Are there really such things as best practices for computer forensics? Opinions vary, but there are generally accepted practices in the profession which provide terrific guidance. We’ve attempted to assemble procedures from varying practitioners considered to be the best in the industry.

What is Computer Forensics?

Let’s start at the beginning. According to Wikipedia, computer forensics is the art and science of applying computer sciences to aid the legal process.1 Think CSI with computers and other electronic media. Computer forensics is the acquisition, authentication, analysis and presentation of electronic evidence. It is deeply rooted in scientific process and generally accepted practices of the computer forensic community. From a legal perspective, it is critical that the computer forensic process and presented evidence be repeatable using various tools and that the outcome be accepted as part of a peer review process (Daubert opinion2).

Chain of Custody

Of utmost importance in the handling of any evidence is preserving the chain of custody. Hopefully, “chain of custody” is not a foreign term to you. If so, we’ll explain what chain of custody is and why it is so important.

Chain of custody pertains to the integrity and handling of evidence as it relates to legal matters.3 This means that the chain of custody will document the seizure, custody, control, transfer, analysis and disposition of evidence. The evidence may be in paper or electronic form and is subject to the same tracking and documentation irrespective of form. The primary purpose for this documentation is to verify that the evidence was not tampered with and that only authorized persons had access to the evidence.

Depending on the circumstances, each individual that had access to the evidence may have to testify as to what occurred while the evidence was in their control. The completed documentation should be chronologically recorded so as to withstand any legal challenges to the authenticity of the evidence. The documentation should show the date and time of evidence transfer, the identity of the evidence handlers and the duration of the

---

Evidence custody (this can be specifically logged or calculated from the “evidence return” time). Sometimes the documentation will also indicate the purpose of the transfer. Examples of purpose could include things such as “analysis review” or “copy made for opposing party.”

Evidence Seizure

A few words need to be said concerning the seizure of electronic evidence. For the most part, you will not be involved in actually performing a first response seizure, but the following points will help maintain the integrity of the electronic evidence and maximize retrievable data, while minimizing data destruction.

If the evidence is a computer and it is not turned on, do NOT power it up. Document the entire installation and use digital photographs if required. If the device to be seized is a running server, do an orderly shutdown using the appropriate commands for the operating system. Once the server is shut down, pull the power plug and do not power it up again. If the server contains a RAID\(^4\) (Redundant Array of Independent Disks), document the RAID configuration prior to shutting the server down. This information will be useful should the RAID array have to be virtually “reassembled” during a forensic examination.

Besides computers, consider other electronic devices and media that may contain electronic information. These can include such items as PDAs, flash memory drives, external USB hard disks, cell phones, iPods\(^5\), CD-ROMs, DVDs and other storage devices. No matter what type of electronic device is seized, make sure that the power adapters and communications cables are also seized. These power adapters may be needed to keep the device charged so as not to lose any data that may be stored in volatile memory. The communications cables are important, especially if the device utilizes a proprietary connection.

Evidence Documentation

Once you receive the evidence, it needs to be documented. There are several steps to the documentation, depending on what form the evidence is in.

1. In all cases, digital photographs should be taken of all pieces of evidence to document the original condition and configuration (in the case of computers, etc.). Be sure to photograph from multiple angles and even the inside of devices that need to be disassembled.

---


\(^5\) iPod is a device sold by Apple Computer, Inc. that stores music, video and sound files for portable playback. The use of iPod refers to any portable device similar in nature to the actual iPod itself.
2. Assign an evidence number to each piece of equipment or media. These numbers should be sequential (to help identify if there are any gaps or missing evidence) and be assigned to each unique piece of information. If there are multiple hard disks in a machine (e.g. RAID array in a server), each drive should have its own unique evidence number.

3. Document each piece of evidence by describing it and noting any unique configuration or items making up the evidence. As an example, when documenting a computer, be sure to list the form factor (e.g. server, desktop, laptop, tower computer, etc.), the manufacturer, model number, serial number and included interfaces and devices. These interfaces and devices would include such things as USB\(^6\), FireWire\(^7\), serial communications interface\(^8\), CD-ROM, DVD, CD-RW\(^9\) (for compact disc, rewritable), floppy disk, NIC\(^10\) (network interface card), wireless NIC\(^11\), integrated sound, modem\(^12\), video card and any other unit or peripheral interface.

4. Document each hard disk that is included in the computer. Documentation should include the manufacturer, model number, serial number, disk capacity and interface\(^13\) (e.g. SCSI\(^14\), IDE\(^15\), Serial ATA\(^16\), etc.). Don’t forget to take a digital picture of the disk too and make sure to capture how the disk is pinned or strapped.

5. Tag loose media with unique identifiers. As an example, you may have a stack of unlabeled floppy disks as part of your electronic data media. Since they don’t have labels, you have to find a way to uniquely identify each disk so that you can

go back to the proper piece of media should relevant information be found during the analysis. One way is to place your own label (sticky paper type) on each disk with the evidence number clearly identified. This same process can also be done for CDs that are not labeled. In either case, make sure that the documentation and chain of custody form state that the evidence was received in an unlabeled state.

**Forensic Acquisition**

Now that the electronic evidence has been “seized,” what are the next steps? A forensic image needs to be created. The forensic image is a bit-by-bit verified, duplicate of the original evidence. The following sections will identify some concerns and considerations in obtaining and processing the forensic image.

