Integrating Forensic Investigation Methodology into eDiscovery

GIAC (GCFA) Gold Certification

Author: Colin Chisholm, chisholm.colin@gmail.com
Advisor: Jeff Groman

Accepted: January 5th, 2010

Abstract

The legal process of Discovery was changed in 2006 with the introduction of rules specifically dealing with electronically stored information (ESI), creating the process of eDiscovery. The application of forensic investigation methodology to the eDiscovery process can help both legal and technical professionals meet the goals of preserving and collecting data in a manner that is legally defensible and forensically sound.

1. Introduction

The intent of this paper is twofold; to provide a primer on the eDiscovery process for forensic analysts and to provide guidance on the application of forensic investigative methodology to said process.

Even though security practitioners such as forensic analysts operate in the legal vertical, they necessarily view and approach eDiscovery from a different perspective than legal professionals. This paper proposes that both parties can benefit when they integrate their processes; forensic tools and techniques have been used in the collection, analysis and presentation of evidence in the legal system for years. The history, and precedent, of applying forensic science to the legal process can be leveraged into the eDiscovery process. This paper will also detail how the scope and work for a forensic investigator during the eDiscovery process differs from a typical forensic investigation.

2. Disclaimer

Although this paper deals with aspects of the American legal system and discusses a range of legal and technical topics, it should not be construed as legal advice or used as the basis of a pending eDiscovery case. Consult legal counsel and technical staff to develop appropriate policies, standards and procedures for your organization regarding eDiscovery and forensic investigations.

3. Audience

The intended audience for this paper is primarily forensic analysts and other security practitioners. Legal professionals may also benefit in their interactions with technical personnel by viewing the eDiscovery process from a technical perspective.

4. Scope and Assumptions

This paper's scope is limited to the preservation/collection phases of eDiscovery and the forensic investigation methodologies that can be applied to those phases. The goal is to ensure that data preserved/collected is legally defensible and forensically sound.

From a legal perspective the scope of the paper will be limited to the United States court system and the requirements imposed by the Federal Rules of Civil Procedure. Variances in state rules of civil procedure cannot realistically be addressed here, although rulings in state courts may be used for illustrative and analytical purposes. The assumption is made that all actions taken during the eDiscovery process will occur within the borders of the United States.

From a practical and logistical perspective the assumption is made that the party or parties performing the preservation/collection phases of eDiscovery have leadership approval as well as unfettered access to the systems in question. The term "organization" will be used throughout this paper, defined as "`a body of persons (such as a society or corporation) formed for a common purpose`" (Garner, B. (2005). *Black's law dictionary).* Although individuals can be subject to litigation the assumption will be made that the audience for this paper will be associated with organizations.

5. Legal Primer

Forensic analysts would be well served to have a fundamental education in the legal processes associated with eDiscovery to understand their role and what is likely to be expected of them. To that end, what follows is a primer on the relevant aspects of the U.S. legal system.

5.1. Common Law

The United States operates under a Common Law legal system. Common Law is a system of jurisprudence that originated in the United Kingdom as unwritten laws and has evolved into a formalized, written system of case law. Case law establishes precedent, which is referenced in order to determine how the law applies to a similar case. Changes to precedent, either minute or substantial, are made reluctantly and on a case-by-case basis under the principle of "stare decisis." Judges and courts generally will not overturn precedent established in the common law system. The common law system is characterized by providing an extremely stable collection of laws that provide continuity and reasonable expectations for all parties when entering the legal system.

5.2. Litigation

The litigation process begins when one party (the plaintiff), files a complaint against another party (the defendant) in court. The complaint will state allegations of harm, explain what action(s) or inactions were taken by the defendant to cause the alleged harm, and detail the legal basis of the complaint. Complaints are then filed with the court and served on the defendant.

In response, the defendant will file a responsive pleading with the court, generally by way of an answer to the complaint or a motion directed at the complaint. A motion is a "written or oral application requesting a court to make a specified ruling or order" (Garner, B. (2005). *Black's law dictionary).* The intricacies of motion practice are beyond the scope of this paper but can be generalized as the requests made by either party to the judge for a ruling on a legal issue during the litigation process.

When an organization can reasonably anticipate or initiate litigation, there is a duty to preserve relevant information. This "legal hold" requires the preservation of this information which supersedes the organization's records and data management processes and any document destruction policy. Data management typically dictates the destruction of records and the reuse of media based on an established criteria. For example, certain records and data may no longer be necessary for daily operations and consume resources necessary for additional operations. In such situations, the organization may have an established document destruction policy which dictates that unnecessary information be deleted or archived. ESI destroyed under an organization's document destruction policy prior to, or reasonably anticipated litigation is exempted under the FRCP 37(e) Safe Harbor provision (Gippetti v. UPS, Inc., 2008 WL 3264483 (N.D. Cal. Aug. 6, 2008)).

Regulatory and/or compliance requirements may also dictate the destruction, relocation or transfer of records and data. However, when litigation is pending or can be reasonably anticipated, an organization must preserve relevant information and separate it from the active environment. The destruction or alternation of this information (intentional or unintentional) is classified as "spoliation" of evidence and may lead to serious sanctions against the organization by the court. An example includes monetary sanctions and the requirement of the defendant to pay the plaintiff's reasonable costs and attorney fees required to identify and respond to the spoliation (Rimkus Consulting Group, Inc. v. Cammarata, 2010 WL 645253 (S.D. Tex. Feb. 19, 2010)).

5.3. 26(f) Conference

One of the requirements implemented into the Federal Rules of Civil Procedure concerning eDiscovery is the so-called 26(f) conference. This term is derived from Rule 26(f) in the Federal Rules of Civil Procedure. Attorneys from both sides are required to attend this conference within 21 days of a scheduling conference with the court, which generally occurs within 90-120 days of the appearance or service of the complaint on the defendant.

The 26(f) conference represents the first, potential point of contact between legal counsel and the forensic analyst (or other technical personnel) beyond the initial legal hold and requirement to preserve information. A consultation between these parties, legal and technical, can be used to help clarify scope for systems and data.

