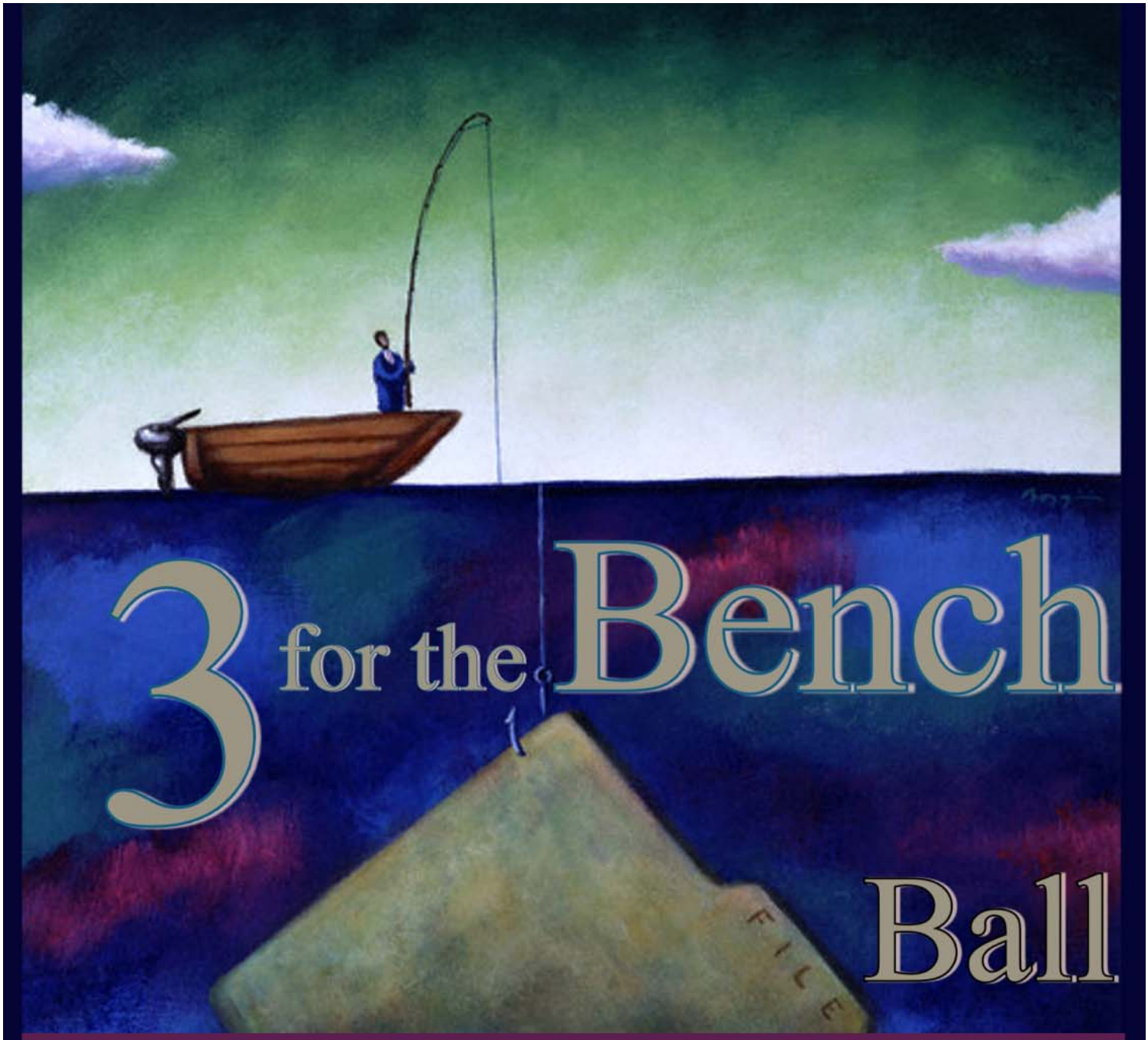
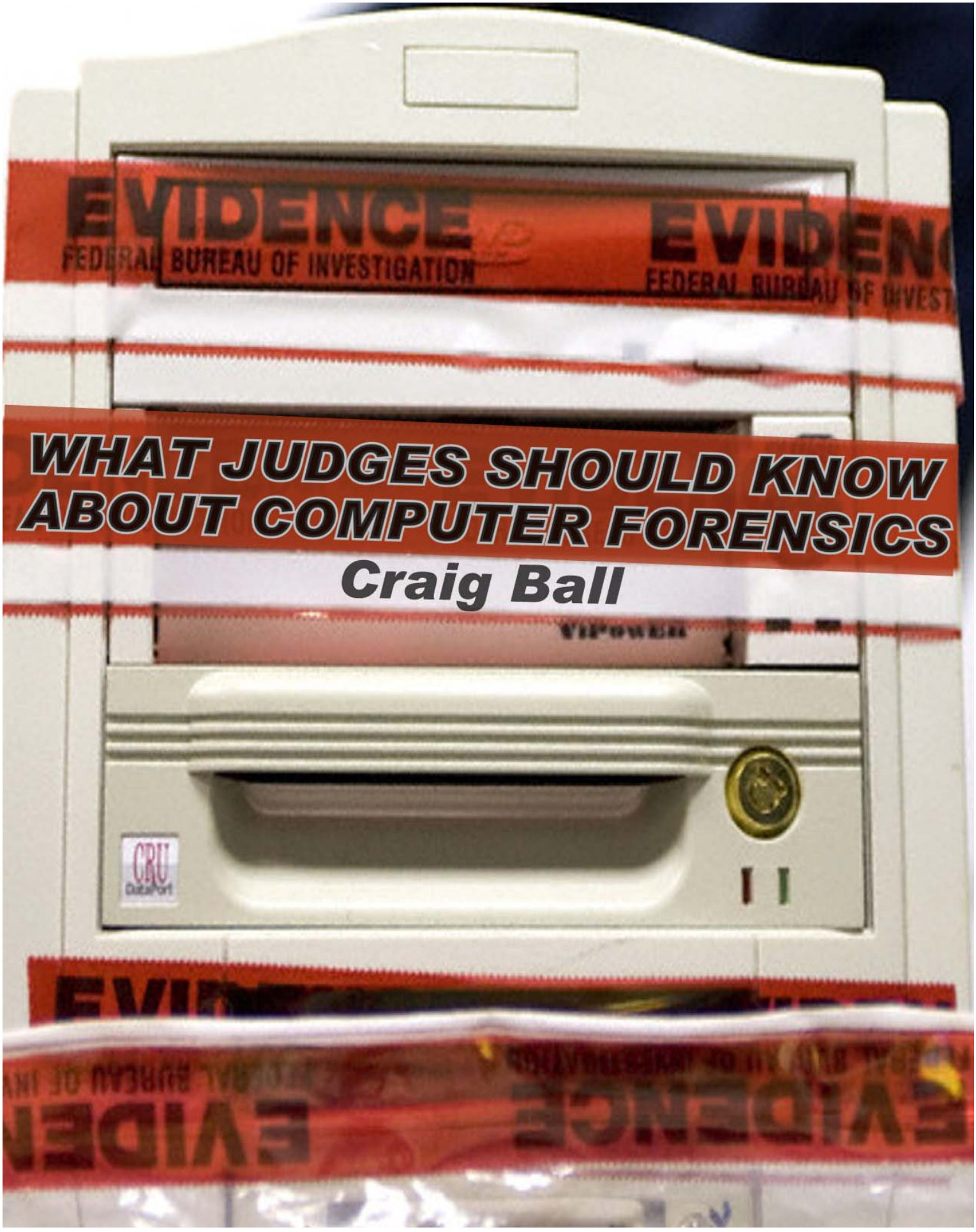