**On-site Acquisition**

Obtaining a forensic image of a hard disk on-site can be one of the slowest and most expensive of the options. The equipment transported to the field is typically less powerful than that found in the forensic examiner’s lab. This is due to the portability requirements for the equipment. As an example, most laptops are less powerful than a desktop computer even though there are some pretty powerful laptops these days. The portable computers also tend to have less RAM\(^\text{17}\) (Random Access Memory) available, which also impacts on the performance. Recently, hardware imagers (e.g. HardCopy III) specifically designed for creating forensic images are taking over the on-site role. These devices are very fast (faster than laptops) and have additional features such as logging and automatic verification. They are a dedicated device designed solely for on-site forensic imaging and perform no other function. They cost several thousands of dollars, but are a good alternative over using a laptop for acquisitions.

Another cost impact is that the examiner has to be on-site with the equipment during the entire time that the image is being generated. This may be a very time consuming process, depending on the amount of data that is being imaged. Since the examiner is “baby sitting” the process, the clock continues to run at the examiner’s hourly rate.

On-site acquisitions can be risky if the computers to be acquired are not well documented or if the components are not well know. A good forensic examiner will have a lot of adapters, software and methods in which to make a forensic image, but it is impossible to travel with every piece of equipment and software method. Make sure that the examiner is fully aware of the computer configurations and included peripherals. Consider allowing the examiner to make a “dry run” visit if you are unsure of the components in the computer(s) to be

imaged. This pre-visit can conserve valuable time and ensure a successful forensic image is possible, prior to taking the computers off-line and out of business.

Continuation of the business operation is a prime concern when obtaining forensic images. Under normal circumstances the computer that is to be imaged is taken “off-line” during the process. This can be particularly disruptive where servers are concerned, especially if there is only one server providing the specific function such as an e-mail server. Powering off the server and shipping it to a location for imaging probably isn’t practical, hence the requirement for an on-site acquisition.

Since we want to minimize business impact, speed of the acquisition is another major concern. There are several things that can be done to make the acquisitions go as fast as possible. Doing as many simultaneous acquisitions as possible will reduce the overall down time of the computers. When creating a forensic image, it is possible to compress the data so that the target storage size is a reduced size as compared to the original, however, compression adds to the acquisition time. On-site acquisitions should be done with no compression, thereby acquiring the data as fast as possible. The images can always be compressed at a later time in a controlled lab environment, where time is not a major concern. Remember that the forensic acquisition will take whatever time it takes. This is a matter of physics and you cannot not rush or accelerate the process beyond what physics allows. Once started it must be allowed to go to completion.

The discussion of on-site forensic acquisitions has revolved around shutting down computers and removing them from service. This is the normal way of achieving an on-site forensic image. However, there are now software methods that allow for a forensic image to be obtained on a live, operating system. A piece of software called a servlet18 or agent19 is installed on the machine to be acquired. This servlet/agent acts as a “hook” to the system, allowing a forensic image to be taken of a “live” system. It is noted that this type of forensic acquisition is a much more complicated and costly method. Additional hardware and software are required to accomplish the acquisition and licensing costs begin in the hundreds of thousands of dollars.

A recent technological development by one vendor allows the “live” acquisition to take place in a more cost effective and less intrusive manner. A server is installed at the client site to act as the repository for the forensic images. The data from the imaged machine is cached so that the machine does not come to a grinding halt due to the processor overhead. This allows for a more covert

acquisition, where the user doesn’t even know that a forensic image is being created. It is anticipated that the cost for this solution will begin in the tens of thousands of dollars.

Finally, several vendors now provide the ability to create logical evidence files in a forensic manner. This is more a method of preserving the logical files and does not truly capture the unallocated spaces. The files dates, times, attributes, etc. are preserved. While logic evidence files can be acquired from a running system, they will not correctly preserve any active databases such as that used by Exchange. The Exchange database must be dismounted in order to properly preserve it using this method, otherwise the database will be corrupted when trying to work with the logical evidence file.
**Off-site Acquisition**

This type of acquisition is done in the examiner's lab. The forensic lab contains equipment that is much more powerful and flexible as compared to the mobile field equipment. Also, multiple drive adapters are normally available to handle the various types of hard disk interfaces. As an example, a single computer may have IDE, SCSI and SATA\(^{20}\) hard drives. This would require a different adapter to deal with the SATA drive and a different process to acquire the SCSI drive, all of which is easily accomplished in a lab environment.

The other advantage of off-site acquisitions is the reduction in cost. There is no need to "baby sit" the acquisition process once it is started in a lab environment. The lab can be secured and the acquisition left to run to completion. Normally, this means a much lower cost since the examiner can go off and work on another case while your evidence is being acquired. As a result, many computer forensic companies will charge a flat fee for each hard drive acquisition and not a time-based fee as they would for on-site work.

**Image Copies**

On occasion, you may be able to save even more money by getting a copy of the forensic image from the other party. This saves time and money, especially since the forensic image can be validated and a proper chain of custody maintained. So what do you ask for if requesting a copy of a forensic image. There are several formats that are typically requested. One image format is an EnCase\(^{21}\) evidence file format. EnCase is a computer forensic software application developed by Guidance Software (http://www.guidancesoftware.com). The EnCase file format is a proprietary format and is commonly used by commercial and law enforcement entities. There are third party products that will generate EnCase evidence files, but there usage is not recommended. The third parties have reversed engineered the EnCase file format and have not directly licensed usage. The generation of EnCase files by a non-Guidance Software product just invites a court challenge to the authenticity and accuracy of the data.

Another file format is that which is produced by the SafeBack\(^{22}\) software, which is a product of New Technologies, Inc. (NTI), a subsidiary of Armor Holdings, Inc. The SafeBack file is also a proprietary format and is not as commonly used. Version 2.x of the SafeBack form is supported by a large number of computer forensic software packages. However, since the introduction of version 3.x, few in the industry have upgraded or continue to use SafeBack. Hence, SafeBack files are rarely seen except for much older, legacy cases.