During the 26(f) conference attorneys are required to discuss "`any issues about preserving discoverable information; and develop a proposed discovery plan`" (The Committee on the Judiciary House of Representatives. (2007). *Federal rules of civil procedure).* The effort to develop a discovery plan must be conducted in good faith and an outline submitted to the court with 14 days after the conference (The Committee on the Judiciary House of Representatives. (2007). *Federal rules of civil procedure).* Topics likely to be discussed at the 26(f) conference that may involve forensic analysts or other technical personnel may include:

- Identification of key data custodians
- Location of relevant network, file and email servers
- Identification of relevant user-level systems (i.e. laptops, smart phones)

- Identification of third-party or offsite systems
- Backup/Restore and media retention processes (for all systems)
- Media sanitization processes
- Technical controls to limit data storage and transmission
- Document retention policy and processes
- Reasonable accessibility of discoverable information
- Expert witnesses intended for testimony

5.4. Information, Records & Evidence

Three terms that are used throughout the eDiscovery process are "information," "records" and "evidence." These terms are used with a degree of interchange among technical personnel however working in the legal vertical one should endeavor to use the correct terms in their correct context.

Information is the quantity required for preservation/collection. Information is an element of an organization's regular operations.

Records are a form of information, regardless of the medium or format, that have value to an organization. Collectively the term is used to describe both documents and recorded data. Records are also an element of an organization's regular operations

Evidence can be anything (testimony, documents, tangible objects) that tends to prove or disprove the existence of an alleged fact .

Regardless of the specific terminology, for purposes of eDiscovery (and in fact discovery generally), obligations of the producing party are broad and include "discovery regarding any nonprivileged matter that is relevant to any party's claim or defense . . . Relevant information need not be admissible at the trial if the discovery appears reasonably calculated to lead to the discovery of admissible evidence" (The Committee on the Judiciary House of Representatives. (2007). *Federal rules of civil procedure)*.

5.5. Forensic Evidence in Court

The ultimate goal of preserving and collecting information and data in the context of eDiscovery is admission and presentation in court. One fundamental requirement for admissibility is that the evidence be relevant. Relevant evidence is that which tends "to prove or disprove a matter in issue" (Garner, B. (2005). *Black's law dictionary).* Another requirement for admissibility is authenticity; assurance that the evidence is what it claims to be.

The same requirements for authenticity applies to all records submitted as evidence, regardless of whether they are electronic or physical. The possibility of alteration does not render electronic evidence inherently unreliable. It has been ruled that "the fact that it is possible to alter data contained in a computer is plainly insufficient to establish untrustworthiness" (*United States v. Glasser,* 773 F.2D 1533, 1559 (11th Cir. 1985). Data that has been altered by a tool during the preservation/collection process may still be presented in court so long as the alterations are not significant to the issues in the case. Unless the opposing party is capable of providing proof that electronic evidence has been altered, only its weight, not admissibility, can be affected.

The "best evidence" rules states that the "original writing, recording, or photograph is required" (The Committee on the Judiciary House of Representatives. (2007). *Federal rules of civil procedure)* to be admitted as evidence. This statement may appear incompatible given the ease of duplication of digital evidence. This requirement is more relevant to physical evidence, a realm in which perfect copies are difficult than in the electronic realm. The Federal Rules of Evidence accommodate ESI by providing exceptions for duplicates ("a duplicate is admissible to the same extent as an original") (The Committee on the Judiciary House of Representatives. (2007). *Federal rules of civil procedure)* so long as the duplicate accurately reflects the data.

The ultimate role of the forensic analyst or security practitioner within the eDiscovery process may be testimony. The acceptance of expert testimony requires that it be based on specific knowledge. Expert testimony which "meets federal requirements for relevance and reliability" (Garner, B. (2005). *Black's law dictionary)* is determined by a pretrial "Daubert Hearing" conducted by federal district courts. The Daubert Hearing specifically relates to Federal Rule of Evidence 702, requiring that expert testimony must be based on "scientific, technical, or other specialized knowledge" (Garner, B. (2005). *Black's law dictionary)* which guides the forensic analyst in understanding and interpreting the evidence. The methodology and tools used in the field of computer forensics have been explicitly designed to help meet these criteria.

6 eDiscovery

This is the point at which discovery, defined as "`compulsory disclosure, at a party's request, of information that relates to the litigation`" (Garner, B. (2005). *Black's law dictionary)* is conducted "`before trial to reveal facts and develop evidence`" (Garner, B. (2005). *Black's law dictionary).* In addition to compulsory disclosure “at a party’s request” Rule 26(a) of the Federal Rules of Civil Procedure also requires automatic initial disclosure of certain information set forth in the rule.

eDiscovery is a related term that has gained use to incorporate electronically stored information (ESI) into the discovery process. From a technical perspective, ESI will generally be referred to as data from this point forward. Although this paper is primarily focused on the role of eDiscovery during the initial stages of litigation, post-judgment discovery "conducted after judgment has been rendered" (Garner, B. (2005). *Black's law dictionary)* is also a possibility; the methodologies discussed here are also applicable to post-judgment discovery.

Each party (plaintiff, defendant, third parties, etc.) has a duty to disclose and provide copies or descriptions of "`all documents, electronically stored information, and tangible things`” (The Committee on the Judiciary House of Representatives. (2007). *Federal rules of civil procedure)* that may be used to support their claims or defenses. An astute observer may point out that the discovery process must have included ESI prior to December of when the Federal Rules of Civil Procedure were changed. Although electronic information was addressed for the first time in the early 1980's (Benson, J. (2008, August). *When Lawyers attack: dealing with the new rules of electronic discovery*), specific rules did not exist to address the scope of relevant information.

6.1. Authenticity and Chain of Custody

Authentication is defined as "the act of proving that something (as a document) is true or genuine, especially so that it may be admitted as evidence" (Garner, B. (2005). *Black's law dictionary).* Forensic investigative methodology can provide both processes and technical means to assure the authentication of evidence from acquisition to presentation in court. The chain of custody can play a large role in assuring authenticity.

Chain of custody is a process of handling evidence which creates and maintains a transaction record for each individual who assumes or releases possession of the evidence. From a technical perspective one can use hashing to ensure the integrity of the evidence as well as any ESI associated with the eDiscovery process (such as the chain of custody forms and analyst notes). The chain of custody record provides both a continuity of custody from acquisition to presentation and a list of individuals who have had possession of the evidence. This creates a web of accountability composed of records, information, and individuals available to the court if the relevance of evidence is questioned.

The relevance of evidence may be affected if the chain of custody is broken, suspect, or inconsistent. A break in the chain of custody does not automatically render evidence inadmissible so long as there is sufficient additional proof that the evidence is what it purports to be. However, the "weight" of the evidence or "the persuasiveness of some evidence in comparison with other evidence," (Garner, B. (2005). *Black's law dictionary)* can be affected if integrity is in question. The chain of custody will also help ensure that the evidence has not been in the possession of any individual who is not subject to questioning.

