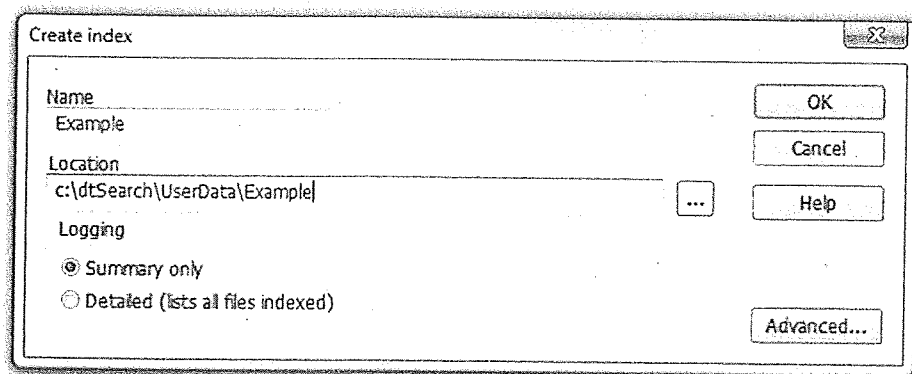
party application as the underlying search engine. The fact that dtSearch Desktop is widely used in such a fashion is further evidence that it is a tool that can meet your searching needs, whether your collection is a few hundred files, a few thousand files, or many more.

Assume for purposes of our discussion on dealing with e-discovery in small cases that you receive a DVD with 1 GB of miscellaneous computer files. The files of interest include a couple of Microsoft Outlook PST files containing e-mail, multiple Word and WordPerfect documents, some simple text files, some Excel spreadsheets, and multiple PDF files. There are also miscellaneous system files included on the DVD that are really of no concern. You want to take these raw files and create a searchable collection of documents.

To start the process you must copy the data on the DVD to the hard drive of the workstation on which dtSearch Desktop is installed. Next, you must open the dtSearch Desktop program. From there, you must index your collection. This is the most time-consuming process, but it prepares the collection of documents for subsequent use by indexing the location of each word in each file. Once the collection is indexed, searching is nearly instantaneous.

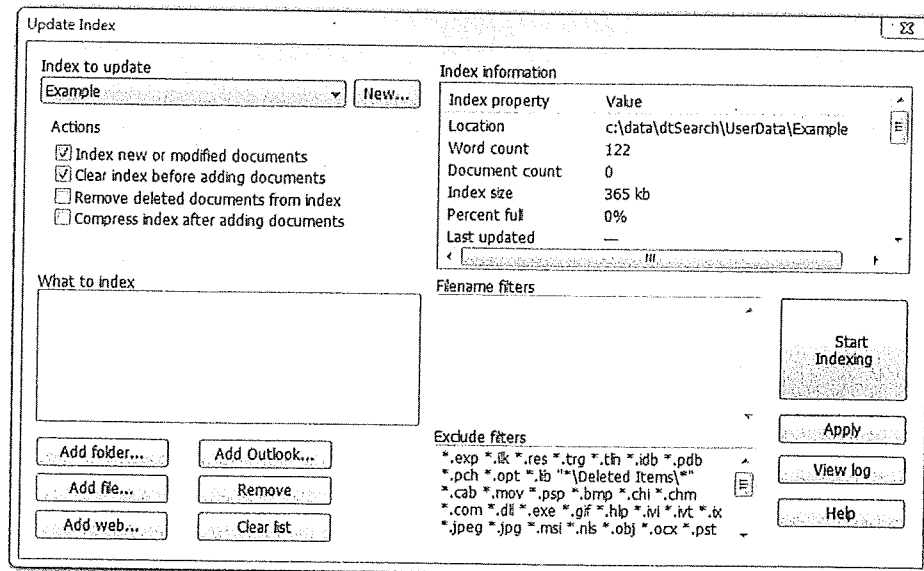
To index the documents, select "Create index" from the Index drop-down menu. This opens a Create index dialog box shown as Figure 7.1 in which you name the database, browse to the location where you want the index stored, and determine whether a summary or detailed list of indexed files is required.

Figure 7.1



After you click "OK," you will be asked if you want to add documents to the index. Click "Yes," then go to the Update Index dialog box, shown as Figure 7.2.

Figure 7.2



Using this dialog box, you can determine whether to index new or modified documents, clear the existing index before adding documents, remove deleted documents from the index, or compress the index after adding documents. Select the appropriate actions by checking those options you wish to activate. Next, add materials by using the features in the “What to index” portion of the dialog box. Using the available buttons, you can add folders, add individual files, add information from the Web, and even add folders from Microsoft Outlook to the index. Clicking on the appropriate button will enable you to browse to the location of the items you wish to add to the index. Note that in the “Exclude filters” section of the dialog box, it is possible to specify by extension those types of files you wish to exclude from the indexing process, such as the miscellaneous system files that were included on the DVD that are not relevant to our hypothetical matter at hand. Once all of the necessary selections are made, click “Start Indexing,” and the indexing process will begin. Once the indexing process is complete, every word in the collection is available for searching.

Upon completion of the indexing process, a report can be generated, shown as Figure-7.3, that identifies the date and start/stop times the indexing process was run, the total number of words in the index, the total number of documents contained in the index, the number of documents indexed, the bytes of indexed data, and the documents that were removed from the index. For documents that were unable

Figure 7.3

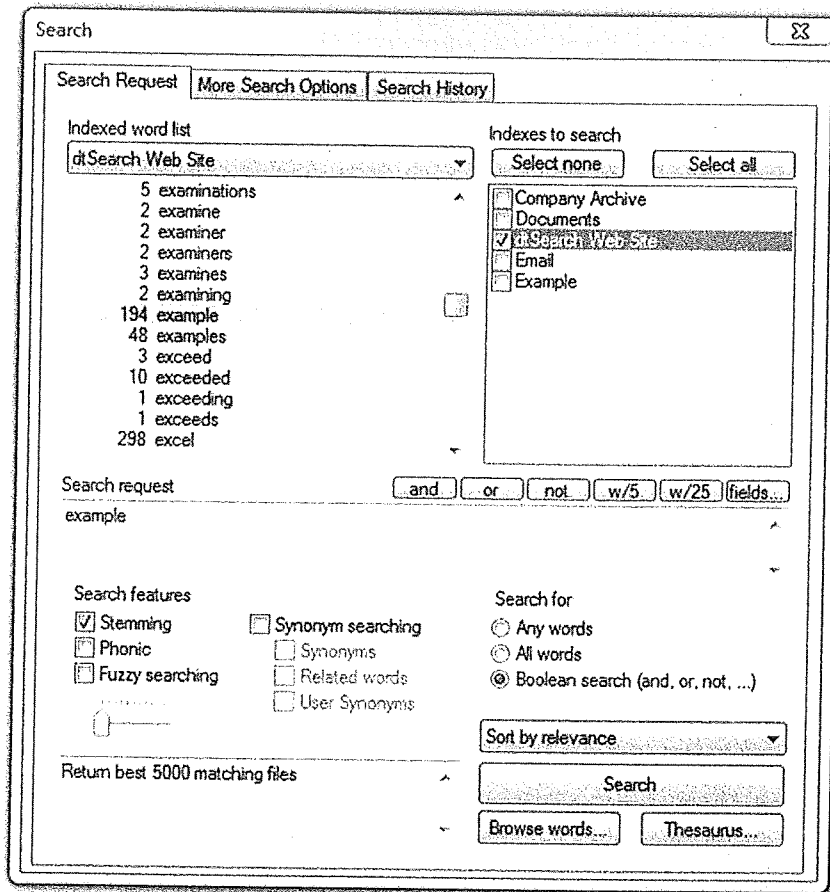
Index Update Report	
C:\Users\Bruce\AppData\Local\dtSearch\Sample Index	
Actions:	Create Add Compress RemoveDeleted
Result:	OK
Start	9/27/2011 10:16:11 AM
End	9/27/2011 10:16:33 AM
Documents indexed	
Total words in index	32,491
Total documents in index	1,109
Documents indexed	1,109
Bytes indexed	231,710 kb
Documents removed	0
Documents not indexed	
Documents not indexed	0
Encrypted files	0
Unreadable files	0
Skipped as "binary"	0
Documents partially indexed	
Partially encrypted files	0
Partially unreadable files	0

to be indexed, the report identifies the number of documents that did not process and indicates those that were encrypted or unreadable or that were skipped as binary files. Partially indexed files are identified as well. Finally, you can generate a list of the documents that were not indexed if needed.

To begin to search the indexed files, click on the "Search" button on the button bar or select "Search" from the Search drop-down menu to open the Search dialog box, shown as Figure 7.4.

The top right of the dialog box shows a list of the indexes that are available for searching. You can search in one or more than one index at a time. The "Indexed word list" is located at the top left of the dialog box, and it shows a list of all the words that are currently available for search-

Figure 7.4



ing in the index. In the example above, the word "example" appears 194 times in the dtSearch Desktop Web site index. You can insert the word "example" into the "Search request" box, and it will return the 194 hits identified in the word list. It is possible to expand the search results on the word "example" if you enable the "Stemming," "Fuzzy searching," or "Synonym searching" search features. Stemming, for example, would return all incidences of the word "example" and "examples." Depending on the sensitivity you set, fuzzy searching would return additional hits such as "examining" and "examines." It also helps by accommodating typos. Synonym searching would return those documents that have the word "example" in them, along with documents containing synonyms such as "sample" or "exemplar."