A third and very common evidence file format is a bit-stream format. This is a non-compressed complete bit-by-bit image of the original evidence. The files are also known as ‘dd’ images. This is because the ‘dd’\textsuperscript{23} (data dump) command of Unix is used to create the files. These days Linux is a very common operating system used to generate a dd image. The dd image is also known as a raw image. The hardware-based forensic imagers typically output a raw image file.

An open source product that is used to create a variant of the ‘dd’ files is called ‘dcfldd’ and is available at http://dcfldd.sourceforge.net. Dcfldd is an enhanced GNU ‘dd’ with features useful for security and forensics.\textsuperscript{24} Some of the useful features include the ability to do on-the-fly hashing, a progress bar of data acquired, verification of the data and the ability to split the bit-stream file into smaller chunks. The ability to split the data file into multiple smaller files means that they can be copied to alternate media such as CDs or DVDs. The military investigation units tend to use ‘dcfldd’ as their standard imaging software.

So what does all of this mean? It means that if someone has already created a forensic image, you don’t have to go through the work all over again and can merely get a copy of the evidence files. It is not uncommon for an examiner to copy EnCase evidence files to a hard disk and give them to the other side or for the FBI to copy dd images to a hard disk to give to the defense expert. This saves a tremendous amount of money since forensic images do not need to be reacquired. When receiving copies of evidence files, make sure you also obtain the hash values for the original drives or other acquired media so that you can validate the evidence.

**Hash Sets**

How do you definitively identify data or specific information contained in your electronic evidence? One method is to use a process called hashing\textsuperscript{25}. Hashing is the transformation of characters into a shorter fixed-length value. These values facilitate a much faster searching than to interrogate the original, longer length data. The shortened hashed value can be viewed as an index to the original data.

You can create a hash set (or a single hash value) for any desired data. As an example, a hash set could include the values for all the operating system files included in a specific version of Windows. Obviously, you want to hash only the files that do not change and not those that are modified as a result of usage. Once you have identified the non-changing Windows system files, a hash set can be created to compare to future electronic evidence files.

Normally, you are not concerned with operating system files. User data files are of much more import during an analysis. Therefore, if you can identify the operating system files and filter them out, the amount of data to analyze is reduced. This speeds up any additional searching and analysis.

A good source for operating system file hash sets and hash sets for specific software applications is the National Software Reference Library (NSRL) project. The NSRL is designed to collect software from various sources and incorporate them into a Reference Data Set (RDS). Currently, the NSRL hash files are delivered on four (4) CDs and provide over 16 million unique SHA-1, MD5 and CRC32 values for over 52 million files. The files are in a compressed form and will expand to several gigabytes in size.

Another purpose for creating a hash set is to specifically identify user data. As an example, you could hash the files that represent specific company data. This database of hash values could then be used to compare against the opponents computer systems to determine if your company data exists on their machines.

Finally, the hashing process is also used to authenticate and validate a forensically acquired image. Typically, a value called an MD5 is used to calculate a value for the original evidence and compared to the forensic image that is acquired. The MD5 is an algorithm that is used to calculate a 128-bit value for the specific data being hashed. In the case of forensic images, the MD5 is calculated for every sector of the original hard disk. Once the forensic image is created, another MD5 is calculated for every sector of the imaged evidence. Both MD5 values should be identical, which indicates a forensically sound image. The MD5 value is also called the “digital fingerprint” for the data being hashed. As previously mentioned, you want to obtain the hash value for the original evidence if you are receiving copies of evidence. This is so you can validate that the copy is a true and accurate representation of the originating evidence.

**Child Pornography**

It is very unfortunate that we have to be concerned with the existence of child pornography when considering electronic evidence. Technically, the mere possession of one image of child pornography is a violation of Federal law and a criminal offense. When acquiring any electronic evidence, it must be scanned for

---

27 id
the existence of child pornography. There are two ways to scan the evidence for child pornography. One is to use a database of known hash values for child pornography. There are several sources for the hash sets, but there is limited access to them for the private sector. The major source for hash set information regarding child pornography is NCMEC\(^{29}\) (National Center for Missing and Exploited Children), which is restricted to law enforcement only.

The electronic evidence is compared to any of the child pornography hash sets and the MD5 hashes are reviewed to see if there are any positive hits. A hit is where the MD5 hash of a file matches a known value for child pornography. A positive match of an MD5 value means that the file is known child pornography or suspected child pornography and must be reported to the authorities.

The second method is to visually scan each image and video file. This is a very time consuming process. Several of the computer forensic analysis software applications provide the ability to place the files into a thumbnail view so that multiple image files are presented in a gallery fashion. This allows for viewing multiple files at one time. It is still time consuming as a single hard drive can contain tens of thousands of image files. The video files (e.g. MOV, AVI, MPEG, etc.) must also be reviewed individually as they could also contain child pornography.

**Virus/Spyware Scan**

Another preliminary step in the forensic acquisition process is to scan the evidence for the presence of viruses, worms, Trojans, spyware, keystroke loggers, etc. This scan process will help identify if the subject’s computer was possibly being controlled by an external entity or if the computer activity may be due to other causes not user generated. If you are working with the original evidence, the simplest way to accomplish the scan is to leave it connected to the write blocking device and launch a scan against the evidence. Normally, there are two types of scans and appropriate software that are used to accomplish these tasks.

Configure your antivirus software to do a log only scan and select the target as the evidence drive. Once the scan is completed, save the log file to document any existence of viruses, worms, Trojans, etc. that may exist. A separate scan is accomplished using anti-spyware software. Configure the software to do a logging only scan and save the log file after the scan completes. Recent versions of antivirus software also contain the ability to scan for malware so a separate scan would not be required.