Three Articles for District Court Judges

National Workshop for District Judges

What Judges Should Know About Computer Forensics
What Judges Should Know About Discovery from Backup Tapes
Musings on E-Discovery: Ball in Your Court

Craig Ball


EVIDENCE
FEDERAL BUREAU OF INVESTIGATION
WHAT JUDGES SHOULD KNOW
ABOUT COMPUTER FORENSICS
Craig Ball

What Judges Should Know About Computer Forensics

Craig Ball[1]

Courts increasingly see motions by litigants seeking access to an opponent's computers for the purpose of conducting a computer forensic examination. The impetus may be allegations of discovery abuse, stolen intellectual property, spoliation, forgery, network intrusion, child pornography, piracy, discrimination or a host of other claims.

When bits and bytes are involved, it can be hard to know if the proposed examination is reasonable and necessary or an abusive fishing expedition.

This article looks at some of the fundamentals of computer forensics to help judges weigh the need and burden of acquisition and examination. It addresses, *inter alia*, what computer forensics can and cannot accomplish and flags common errors made by parties and the courts in ordering such examinations.

Table of Contents

[1] The author gratefully acknowledges the invaluable editorial contributions of his spouse, Diana Ball, and of esteemed colleagues, Sharon Nelson and John Simek of Sensei Enterprises, Inc., for their helpful suggestions.

What is Computer Forensics?

A computer's operating system or **OS** (e.g., Windows or Vista, Mac or Linux) and installed software (**applications**) generate and store much more information than users realize. Some of this unseen information is **active data** readily accessible to users, but requiring skilled interpretation to be of value in illuminating human behavior. Examples include the data *about* data or **metadata** tracked by the OS and applications, but not displayed onscreen. For example, Microsoft Outlook records the date a Contact is created, but few of us customize the program to display that "date created" information.

Other active data reside in obscure locations or in coded formats less readily accessible to users, but enlightening when interpreted and correlated. Log files, hidden system files and information recorded in non-text formats are examples of **encoded data** that may reveal information about user behavior.

Finally, there are vast regions of hard drives and other data storage devices that hold **forensic data** even the operating systems and applications can't access. These "data landfills," called **unallocated clusters**[2] and **slack space**[3], contain much of what a user, application or OS discards over the life of a machine. Accessing and making sense of these vast, unstructured troves demands specialized tools, techniques and skill.

Computer forensics is the expert acquisition, interpretation and presentation of the data within these three categories (**Active**, **Encoded** and **Forensic** data), along with its juxtaposition against other available information (e.g., credit card transactions, keycard access data, phone records and voice mail, e-mail, documents and instant message communications and texting).

In litigation, computer forensics isn't limited to personal computers and servers, but may extend to all manner of devices harboring electronically stored information (**ESI**). Certainly, external hard drives, thumb drives and memory cards are routinely examined. *When relevant*, information on cell phones, cameras and even automobile navigation systems and air bag deployment modules may be implicated. The scope of computer forensics—like the scope of a crime scene investigation—should be reasonably tailored to the available evidence and issues before the court.

How Does Computer Forensics Differ from Electronic Discovery?

Computer forensics is a non-routine subcategory of "e-discovery." In simplest terms, electronic discovery addresses the ESI accessible to litigants; computer forensics addresses the ESI accessible to forensic experts. However, the lines blur because e-discovery often requires litigants to grapple with forms of ESI—like backup tapes—traditionally regarded as inaccessible, and computer forensics relies on information readily accessible to litigants, such as file modification dates.

The principal differentiators are **expertise** (computer forensics requires a unique skill set), **issues** (most cases can be resolved without resorting to computer forensics, though some will hinge on matters that can only be resolved by forensic analysis) and **proportionality** (computer forensics injects issues of expense, delay and intrusion). Additionally, electronic discovery tends to address evidence as discrete information items (documents, messages, databases), while computer forensics takes a more systemic or holistic view of ESI, studying information items as they relate to one another and in terms of what they reveal about what a user did or tried to do. And last, but not least,

[2] Unallocated clusters are storage areas flagged by the file system as available to hold data. When these have been previously used for data storage, their former contents linger until overwritten by new data.

[3] File slack space is the excess storage space between the end of a file and the end of the final cluster in which the file is stored. Slack space may hold fragments of deleted files.

electronic discovery deals almost exclusively with existing ESI; computer forensics tends to focus on what's gone, how and why it's gone and how it might be restored.

When to Turn to Computer Forensics

Most cases require no forensic-level computer examination, so courts should closely probe whether a request for access to an opponent's machines is grounded on a genuine need or is simply a fishing expedition. When the question is close, courts can balance need and burden by using a neutral examiner and a protective protocol, as well as by assessing the cost of the examination against the party seeking same until the evidence supports reallocation of that cost.

Certain disputes fairly demand forensic analysis of relevant systems and media, and in these cases, the court should act swiftly to support appropriate efforts to preserve relevant evidence. For example, claims of data theft may emerge when a key employee leaves to join or become a competitor, prompting a need to forensically examine the departing employee's current and former business machines, portable storage devices and home machines. Such examinations inquire into the fact and method of data theft and the extent to which the stolen data has been used, shared or disseminated.