While the chain of custody can be used to effectively provide assurance of authenticity, it cannot however provide assurance of integrity. That is to say, it cannot prove that the evidence has not been altered while in the possession of any particular party. The use of a technical control such as hashing can be used to help ensure integrity.

6.2. Integrity and Hashing

The integrity of data is defined as "a requirement that information and programs are changed only in a specified and authorized manner" (National Research Council, *Computers at Risk*, (Washington. DC: National Academy Press, 1991). The unauthorized alteration of data may occur intentionally or unintentionally, either of which can affect the weight of the evidence in court.

Hashing is defined as the process of taking an amount of data (such as a file or the bit stream image of a hard drive) and applying a "complex mathematical algorithm to generate a relatively compact numerical identifier (the hash value) unique to that data" (Salgado, R. (2005). *Fourth amendment search and the power of the hash)* As an example, a hash value generated against a document will be unique; if the document is changed the subsequent hash value will be different.

Several technical points on hashing algorithms are relevant. The algorithms ("iterative, one-way hash functions") (National Institute of Standards and Technology (2001). *Secure hash standard* (FIPS 180-2) that are used to generate the hash value cannot be reversed to reconstruct the original data. Also, the odds of two non-identical pieces of data generating the same hash value are remote (Schneier, B. (1996). *Applied cryptography)* The MD5 algorithm can generate more than 340 billion, billion, billion, billion values; the SHA-1 algorithm can generate a range of values over four billion larger than MD5 (Schneier, B. (1996). *Applied cryptography)*.

The use of hashing to ensure the integrity of evidence is not a legal requirement. It is, however, regarded as a best practice that is "used in nearly every examination of seized digital media" (Salgado, R. (2005). *Fourth amendment search and the power of the hash)*. Hashing should be viewed as one element to ensure evidence admissibility in combination with chain of custody. Speaking to its prevalence, "it is clear that hashing has become an important fixture in forensic examinations" (Salgado, R. (2005). *Fourth amendment search and the power of the hash)*.

The integrity of evidence can be assured with the use of hashing, proving that evidence has not been changed in an unaccountable manner. Examples of legitimate changes to evidence would include extracting a partition from a disk image or summarizing large amounts of information that would be impractical to present in their native format (e.g. system logs, firewall rules, etc).

Many forensic tools that acquire data are capable of generating a hash value at the time of capture. This initial hash value should be recorded into the chain of custody and verified at each stage of access or transfer. This is especially important because it is unlikely that the original, acquired data will be presented in court. It is most likely that a copy of the original will be made, against which a hash value will also be generated. The hash value should be verified at each step of the chain of custody to ensure that the data has not been contaminated or misidentified.

6.3. Data Types

Volatile data, that which is in use by a system but not written to media is subject to the acquisition process under eDiscovery (Columbia Pictures, Inc. v. Bunnell, 245 F.R.D. 443 (C.D. Cal. 2007)). Volatile data is subject to significant change within seconds, capturing a snapshot of relevant volatile data within the window of activity is unlikely and difficult. Examples of volatile data would include data in memory, network status, and connections and running processes.

Logical Backups of data from "active" sources are likely to be a primary source of relevant data. These would include file servers, email servers, desktops, laptops and smart phones. These systems are likely in use (or recently retired) and likely contain data that can be preserved/collected during regular operations. Examples of data that could be required from logical backups include office files (documents, spreadsheets), images, source code, or years-old email. The methodology section of this paper details how best to obtain this type of data while preserving integrity and authenticity.

Archives should not be overlooked as a source of relevant data. Archiving is the process of moving data out of systems (such as email) into another system (such as to a file server). A common example is of users archiving email so they can move it from an email server to a file server. The archive remains available to the user without consuming valuable space on the email server.

System and data backups generated as part of an organization's regular operations are a relatively straightforward source of data for preservation/collection. They are likely to have associated written processes and technical logs and be performed on a regular schedule. Most organizations rotate their backup media through an on-site/off-site storage schedule. Depending on the timeframe in which the data is required it may be necessary to retrieve offsite backups.

6.4. System Types

The growing number of systems in the consumer electronics market that can process, store or transmit data will necessarily be reflected in the scope of systems subject to eDiscovery. Given that litigation will reflect the "real world" more than a controlled corporate or business environment, the scope of systems may be large. Law enforcement agencies and their experience as first responders can be a valuable resource for determining the scope of system types. As a general guide, the U.S. Department of Justice provides a list of systems in the *Electronic Crime Scene Investigation: A Guide for First Responders, Second Edition*. The systems detailed in this guide represent a front-line view of the range of systems that may come into play during the 26(f) conference. Examples include:

- Computer Systems (desktops, laptops, servers)
- Storage Devices (hard drives, external hard drives, removable media, thumb drive, memory cards)
- Handheld Devices (cell phones, smart phones, PDAs, IM devices)
- Peripheral Devices (keyboard, mouse, microphones, USB/Firewire hubs,
- Web cameras, memory card readers, VoIP device)
- Other (data storage tape drives, surveillance equipment, digital cameras, video cameras, digital audio recorders, digital video recorders, mp3 players, satellite audio, video receiver and access cards, video game consoles, computer chat headsets, KVM, SIM card readers, GPS', thumb print readers)

6.5. Forensic Investigation versus eDiscovery

The processes followed for forensic investigation and eDiscovery are similar; this reflects their role in the legal vertical and the relative similarity of their goals. As illustrated, the phase in common for both processes involves the preservation, collection and acquisition of either information or evidence.

Scope of work is a primary difference for the forensic analyst between a forensic investigation and the eDiscovery process. A typical forensic investigation that follows the forensic investigation methodology will begin with an incident and follow through to data acquisition, analysis and reporting. That final stage, reporting results, may involve the presentation of findings or testimony in court, depending on the nature of the incident.

For example, an incident the involves the violation of organizational policy, not laws, is unlikely to involve the legal system.

In contrast, the forensic analyst's involvement in the eDiscovery process will likely be limited to technical consultation and the preservation/collection of relevant information. This role may begin as early as the 26(f) conference, assisting legal consul with technical issues and determining the scope of relevant systems and data.