If you do not have the original evidence drive, then you will need to perform the scans using a different method. Some forensic software applications allow you to mount the evidence in a read-only mode and present the data to your analysis machine as if it is another disk drive. In this way, you merely select the

appropriate disk drive as the target for the scans. Another option would be to load the evidence into a virtual environment using a product such as VMware Workstation.\(^3\) This product allows you to load your evidence into a virtual area of your analysis machine, thereby observing the subjects computer exactly as they would. Not all operating systems are supported in virtual machines so this may not be an option for some pieces of evidence.

**Write Blockers**

As previously mentioned, the use of a write blocker is critical to maintaining data integrity of the evidence. A fundamental requirement of computer forensics is that original data is not modified in any way from its original condition. Write blockers maintain this data integrity. There are several ways to write block the original evidence.

One of the simplest ways is through the use of hardware devices, specifically designed to prevent writing to the original evidence. Just connecting an evidence drive to a Windows-based computer will modify the original contents. It may be something as simple as creating a temporary file on the drive to scanning the entire contents for viruses and changing all the file access times. There are several vendors of hardware write blockers. Guidance Software provides their FastBloc hardware write blockers in several forms.\(^3\) One model is intended to be permanently installed in a desktop computer and the other is a portable unit with FireWire and USB connectivity. The advantage of the FastBloc hardware over the other vendors is the integration with the EnCase forensic software, which is the most popular forensic analysis software in use today. Using FastBloc with EnCase will show that the write blocker was used when the EnCase reports are generated. Using other vendors with EnCase will protect the evidence, but the existence of the write block usage will not show up in any EnCase report unless manually documented, whereas FastBloc will automatically identify itself in the reports.

Another popular write block device is the FireFly\(^3\) by Digital Intelligence. They provide one model that will write block IDE drives and one model that will write block SATA drives. SATA drives are pretty much the norm these days on new computers. As technology changes so must the methods for protecting the original evidence from modification.

Tableau\(^3\) manufactures forensic bridges, which provide read only and read/write capability for a number of drive interfaces. They are very popular and have native

---

SATA and IDE interfaces. They are very flexible and can connect to your computer via eSATA, FireWire 800, FireWire 400 or USB connection.

Wiebetech\textsuperscript{34} also produces several hardware write blocking systems that are used in the industry. Their field kit is extremely popular as it contains all sorts of adapters to deal with the various types of drive interfaces that may be encountered in the field.

Write blocking can also be achieved through software means. Guidance Software has an EnCase add-on module for achieving write blocking via software means. The FastBloc Software Edition\textsuperscript{35} protects the original evidence when connected to specifically supported interface cards. There is an additional charge for this module and it requires having the EnCase license too. Another software write blocker from ForensicSoft, Inc.\textsuperscript{36} (SAFE Block XP) is available and does not require any additional licenses or products.

Finally, a large number of forensic examiners are achieving write blocking through the use of Linux. Linux is used to manually mount the evidence in a read-only mode. Following the read-only mount, the dd function is used to generate the forensic image. Linux is becoming very popular for imaging as it is a low or no cost solution, however, the command line operation for forensic acquisitions makes its usage more complicated and better suited for the technically capable.

In addition to the dd function of Linux, there is an open source program called dcfldd and is available at http://dcfldd.sourceforge.net/ and was discussed in the image file format section.

No matter which type of write blocking you choose (hardware, software, Linux, etc.) you should periodically test and verify the effectiveness of the write blocking method. As an example, perhaps a new version of the software write blocking module has a bug when trying to protect a disk connected to a specific vendor’s interface card and actually allows writing to the drive. So how would you test your write blocking methods? The easiest way is to use a hashing methodology that was previously mentioned. Connect your test drive to the write blocker and hash the drive to generate a hash value. Attempt to modify the drive contents while connected to the write blocker. Shut down the entire system and power it back up again. Re-hash the exact same number of sectors on the test drive as you did originally. The hash values better match, which indicates that the write blocking mechanism is working properly. Don’t simply rely on the manufacturer’s statement that the device is properly protecting the evidence. No technology is 100% accurate 100% of the time. Periodic testing will increase your

confidence factor and ensure that you are not modifying the original evidence in any way.

Forensic Software

There are several manufacturers of forensic analysis software. It would be impossible to list every piece of software available. We will identify the more popular packages that you may come across or hear about as you deal with electronic evidence.

As previously mentioned, EnCase is the most popular and widely used of the computer forensic software applications. It is commonly used by law enforcement and private sector examiners alike. EnCase is a complicated software package and requires training and experience to harness its power as a forensic tool. EnCase is a Windows-based software package as are many of the other vendor’s products. There are add-on modules37 for EnCase that allow enhanced functions such as the Virtual File System (VFS), Physical Disk Emulator (PDE), Decryption System (EDS), FastBloc SE (Software Edition) and CD/DVD Module.

A competitor to EnCase and gaining in popularity is Forensic Toolkit (FTK) by Access Data. There are two major features of FTK that are causing its increase in popularity. The first is cost. FTK is much less expensive as compared to EnCase. The second popular feature is speed. FTK uses a pre-processing step that indexes the electronic evidence into various groupings. This indexing helps speed the searching of the data during the analysis phase. FTK is also known for its e-mail handling capability, however, other vendors are improving their ability and individual tools for specific e-mail stores is also used. In response to the speed of searches in FTK, Guidance Software has provided an indexing process in version 6 of EnCase.