Cases involving credible allegations of destruction, alteration or forgery of ESI also justify forensic analysis, as do matters alleging system intrusion or misuse, such as instances of employment discrimination or sexual harassment involving the use of electronic communications. Of course, electronic devices now figure prominently in the majority of crimes and many domestic relations matters, too. It's the rare fraud or extramarital liaison that doesn't leave behind a trail of electronic footprints in web mail, online bank records and cellular telephones. For further guidance on circumstances justifying direct access to an opponent's ESI, see, e.g., *Ameriwood Ind., Inc. v. Liberman*, 2006 WL 3825291 (E.D. Mo. Dec. 27, 2006).

Balancing Need, Privilege and Privacy

A computer forensic examiner sees it all. The Internet has so broken down barriers between business and personal communications that workplace computers are routinely peppered with personal, privileged and confidential communications, even intimate and sexual content, and home computers normally contain some business content. Further, a hard drive is more like one's office than a file drawer. It may hold data about the full range of a user's daily activity, including private or confidential information about others. Trade secrets, customer data, e-mail flirtations, salary schedules, Internet searches for escort services, bank account numbers, medical records and passwords abound.

So how does a court afford access to the non-privileged evidence without inviting abuse or exploitation of the rest? With so much at stake, courts need to approach forensic examination cautiously. Granting access should hinge on demonstrated need and a showing of relevance, balanced against burden, cost or harm. It warrants proof that the opponent is either untrustworthy or incapable of preserving and producing responsive information, or that the party seeking access has some proprietary right with respect to the drive or its contents. Showing that a party lost or destroyed ESI is a common basis for access, as are situations like sexual harassment or data theft where the computer was instrumental to the alleged misconduct.

The parties may agree that one side's computer forensics expert will operate under an agreed protocol to protect unwarranted disclosure of privileged and confidential information. Increasingly, courts appoint neutral forensic examiners to serve as Rule 53 Special Masters for the purpose of performing the forensic examination *in camera*. To address privilege concerns, the information

developed by the neutral is first tendered to counsel for the party proffering the machines for examination, which party generates a privilege log and produces non-privileged, responsive data. Use of a Special Master largely eliminates the risk of privilege waiver in unrelated litigation, a potential addressed in *Hopson v. Mayor of Baltimore*, 232 F.R.D. 228 (D. Md. 2005)

Whether an expert or court-appointed neutral conducts the examination, the order granting forensic examination of ESI should provide for handling of confidential and privileged data and narrow the scope of examination by targeting specific objectives. The examiner needs clear direction in terms of relevant keywords and documents, as well as pertinent events, topics, persons and time intervals. A common mistake is for parties to agree upon a search protocol or secure an agreed order without consulting an expert to determine feasibility, complexity or cost. Generally, use of a qualified neutral examiner is more cost-effective and ensures that the court-ordered search protocol is respected.

Who Performs Computer Forensics?

Computer forensics is a young discipline, so the most experienced examiners may be largely self-taught. Experienced examiners still tend to emerge primarily from law enforcement, but this is changing as a host of computer forensics certification courses and even college degree plans have appeared. Unfortunately, though the ranks of those offering computer forensics services are growing rapidly, there is inadequate assessment or regulation of the profession. No universally recognized standard exists to test the training, experience and integrity of forensic examiners. A few states require computer forensic examiners to obtain private investigation licenses, but don't demand that applicants possess or demonstrate expertise in computer forensics.

Computer experts without formal forensic training or experience may offer their services as experts, but just as few doctors are qualified as coroners, few computer experts hold forensic qualifications. Programming skill has little practical correlation to skill in computer forensics.

Selecting a Neutral Examiner

Ideally, the parties will agree upon a qualified neutral. When they cannot, the court might:

1. Require the parties to designate examiners they deem qualified, then have the partisan examiners agree upon a third party neutral examiner;
2. Seek recommendations from other judges before whom well-qualified examiners have appeared; or,
3. Review the *curriculum vitae* of examiner candidates, looking for evidence of training, experience in court, credible professional certification, publications, bench references and other customary indicia of expertise. Checking professional references is recommended, as CV embellishment is a great temptation in an unregulated environment.

A computer forensic analyst must be able to grasp the issues in the case and, where indicated, possess a working knowledge of privilege law.

What Can Computer Forensics Do?

Though the extent and reliability of information gleaned from a forensic examination varies, here are some examples of the information an analysis can uncover:

1. Manner and extent of a user's theft of proprietary data;
2. Timing and extent of file deletion or antiforensic (e.g., wiping software) activity;
3. Whether and when a thumb drive or external hard drive was connected to a machine;
4. Forgery or alteration of documents;

5. Recovery of e-mail and other ESI claimed not to exist or to have been deleted;
6. Internet usage, online research and e-commerce transactions;
7. Intrusion and unauthorized access to servers and networks;
8. Clock and calendar manipulation;
9. Image manipulation; and
10. Second-by-second system usage.

What *Can't* It Do?

Notwithstanding urban legend and dramatic license, there are limits on what can be accomplished by computer forensic examination. To illustrate, an examiner generally cannot:

1. Recover any information that has been completely overwritten—even just once—by new data;
2. Conclusively identify the hands on the keyboard if one person logs in as another;
3. Conduct a thorough forensic examination without access to the source hard drive or a forensically-sound image of the drive;
4. Recover data from a drive that has suffered severe physical damage and cannot spin;
5. Guarantee that a drive won't fail during the acquisition process; or
6. Rely upon any software tool to autonomously complete the tasks attendant to a competent examination.

Supervision of Examination

A party whose systems are being examined may demand to be present throughout the examination. This may make sense and be feasible while the contents of a computer are being *acquired* (duplicated); otherwise, it's an unwieldy, unnecessary and profligate practice. Computer forensic examinations are commonly punctuated by the need to allow data to be processed or searched. Such efforts consume hours, even days, of "machine time" but not examiner time. Examiners sleep, eat and turn to other cases and projects until the process completes. However, if an examiner must be supervised during machine time operations, the examiner cannot jeopardize another client's expectation of confidentiality by turning to other matters. Thus, the "meter" runs all the time, without any commensurate benefit to either side except as may flow from the unwarranted inflation of discovery costs.

One notable exception is the examination of machines believed to house child pornography. As possession of child pornography is itself a crime, the government requires that examinations be conducted on government premises and under close supervision.; refusing to allow data to be processed in the examiner's lab.