The respective timelines of a forensic investigation versus the eDiscovery process can be substantially different. A full forensic investigation may be tied to incident response, requiring rapid action to acquire a "snapshot" of systems and data. This incident response may be measured in minutes and hours; the eDiscovery process is likely to be slower. Need to preserve and collect information is in response to a legal requirement, a slower process than incident response, measured in the weeks and months if not years.

6.6. Ad-Hoc versus Forensic Methodology

The use of an ad-hoc process to preserve and collect information can have significant ramifications later in the litigation process. Courts have ruled that expert opinion on a methodology that diverges "significantly from the procedures accepted by recognized authorities in the field ... cannot be shown to be 'generally accepted as a reliable technique'" (*United States v. Solomon*, 753 F.2d 1522, 1526 (9th Cir. 1985). The credibility of an expert (such as the forensic analyst) can be affected by not following peer-accepted procedures. Using an established forensic investigative methodology that has been employed in prior cases, in combination with proven tools, will help ensure that data is preserved and collected in a manner that is legally defensible and forensically sound.

6.6.1. Stage One: Verification *(Forensic Investigation)*

Verification is concerned with confirming that an incident or action has occurred that warrants the initiation of a forensic investigation. There is no directly comparable stage in the eDiscovery process. The criteria for verification can come from multiple sources; an organization's internal policies, local, state, or federal law, an organization's internal policies or a triggering event. The security practitioner may or may not be involved in this stage, depending on size of the organization and the scope of the incident.

6.6.2. Stage Two: Identification *(eDiscovery)* and System Description *(Forensic Investigation)*

From an eDiscovery perspective this is the point at which the triggering event has occurred and been verified (in stage one). The trigger event can be a judicial order, a discovery request, or mere knowledge of a pending or future legal proceedings likely to require the data (*The Electronic discovery reference model)*. Potential sources of data (such as systems), subject matter experts (such as forensic analysts) and other required resources will be identified and allocated at this point. From a forensic investigation perspective this is the point at which detailed descriptions of the systems in scope will be collected by the forensic analyst or security practitioner. This description is a crucial first step in evaluating systems in scope and to plan the remainder of the investigation. The priority of a system, its role in the organization and the type of data it contains will affect this stage. For example, workstations will be treated differently than mission-critical servers when it comes to preservation, collection and evidence acquisition.

6.6.3. Stage Three: Preservation, Collection *(eDiscovery)* and Evidence Acquisition *(Forensic Investigation)*

The preservation, collection and evidence acquisition phases are concerned with nearly identical goals between eDiscovery and forensic investigation. Both processes are concerned with acquiring relevant data in scope in a manner that minimizes data loss in a manner that is legally defensible, auditable, proportionate, reasonable and efficient.

Forensic investigation is primarily concerned with the following four principles; the minimization of data loss, the recording of detailed notes, the analysis of collected data and reporting findings (Lee, R. (2008, December). *Forensic and Investigative Essentials)*. The eDiscovery process is similarly concerned with the reservation/collection of relevant information, followed by processing, review and analysis.

A tailored preservation strategy should be developed based on the details of the incident and leveraging the information gathered during the prior Identification and System Description phase. Systems and data in scope should be handled in such a way to avoid data destruction and a preservation plan should be developed prior to execution. As for the hands-on process for preservation/collection/acquisition, a technical example follows in section 7.2.4. All system types are at play in this stage.

Stage Three (Guidelines)

The ultimate role of the information preserved and collected during the eDiscovery process is not to meet a binary, mathematical, yes-or-no technical requirement. The role of this material is ultimately in court, an environment in which evidence is used to "prove" or "disprove" facts in the context of competing stories presented by the attorneys for each side. The human factor is key, requiring persuasion in an environment where hard facts alone will not necessarily win a case.

Consistency of Process

In the legal context, the use of a consistent process and procedure carry more weight than any particular toolset or technology. Failure to follow a recognized, consistent process can have significant ramifications in court. Organizations should perform forensic investigations using a consistent process to preserve/collect data in a legally defensible manner.

Use Forensic Toolkits

Forensic toolkits are specialized tools designed to meet the criteria for forensic investigations; to access and acquire data in a manner that makes as few changes to the target media as possible. Toolkit options run the gamut from freely distributed applications and LiveCDs (bootable operating systems on removable media) to enterprise-level commercial applications.

"Live" Acquisition is Best

When possible, the acquisition of a "live" system (including volatile data) should be achieved. This type of acquisition will capture a snapshot of the system in question including the contents of memory, running processes, and network connections as well as allocated and unallocated (deleted) files. Given the relative timelines of forensic investigation versus eDiscovery, this may not be feasible but should not be overlooked as a best case scenario.

Bit Stream Imaging is the Gold Standard

Forensic data acquisition can be divided into two methods: bit stream imaging and logical backup. Bit stream imaging (a.k.a. disk imaging) generates a bit-for-bit copy of the original media including unallocated (deleted) data. Bit stream imaging requires access to the entire volume, a requirement more easily accomplished with low-capacity systems (such as laptops) and storage devices (such as memory cards and disc-based media). The acquisition of high-capacity systems such as servers and storage arrays through bit stream imaging is possible, but may be prohibitive given the logistical difficulty of storing and copying such a large amount of data.

Logical backups copy the directories and files from a volume such as a storage device or network file share. This process does not capture additional data from the media, such as deleted files or residual data stored in slack space. The use of hashing to ensure integrity and authenticity is essential when performing logical backups.

Make No Changes

During the preservation/collection process do not alter, delete or add data within reason. The use of forensic toolkits will help reduce the impact of acquisition and collection on the target media. These tools are designed to access media in a read-only state and will not create or modify files on target systems unless absolutely necessary. These tools typically publish a "modify" list of what files they modify if used on a "live" system.

Hash Everything

The use of hashing can help ensure the authenticity and integrity of data throughout the investigation and litigation process. All electronic data should be hashed at the point of acquisition, transfer of custody and modification. The hash values should be recorded in multiple locations such as the analyst's log and chain of custody forms.

Log Everything

Forensic analysts should keep detailed logs of the actions they perform through the acquisition and collection process. These logs can be used to prove that a consistent practice was followed. Analysts may also use these logs to refresh their recollection during testimony (The Committee on the Judiciary House of Representatives. (2006). *Federal rules of evidence*). Logs can be created and maintained either on paper or in electronic form. This may be a matter of personal preference for the forensic analyst, but each option can have an impact on the process.

Paper logs can be perceived as less susceptible to tampering or alteration which may be an issue if litigation continues to court. On the other hand, paper logs are less likely to contain detailed technical information since that information would need to be entered "by hand."