X-Ways Forensic38 by X-Ways Software Technology AG is even less expensive than FTK. X-Ways is used by some examiners, primarily as a validation tool. Like FTK and EnCase, X-Ways Forensic is a Windows-based product. Its presentation and interface is similar to other forensic applications. X-Ways handles some forms of e-mail much better than FTK. The major advantage is the ability to preserve the linkage of an e-mail message to the body contents. It does require some time to pre-process in order to maintain this relationship, but is well worth it in the long run.

The use of EnCase is beginning to fall out of favor within the computer forensics community. Guidance Software appears to be targeting very large organizations and companies for their products, abandoning those loyal EnCase users of the early days. Access Data has had

---

trouble with their introduction of FTK 2.x and examiners are steering away from that product line as well. The hardware requirements alone make an FTK 2.x installation very impractical for most examiners. Those with FTK 1.x licenses are not upgrading or investing in further EnCase versions. Instead, we are seeing a migration among computer forensic examiners to the X-Ways product line.

No single piece of software or hardware will address all of the situations encountered with electronic evidence. Many different tools are needed to effectively analyze the electronic evidence of today. As an example, a specialized tool may be needed to analyze the data on a multisession CD. CD/DVD Inspector is an example of a special tool for analyzing optical media. The increased use of cellular phones is also requiring special software to deal with text messages, call logs and phone book entries on a cell phone. One popular package for analyzing cell phones is Device Seizure by Paraben Forensics. Another challenge with cell phones is the large number of power supplies and data interconnection cables that are required for interfacing the phone to the analysis computer. Prior to expending any significant time and expense in dealing with cell phone forensics, make sure that the target cell phone is supported by the analysis software, that a power source is available and that a data cable is available for the phone.

Besides the core analytical software used to investigate the electronic evidence, other software may be needed to deal with specific aspects of the data. Software may be needed to view specific image file formats if it is not supported in the forensic software package. As an example, the forensic software may not be able to view the image format used with AOL’s software; therefore the images will have to be viewed outside of the forensic software using an application that can view AOL images. Perhaps the relevant e-mail messages are accessed using the Lotus Notes application, which is not supported by the forensic analysis software. Again, a separate software package would be needed to analyze the Notes information outside of the forensic software application. This is not an unusual situation as no single software tool can handle all of the formats of electronic evidence you will encounter.

**Computer Forensic Consultants**

What do you look for when selecting a computer forensic consultant or firm? Probably the most effective method is through referrals. What better way to learn about a company and its personnel than through someone that has already utilized their services? You should also review the CVs of the experts. They should have technical certifications such as CNE (Certified Novell Examiner), MCSE (Microsoft Certified Systems Engineer), CCNA (Certified Cisco Network Administrator) and the like. In addition to the technical certifications, look for computer forensic specific certifications. As of this writing the

---


EnCE (EnCase Certified Examiner) and CCE (Certified Computer Examiner) are the two most prevalent and respected certifications.

Besides actual credentials, see if the personnel has ever testified and/or been qualified as an expert in computer forensics. You probably won’t feel very comfortable getting a “rookie” technologist that has never qualified. If the forensic technologist is going to testify or be deposed, can they take complex electronic evidence concepts and present them with a clear English description? The inability to present simple and clear analogies to a jury or judge is a huge anchor that can sink your case in a heart beat.

Have they been published? Do they speak on computer forensics? If so, their testimony is likely to be seen as more credible by a judge or jury. Make sure you speak with the expert yourself and ask a few questions to get a sense of whether you can work comfortably with this person. Don’t hesitate to “grill” them a little – a true expert won’t mind in the least – and you’ll know if they will be comfortable make on the fly responses from the witness stand.

**Evidence Sampling**

So how do you contain costs and maximize your chances of relevant evidence “harvesting?” All too often, forensic companies are advising clients to acquire all of the electronic information available. This is an extremely costly endeavor. It would be far more cost effective to only acquire the computers and/or media that contained relevant information, but how do you know which computers to acquire? A technique called sampling can be used to determine where the relevant information may reside. Select several computers for some of the key players in the case and forensically acquire them. Analyze the “subset” to determine if there is enough data to support your claims. Choosing to first analyze a smaller amount of data is less expensive and can quickly determine if relevant data may exist in electronic form.

Perhaps a real world example can better explain sampling and the resulting effect.

*Case Scenario*:  
The defendant is accused of continued use of plaintiff’s proprietary database application following a fallout in the contract negotiations. The defendant claims to have deleted or returned all of the data associated with the database application. The plaintiff contends otherwise. The plaintiff desires to perform a forensic acquisition of all the defendant’s computers in order to determine the existence of the proprietary data. The defendant contends that performing the acquisitions is overly burdensome and would significantly impact their business operation,

---

effectively shutting them down for several days. The judge allows for a sampling of the main server only for up to six hours over an agreed upon weekend.

Prior to actually visiting the defendant’s site for the sampling effort, a hash set of the plaintiff’s database application was created. The hash set represents a digital value for each file comprising the database application, effectively presenting a “digital fingerprint” for each piece of digital data contained in the plaintiff’s application. While on-site at the defendant’s place of business, the hash values for files on the server were compared to the hash set, which represented the plaintiff’s proprietary database application. There were over 900 positive hits, which represents that files from the plaintiff’s database application were still present on the defendant’s server.

The above referenced case is illustrative of a practical sampling technique. This process saved a significant amount of money since only one computer (a server) was sampled instead of every one of the defendant’s computers. This is similar to the sampling that occurred in Zubulake v. UBS Warburg, where Judge Scheindlin ordered the restoration of a small number of backup tapes to determine if any relevant e-mails would be found on the tapes.