Forensic Acquisition & Preservation

Courts are wise to distinguish and apply different standards to requests for forensically-sound *acquisition* versus those seeking forensic *examination*. Forensic *examination* and analysis of an opponent's ESI tends to be both intrusive and costly, necessitating proof of compelling circumstances before allowing one side to directly access the contents of the other side's computers and storage devices. By contrast, forensically duplicating and preserving the status quo of electronic evidence is relatively low-cost and can generally be accomplished without significant intrusion upon privileged or confidential material. Accordingly, the court should freely allow forensic preservation upon a bare showing of need.

Acquisition guards against both intentional spoliation and innocent spoliation engendered by continued usage of computers and intentional deletion. It also preserves the ability to later conduct a forensic examination, if warranted.

During the conduct of a forensic acquisition:

1. Nothing on the evidence media may be altered by the acquisition;
2. Everything on the evidence media must be faithfully acquired; and,
3. The tools and processes employed should authenticate the preceding steps.

These standards cannot be met in every situation, but the court should require the party deviating from the accepted criteria to justify the departure.

Exemplar Acquisition Protocol

An exemplar protocol for acquisition follows, adapted from the court's decision in *Xpel Techs. Corp. v. Am. Filter Film Distribs*., 2008 WL 744837 (W.D. Tex. Mar. 17, 2008):

> The motion is GRANTED and expedited forensic imaging shall take place as follows:
>
> A. The Forensic Examiner's costs shall be borne by the Plaintiff.
>
> B. Computer forensic analysis will be performed by _____ (the "Forensic Examiner").
>
> C. The Forensic Examiner must agree in writing to be bound by the terms of this Order prior to the commencement of the work.
>
> D. Within two days of this Order or at such other time agreed to by the parties, defendants shall make its computer(s) and other electronic storage devices available to the Forensic Examiner to enable him to make forensically-sound images of those devices, as follows:
>
> i. Images of the computer(s) and any other electronic storage devices in Defendants' possession, custody, or control shall be made using hardware and software tools that create a forensically sound, bit-for-bit, mirror image of the original hard drives (e.g., EnCase, FTK Imager, X-Ways Forensics or Linux dd). A bit-stream mirror image copy of the media item(s) will be captured and will include all file slack and unallocated space.
>
> ii. The Forensic Examiner should photographically document the make, model, serial or service tag numbers, peripherals, dates of manufacture and condition of the systems and media acquired.
>
> iii. All images and copies of images shall be authenticated by MD5 hash value comparison to the original hard drive(s).
>
> iv. The forensic images shall be copied and retained by the Forensic Examiner in strictest confidence until such time the court or both parties request the destruction of the forensic image files.
>
> v. Without altering any data, the Forensic Examiner should, as feasible, determine and document any deviations of the systems' clock and calendar settings.

> E. The Forensic Examiner will use best efforts to avoid unnecessarily disrupting the normal activities or business operations of the defendants while inspecting, copying, and imaging the computers and storage devices.
>
> F. The Defendants and their officers, employees and agents shall refrain from deleting, relocating, defragmenting, overwriting data on the subject computers or otherwise engaging in any form of activity calculated to impair or defeat forensic acquisition or examination

Forensic Examination

There is no more a "standard" protocol for forensic examination than there is a "standard" set of deposition questions. In either case, a good examiner tailors the inquiry to the case, follows the evidence as it develops and remains flexible enough to adapt to unanticipated discoveries. Consequently, it is desirable for a court-ordered protocol to afford the examiner discretion to adapt to the evidence and apply their expertise.

Although the goals of forensic examination vary depending on the circumstances justifying the analysis, a common aim is recovery of deleted data.

The Perils of "Undelete Everything"

Even if the parties agree, be wary of issuing an order directing the examiner to, in effect, "undelete all deleted material and produce it." This was the court's approach in *Ameriwood Ind., Inc. v. Liberman*, 2006 WL 3825291 (E.D. Mo. Dec. 27, 2006), where the forensic examiner was ordered to recover:

> [A]ll available word-processing documents, incoming and outgoing email messages, PowerPoint or similar presentations, spreadsheets, and other files, including but not limited to those files that were "deleted." The Expert shall provide the recovered documents in a reasonably convenient and searchable form to defendants' counsel, along with, to the extent possible, the information showing when any files were created, accessed, copied, or deleted, and the information about the deletion and the contents of deleted files that could not be recovered. Id. at 13.

Although it may seem sensible at first blush, a directive that speaks in terms of all "other" files and all "deleted" files—especially one that seeks detailed information about "the contents of deleted files that could not be recovered"--creates unrealistic expectations and invites excessive cost. Here's why:

Historically, libraries tracked books by noting their locations on index cards in a card catalog. A computer manages its hard drive in much the same way. The files are the "books" and their location is tracked by a card catalog-like index called the **file table**. But there are two key differences between libraries and computer file systems. Computers employ no Dewey decimal system, so electronic "books" can be on any shelf. Further, electronic "books" may be split into chapters and those chapters stored in multiple locations across the drive. This is called "**fragmentation**." Computers depend on their file tables to keep track of all those file fragments

When a user hits "Delete," nothing happens to the file targeted for deletion; only the file table changes. It's as if someone tore up a card in the card catalogue. Like its literary counterpart, the deleted file is still on the "shelf," but now it's a needle in a haystack, lost among millions of unallocated clusters.

To recover deleted files, a computer forensic examiner employs three principal techniques:

1. File Carving by Binary Signature

Because most files begin with a unique digital signature identifying the file type, examiners run software that scans each of the millions of unallocated clusters for particular signatures, hoping to find matches. If a matching file signature is found and the original size of the deleted file can be ascertained, the software copies or "carves" out the deleted file. If the size of the deleted file is unknown, the examiner designates how much data to carve out. The carved data is then assigned a new name and the process continues.

Unfortunately, deleted files may be stored in pieces as discussed above, so simply carving out contiguous blocks of fragmented data grabs intervening data having no connection to the deleted file and fails to collect segments for which the directory pointers have been lost. Likewise, when the size of the deleted file isn't known, the size designated for carving may prove too small or large, leaving portions of the original file behind or grabbing unrelated data. Incomplete files and those commingled with unrelated data are generally corrupt and non-functional. Their evidentiary value is also compromised.

File signature carving is frustrated when the first few bytes of a deleted file are overwritten by new data. Much of the deleted file may survive, but the data indicating what type of file it was, and thus enabling its recovery, is gone.