Electronic logs can contain more detailed technical information but may be perceived as less tangible, more susceptible to alteration, and therefore have less "weight." To counteract this, the integrity of electronic logs should be ensured with hashing. In civil and criminal cases, the notes and activity logs created by the forensic analyst are subject to discovery and can therefore be entered as evidence.

Record and Preserve Chain of Custody

The chain of custody, in combination with hashing, is essential to assuring the authenticity and integrity of evidence. The chain of custody should begin with data acquisition, and be maintained until acceptance as evidence.

Make Copies

After acquisition, bit stream images and logical backup data should be duplicated and stored on clean, preferably unalterable media such as DVD-ROM. Hashing should be employed to ensure that duplicate copies of this data are identical the original and should be reflected in the chain of custody. At least two copies of collected evidence should be available (one for storage and one for analysis and/or presentation).

Stage Three (Scenario & Example)

In this scenario a USB storage device has fallen into scope for eDiscovery. The defendant has been accuses of making an unauthorized copy of a research paper, representing the intellectual property of their employer. It is alleged that a copy of this research paper resides on this USB storage device which was in the defendant's possession upon arrest.

The toolset used in this acquisition is the SIFT Workstation, a forensic investigation workstation configured with freely available open-source tools. Both logical backup and bitstream backup (imaging) will be demonstrated.

a. Connect SD Card to Forensic Workstation

b. Escalate privileges

```
[root@SIFT-Workstation ~]# sudo su
```

c. View Partition Tables

The SIFT workstation contains two hard disks (`/dev/sda` and `/dev/sdb`). The target media (the USB device) is `/dev/sdc`. Running the `fdisk` application confirms that the size of the volume is 128MB and indicates that the filesystem is FAT32.

```
[root@SIFT-Workstation ~]# sudo fdisk -l
Disk /dev/sda: 32.2 GB, 32212254720 bytes
255 heads, 63 sectors/track, 3916 cylinders
Units = cylinders of 16065 * 512 = 8225280 bytes
Disk identifier: 0x0007eca3
Device Boot Start End Blocks Id System
/dev/sda1 * 1 30 240943+ 83 Linux
/dev/sda2 31 279 2000092+ 82 Linux swap / Solaris
/dev/sda3 280 3916 29214202+ 83 Linux
Disk /dev/sdb: 32.2 GB, 32212254720 bytes
255 heads, 63 sectors/track, 3916 cylinders
Units = cylinders of 16065 * 512 = 8225280 bytes
Disk identifier: 0x000c3c7e
Device Boot Start End Blocks Id System
/dev/sdb1 1 3916 31455238+ 83 Linux
Disk /dev/sdc: 131 MB, 131072000 bytes
255 heads, 63 sectors/track, 15 cylinders
Units = cylinders of 16065 * 512 = 8225280 bytes
Disk identifier: 0x00000000
Device Boot Start End Blocks Id System
/dev/sdc1 1 16 127999+ b W95 FAT32
```

```
Partition 1 has different physical/logical beginnings (non-Linux?):
phys=(1023, 254, 63) logical=(0, 0, 2)
Partition 1 has different physical/logical endings:
phys=(1023, 254, 63) logical=(15, 238, 31)
```

d. Create destination directories

Several directories need to be created to receive data generated through the preservation/collection process:

```
/media/128mb (mount point)
/images/128mb (destination directory)
/images/128mb/files (destination subdirectory)
```

```
[root@SIFT-Workstation ~]# mkdir /media/128mb
[root@SIFT-Workstation ~]# mkdir /images/128mb
[root@SIFT-Workstation ~]# mkdir /media/128mb/files
```

e. Mount target media in read-only mode

Using the read-only option for mounting the target media assures that no changes can be made during the acquisition process. Using the mount command, specify read-only (`-o ro`) and a FAT filesystem (`-t vfat`). The source is the FAT volume of the USB device (`/dev/sdc1`) and the destination is a directory on the SIFT workstation (`/media/128bm`).

```
[root@SIFT-Workstation ~] mount -o ro -t vfat /dev/sdc1 /media/128mb
```

f. Logical Backup - Hash contents

In the case of a logical backup, it is essential to hash the target data to ensure integrity and authenticity before making a copy. The `md5deep` application can calculate hash values for multiple files in sequence and output the results to a log file. This will provide both a complete listing of all the (allocated) files on the USB device and a corresponding MD5 value.

```
[root@SIFT-Workstation ~]# md5deep -r /media/128mb > /images/128mb/md5deep.txt
[root@SIFT-Workstation ~]# less /images/128mb/md5deep.txt
83f2ab701e51953f753d936f2aa9df61 /media/128mb/.fseventsd/fseventsd-uuid
88ebc367b9b76ac43b2441bc87df8f06 /media/128mb/.fseventsd/636573319d0768aa
4df8c31c0019eb6e8be4153b29b07d1c /media/128mb/documents/secret-paper193.pdf
0a69d98a73a277a181058155226bb316 /media/128mb/documents/Becoming a Forensic
Investigator.pdf
ddccbcd4734a0cd23f9e9f325c0d736a /media/128mb/documents/Forensic Analysis of a
```

```
Compromised Intranet Server.pdf
d83a0b8d03044e99a0bad8560f96bd0b /media/128mb/documents/Remotely Accessing
Sensitive Resources.pdf
57a49ce3fc0c8978c4ac4ccbd0f84912 /media/128mb/wallpaper/
01875_horseheadnebula_2560x1600.jpg
ad0c9430a9abe768fcd81a214f894998 /media/128mb/wallpaper/
01881_orionnebula_2560x1600.jpg
d788f3a0db5369bd96b16f4f2645e5ef /media/128mb/wallpaper/
02059_bridge_2560x1600.jpg
bc444c3b2baae36ba8bccf570ad62674 /media/128mb/wallpaper/
02080_oldstone_2560x1600.jpg
```

g. **Logical Backup - Copy Files**

Now that a hash has been calculated for the source files contained from the read-only mount point (`/media/128mb`), the files in question will be copied to a local directory (`/images/128mb/files`).

```
[root@SIFT-Workstation ~] cp -R /media/128mb /images/128mb/files
```