Note that to the right, just above the "Search request" box, there are a series of buttons that can be used to build a Boolean string search if you

want to construct a more advanced search. These buttons allow you to use the “and, or, not, and within a specified range” commands to refine your search. There are other searching techniques that can be used in Boolean searching. For example, you can use quotation marks around phrases to search for entire phrases. You can insert “+” in front of any word or phrase to require it in the search results, or insert “-” in front of the word or phrase to exclude it. Other more advanced searching techniques are also available.

In Figure 7.5, under the “Search Request” tab, a search request was run on the word “eDiscovery” in the Testdb database. Stemming was selected, and a limitation of the best 5,000 matches was set.

Figure 7.5

The screenshot shows a software interface for searching. At the top, there are three tabs: "Search Request" (selected), "More Search Options", and "Search History".

Indexed word list: A list of words with their frequencies. "eDiscovery" has the highest frequency at 4,764. Other words include edis6m (2), edisc (2), ediscactnovdraft (5), ediscocguidelinesfinal (9), ediscov1 (4), ediscove (35), ediscovery11x17final (2), ediscovery1_001 (8), ediscovery2_001 (8), ediscovery_rlist (1), ediscovery_w_notes (7), and ediscoverybytes (7).

Indexes to search: A list box containing "Testdb" with a checked checkbox. Above the list are "Select None" and "Select All" buttons.

Search request: A text box containing "eDiscovery". Above it are buttons for "and", "or", "not", "w/5", "w/25", and "fields...".

Search features: A section with checkboxes for "Stemming" (checked), "Phonic searching", "Fuzzy searching", "Synonym searching", "Synonyms", "Related words", and "User synonyms".

Search for: Radio buttons for "Any words" (selected), "All words", and "Boolean search (and, or, not, ...)".

Sort by relevance: A dropdown menu set to "Sort by relevance".

Return best 5000 matching files: A text box containing "5000".

Buttons: "Search", "Browse Words...", and "Thesaurus..." are located at the bottom right.

To further refine the search, the "More Search Options" tab was selected, shown in Figure 7.6. In addition to setting the search results options to return a maximum of 5,000 files and to stop searching after 25,000 files, the "File modification date" drop-down menu is highlighted in Figure 7.6 to show how you can select a date range based on the last modified date. Additional search options are also demonstrated.

Figure 7.6

The screenshot displays the 'More Search Options' tab in the dtSearch Desktop application. The interface is organized into several sections:

- Search Request | More Search Options | Search History** (Tabs)
- Search results options**
 - ☒ Limit search results to the best matching files
 - Number of files to return: 5000
 - Stop search after: 25000 files
- File filters**
 - Name matches: _____
 - Name does not match: _____
- File modification date** (Highlighted section)
 - Drop-down menu: All files, regardless of date (with a list of options including 'All files, regardless of date', 'File last modified before', 'File last modified after', and 'File last modified between' and 'and')
 - Clear Filters button
- Unindexed search options**
 - Search type: Index search only (drop-down menu)
 - What to search (unindexed search): _____ (text input field)
 - Buttons: Add Folder..., Add Files..., Remove, Clear List

The search results screen in the form of a spreadsheet report is set forth in Figure 7.7. Retrieved documents will sort by relevance as the default; however, you can sort by date or search by number of hits in each document by clicking the column headers to change the sort order. Additional information regarding each document is set forth in the spreadsheet. If desired, a report of the spreadsheet results can be saved or printed.

In Figure 7.7, a split screen is shown with the top-half of the spreadsheet containing the search results and the bottom-half containing a rendering of the highlighted document that was selected. The search term is highlighted in yellow. There are right and left arrow buttons on the button bar that allow you to move forward or backward from hit to hit. Buttons also exist to help you navigate from document to document. If you wish, you can select the button on the button bar containing the pencil and note, or hit "F8" on your keyboard, and the selected item will open in its native format.

Figure 7.7

The screenshot shows a search application window with a menu bar (File, Edit, Search, Index, View, Options, Help) and a toolbar. The main area is split into two panes. The top pane displays a list of search results in a table format with columns: Name, Score, Hits, Location, Date, Size, and Index. The bottom pane displays a preview of the selected document, which is titled "Everything You Need to Know about Records Management, Identification and Preservation" by Bruce A. Olson. The preview shows the introduction of the document, discussing the role of an attorney in the new era of electronic information and the obligation to preserve electronically stored information (ESI).

Name	Score	Hits	Location	Date	Size	Index
1 Ediscovery%20Workflow%20White%20Paper%20-%20Final%20-%20BAO0003).PDF	100%	282	C:\BAO\Workflow\Files\SPEAKING	11/15/2004	1,273,240	Testab Downloadable White Paper
2 Ediscovery%20Workflow%20White%20Paper%20-%20Final%20-%20BAO0003).PDF	100%	282	C:\BAO\Workflow\Files\LIT-SPEAKING	11/15/2004	1,273,240	Testab Downloadable White Paper
3 Ediscovery - eDiscovery Workflow White Paper (EDISCO-1).PDF	84%	230	C:\BAO\Workflow\Files\COM	7/21/2003	1,739,571	Testab Microsoft Word - eDiscovery Work
4 Ed Review feature article (03197872).DOC	80%	1	C:\BAO\Workflow\Files\TECHNO LAWYER	10/3/2007	32,768	Testab Ediscovery: Everything You Need
5 Ed Analysis feature article (03197872).DOC	80%	1	C:\BAO\Workflow\Files\TECHNO LAWYER	10/3/2007	32,768	Testab Ediscovery: Everything You Need
6 RE Version 2.5 (00051384).MSG	70%	320	C:\BAO\Workflow\Files\COM	9/11/2003	7,072,250	Testab Ediscovery 11x17Final
7 Ediscovery - The eDiscovery Console (EDISCO-2).PDF	35%	24	C:\BAO\Workflow\Files\SPEAKING	7/21/2003	2,499,110	Testab Ediscovery 11x17Final
8 20061130 Darsinger Juhvika Bomber 3m Ediscovery Show (03154612).PPT	34%	16	C:\BAO\Workflow\Files\SPEAKING	12/11/2003	1,148,416	Testab Using the New Federal Ediscovery
9 20061130 Darsinger Juhvika Bomber 3m Ediscovery Show (03154612).PPT	34%	16	C:\BAO\Workflow\Files\SPEAKING	12/11/2003	1,148,416	Testab Using the New Federal Ediscovery
10 Tiberg v. Hunt Mgmt. Co., 2005 WL 2759600 (S.D.N.Y. Oct. 24 (03149257)).pdf	33%	7	C:\BAO\Workflow\Files\EDISCOVERY	11/6/2004	47,937	Testab Court Expands Ediscovery where C
11 Tiberg v. Hunt Mgmt. Co., 2005 WL 2759600 (S.D.N.Y. Oct. 24 (03149257)).pdf	33%	7	C:\BAO\Workflow\Files\EDISCOVERY	11/6/2004	47,937	Testab Court Expands Ediscovery where C
12 Fiat EDRI.doc (00216791).DOC	26%	17	C:\BAO\Workflow\Files\TECHNO LAWYER	2/25/2008	61,952	Testab Ediscovery: Everything You Need
13 Ediscovery - The Extended DM (EDISCO-1).PDF	25%	57	C:\BAO\Workflow\Files\COM	7/21/2003	906,445	Testab Microsoft Word - Extensions to DII
14 Ediscovery - The Extended DM (EDISCO-1).PDF	25%	57	C:\BAO\Workflow\Files\COM	7/21/2003	906,445	Testab Microsoft Word - Extensions to DII
15 MLR_LF_FRCPeDiscovery_17apr03 (03142257).PDF	21%	9	C:\BAO\Workflow\Files\SPEAKING	9/11/2003	155,830	Testab MLR_LF_FRCPeDiscovery_17apr0
16 MLR_LF_FRCPeDiscovery_17apr03 (03142257).PDF	21%	9	C:\BAO\Workflow\Files\SPEAKING	9/11/2003	155,830	Testab MLR_LF_FRCPeDiscovery_17apr0
17 Appendix (03277222).DOC	20%	25	C:\BAO\Workflow\Files\SPEAKING	11/10/2003	223,232	Testab THE E-DISCOVERY PROCESS E
18 Appendix (03277222).DOC	20%	25	C:\BAO\Workflow\Files\SPEAKING	11/10/2003	223,232	Testab THE E-DISCOVERY PROCESS E
19 Technical eDiscovery: Everything You Need to Know About (BAO0590).MSG	20%	13	C:\BAO\Workflow\Files\TECHNO LAWYER	7/31/2008	91,618	Testab Sender: TechnoLawyer Spoof@w
20 Philadelphia program outline (03094442-2).DOC	18%	5	C:\BAO\Workflow\Files\TECHNO LAWYER	8/21/2005	118,736	Testab WORKING COPY OF BAO 6/21/05
21 Technical eDiscovery: Everything You Need to Know About (BAO0591).MSG	17%	13	C:\BAO\Workflow\Files\TECHNO LAWYER	7/31/2008	114,176	Testab Sender: TechnoLawyer Spoof@w

Everything You Need to Know about Records Management, Identification and Preservation
by Bruce A. Olson

INTRODUCTION

In this new era of **EDISCOVERY**, the role of an attorney is much different than it used to be. You no longer have the luxury of waiting until a client has decided to sue or has been sued to determine what electronic information about a particular disputed matter exists, or where such information is located. The obligation to preserve electronically stored information (ESI) arises when you or your client have a reasonable expectation that litigation will ensue. What that "reasonable expectation" is varies from case to case. In some businesses that expectation is for all intents and purposes always present. If you represent a potential plaintiff the obligation to preserve may well arise before you are even sure you have a case, or your client has decided it actually wants you to commence in litigation. You can't start sending out demand letters or conducting settlement negotiations just to see how far you can get short of filing a lawsuit without having taken appropriate steps to preserve your client's ESI. On the flip side, if you receive a demand letter you can't just wait and see what happens, or do nothing until your client is actually sued before you initiate a litigation hold. If you do wait to see what happens you should probably expect a claim of spoliation of evidence being raised somewhere along the line.