**Searching the Evidence**

At the heart of evidence analysis is the function of searching. Searching can mean trying to find data that exists in a number of different forms. The desired data is typically found in the form of documents (word processing, spreadsheets, etc.), financial information, images (pictures, movies, etc.), messages (e-mail, IM, chat rooms, etc.) and artifacts of electronic data.

The search methods vary depending upon the data being searched. Keyword searches are the most common form of search criteria. The term keyword is used generically to identify words or phrases. In the simplest form, a single word is used to search against the evidence. The data is then scanned for the existence of the word selected in the search. As an example, selecting for the word Frank will return the word and all of its various forms. So ‘frankly’ and ‘Frankfurt’ would also be identified as search hits. Remember that a hit is a positive result where the search term exists.

Selecting common terms for search words will return thousands and thousands of hits, making it almost impossible to review the search results in an efficient manner. Therefore, it is important to select appropriate search terms and/or use different searching methods to maximize the amount of relevant data that may be returned. A first step is to select very narrow search terms or phrases so that a lot of “noise” data is not return. The problem with simple searches is that all variants of the selected term are returned as was previously described. A secondary problem is that an EXACT match is required to return a search hit. So why is this a problem? Not everyone types or enters
data correctly on every occasion. If you are searching for the keyword ‘shipper’ you won’t find a document or e-mail message if someone typed it as shopper (one letter away on a QWERTY\textsuperscript{43} keyboard). The key document to crack open your case may have just been missed.

If you use simple search methods (exact matches) what can you do to return more pertinent results? Avoiding common terms has already been discussed. Perhaps making the keyword case sensitive would help reduce the “noise” hits too. Another “trick” is to search on the exact keyword and not accept the various forms of the word.

Some search engines are capable of “fuzzy” searches. Fuzzy searching\textsuperscript{44} promises to find results for misspelled and similar words. The fuzzy search engines work on similarities, but we’re still beholden to the specific logic that the computer programmer built into the engine. This “fuzzy” logic may not be what you intended or desire, but you have no choice but to accept the results. Some of the more sophisticated search methods contain “artificial intelligence” and claim to “learn” from the user’s actions. In these cases, a very large amount of data is needed to even get something close to acceptable results. The user enters a keyword and views the results. If the user “discards” all of the results then the search engine “knows” to try something else on the next round. If the user selects some of the results as relevant, the search engine factors in the selected characteristics for the next round of searching. Like the fuzzy search logic, the “artificial intelligence” engine results are the output of the computer program logic and how to handle the user inputs.

Rather than merely searching for specific words, the use of phrases can significantly reduce the amount of false positives returned. In addition, the use of Boolean\textsuperscript{45} searches helps refine the search results. A Boolean search is where you use multiple keywords and conditionals. As an example, you may define a Boolean search for “General Electric and 747” hoping to get documents relating to the jet engines manufactured by GE that are used on a Boeing 747 aircraft. Boolean searches can be as simple or as complex as you want. Many search engines allow for Boolean operators such as AND, OR, NOT or W/7 (within 7 words). Make sure you are familiar with the search string operators for your particular application and/or search engine. How does the software handle wildcard values? Are hyphens indexed or ignored?

When doing the searches, it is important to understand the form of the data and how it is stored on the computer media. Not all data is stored in clear text so you may have to create your keyword searches using hexadecimal values or some other form. One very common form of writing the data to disk is using Unicode\textsuperscript{46}. Unicode provides a method to uniquely represent every character no matter what the platform, program or language. As an example, the capital letter L is represented as a hexadecimal 004C in the Basic

Latin language. This would be true for multiple operating systems (e.g. Windows NT, Windows 2000, Sun Solaris, Mac OS 9.2, Mac OS X 10.1, etc.) and multiple programming languages, which is consistent with the Unicode standard. Therefore, you need to define your keyword search using Unicode characters when working in a Unicode environment. The good news is that you don’t have to know the entire Unicode table or even look it up. The search engine will provide a selection for Unicode and do the hexadecimal conversion for you.

Perhaps your search involves a foreign language in addition to English. Again, you need to structure your keywords according to the foreign language spelling for the words you desire. This can increase the review process, especially if you are search for the same keyword in multiple languages. Fortunately, most cases do not involve foreign languages so the keywords should be fairly straight forward.

As previously mentioned, you need to know how the target data is stored on the computer. We have already discussed Unicode, which is going to be the most common form of storage for clear text data. What about search pictures? Obviously you cannot do a keyword search for data contained in an image. Fortunately, a large number of the common image file formats contain a pre-defined header within the file. This means that each file starts with the same hexadecimal value. As an example, A JPG file will have the values of JFIF in byte position 6 through 9. This makes it easy to search for JPG files since their header is uniquely defined. Not all files have unique header information so determining the file type is nearly impossible. Also, when searching through electronic evidence, especially unallocated space\(^{47}\) (disk area that is available for overwriting that contains data from previous usage), for files with known header information, it is rare that there will be a clearly defined footer so that you can determine where the file ends. What this means in the practical sense is that you normally only get portions of files when searching through the unallocated space.

Perhaps the data is stored in a compressed format. Searching this type of data requires that you know the compression algorithm in order to determine if there is any relationship to the stored hexadecimal characters and the clear text that you are looking for. Normally, you would use the native application and view the files in their native environment or export the data to a different format (e.g. text file) so that it can be searched more easily. There are exceptions to this conversion requirement. Some searching software is programmed to “understand” or decode the file contents of some of the more popular applications. (e.g. Word, Excel, Outlook PST, PDF, etc.) This is extremely valuable when reviewing the electronic evidence. In addition, we are beginning to see products that can search compressed and compound files (e.g. ZIP). If you are using an electronic evidence management software package such as Summation or Concordance, make sure you know which file formats are supported. You may need to have your forensic technologist convert the data prior to your review. Probably one of the most common conversions is to take the Exchange mail server data (.EDB files) and convert them to .PST\(^{48}\) files. A .PST file is a personal folder file and easily viewed using Outlook.