h. **Logical Backup - Verification**

Running `md5deep` against the copied contents of the USB device will both provide a complete file list and a confirmation that the hash values of the source files (`/media/128mb`) have not changed after the files have been copied to a new location (`/images/128mb`).

```
[root@SIFT-Workstation ~]# md5deep -r /media/128mb > /images/128mb/md5deep.txt
[root@SIFT-Workstation ~]# less /images/128mb/md5deep.txt
83f2ab701e51953f753d936f2aa9df61 /media/128mb/.fseventsd/fseventsd-uuid
88ebc367b9b76ac43b2441bc87df8f06 /media/128mb/.fseventsd/636573319d0768aa
4df8c31c0019eb6e8be4153b29b07d1c /media/128mb/documents/secret-paper193.pdf
0a69d98a73a277a181058155226bb316 /media/128mb/documents/Becoming a Forensic
Investigator.pdf
ddccbcd4734a0cd23f9e9f325c0d736a /media/128mb/documents/Forensic Analysis of a
Compromised Intranet Server.pdf
d83a0b8d03044e99a0bad8560f96bd0b /media/128mb/documents/Remotely Accessing
Sensitive Resources.pdf
57a49ce3fc0c8978c4ac4ccbd0f84912 /media/128mb/wallpaper/
01875_horseheadnebula_2560x1600.jpg
ad0c9430a9abe768fcd81a214f894998 /media/128mb/wallpaper/
01881_orionnebula_2560x1600.jpg
d788f3a0db5369bd96b16f4f2645e5ef /media/128mb/wallpaper/
02059_bridge_2560x1600.jpg
bc444c3b2baae36ba8bccf570ad62674 /media/128mb/wallpaper/02080_oldstone_2560x1600.jpg
```

i. Logical Backup - Additional Hash Verification

As an additional verification of the validity of hash values, simply opening a file and re-saving it or renaming a file will not affect the hash value. Both of these changes occur at the metadata layer (illustrated in Appendix A) and not the data layer. As a proof of concept, make a copy of a single file, assign it a new name, and run MD5 against the copy and compare to the original. The hash values are identical, providing assurance that integrity has been preserved.

```
[root@SIFT-Workstation ~]# cp /images/128mb/files/128mb/documents/secretpaper193.
pdf
[root@SIFT-Workstation ~]# md5 /images/128mb/files/128mb/documents/secretpaper193.
pdf
4df8c31c0019eb6e8be4153b29b07d1c
[root@SIFT-Workstation ~]# md5 /images/128mb/files/128mb/documents/paper.pdf

4df8c31c0019eb6e8be4153b29b07d1c
```

j. Logical Backup - Hash the analyst's log file

The script command has been used to create an electronic log of all activity in the terminal window during this process. Now that the data has been acquired and hash values calculated to verify integrity and authenticity the analyst log file itself will be hashed. The log file contains the output of al hash values generated thus far; creating a hash value log for the log file help ensure the integrity for these file from the acquisition.

```
[root@SIFT-Workstation ~]# md5 /root/logical-backup.txt > /root/logicalbackup.
md5.txt
[root@SIFT-Workstation ~]# less /root/logical-backup.md5.txt

a3f00f522a74bb95a3c8812c92324569 /root/logical-backup.txt
```

k. Bit Stream Backup - Acquisition

Bit stream backup (or disk imaging) allows the forensic analyst to acquire the complete contents of the target media, including slack space and unallocated (deleted) files. In this example we will acquire the entirety of the 128MB USB storage device. The assumption is made that the partition information and mount points from the logical backup process above have not changed. The `dc3dd` application will be used for acquisition and creating a hash value of the data. This time the source is the full volume (`/dev/sdc`) not just the FAT data partition (`/dev/sdc1`) in the logical backup demonstration. The destination is a single image file copied to the forensic workstation (`/images/128mb.img`). The hash value will be written to a file (`/images/128mb.md5`).

```
[root@SIFT-Workstation ~]# dc3dd if=/dev/sdc of=/images/128mb.img hashwindow=1
hashlog=/images/128mb.md5
warning: sector size not probed, assuming 512
dc3dd 6.12.3 started at 2010-04-02 10:02:20 -0400
command line: dc3dd if=/dev/sdc of=/images/128mb.img hashwindow=1 hashlog=/
images/128mb.md5
compiled options: DEFAULT_BLOCKSIZE=32768
sector size: 512 (assumed)
256000+0 sectors in
256000+0 sectors out
131072000 bytes (125 M) copied (??%), 554.887 s, 231 K/s
dc3dd completed at 2010-04-02 10:11:35 -0400
```

l. Bit Stream Backup - Verification

After the acquisition is complete the hash values are verified using `md5`.

```
[root@SIFT-Workstation ~]# less /images/128mb.md5
83312644be72b935bf3b75571842c84a
[root@SIFT-Workstation ~]# md5 /home/sansforensics/bitstream-backup.txt > /
home/sansforensics/logical-backup.md5.txt
[root@SIFT-Workstation ~]# md5 /images/128mb.md5
83312644be72b935bf3b75571842c84a
```

m. Bit Stream Backup - Hash the analyst's log file

Again, the `script` command has been used to create an electronic log of all activity in the terminal window during this process.

6.6.4. Stage Four: Processing/Review/Analysis *(eDiscovery)* and Media Analysis, String/Byte Search, Timeline Analysis, Data Recovery *(Forensic Investigation)*

In both methodologies this stage involves analyzing the data that has been collected during the preservation/collection/acquisition phase. Processing this data is the most time consuming portion of the eDiscovery/Forensic Investigation process. This involves applying the details of the incident, the verification and system description information and making educated decisions when examining this "snapshot" of data for relevant evidence. Timelines should be created, relevant data in scope should be identified and analyzed, metadata should be taken into account and data recovery may be required. Data types and metadata are at play in this stage.

6.6.5. Stage Five: Production *(eDiscovery)*

The production stage is unique to eDiscovery and involves the preparation and production of ESI in a format that has already been agreed to by the parties during the 26(f) conference. The security practitioner may be involved in this stage as well to ensure that, from a technical perspective, all of the ESI is made available in the agreed upon formats. It cannot be assumed that original data formats will be usable among all parties, particularly if one party uses legacy or proprietary systems. Is is not required that ESI be presented in its native electronic format (Schmidt v. Levi Strauss & Co., 2007 WL 2688467 (N.D. Cal. Sept. 10, 2007)). In order to meet the requirement for ESI to be presented in an agreed upon and usable format, ESI may need to be converted into a standardized, searchable format. Data formats such as text, images, numerical data, databases and email may need to be presented in near-native formats, summary form or on physical paper.