In other words, **EDISCOVERY** does not just affect litigators. Planning for and dealing with **EDISCOVERY** is part and parcel of any good corporate, employment, public sector, private sector or any other type of legal practice where you offer ongoing client advice. If you are not a litigator and think you can ignore this burgeoning area of the law, you might as well put your malpractice carrier on notice now. In fact, a new breed of **EDISCOVERY** advice lawyer is beginning to emerge. That lawyer blends a thorough knowledge of the litigation process with knowledge about electronic records management, general records management, industry specific regulatory compliance, and comprehensive IT knowledge about hardware, software, e-mail, Internet usage, networks, storage devices, and the many other technologies that are now part of our normal day to day business operations.

Finally, as shown in Figure 7.8, the "Search History" tab is available on the Search dialog box to generate a list of each search that was run. The history provides the search term, the number of hits, and the date the search was run. This is a useful feature that allows you to repeat a prior search and generate the same search results by simply selecting and double clicking on the desired search term or by highlighting a search and then clicking the "Open" button.

Figure 7.8

The screenshot displays the dtSearch Desktop application window. At the top, there are three tabs: "Search Request", "More Search Options", and "Search History". The "Search Request" tab is active, showing a table with three columns: "Search Request", "Files", and "Date". The table contains three rows of data. To the right of the table are three buttons: "Open", "Delete", and "Delete All". Below the table, there is a section labeled "Search request" with a text input field containing "eDiscovery" and an "Insert" button. At the bottom, there is a section labeled "Files retrieved" with a table that has three columns: "Name", "Hits", and "Location". The table is currently empty.

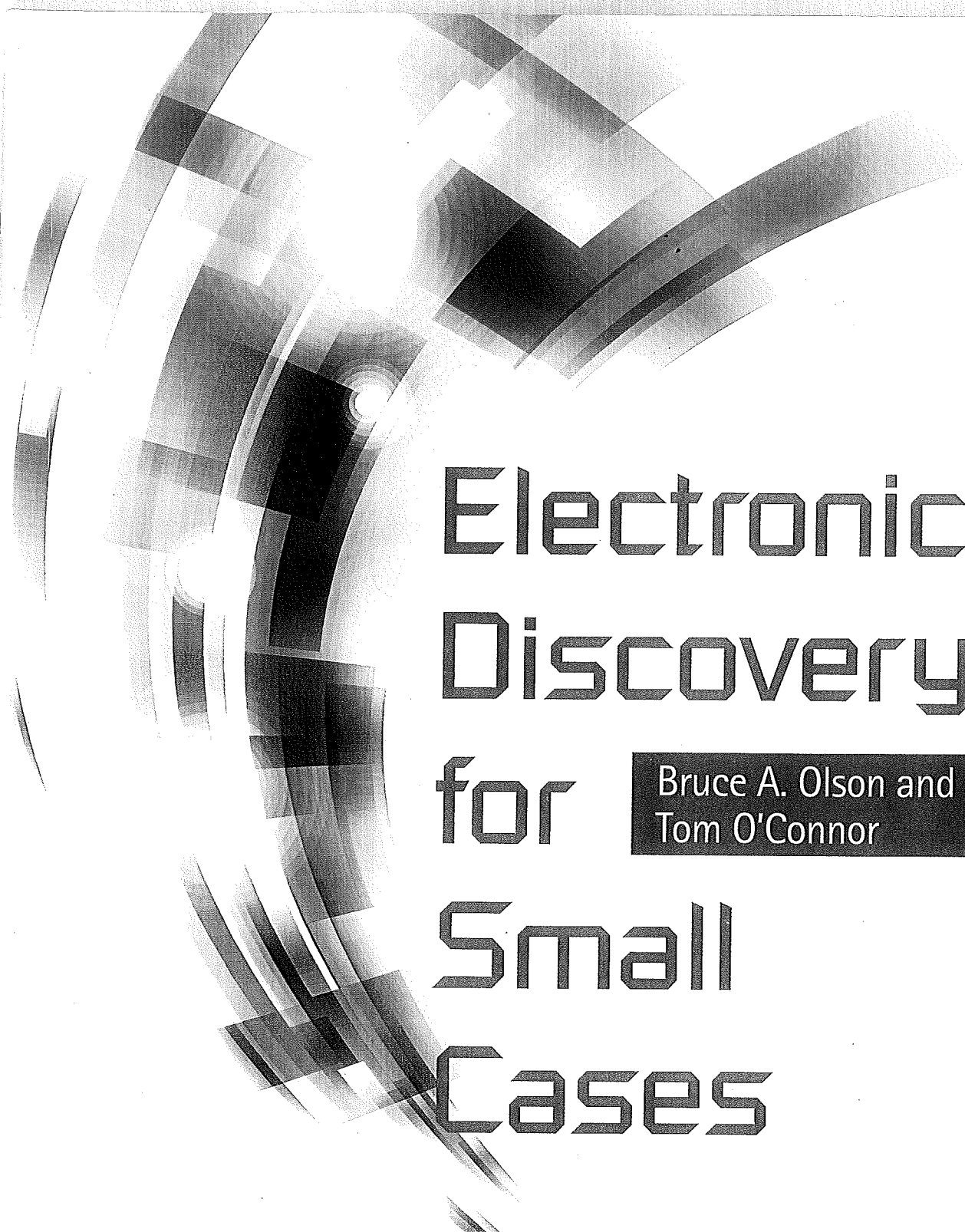
Search Request	Files	Date
S3 eDiscovery	642	9/27/2011 10:...
S2 eDiscovery	642	9/27/2011 10:...
S1 eDiscovery	642	1/17/2010 4:0...

Search request
eDiscovery

Files retrieved

Name	Hits	Location
------	------	----------

All of the common file formats you are likely to encounter are searchable using dtSearch Desktop. It is a logical solution to managing smaller amounts of ESI commonly received in small cases. It enables you to search and view everything in the collection without the need to purchase more expensive litigation support alternatives. Its use is not limited to small cases, but for those who typically deal with small cases, dtSearch Desktop is an affordable and robust solution.



Electronic Discovery for Small Cases

Bruce A. Olson and
Tom O'Connor

Managing Digital Evidence and ESI



LAW PRACTICE MANAGEMENT SECTION
MARKETING • MANAGEMENT • TECHNOLOGY • FINANCE

CHAPTER EIGHT

A Processing, Review, and Production Tool All in One

AS E-DISCOVERY BECOMES A normal part of every litigated matter, including small- to mid-sized cases where the potential cost of e-discovery can rival the value of the litigated claim, the need for an economical way to process, analyze, review, and produce the relevant ESI is critical. The terms “e-discovery” and “bargain” rarely are found together in the same sentence. However, in the case of Digital WarRoom Pro (DWR Pro), they really do belong together. At a cost of only \$895 for a single-seat installation when this book was written, DWR Pro offers a feature-rich, PC-based, e-discovery application for rapid processing, review, and production of small- to medium-sized document collections (typically in the range of 500,000 documents or less). Multiuser, network-based solutions are also available for additional cost, and the program can be deployed just as successfully against much larger collections in a team environment.

Digital WarRoom Pro

DWR Pro offers an end-to-end solution to deal with ESI from identification through processing, analysis, review, and production. Collected custodian ESI, including PST and ZIP files as well as native files, can be quickly processed for review. This includes the option of de-NISTing to remove application software, .EXE files, and other files not relevant to e-discovery. Container files and e-mail files are automatically extracted during processing. The entire collection is then indexed and hashed, and a document database is created for analysis, review, and production purposes.

Installation requires a two-step process, and you must understand up front that some additional software is required for DWR Pro to function properly. First, you need an installation of Microsoft Office Professional 2007 or later on the PC on which you install DWR Pro. You also need Internet Explorer 6.0 SP1 set as the default browser and Microsoft .NET Framework 4.0 or later. The program will run on Windows XP Professional SP2, but Windows 7 Professional with 4 GB RAM and at least 160 GB of hard drive space is recommended. You must also purchase and install Quick View Plus to serve as a viewer (\$49). You must first install SQL Server Express (at no additional cost) prior to installing DWR Pro, but once that is done, the balance of the program installation is simple and straightforward.

A typical Windows application interface is used. If you are familiar with Microsoft Outlook, you will easily be able to navigate the program menus and controls in the frames on the left side of your screen. As set forth in Figure 8.1, the document “Review” window in the main workspace provides a document list in grid format similar to Microsoft Excel.