File signature carving requires that each unallocated cluster be searched for each of the file types sought to be recovered. When a court directs an examiner to "recover all deleted files," that's an exercise that could take weeks, followed by still more weeks spent culling corrupted files. Instead, the protocol should specify the *particular* file types of interest based upon how the machine was used and the facts and issues in the case.

2. File Carving by Remnant Directory Data

In some file systems, residual file directory information revealing the location of deleted files may be strewn across the drive. Forensic software scans the unallocated clusters in search of these lost directories and uses this data to restore deleted files. Here again, reuse of clusters can corrupt the recovered data. A directive to "undelete everything" gives no guidance to the examiner respecting how to handle files where the metadata is known but the contents are suspect.

3. Search by Keyword

Where it's known that a deleted file contained certain words or phrases, the remnant data may be found using keyword searching of the unallocated clusters and slack space. Keyword search is a laborious and notoriously inaccurate way to find deleted files, but its use is necessitated in most cases by the enormous volume of ESI. When keywords are not unique or less than about 6 letters long, many false positives ("**noise hits**") are encountered. Examiners must painstakingly look at each hit to assess relevance and then manually carve out responsive data. This process can take days or weeks for a single machine.

Better Practice than "Undelete" is "Try to Find"

The better practice is to eschew broad directives to "undelete everything" in favor of targeted directives to use reasonable means to identify specified types of deleted files. To illustrate, a court

might order, “Examiner should seek to recover deleted Word, Excel, PowerPoint and PDF files, as well as to locate potentially relevant deleted files or file fragments in any format containing the terms, ‘explosion,’ ‘ignition’ or ‘hazard.’ If the examiner finds evidence of deletion of other files satisfying these criteria but which prove unrecoverable, the examiner shall, as feasible, identify such files and explain the timing and circumstances of their deletion.”

Eradication Challenges

When confidential or proprietary data ends up where it doesn’t belong, courts may order a computer forensic expert to eradicate it. Redaction of data from a hard drive is more challenging than most lawyers and judges appreciate. In fact, it's harder than some forensic examiners realize.

Files sought to be deleted may exist in multiple iterations, versions and fragments within the active and/or the deleted areas of the hard drive. There's rarely just one copy of any file that must be found and destroyed. As discussed above, the target file may be fragmented (stored in segments separated by unrelated information). If fragmented files were deleted, the pointers to the fragments may be lost, complicating the examiner’s ability to gather all the pieces needing to be erased. Reduced to manual examination of thousands or even millions of clusters, the task quickly becomes infeasible.

When framing an eradication protocol, the court should assess the lengths a party may go to in an effort to recover and use the data. Relegating accessible copies and drafts of files irrevocably to the digital trash heap may be sufficient to forestall use by ordinary users. In other circumstances, the sensitivity of the data or the sophistication and resources of the user may dictate the data be eradicated in a manner impervious to forensic recovery.

There are three principal areas where the data resides: Allocated Clusters (active data), Unallocated Clusters and File Slack Space. Importantly, the last two are not needed by users for proper function of their machines so selective eradication within these forensic areas is wasted effort (except insofar as may be required to gather evidence of the misconduct or establish damages). Instead, a sufficient eradication protocol may require the examiner to first overwrite or “double delete” (deletion followed by emptying of the recycle bin) the contraband data and then to thoroughly overwrite the entire contents of unallocated clusters and slack space. Alternatively, the protocol may provide for the examiner to relocate only benign data to a new drive and destroy or sequester the drive holding the contraband data. Either method is reasonably effective in preventing the user from regaining access to the contraband data using the sterilized drive.

Before any changes are made to the evidence drive to effect data eradication, the court should assess whether the contents of the drive with contraband data must first be preserved by forensic imaging for use as evidence in the case.

Exemplar Examination Protocol

Computer forensics examinations are often launched to resolve questions about the origins, integrity and authenticity of electronic documents. The processes employed are specialized and quite technical. Following is a list of exemplar steps that might be taken in a forensic examination to assess the alleged authoring dates of particular Excel and Word documents and e-mail:

1. Load the authenticated image into an analysis platform and examine the file structures for anomalies.

2. Assess the integrity of the evidence by, e.g., checking Registry[4] keys to investigate the possibility of drive swapping or fraudulent reimaging and looking at logs to evaluate BIOS date manipulation.
3. Look at the various creation dates of key system folders to assess temporal consistency with the machine, OS install and events.
4. Look for instances of applications that are employed to alter file metadata and seek to rule out their presence, now or in the past.
5. Gather data about the versions and installation of the software applications used to author the documents in question and associated installed hardware for printing of same.
6. Seek to refine the volume snapshot to, e.g., identify relevant, deleted folders, applications and files.
7. Carve the unallocated clusters for documents related to Excel and Word, seeking alternate versions, drafts, temp files or fragments.
8. Look at the LNK[5] files, TEMP directories, Registry MRUs[6] and, as relevant, Windows prefetch area[7], to assess usage of the particular applications and files at issue.
9. Look at the system metadata values for the subject documents and explore evidence, if any, of alteration of the associated file table entries.
10. Run keyword searches against the contents of all clusters (including unallocated clusters and file slack) for characteristic names, contents of and misspellings in the source documents, then review same.
11. Sort the data chronologically for the relevant Modified, Accessed and Created (MAC) dates to assess the nature of activity proximate to the ostensible authoring dates and claimed belated authoring dates.
12. Run a network activity trace report against, inter alia, the index.dat[8] files to determine if there has been research conducted at pertinent times concerning, e.g., how to change dates, forge documents and the like.
13. Examine container files for relevant e-mail and confirm temporal consistency. If web mail, look at cache data. If not found, carve unallocated clusters in an effort to reconstruct same.
14. Gather the probative results of the efforts detailed above, assess whether anything else is likely to shed light on the documents and, if not, share conclusions as to what transpired.

Problematic Protocols

Though the preceding is actually a simplified and focused examination protocol, it details activities clearly beyond the ken of most lawyers and judges. Effective protocols demand technical expertise to design and describe. Not surprisingly, court-ordered examination protocols seen in reported cases are frequently forensic examinations in name only or simply gloss over the actions permitted to the examiner. ***See*** Appendix A: An Illustration of Problematic Examination Protocols for a discussion of two recent decisions that exemplify the multitude of problems that can result from misguided examination protocols.