Encryption is a killer for computer forensics and its impact on searching is equally deadly. You cannot search encrypted data in its native form. This means that all of those encrypted e-mail messages will not “bubble up” with even the narrowest of keywords. The e-mail messages must be decrypted prior to running the search. This is another step in the evidence processing and will increase the cost of analysis and review. Pretty Good Privacy (PGP) is a very popular encryption application and is used for encrypting e-mail messages among other things. There is no way to bulk decrypt the messages without writing specialized software or scripting routine. Needless to say, perhaps dealing with the unencrypted messages first will reveal the information needed for the case without having to go through the time and expense of decrypting messages.

Searching a database is also not a straightforward process. Even if the data is stored in clear text, what is the data relationship to other data in the database record? Products such as the popular dtSearch\(^\text{49}\) have APIs\(^\text{50}\) or other methods to facilitate the searching of databases. Effectively, the search engine is reading the data from each record of the database and adding relevant data to its index. Again, make sure you know which database formats and versions are supported by the search engine or the searching ability within your electronic document review application. Custom queries may have to be run against the database to extract the desired information in a report format. Searching through databases is specific to the database structure and type. It is best to see if relevant information is available in other places that are more easily accessible before trying to search and extract data from databases.

Finally, there is another search methodology known as concept searching. This type of search returns documents based upon what they represent rather than the raw keywords themselves. As an example, a concept search for “GPS” should return results related to “Global Positioning System,” “Satellite Navigation” and “NAVSTAR.” Concept searching is the latest “rave” among the electronic data discovery companies. Like the other innovative searching techniques, concept searching is dependent on the computer programmer(s) and the logic placed in the search engine. As a result, there is a tendency to miss relevant data. One very valuable use of concept searching is to help determine what “raw” keywords can be used for traditional searches. This leaves the conceptual searching technique as a supplement search methodology and not as a data filtering technique. Depending on the concept search to filter data is a dangerous approach. As the technology improves, searching results will improve, but leave the concept searching to the outside of your core evidence review.

**Forensic Reports**


At some point during the handling of electronic evidence you will have the opportunity to see a computer forensic report. When involved in criminal defense work, the forensic report will be generated by law enforcement. The quality and contents of these reports varies widely and is not dependent on the source. There are good and bad computer forensic reports generated by the private sector and law enforcement too. There are times when you will not want the results of the computer forensic examination committed to paper. After all…the contents of the report may be discoverable and not protected by any client/attorney privilege or work product. Assuming that a formal report is requested or required, there are certain elements that should be present in the report.

1. **Evidence Documentation** – A section of the report should identify the evidence that is being analyzed. The manufacture, model number and serial number should be listed. The hash value (typically MD5) of the original and of the forensically duplicated evidence should be shown. Although not necessary, it is helpful to have a listing of the files and folder structure prior to any attempt to recover deleted information.

2. **Scan Results** – The results of the virus and spyware scans are to be documented. This information may show the existence of keystroke loggers or a Trojan horse, which could impact whether the user had knowledge of certain events occurring on the computer.

3. **Date & Time Validation** – The date and time setting for the computer at the time of forensic acquisition is to be documented. It is amazing how many forensic reports are missing this very critical piece of information. The computer clock setting is important as it establishes a baseline for the computer activity. If the computer clock is not accurate then the examiner will make incorrect conclusions as to when certain events occurred on the computer. Don’t forget the time zone setting as well. The analysis may have to account to time variations due to the configured time zone.

4. **Analysis Results** – This is the meat of the report. This section will have statements of fact that are supported by the electronic evidence. As an example, there may be a statement that a particular software application was installed on the computer at a particular time. There may be Windows registry entries, file creation dates and a log file that all support the statement about the installation. When specific files or data are being referenced in the report, the relevant metadata should be shown along with the starting physical sector where the information can be found. This physical sector will help any other examiner find the same information since computer forensics is a repeatable science. The relevant metadata is typically the Windows file attributes like creation date, last modified date, last accessed date and last written date for NTFS formatted drives. It makes no sense to try and
present any dates associated with data coming from the unallocated sectors. No date attributes are available for information in the unallocated sectors, but many a rookie examiner will try to tag dates since that’s what they do for all files identified in the report.

5. **Summary/Conclusion** – A summary or conclusion paragraph should identify the primary findings of the report. The really bad computer forensic reports are nothing more than a regurgitation of the evidence. There is just a listing of files with the file attributes and no conclusion as to the significance of the files.

Reports may be on paper or optical media. Lawyers love paper and are always asking to have their electronic evidence printed. Unfortunately, all but the simplest computer forensic reports do not lend themselves well to paper. Generally the report will be provided on CD in a word processing document or in HTML\(^{51}\) code to be accessed with a web browser. The nice feature with getting the report on CD is that you can get access to the actual file that is being referenced in the report. Typically, the report will have hyperlinks to the actual file. This means the native file is available for viewing with whatever software the reviewer desires.

When dealing with child pornography cases, images, movies, etc. will not be included in the report as that could constitute redistribution. The file names will be listed, but no image representation can be committed to paper or optical media. Even though there may be special circumstances, it is always better to proceed with caution and get a court order that outlines the specific procedures and abilities when dealing with child pornography.