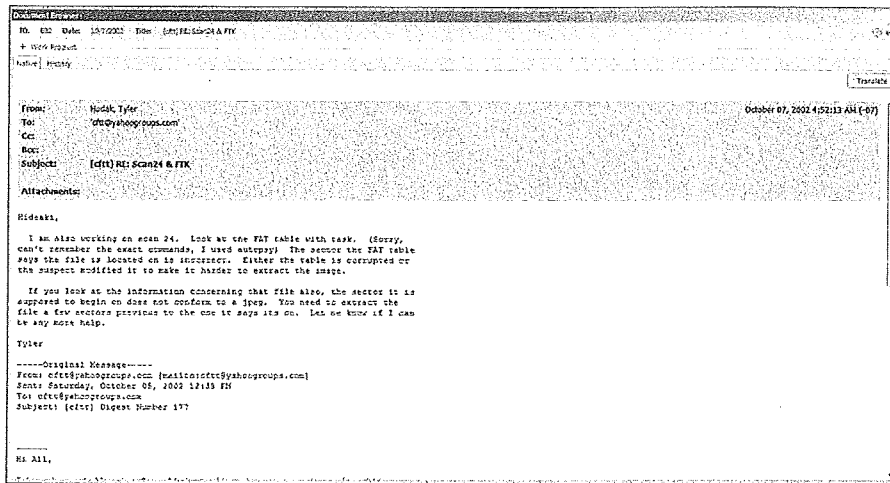
Figure 8.1

Microsoft Word 2003 interface showing a document titled "Digital Workbook Pro - Easy Collection #1". The document is a table of contents with columns for page number, chapter, title, and page number. The table lists chapters 1 through 10, including topics like "Introduction", "Getting Started", "Using the Workbook", "Using the Workbook", "Using the Workbook", "Using the Workbook", "Using the Workbook", "Using the Workbook", "Using the Workbook", and "Using the Workbook".

Page Number	Chapter	Title	Page Number
1	1	Introduction	1
2	2	Getting Started	2
3	3	Using the Workbook	3
4	4	Using the Workbook	4
5	5	Using the Workbook	5
6	6	Using the Workbook	6
7	7	Using the Workbook	7
8	8	Using the Workbook	8
9	9	Using the Workbook	9
10	10	Using the Workbook	10

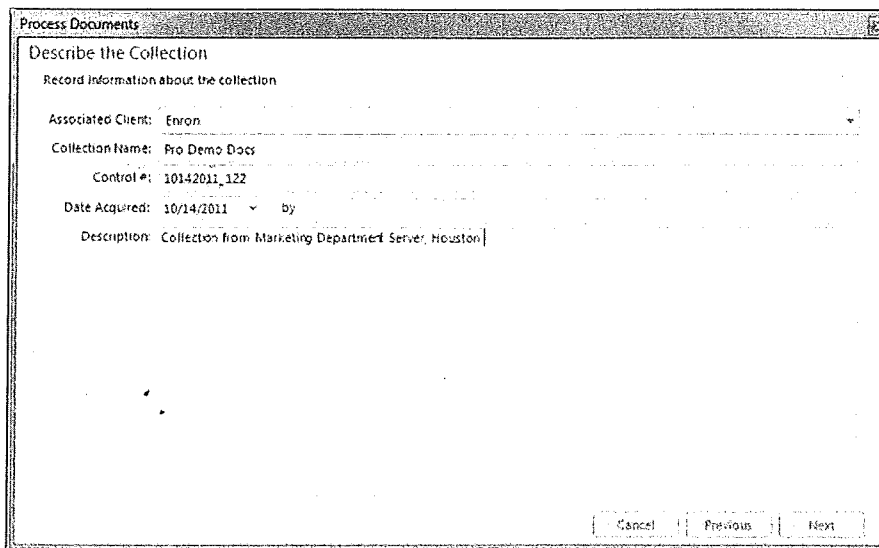
The document browser pane (Figure 8.2), where a selected document is reviewed, can be positioned in any quadrant of the main screen, the same screen, or dragged onto a second monitor allowing you to use dual monitors for an expanded view.

Figure 8.2



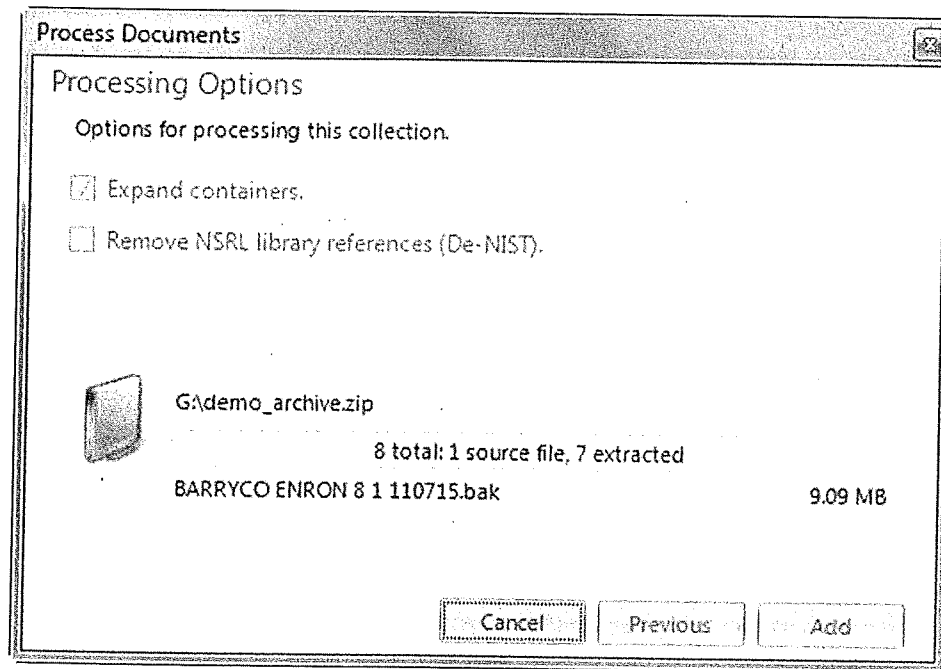
Adding data to the program is simple. Activate the "Process Documents" menu to either browse through the explorer tree to the collections of processed data, or drag and drop new folders of custodian data onto the processing window (Figure 8.3). Each folder of data is profiled using the date of acquisition, the name of the collection, the name of the person who acquired the data, the client, and any relevant notes as descriptors.

Figure 8.3



You have the option at this point of extracting ZIP files, PST files, and other container files automatically. You can also remove operating system and application files using a de-NIST process. The "Admin" tab provides a link to download the current NSRL library from the DWR Pro Web site. See Figure 8.4.

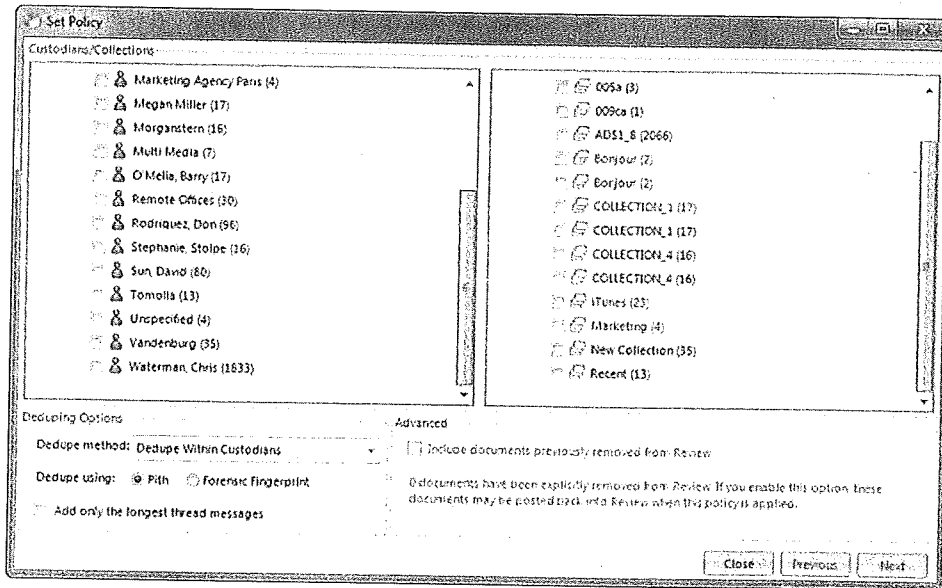
Figure 8.4



When adding ESI collections from custodians, you have the ability to import the data and associated load files to automatically code fields of information in the database. An exception report showing any items that were password protected, corrupt, or failed to process properly is automatically generated. A collections report and processing history log are also created and automatically updated as new items are added to the collection.

As shown in Figure 8.5, policies can be applied that allow deduping within custodian or across the matter, or to retain duplicate documents. Date-range limitations and file-type limitations are available to provide early culling. For example, you can set the parameters to process content created on or after June 1, 2008, that include documents, spreadsheets, and e-mail. The balance of the files in the raw data will not be processed.