The usable requirement in the presentation of ESI can involve factors such as searchability, fielded data, redaction, metadata and summaries. Searchable text can contain readable text as well as metadata. This is likely in a common text-based format (such as ASCII text) that has been extracted or converted from original or near-native formats.

Fielded data refers to additional data that relates to the original content data such as metadata. Defined literally as “data about data” it can loosely be thought of as additional information attached to data. Examples could include the formulas used to calculate values in a spreadsheet, edit history statistics for a document or hardware details about a camera used to take a digital photograph. An analog equivalent for metadata would be the table of contents or index in a printed book. This use of the term metadata should not be confused with the metadata layer when discussing file system forensic analysis.

The establishment of metadata’s scope in the eDiscovery process is typically addressed by the parties in a Rule 26(f) conference. Precedent has been established that metadata is as subject to discovery as original data. The Supreme Court of Arizona has ruled that "`the metadata in an electronic document is part of the underlying document; it does not stand on its own.`" (Lake v. City of Phoenix, 222 Ariz. 547, 218 P.3d 1004 (2009)). Redaction is also a requirement for ESI at the presentation stage, even after the process of limiting the scope of data and systems in the prior stages and the 26(f) conference.

Summaries of ESI such a voluminous writings, recordings or photographs may be presented as a "`a chart, summary or calculation`" (The Committee on the Judiciary House of Representatives. (2007). *Federal rules of civil procedure).* They allow for large amounts of information such as to be presented efficiently in material form, possibly in combination with testimony. Any ESI that is subject to any form of conversion or alteration from its original form in preparation for presentation should be subject to rigorous quality control to ensure that content has not been changed. All ESI should be subject to thorough hashing to ensure the integrity of data.

6.6.6. Stage Six: Presentation *(eDiscovery)* and Reporting Results *(Forensic Investigation)*

At this final stage of the eDiscovery and Forensic Investigation process, the ESI that has been identified, acquired, analyzed and prepared will be displayed to an audience. For forum for this audience can be in the form of a deposition, hearing or trial. The goal of this stage is to provide targeted evidence to prove or disprove statement of facts in the overall context of eliciting further information, validate existing facts or positions, or persuade an audience. (*The Electronic discovery reference model*.)

7. Pre-Discovery Planning

The requirements to preserve & collect information under eDiscovery can impose a financial and resource burden on an organization. Building on the legal primer and forensic investigation methodology already covered, what follows is a broad guide to pre-discovery planning.

Acknowledge Risk

Organizational leadership should understand the eDiscovery process and recognize the potential risks and requirements in the event of litigation. Some form of leadership buy-in would send a clear signal to the organization that eDiscovery is a significant reality that carries real consequences and risks. Expressing the importance of a response plan and the allocation of resources to develop such a plan is a key first step. Once there has been an official recognition of the importance and impact of eDiscovery, a policy should be developed and published. The policy should contain senior management's directive, establish goals, and assign resources to accomplish those goals.

Documented Processes

The organization should develop, publish (internally) and maintain documented processes related to eDiscovery. Clear guidelines and procedures should be developed regarding forensic investigation activities. These guidelines and procedures should be subject to ongoing review by legal council. A balance should be achieved between legal contracts, privacy policies, statues, regulations, and existing internal policies. Coordination should occur between leadership, business managers, technical managers, human resources and legal counsel to establish and maintain this balance.

Allocate Resources

Resources should be determined and allocated prior to potential litigation. A published policy document and the associated leadership buy-in will help clear the way for these resources from a financial and workload perspective.

Staff are the primary component of any organization’s resources regarding eDiscovery. These security practitioners (including forensic analysts) should have access to appropriate training and other educational resources. They should develop and maintain a strong working familiarity with the tools that will be used in the preservation/collection process. Staff should also have a reasonably comprehensive technical knowledge of computer forensic science as well as an approved, standardized toolkit.

Adequate storage media for the preservation, collection and duplication of data should be available in advance of litigation. Write-blocking hardware devices should be available to ensure read-only access to media during the acquisition process. Chain of custody forms (paper or electronic) should be prepared and available in combination with evidence bags to assist with assuring the authenticity and integrity of media.

Data Classification & Management

Organizations that maintain a data classification and management process will have a distinct advantage in the event of eDiscovery. Having an ongoing process which can identify the location (both physical and logical) of data and records will allow an organization to react quickly and efficiently. Mapping this relationship between data and systems will also permit the organization to more efficiently separate the relevant information from the active environment.

Test Backup and Restore Processes

In addition to the data classification and management effort, an organization should ensure that the backup and restore processes currently in use are effective. Most organizations have a backup process for important records and servers to ensure that data can be retrieved in the event of an unforeseen occurrence. These backups should be tested regularly by randomly restoring select data.

Legal Concerns

An understanding should be communicated to relevant personnel that the forensic investigation process will almost certainly involve the preservation/collection of sensitive data. Examples of sensitive data that may be involved include social security numbers, passwords, credit card information, and health care information. Criminal material such as illegal software and child pornography may also be revealed. Human Resources and legal counsel should be involved in reviewing processes and procedures to handle any potential issues and ensure the confidentiality of said data to those directly involved in the eDiscovery process.

8. Conclusions

In conclusion, both the technical security practitioner and the legal professional can benefit from having an understanding of their respective roles in the eDiscovery process. With this understanding, and working together, the financial and resource costs to an organization can be minimized. The use of a consistent and peer-accepted methodology can yield a significant benefit for legal counsel in the litigation process. Combining written logs, witnesses, testimony and ESI can equal legally defensible and forensically sound evidence. Conversely, not following these processes can result in questionable evidence which can have both financial and legal implications for the organization.

9. References

Benson, J. (2008, August). *When Laywers attack: dealing with the new rules of electronic discovery*. Presented at Black Hat Technical Security Conference, Las Vegas, NV

Carrier, B, & Spafford, E. (2003). Getting physical with the digital investigation process. *International Journal of Digital Evidence*, *2*(2), Retrieved from http://www.ijde.org/

Carrier, B. (2002, October). *Open source forensic tools: the legal argument* (PDF). Retrieved from http://www.digital-evidence.org/papers/opensrc_legal.pdf

Columbia Pictures, Inc. v. Bunnell, 245 F.R.D. 443 (C.D. Cal. 2007). Retrieved from http://lctjournal.washington.edu/vol5/a23Hall.html

The Committee on the Judiciary House of Representatives. (2006). *Federal rules of evidence* (PDF), Retrieved from www.uscourts.gov/rules/Evidence_Rules_2007.pdf

The Committee on the Judiciary House of Representatives. (2007). *Federal rules of civil procedure* (PDF), Retrieved from www.uscourts.gov/rules/civil2007.pdf

Denny, W. (2005). *Electronic discovery: understanding preservation obligations, the potential for cost-shifting, and current developments*. Retrieved from http://library.findlaw.com/2005/Feb/6/133662.html

Garner, B. (2005). *Black's law dictionary, abridged eighth edition*. West Publishing Company.