To safeguard against time- and money-wasting examinations, the court and counsel must either avail

[4] The system Registry is a complex database used by the Windows operating system to record configuration and other data pertaining to the file system and installed applications..
[5] LNK (pronounced “link’) files are shortcut files which Microsoft Windows automatically creates for the operating system’s use each time a user accesses a file or storage device.
[6] MRU stands for “Most Recently Used.” Entries called “keys” within the system Registry record the files used most recently by applications.
[7] To optimize performance, Microsoft Windows stores data revealing program usage patterns.
[8] Index.dat files store records of a user’s Internet activity, even if the user has deleted their Internet history.

themselves of expert assistance or become conversant about the technical issues presented in order to craft examination protocols that are feasible, cost-effective and calculated to achieve the desired ends. Proposed protocols should be drafted by an expert, or at least reviewed by one before maturing into an order.

Crafting Better Forensic Examination Orders

In framing a forensic examination order, it's helpful to set out the goals to be achieved and the risks to be averted. By using an aspirational statement to guide the overall effort instead of directing the details of the expert's forensic activities, the court reduces the risk of a costly, wasteful exercise. To illustrate, a court might order: "The computer forensic examiner should, as feasible, recover hidden and deleted information concerning [relevant issues and topics] from Smith's systems, but without revealing to any person(s) other than Smith's counsel (1) any of Smith's personal confidential information or (2) the contents of privileged attorney-client communications."

The court issued a clear, succinct order in **Bro-Tech Corp. v. Thermax, Inc., 2008 WL 724627 (E.D. Pa. Mar. 17, 2008)**. Though it assumed some existing familiarity with the evidence (e.g., referencing "the Purolite documents"), the examiner should have had no trouble understanding what was expected and conducting the examination within the confines of the order:

> (1) Within three (3) days of the date of this Order, Defendants' counsel shall produce to Plaintiffs' computer forensic expert forensically sound copies of the images of all electronic data storage devices in Michigan and India of which Huron Consulting Group ("Huron") made copies in May and June 2007. These forensically sound copies are to be marked "CONFIDENTIAL--DESIGNATED COUNSEL ONLY";
>
> (2) Review of these forensically sound copies shall be limited to:
> (a) MD5 hash value searches for Purolite documents identified as such in this litigation;
> (b) File name searches for the Purolite documents; and
> (c) Searches for documents containing any term identified by Stephen C. Wolfe in his November 28, 2007 expert report;
>
> (3) All documents identified in these searches by Plaintiffs' computer forensic expert will be provided to Defendants' counsel in electronic format, who will review these documents for privilege;
>
> (4) Within seven (7) days of receiving these documents from Plaintiffs' computer forensic expert, Defendants' counsel will provide all such documents which are not privileged, and a privilege log for any withheld or redacted documents, to Plaintiffs' counsel. Plaintiffs' counsel shall not have access to any other documents on these images;
>
> (5) Each party shall bear its own costs;

Of course, this order keeps a tight rein on the scope of examination by restricting the effort to hash value, filename and keyword searches. Such limitations are appropriate where the parties are seeking a small population of well-known documents, but would severely hamper a less-targeted effort.

Hashing

In the order just discussed, the court referenced MD5 hash value searches. Hashing is the use of mathematical algorithms to calculate a unique sequence of letters and numbers to serve as a "fingerprint" for digital data. These fingerprint sequences are called "message digests" or, more commonly, "hash values." It's an invaluable tool in both computer forensics and electronic discovery, and hashing is deployed by courts with growing frequency.

The ability to "fingerprint" data enables forensic examiners to prove that their drive images are faithful to the source. Further, it allows the examiner to search for files without the necessity of examining their content. If the hash values of two files are identical, the files are identical. This file-matching ability allows hashing to be used to de-duplicate collections of electronic files before review, saving money and minimizing the potential for inconsistent decisions about privilege and responsiveness for identical files.

These are the most important things for a jurist to know about hashing:

1. Electronically stored information of any type or size can be hashed;
2. The algorithms used to hash data are not proprietary, and thus cost nothing to use;
3. No matter the size of the file that's hashed, its hash value is *always* a fixed length;
4. The two most common hash algorithms are called MD5 and SHA-1;
5. No one can reverse engineer a file's hash value to reveal anything about the file;
6. The chance of two different files having matching MD5 hash values is one in 340 *trillion trillion trillion*.

A court may order the use of hash analysis to:

1. Demonstrate that data was properly preserved by recording matching hash values for the original and its duplicate;
2. Search data for files with hash values matching hash values of expropriated data alleged to be confidential or proprietary;
3. Exclude from processing and production files with hash values matching known irrelevant files, like the Windows operating system files or generic parts of common software; or,
4. Employ hash values instead of Bates numbers to identify ESI produced in native formats. Much ESI no longer lends itself to printable, page-like forms. Hash values offer a low-cost, reliable way to uniquely identify and authenticate these new forms.

Hashing is often a pivotal tool employed to conclusively identify known contraband images in prosecutions for child pornography.

Although hashing is a useful and versatile technology, it has a few shortcomings. Because the tiniest change in a file will alter that file's hash value, hashing is of little value in finding contraband data once it's been modified. Changing a file's name won't alter its hash value (because the name is generally not a part of the file), but even minimally changing its contents will render the file unrecognizable by its former hash value. Another limitation to hashing is that, while a changed hash value proves a file has been altered, it doesn't reveal how, when or where within a file changes occurred.

Frequently Asked Questions about Computer Forensics

How do I preserve the status quo without ordering a party to stop using its systems?

The ongoing use of a computer system erodes the effectiveness of any subsequent computer forensic examination and presents an opportunity to delete or alter evidence. Where credible allegations support the need for forensic examination, the best course is to immediately require that a forensically sound image of the machine or device be secured by a qualified technician and authenticated by hash value calculation. Alternatively, the party in control of the machine may agree to replace the hard drive and sequester the original drive such that it will not be altered or damaged.

A party wants to make "Ghost" images of the drives. Are those forensically sound?

No. only tools and software specially suited to the task collect every cluster on a drive without altering the evidence. Off-the-shelf software, or the failure to employ write protection hardware devices, will make changes to the evidence and fail to collect data in all of the areas important to a thorough forensic examination.

The use of Ghost imaging methods may be entirely sufficient to meet preservation duties when issues requiring computer forensics issues aren't at stake.

Do servers need to be preserved by forensically sound imaging, too?

Though forensic examiners may differ as to when exceptions apply, as a general rule, forensically sound imaging of servers is unwarranted because the manner in which servers operate makes them poor candidates for examination of their unallocated clusters. This is an important distinction because the consequences of shutting down a server to facilitate forensic acquisition may result in severe business interruption consequences to a party. Live acquisition of the server's active data areas is usually sufficient and typically doesn't require that the server be downed.