Figure 8.5



Once the data collections are loaded and indexed, a number of early case assessment tools are available to help you quickly determine the size and composition of the case database. As shown in Figure 8.6, a "What If" report can be analyzed in grid view using a variety of comparative factors, including custodian, collection, file extension, and key word. For exam-

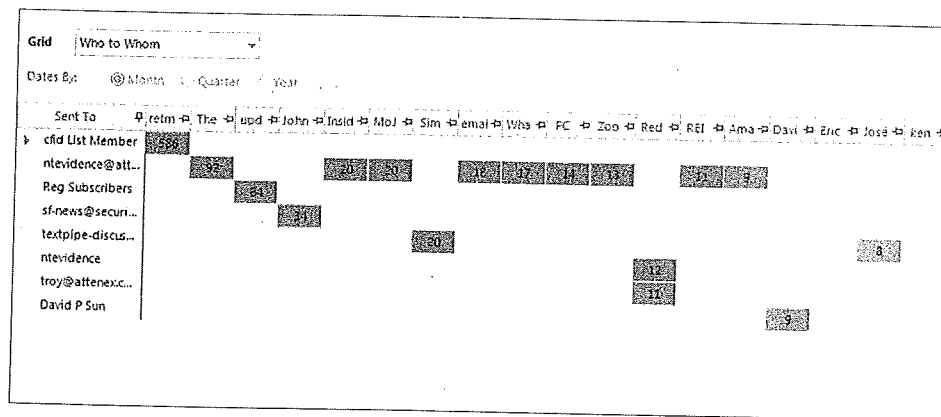
Figure 8.6

Grid	Date	by	Ext	Current Docs: 2215
Dates By:	Collection	Year		
	Custodian			
2002	EM			
2002	Keywords			
2002	Mark			
2002	Who to Whom			
2002 Oct				
2002 Sep				
2003 May				
2003 Sep				
2004 Apr				
2004 Aug				
2004 Dec				
2004 Feb				
2004 Jan				
2004 Jul				
2004 Jun				
2004 Mar				
2004 May				
2004 Nov				
2004 Sep				
2005 Apr				
2005 Aug				
2005 Dec				

ple, you can generate a report that shows each category of data by extension type for each individual collection of documents in the total collection to quickly determine where the bulk of the case data resides and what type of file types are involved. This helps you determine the number of files you must review to help develop a review budget.

As shown in Figure 8.7, a "Who to Whom" analysis of e-mail is another early case assessment tool that provides a graphical view of the number of e-mail contacts one individual had with another individual. This allows you to determine early on whose e-mail should be reviewed initially.

Figure 8.7



Once you have loaded the data and indexed it, the full collection is available for review (see Figure 8.8). The spreadsheet view of the collection can

Figure 8.8

be sorted by column, the columns can be dragged and dropped to a new position or dragged off the screen, and you can select which columns to view from a pick list.

As shown in Figure 8.9, the amount of information visible in the spreadsheet can be further filtered by file extensions, custodians, key words, or other more advanced filters that use Boolean search terms to create more complex search parameters.

Figure 8.9

Digital WarRoom® Pro - BarryCO:Enron_8.1

Matter Settings Review Productions Admin Help

Review

Date Range:

Start Date: 1/1/1970

End Date: 10/14/2011

Extensions:

No Extension Filters

	EXT	COUNT
☐	email	1242
☐	c	46
☐	ix	36
☐	pdf	26
☐	doc	24
☐	txt	18

Other Attributes:

☐ Only unindexed files

Custodians

Keywords

Attributes

Marks/Issue Codes

Collections

Advanced

In the review process, you can mark items for privilege, responsiveness, and production, as well as issue code them. Coding can be done on a per-document basis using a work product palette containing the check box "Marks/Issue Codes." See Figure 8.10.

Figure 8.10

Document Browser

ID: 580 Date: 8/30/2002 Title: [CFID] USB flash hard drive forensics {01}

☐ Work Product:

Marks:

☐ Non-Responsive	☐ Tech
☐ Privileged	☐ Don'
☐ Produce	

Issue Codes:

☐ I. Hot	☐ BB. Bi
☐ P. Patent Records	☐ CR. C
☐ 46. 461 Patent	☐ KK. Ki

Comments:

Native History

From: retmail@infobin.org August 30, 2002 12:47:22 PM (-07)

To: CFID List Member

Cc:

Bcc:

Subject: [CFID] USB flash hard drive forensics {01}

Attachments:

{* The following is provided for educational purposes, no claims of *}
 {* copyright are made and opinions are those of each author only & *}
 {* may not reflect the position of his/her organization. *}
 {* Your reply to this message will be redirected to the listserv. *}

From: Byron.S.Collie@WellsFargo.COM

In the Settings menu, you can create user-defined issue codes specific to the case. See Figure 8.11.

You can also right click on an item to obtain standard marking options when working in the spreadsheet view. As shown in Figure 8.12, you can select multiple items in the spreadsheet view and bulk mark them as well.

The viewer will provide a rendered view of each document you select as it appears in its native format, as set forth in Figure 8.13.

Figure 8.11

Issue Codes

CODE	ISSUE	PRODUCE
!	Hot	☐
TH	Theft of IP	☐
46	461 Patent	☐
KO	Korean	☐
PT	Partnership Agmt	☐
P	Patent Records	☐
TR	Trademark	☐
CR	Copyright #1	☐
BB	Biz issue	☐
KK	Kryon	☐
* <Unused>	<Unused>	☐

Cancel OK

Figure 8.12

Scope of Search: Customized Address Drawings, Lay, Ken Morganstein, Khalil Media, O'Maha, Barry, Remote Offices, Rodriguez, Don, Skonenski, Slope, Sun David, Terriola...

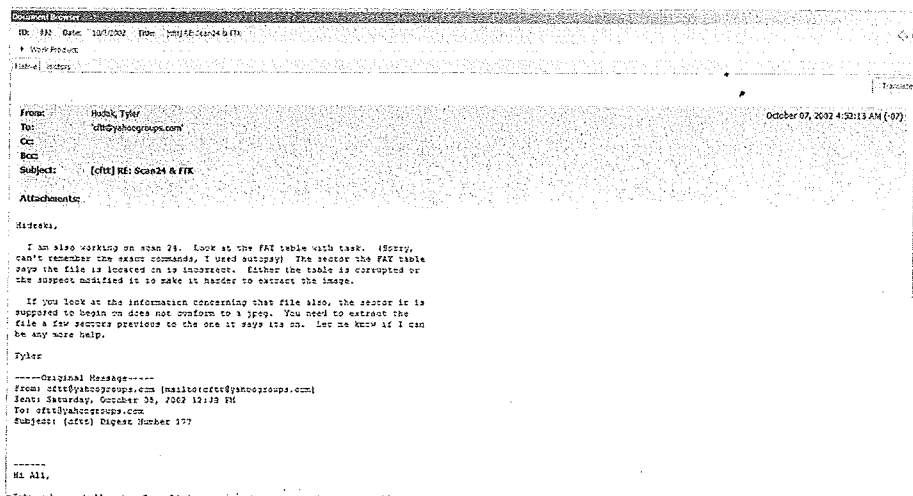
Search Parameters: KEYWORDS (Unk) including Family Members

Current User: 181

Estimated Pages: 2,215

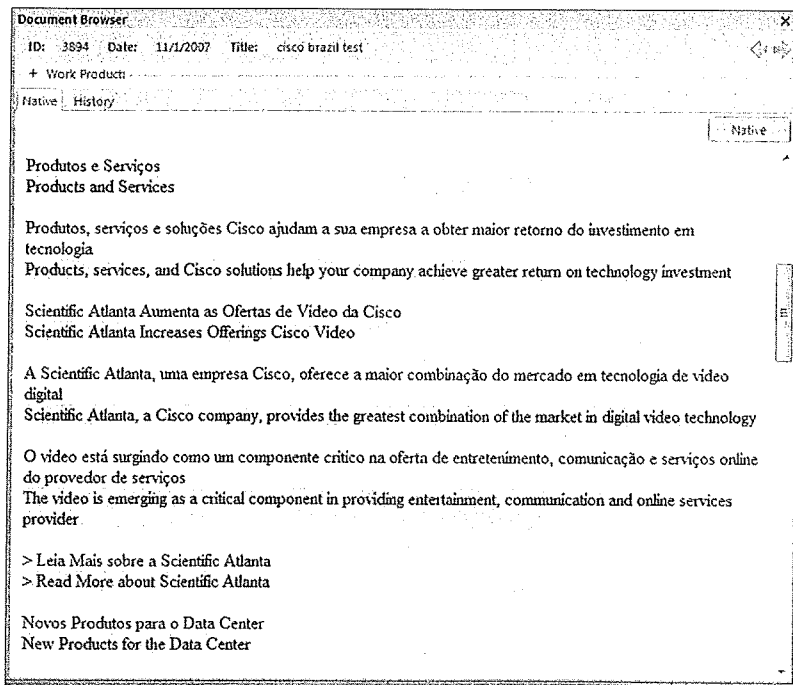
ESTD	MARK	DATE	EXT	TITLE	FAMILY	ID	SCORE	PATH
1		09/30/2002	email	Red Hat Linux 8.0 now available on Red Hat Network	1364	1364	26	001(Chris Waterman)Den
2	Non-Responsive	2	email	RHN Errata Alert: Updated version of GCC 2.96-RH now availa	1368	1368	3	001(Chris Waterman)Den
3	Privileged	2	email	RHN Errata Alert: Updated glibc packages fix vulnerabilities in	1372	1372	3	001(Chris Waterman)Den
4	Produce	2	email	RHN Errata Alert: Updated ypinit packages fixes memory leak	1374	1374	7	001(Chris Waterman)Den
5	Technical Difficulty	2	email	RHN Errata Alert: Updated Mozilla packages for security vulne	1376	1376	3	001(Chris Waterman)Den
6	Don't Know	2	email	RHN Errata Alert: Updated Mozilla packages for security vulne	1378	1378	3	001(Chris Waterman)Den
7	Unmark	2	email	RHN Errata Alert: Updated xinetd packages fix denial of servic	1380	1380	3	001(Chris Waterman)Den
8	Issues	2	email	RHN Errata Alert: Command execution vulnerability in drips	1382	1382	3	001(Chris Waterman)Den
9	Add to binder...	2	email	RHN Errata Alert: Command execution vulnerability in drips	1384	1384	3	001(Chris Waterman)Den
10	Add to production...	2	email	RHN Errata Alert: Updated fetchmail packages fix vulnerability	1390	1390	3	001(Chris Waterman)Den
11	Flag for redaction	2	email	RHN Errata Alert: Updated glibc packages fix vulnerabilities in	1392	1392	3	001(Chris Waterman)Den
12	Remove from Review	2	email	RHN Errata Alert: Updated glibc packages fix vulnerabilities in	1394	1394	3	001(Chris Waterman)Den
13	Comments...	2	email	RHN Errata Alert: Updated unzip and tar packages fix vulnera	1396	1396	7	001(Chris Waterman)Den
14	Copy value	2	email	RHN Errata Alert: Updated xinetd packages fix tempfle	1398	1398	7	001(Chris Waterman)Den
15		07/31/2002	email	RHN Errata Alert: Updated xinetd packages fix temporary file ha	1400	1400	7	001(Chris Waterman)Den
16	5	10/30/2002	email	[red] SWGDE Draft (R3)	1476	1476	3	001(Chris Waterman)Den
17	5	11/15/2002	email	Today's Headlines from NYTimes.com Friday, November 15, 2	1548	1548	3	001(Chris Waterman)Den
18	1	12/08/2002	email	[CJIO] Files with a SPI extension [01]	1488	1488	3	001(Chris Waterman)Den
19	1	11/15/2002	email	Re: [JTT] Questions from newbie - Pls pardon my ignorance	1168	1168	5	001(Chris Waterman)Den
20	3	11/13/2002	email	Increase Storage Capacity with NAS Devices	1292	1292	7	001(Chris Waterman)Den
21	3	10/02/2002	email	How Web Digital Identity Programs Affect Your Business?	1298	1298	3	001(Chris Waterman)Den
22	3	10/18/2002	email	First Look: Best Buy vpr Matrox 110-200AS, AOL 8.0, MSN 8.0	1312	1312	3	001(Chris Waterman)Den
23	3	08/16/2002	email	First Look: Microsoft Money 2003 Deluxe vs. Intuit's Quicken	1326	1326	3	001(Chris Waterman)Den
24	7	11/07/2002	email	Under the Brim Red Hat - November 2002	1354	1354	100	001(Chris Waterman)Den