Gippetti v. UPS, Inc., 2008 WL 3264483 (N.D. Cal. Aug. 6, 2008). Retrieved from http://www.ediscoverylaw.com/2008/08/articles/case-summaries/courtdenies-spoliation-sanctions-for-destruction-of-esi-pursuant-to-documentretention-policy-citing-frcp-37e-safe-harbor-provision/

Lake v. City of Phoenix, 222 Ariz. 547, 218 P.3d 1004 (2009). Retrieved from http://www.ediscoverylaw.com/2009/10/articles/case-summaries/supreme-court-of-arizonaholds-
metadata-is-subject-to-public-records-requests/

Lee, R. (2008, December). *Forensic and Investigative Essentials*. Presented at SANS Computer Forensics, Investigation, and Response training, Washington, DC

Lodge, M. (2009, September 01). Archiving is for e-discovery; backup is for recovery. *The Metropolitan Corporate Counsel*, *17*(9), Retrieved from http://www.metrocorpcounsel.com/current.php?artType=view&artMonth=September&artYear=2009&EntryNo=10056

National Institute of Standards and Technology. (2006). *Guide to integrating forensic techniques into incident response* (Special Publication 800-86). Retrieved from http://csrc.nist.gov/publications/PubsSPs.html

National Institute of Standards and Technology (2001). *Secure hash standard* (FIPS 180-2) (PDF). Retrieved from http://csrc.nist.gov/publications/fips/fips180-2/fips180-2.pdf

National Research Council, *Computers at Risk*, (Washington. DC: National Academy Press, 1991)

Packard, N. (2007, September 20). *Sedona commentary on legal holds*. Retrieved from http://www.thesedonaconference.org

Rimkus Consulting Group, Inc. v. Cammarata, 2010 WL 645253 (S.D. Tex. Feb. 19, 2010). Retrieved from http://www.ediscoverylaw.com/uploads/file/Westlaw_Document_%20Rimkus.doc

Salgado, R. (2008, December). *Legal issues in computer & network investigations*. Presented at SANS Computer Forensics, Investigation, and Response training, Washington, DC

Salgado, R. (2005). *Fourth amendment search and the power of the hash*. Retrieved from http://www.harvardlawreview.org/issues/119/december05/forum_357.php

Schmidt v. Levi Strauss & Co., 2007 WL 2688467 (N.D. Cal. Sept. 10, 2007). Retrieved from http://www.ediscoverylaw.com/2007/11/articles/case-summaries/defendant-notrequired-

to-reproduce-entire-document-production-in-native-electronic-format/

The Sedona Conference. (2008) *"Jumpstart outline"*. Retrieved from http://www.thesedonaconference.org

The Sedona Conference. (2008) *“Preservation, management and identification of sources of information that are not reasonably accessible”*. Retrieved from http://www.theseconaconference.org

Schneier, B. (1996). *Applied cryptography: protocols, algorithms, and source code in c, second edition*. Wiley

United States v. Glasser, 773 F.2D 1533, 1559 (11th Cir. 1985). Retrieved from http://openjurist.org/773/f2d/1553/united-states-v-glasser

United States v. Solomon, 753 F.2d 1522, 1526 (9th Cir. 1985). Retrieved from http://openjurist.org/753/f2d/1522

U.S. Department of Justice. (2008, April 09).*Electronic crime scene investigation: a guide for first responders, second edition*. Retrieved from *http://www.ojp.usdoj.gov/nij*

The Electronic discovery reference model. Retrieved from *http://edrm.net/*

10. Appendix A - File System Layers (Illustration)

Filename Layer

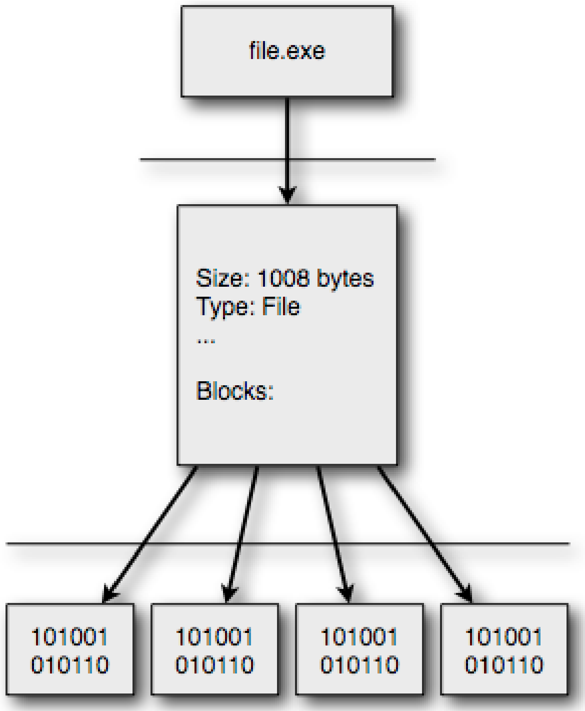
- Filenames stored in:
  - File Metadata - MFT Entry & FAT Directory (Windows only)
  - Directory File - Contains list of files/ directories in that directory

Metadata Layer (Card Catalog)

- Inode, FAT Directory Entry MFT Entry
- Metadata is Addressable
- Metadata is Allocated or Unallocated
- Metadata address = IP address of file
- Contains:
  - Points to Data Layer (list of blocks/ clusters)
  - File type
  - MACTimes
  - Permissions
  - File size
  - Link count

Data Layer

- Clusters (Windows) = Blocks (Unix)
- Clusters/Blocks = 512 sectors
- Clusters/Blocks are Addressable
- Clusters/Blocks are Allocated or Unallocated



FAT	NTFS	EXT2/3
Directory Entry	MFT Entry	Dir. Entry
Directory Entry & FAT	MFT Entry	Inode
Clusters/ Sectors	Clusters	Blocks

Colin Chisholm, chisholm.colin@gmail.com