What devices and media should be considered for examination?

Though computer forensics is generally associated with servers, desktops and laptops, these are rarely the only candidates for examination. When they hold potentially relevant ESI, forensic acquisition and/or examination could encompass external hard drives, thumb drives, media cards, entertainment devices with storage capabilities (e.g., iPods and gaming consoles), online storage areas, optical media, external media (e.g., floppy and ZIP disks), co-located data centers, cell phones, personal digital assistants, automobile air bag modules, incident data recorders ("black boxes"), backup tapes and any of a host of other digital storage devices. Moreover, machines used at home, legacy machines sitting in closets or storage rooms and machines used by secretaries, assistants family members and other persons serving as proxies for the user must be considered as candidates for examination.

How intrusive is a computer forensics examination?

The intrusion associated with acquisition is a temporary loss of access to the computer or other device. To enable an examiner to make a forensically sound image, the user must surrender his or her computer(s) for several hours, but rarely longer than overnight. If a user poses no interim risk of wiping the drive or deleting files, acquisition can generally be scheduled so as not to unduly disrupt a user's activities.

A properly conducted acquisition makes no changes to the user's data on the machine, so it can be expected to function exactly as before upon its return. No software, spy ware, viruses or any other applications or malware are installed.

The intrusion attendant to forensic examination flows from the fact that such examination lays bare any and all current or prior usage of the machine, including for personal, confidential and privileged communications, sexual misadventure, financial and medical recordkeeping, storage of proprietary business data and other sensitive matters. Though it may be possible to avoid intruding on such data within the orderly realm of active data, once deleted, relevant and irrelevant data cannot easily be segregated or avoided. Accordingly, it's important for the court to either impose strict limits on the use and disclosure of such information by the examiner, or the examination should be conducted by a neutral examiner obliged to protect the legitimate discovery and privacy concerns of both sides.

What does it cost?

Though the forensic acquisition and preservation of a desktop or laptop machine tends to cost no more than a short deposition, the cost of a forensic examination can vary widely depending upon the nature and complexity of the media under examination and the issues. Forensic examiners usually charge by the hour, with rates ranging from approximately $200-$500 per hour according to experience, training, reputation and locale. Costs of extensive or poorly targeted examinations can quickly run into five- and six-figures. Nothing influences cost more than the scope of the examination. Focused examinations communicated via clearly expressed protocols tend to keep costs down. Keyword searches should be carefully evaluated to determine if they are over- or under inclusive. The examiner's progress should be followed closely and the protocol modified as needed. It's prudent to have the examiner report on progress and describe work yet to be done when either hourly or cost benchmarks are reached.

Further Reading

Other Articles by Craig Ball

Four on Forensics, available without cost at **http://www.craigball.com/CF4_0807.pdf**
This collection of articles includes, "***Computer Forensics for Lawyers Who Can't Set a Digital Clock*,"** an in-depth but accessible look at the nuts-and-bolts of computer forensics, written for the non-technical reader. Also included are, "*Meeting the Challenge: E-mail in Civil Discovery*," "*Finding the Right Computer Forensics Expert*" and "*Cross-examination of the Computer Forensic Expert*."

Eight on EDD, available without cost at **http://www.craigball.com/EDD8_May_2008.pdf**
This collection focuses on a wide range of electronic discovery topics including:

- **Musings on E-Discovery—Ball in Your Court: April 2005 through June 2008:** The award winning electronic discovery column from Law Technology News.
- **Hitting the High Points of the New e-Discovery Rules**: A thumbnail summary of the e-discovery Amendments to the Federal Rules and some of the ways they change the landscape of litigation.
- **What Judges Should Know About Discovery from Backup Tapes:** The e-discovery wars rage in the mountains of e-mail and flatlands of spreadsheets, but nowhere is the battle so pitched as in the trenches of back up tapes. Here's why, and how not to end up a casualty.
- **The Plaintiff's Guide to Meet and Confer**: Learning to navigate the Rule 26(f) conference and its ilk--asking the right questions and being ready with the right answers—is an essential

advocacy skill. This article instructs requesting parties how to make the most of meet and confer in ways that balance the need for information against the requisite costs and burden.

- **Metadata: Beyond Data About Data**: What's metadata, and why is it so important? It's the electronic equivalent of DNA, ballistics and fingerprint evidence, with a comparable power to exonerate and incriminate. All sorts of metadata can be found in many locations. Some is crucial evidence; some is digital clutter. But because every active file stored on a computer has some associated metadata, it's never a question of whether there's metadata, but what kinds of metadata exist, where it resides and whether its potential relevance demands preservation and production.
- **The Perfect Preservation Letter**: This article looks at what is usually the requesting party's first foray into EDD: the letter demanding preservation of electronic evidence. A well-drafted preservation letter serves as the e-discovery blueprint, and the considerations that go into drafting the "perfect" preservation letter reveal much about the power and perils of EDD. An exemplar letter is included.
- **The Plaintiff's Practical Guide to E-Discovery**: This two-part article focuses on the needs of the requesting party. Part I addresses challenges unique to EDD, elements of a successful e-discovery effort and steps to compel preservation of e-evidence. Part II looks at the pros and cons of production formats, explores common e-mail systems and offers tips for getting the most out of your e-discovery efforts and budget.
- **Discovery of Electronic Mail: The Path to Production**: This article outlines issues and tasks faced in production of electronic mail—certainly the most common and perhaps the trickiest undertaking in electronic discovery. It's a guide to aid attorneys meeting and conferring with opposing counsel, working with e-discovery service providers, drafting production requests and explaining the cost and complexity of e-mail production to clients and the court.

Published by the Federal Judicial Center

Managing Discovery of Electronic Information: A Pocket Guide for Judges by Barbara J. Rothstein, Ronald J. Hedges, and Elizabeth C. Wiggins
Available at **http://www.fjc.gov/public/pdf.nsf/lookup/eldscpkt.pdf/$file/eldscpkt.pdf**

Published by the U.S. Department of Justice, National Institute of Justice

Electronic Crime Scene Investigation: A Guide for First Responders, Second Edition, available at **http://www.ncjrs.gov/pdffiles1/nij/219941.pdf**

Appendix A: Problematic Protocols: Two Recent Decisions

A well-crafted protocol is the key to a successful forensic examination—one that employs the most efficient and cost-effective tools and methods for analyzing the particular type and quantity of ESI presented. Two recent decisions exemplify the problems that flow from inadequate examination protocols.