Figure 8.13



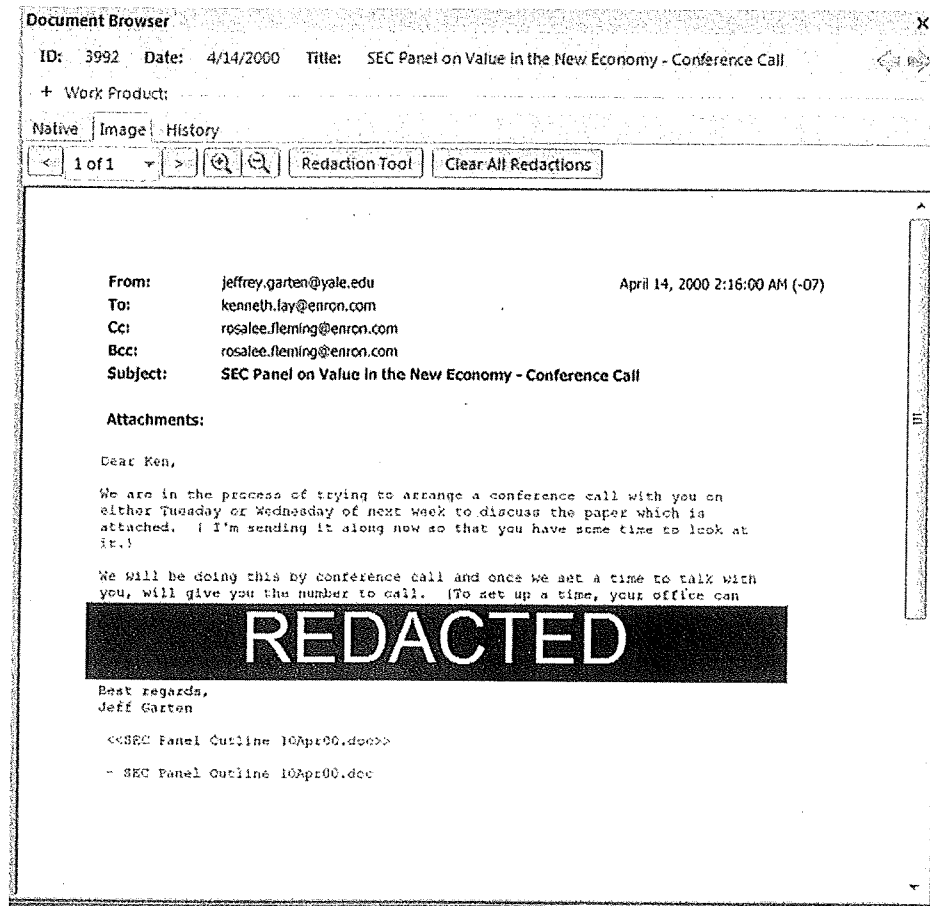
Foreign language content based on the Roman alphabet, Cyrillic, and Asian language character sets can be processed, indexed, and searched within DWR Pro. The Document Browser contains a translation option that will automatically translate non-English content and display the translation, assisting the reviewer in identifying the subject and content of the document. See Figure 8.14.

Figure 8.14



As the collection is reviewed, items that are potentially subject to document production can be flagged on an ongoing basis. They can subsequently be further reviewed as a subset of the total collection, and determinations can be made for purposes of redaction. Redactions appear as shown in Figure 8.17.

Figure 8.17



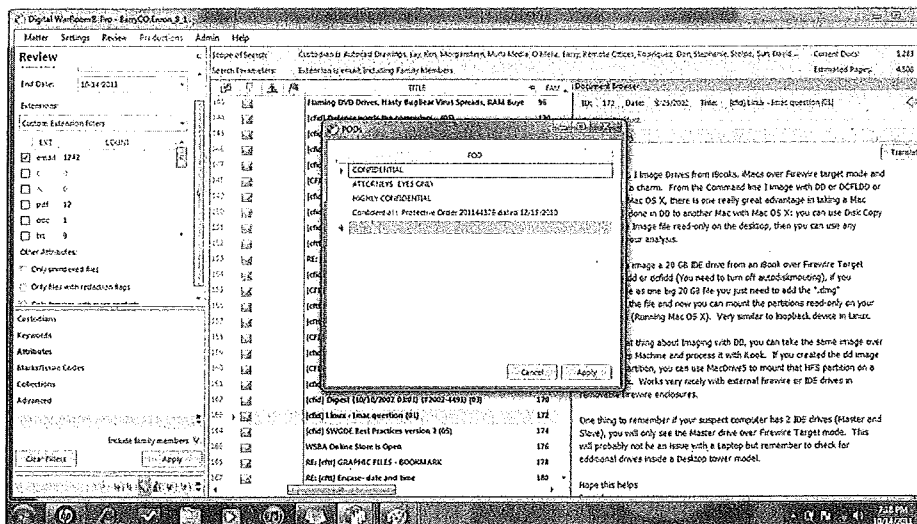
As set forth in Figure 8.18, if further review demonstrates an item should not be produced, it can be right clicked and clawed back from the production.

Right-click functionality also permits the stamping of documents with "Confidential" or "Attorney's Eyes Only." This is a customizable selection to associate a protective order designation with a document. See Figure 8.19.

Figure 8.18

BEG DATE	END DATE	EXT	ID	PHASE	TYPE	MARK	STATE	PAGES	PATH
1	1	email	3962	Produce	Image	Produce	Released	1	463
2	2	email	3964	Produce	Image	Produce	Released	1	740
3	3	email	3978	Produce	Image	Produce	Released	1	713
4	4	email	3976	Produce	Image	Produce	Released	1	1010
5	5	email	3992	Produce	Image	Produce	Released	1	1196
6	6	email					Clawback	1	1089
7	7	email					Clawback	1	908
8	8	email					Released	1	960
9	9	email					Released	1	953
10	10	email					Released	1	353
11	12	email					Released	2	293
13	13	email	4026	Produce	Image	Produce	Released	1	360
14	20	email	4026	Produce	Image	Produce	Released	7	521

Figure 8.19



Other endorsements can be added as needed. As set forth in Figure 8.20, a graphical layout tool helps you design the final production format.

Once the production set has been triaged and finalized, it can be exported as either a native or a TIFF image production set. The production set can then be copied to CD/DVD or placed on a thumb drive or external hard drive for delivery to the opposing side.

During the review phase, the use of subfolders called "binders" is supported. This allows you to create multiple, unique sets of documents for use as discovery productions, for depositions, for distribution to a particular expert, or for trial exhibits. See Figure 8.21.