**Duplicates**

A large amount of the information will be duplicated when dealing with electronic evidence. The human tendency is to keep multiple copies of the same information since it is so easy to store electronically. The multiple instances of the same information significantly adds to the review time and cost. A major decision is going to be whether duplicates are undesirable or if they may indicate significance to your case. Most electronic data discovery companies insist that duplicates are bad and are proud of their various methods for dealing with the duplicate data.

But what exactly is a duplicate and how do you deal with them? Exact duplicates are the easiest to deal with and discover. Remember the hashing algorithms discussed in a prior section? Exact duplicates can be discovered by comparing the hash values for the files. Matching hash values means that the data is identical even if the file name is different. Would you consider this to be a duplicate? Perhaps the same file (as determined by hash value) is in two different locations on a server. You may consider the data to be duplicative since you are only interested in the contents of the file and not the location. However, what if it is important as to where the file was stored on the server. This could

indicate access by someone not normally authorized to use the data. In this case, you would not want to discard the data as a duplicate.

Sometimes we need to classify data as duplicative when it is actually a “near” duplicate. Hash values will not help when trying to cull out “near” duplicates as hashes are too exacting. The concept of “near” duplicates is really concerned with the data contents and not the formatting or metadata associated with the file. This process is normally applied to e-mail messages. As an example, an e-mail message from one person to three other people could exist in four different places. The first would be in the sent items of the originator and perhaps in the inbox of each of the intended recipients. Technically, each message is different since the e-mail header information (routing, recipient, message ID, etc.) will be different for each of the recipients messages. However, you may consider the three recipient messages to be a duplicate of the origin message since the body contents are identical. From the origin message you can determine the recipient e-mail addresses and the contents of the message, therefore, you don’t need to process the other three messages.

One of the steps in reviewing electronic evidence is to address the handling of duplicate and unneeded information. This preprocessing of the data can be done through hashing or other vendor specific software packages. One of the most well known products for dealing with duplicate data is Discovery Cracker by CT Summation and can export to litigation support packages such as Concordance and Summation. No matter which product you use, make sure that you understand what constitutes a duplicate and how the software may be adjusted to include or exclude certain types of information. As with any software processing, you are beholden to the operation of the program logic so make sure you or someone on your team understands how the product works and is prepared to explain to the court why certain data may have been missed or is included.

**Format of Production**

The great debate for the coming years is the issue of format of production. Should you produce the electronically stored information (ESI) in native form or convert it to some other format? Generally, most experts are advising to produce the data in native format.

So what are the choices for formats and why one over the other? For years electronic data discovery companies produced electronic evidence in a TIFF format. TIFF can be considered to be analogous to a digital photo copy. It is an electronic picture of the file contents. As an example, a six page Word document that is converted to TIFF would

---

have six images (or pictures) contained in the file. The advantage of TIFF is that you
don’t need any special software to view the contents. TIFF also makes it very easy to
Bates stamp since each “page” is made up of a single image. So why not produce all of
the electronic data in TIFF format? The primary reason is that a TIFF file is not
searchable. Pictures don’t carry any text to search upon. You would have to perform
OCR\textsuperscript{55} on the TIFF image in order to convert any text in the picture to something
searchable. Another advantage of TIFF is that it strips all of the metadata from the
electronic file. You may elect to produce in TIFF so that the metadata isn’t given to the
other side. Another consideration is cost. EDD companies typically charge on a per page
basis when doing TIFF conversions. This can become quite costly for your case.

An alternative to TIFF production is to convert the files to PDF\textsuperscript{56} format. The pages of a
PDF document can be Bates stamped relatively easily and there is the added bonus of
search ability if the files are converted properly. PDF can convert the pages of a
document to an image (picture), which is very similar to TIFF.

Many EDD vendors are now “migrating” to producing data in TIFF form with metadata
load files and OCR text files in response to the request for native production. The load
files are special files used by Summation and Concordance that contain references to the
actual file. The OCR text files provide a method for searching the TIFF. Think of the
OCR text file as a plain text version of the original document. The metadata load files are
used to restrict the amount and type of metadata that is handed over in production. As
you may imagine from all of this processing, the cost can be very high. This is another
reason why there is such an increased desire for native file production.

When dealing with computer forensics as part of the production phase, there needs to be
an agreement on how to deal with any discoverable information that may reside in the
unallocated space. Remember that unallocated space is the “Wild Wild West” of disk
storage and is totally unstructured. The data is not normally in clear text and consists of
readable characters along with non-printable hexadecimal values. Normally the data will
be manually “carved” (extracted) a predefined amount either side of the “hit” value when
producing something potentially responsive from the unallocated space. As an example,
if the “hit” value where ‘pediatrician’ then perhaps data 80 bytes before and 80 bytes
after ‘pediatrician’ would be extracted as possibly relevant data.

\textsuperscript{55} OCR, SearchSMB.com, 1 May 2003,
\textsuperscript{56} PDF, whatis.com, 10 November 2006,
<http://whatis.techtarget.com/definition/0,289893,sid9_gci214288,00.html>, (26 November 2006).
Checklist for Electronic Media Device Evaluations

Although the exact methodology for the evaluation and analysis of electronic media devices is obviously sophisticated, this general checklist conveys the major components in the identification, isolation, evaluation, and preservation of electronic evidence in a standardized way that will be admissible in court.

Record each media device with a unique identifying number

Write protect each media device

Forensically duplicate each media device to create a true mirror image (note that this does NOT mean copying or Ghosting)

Mathematically verify and validate that the mirror image is identical to the original by using hashing algorithms (MD5, SHA-1, SHA-256)

Scan media devices for viruses and spyware - document the results

Produce directory structure for each media device

Analyze the electronic media and extract relevant information

Secure each media device