Figure 8.20

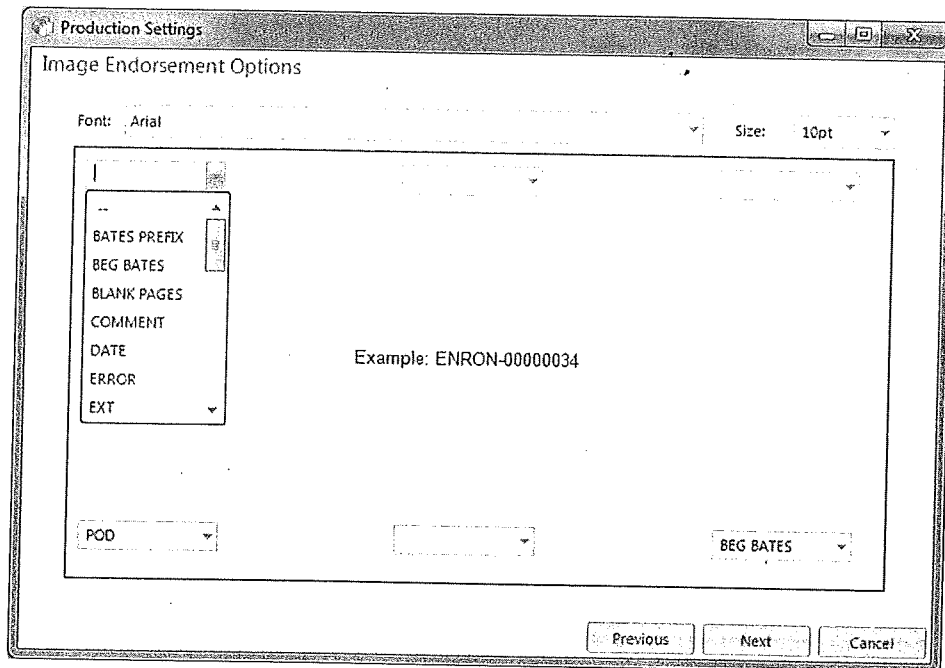


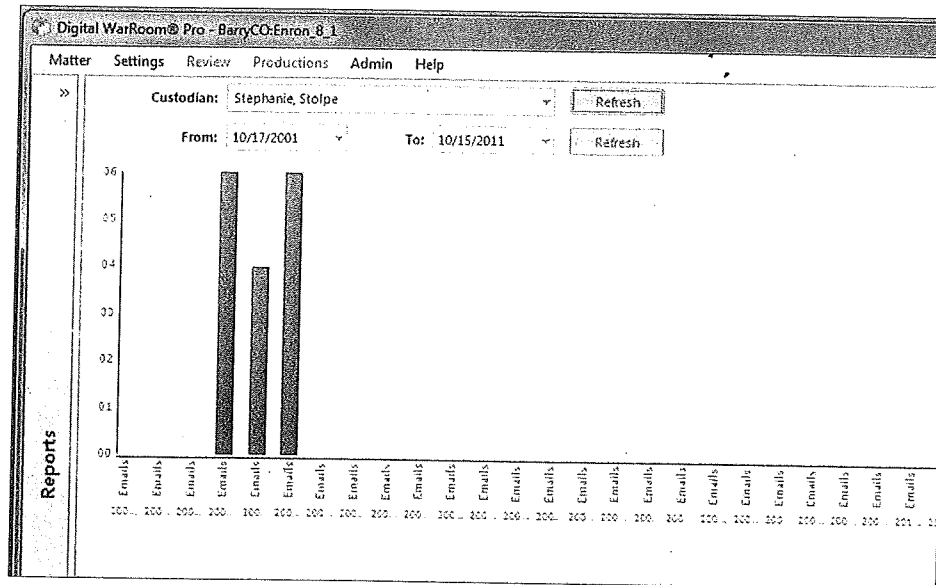
Figure 8.21

	ESTI	MARK	DATE	EXT	TITLE	FAMILY	ID	SCORE	PATH
1	3		11/12/2002	email	Condex Rumors, Tablet PC Details, Yuppie Crack, Roll Your O	52	52	3	001\Chris Waterm
2	4		11/05/2002	email	17 Ways to Foil Hackers, Office 11 Review, Nokia GameBoy Ph	54	54	4	001\Chris Waterm
3	4	Produce	10/01/2002	email	Flaming DVD Drives, Nasty Bugbear Virus Spreads, Palm Buyer's Gu	56	56	3	001\Chris Waterm
4	1	Produce	09/12/2002	email	[cfd] Defense wants the computers... (05)	120	120	3	001\Chris Waterm
5	1	Produce	11/06/2002	email	[cft] Job Opportunity	160	160	4	001\Chris Waterm
6	3	Produce	09/23/2002	email	[cfo] Linux - Inex question (01)	172	172	3	001\Chris Waterm
7	6	Produce	10/30/2002	email	[cfd] SWGDE Best Practices version 2 (03)	174	174	9	001\Chris Waterm
8	2	Produce	09/20/2002	email	[cfd] Running WinWare on a Forensic Analysis PC (02)	192	192	4	001\Chris Waterm
9	3	Produce	08/14/2002	email	Color Office Printers; Dvorak on WinXP Authentication; USB and Fi	202	202	3	001\Chris Waterm
10	2	Produce	08/13/2002	email	[cfo] Forensic Road - long (02)	230	230	3	001\Chris Waterm
11	4	Produce		Non-Responsive	We are exercising something we are willing to die protecting.	250	250	3	001\Chris Waterm
12	22	Produce		Privileged	thyFocus Newsletter #131	366	366	4	001\Chris Waterm
13	18			Produce	thyFocus Newsletter #126	368	368	3	001\Chris Waterm
14	1			Technical Difficulty	List of software tools... (02)	432	432	3	001\Chris Waterm
15	3	Privileged		Don't Know	Defense wants the computers... (05)	592	592	3	001\Chris Waterm
16	1			Unmark	Free Training opportunity (01)	616	616	3	001\Chris Waterm
17	1			Issues	Free Training Opportunities (01)	618	618	6	001\Chris Waterm
18	2			Add to binder...	John Smith Deposition	648	648	4	001\Chris Waterm
19	2	Privileged		Add to production...	Trial Exhibits	658	658	11	001\Chris Waterm
20	4			Flag for redaction	TPS 2 - Info of Interest on SWGDE group (03)	666	666	16	001\Chris Waterm
21	3				SWGDE Best Practices version 2 (03)	672	672	8	001\Chris Waterm
22	2	Non-Respon		Remove from Review	Extracting the data from TPS files (02)	653	653	3	001\Chris Waterm
23	5	Non-Respon		Comments...	ys Headlines from H/Times.com Monday, September 30, 200	302	302	3	001\Chris Waterm
24	5	Non-Respon		Copy value	SWGDE Draft (04)	732	732	8	001\Chris Waterm
25	3		10/27/2002	email	[cfd] I look -- for NOW LEO defense uses (01)	734	734	4	001\Chris Waterm

As shown in Figure 8.22, a binder can be produced and exported to external media with a right-click "Export Binder" command or from binder properties.

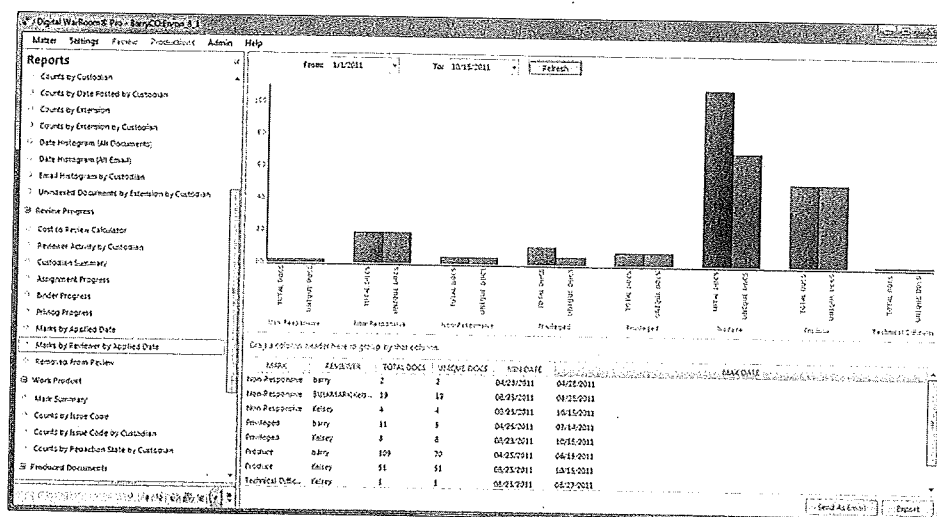
One of the most important features of DWR Pro is its extensive reporting capability. The many built-in report options provide a detailed under-

Figure 8.24



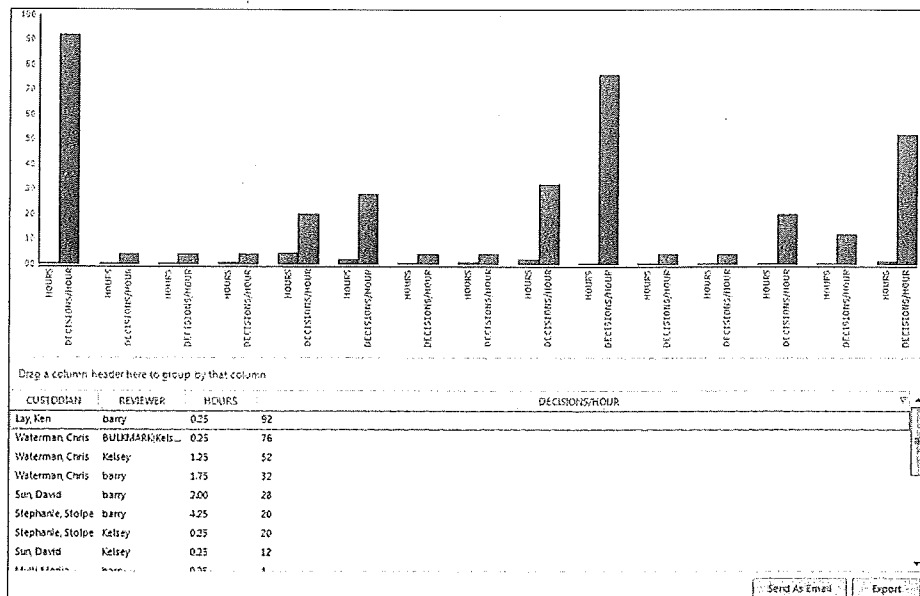
The progress of the review can even be managed using reports that include a cost-to-review calculator, a reviewer activity log, and an assignment progress log. See Figure 8.25.

Figure 8.25



Marks and tags that are made during review are summarized in a report, as shown in Figure 8.26.

Figure 8.26



Production reports can be created, including a production summary, a summary by custodian, a summary by extension, and a Bates gap analysis that is useful in making sure productions are properly Bates stamped in appropriate sequence. See Figure 8.27.

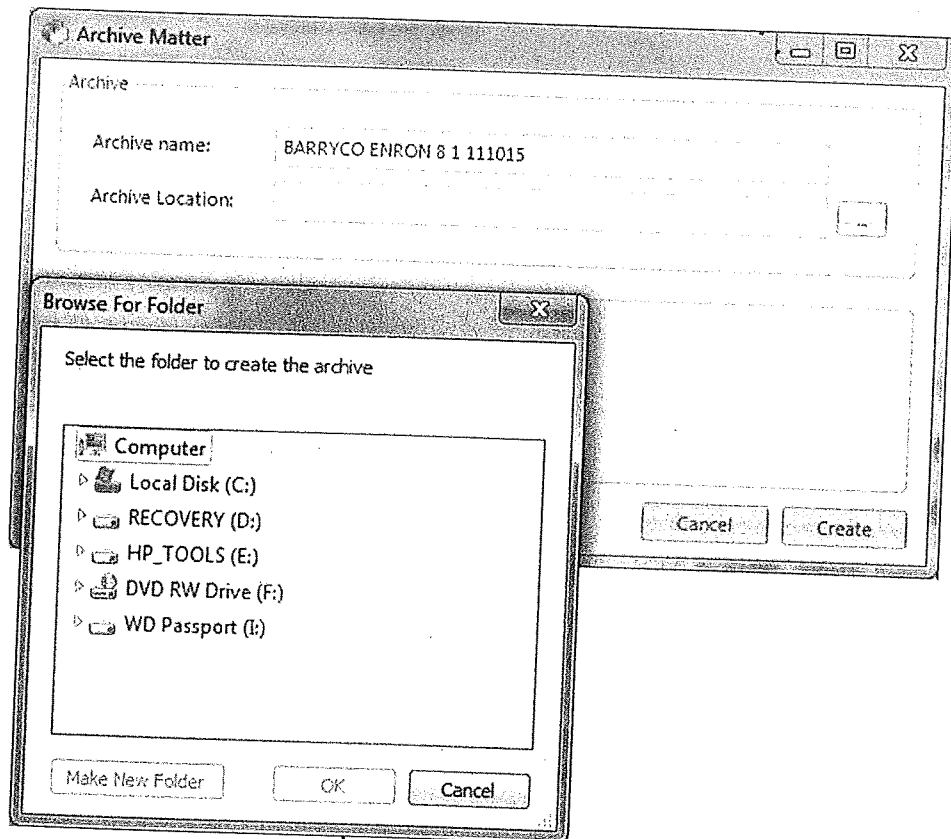
Figure 8.27

FDS	RANGE	STATE	GAP REPORT
FDS501	ENRON-00000001 to ENRON-00000033 (32 pages)	Finalized	Bates gap detected between 5 and 6 Bates gap detected between 31 and 32
FDS502	ENRON-00000034 to ENRON-00000066 (33 pages)	Draft	Bates gap detected between 36 and 66
US DOJ production 1	ENRON-00000067 to ENRON-00000082 (16 pages)	Finalized	Bates gap detected between 67 and 68
FDS507	ENRON-00000083 to ENRON-00000095 (13 pages)	Draft	No Bates gaps detected.
FDS508	ENRON-00000096 to ENRON-00000097 (2 pages)	Draft	No Bates gaps detected.
FDS509	ENRON-00000098 to ENRON-00000099 (2 pages)	Draft	No Bates gaps detected.

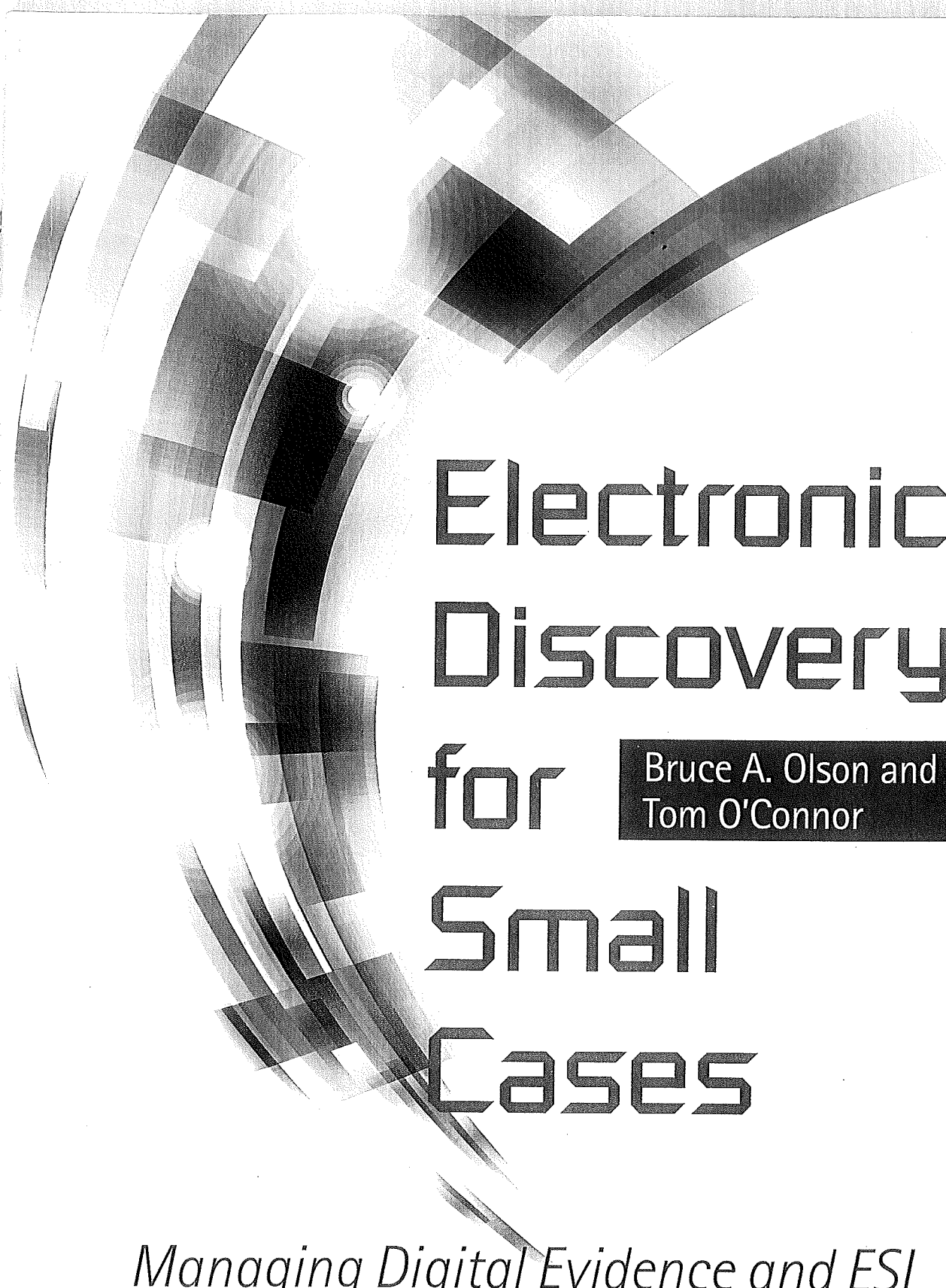
Portable Archive File

If the case grows beyond the size that a single user with DWR Pro can complete, the case transfers easily to the WorkGroup server version for multiuser teams or to DWR Pro's hosted platform. The archive file used to transfer information contains all case data plus document marks, annotations, and all other work product. Similarly, an archive file produced from a hosted or server-based matter may be stored for in-house reference and opened in DWR Pro. See Figure 8.28.

Figure 8.28



DWR Pro is an economical yet full-featured tool that can help you manage the multiple stages of the e-discovery process. It is an ideal solution for managing smaller collections of documents, although it works just as well on large cases. It can serve as your primary review tool as well as your processing tool, or you can create a load file containing a set of relevant documents to be loaded into any other litigation support or review tool you use.



Electronic Discovery for Small Cases

Bruce A. Olson and
Tom O'Connor

Managing Digital Evidence and ESI



LAW PRACTICE MANAGEMENT SECTION
MARKETING • MANAGEMENT • TECHNOLOGY • FINANCE