Consider this protocol from **Ferron v. Search Cactus, L.L.C., 2008 WL 1902499 (S.D. Ohio Apr. 28, 2008):**

> 1. Within seven days of the date of this Opinion and Order, Plaintiff's forensic computer expert shall mirror image both of Plaintiff's computer systems' hard drives and Plaintiff shall preserve this mirror image.
> 2. Plaintiff's forensic computer expert shall then remove only Plaintiff's confidential personal information from the mirror image of Plaintiff's computer systems' hard drives. Plaintiff's expert shall provide Defendants with the protocol he utilized to remove the confidential information.
> 3. Plaintiff shall then provide Defendants' computer forensic expert access to his computer systems' hard drives.
> 4. Defendants' forensic computer expert shall mirror image Plaintiff's computer systems' hard drives in approximately four to eight hours for each system. If the expert finds that this is not enough time, Plaintiff is expected to be reasonable in allowing some additional time. Defendant is expected to be considerate with regard to scheduling times that are less intrusive to Plaintiff and his business.
> 5. Defendants' expert shall review his findings in confidence with Plaintiff prior to making any findings available to Defendants.
> 6. Plaintiff shall identify for deletion any information that is irrelevant and create a specific privilege log of any relevant information for which he claims privilege. The computer forensic expert shall remove the information claimed as privileged and provide all other information to Defendants.
> 7. Defendants' expert shall provide Plaintiff with the protocol he utilized to remove the privileged information.
> 8. Forensic computer experts [omitted] shall act as officers of this Court. Defendants shall be responsible for remunerating [Defendant's expert] and Plaintiff shall be responsible for remunerating [Plaintiff's expert].

It's unclear whether the plan is for plaintiff's expert to sterilize drive images before making *the images* available to the other side's examiner or whether the unsterilized source *hard drives* will be made available to the defendant's expert for imaging and examination. A literal reading supports the latter conclusion, but then what's the point of plaintiff's sterilization effort? Moreover, the order ignores the Herculean challenge faced in thoroughly cleansing a drive of particular confidential information in anticipation of forensic examination. If, for example, the plaintiff's e-mail included both confidential and discoverable messages, the examiner would be obliged to obliterate and reconstitute the plaintiff's e-mail container file. Additionally, the unallocated clusters typically carry tens or hundreds of gigabytes of commingled, undifferentiated data. Finally, the order offers no guidance as to what the examiners are allowed or expected to do, or whether the defendant's expert is limited in what he can share with defense counsel. The order appears detailed, but offers practically no pertinent guidance.

Another recent case, **Coburn v. PN II, Inc., 2008 WL 879746 (D. Nev. Mar. 28, 2008),** exemplifies the difficulties attendant to programming a forensic examination in a court order. The court modeled its order on the protocol in **Playboy Ent., Inc. v. Welles, 60 F.Supp.2d 1050 (S.D. Cal. 1999).** While there is much to commend the order in terms of addressing the choice of expert, privilege concerns, confidentiality and convenience, the order is silent as to whether or how the forensic expert will recover information or analyze the data.

> 1. The parties shall meet, confer and agree upon the designation of a computer expert who specializes in the field of electronic discovery to create a "mirror image" of the relevant hard drives. If the parties cannot agree on an expert, they shall submit suggested experts to the court by April 18, 2008. The court will then select and appoint a computer specialist. The services of the expert will be paid by defendants.
>
> 2. The court appointed computer specialist will serve as an officer of the court. To the extent the computer specialist has direct or indirect access to information protected by the attorney-client privilege, such "disclosure" will not result in a waiver of the attorney-client privilege. Defendants herein, by requesting this discovery, are barred from asserting in this litigation that any such disclosure to the court designated expert constitutes any waiver by Coburn of the attorney-client privilege. The computer specialist will sign a protective order stipulated to by the parties. Lastly, any communications between defendants and/or defendants' counsel and the computer specialist as to the payment of fees and costs pursuant to this order will be produced to Coburn's counsel.
>
> 3. The parties shall agree on a day and time to access Coburn's computer. Defendants shall defer to Coburn's personal schedule in selecting this date. Representatives of both parties shall be informed of the time and date, but only Coburn and her counsel may be present during the hard drive recovery.
>
> 4. After the computer specialist makes a copy of Coburn's hard drives, the "mirror image" (which the court presumes will be on or transferred to a disk(s)) will be given to Coburn's counsel. Coburn's counsel will print and review any recovered documents and produce to defendants those communications that are responsive to any earlier request for documents and relevant to the subject matter of this litigation. Such discovery shall include, but not be limited to, information pertaining to defendants' contention that Coburn misappropriated their trade secrets. While no counterclaim for misappropriation of trade secrets has been made, such information is relevant to, and is reasonably calculated to lead to admissible evidence concerning, Coburn's alleged damages and emotional distress, and for impeachment purposes. All documents that are withheld on a claim of privilege will be recorded in a privilege log.
>
> 5. Coburn's counsel will be the sole custodian of and shall retain this "mirror image" disk(s) and copies of all documents retrieved from the disk(s) throughout the course of this litigation. To the extent that documents cannot be retrieved from Coburn's computer hard drives or the documents retrieved are less than the whole of data contained on the hard drives, Coburn's counsel shall submit a declaration to the court together with a written report signed by the designated expert explaining the limits of retrieval achieved.
>
> 6. The "mirror image" copying of the hard drives, and the production of relevant documents, shall be completed by May 30, 2008.

Notably, all the expert does is duplicate the drive and hand it over to counsel for printing and review of any "recovered" documents; but the expert isn't permitted to *recover* any documents, and presumably Coburn's counsel is ill-equipped to conduct a forensic examination or recover any hidden or deleted data without expert assistance. The "forensic" nature of the process is illusory because the examiner isn't permitted to do more than duplicate the hard drive. The party seeking forensic examination is in no wise aided because the process isn't calculated to expose new evidence. Coburn already had access to the contents of her hard drive, and Coburn's counsel was already under an obligation to search same for responsive ESI. It's an empty, expensive exercise.

What Judges Should Know about Discovery from Backup Tapes

Craig Ball

What Judges Should Know About Discovery from Backup Tapes

By Craig Ball

The electronic discovery wars rage in the mountains of e-mail and flatlands of spreadsheets, but nowhere is the battle so pitched as in the trenches of back up tapes, those vast electronic packing crates at the heart of front page cases like *Zubulake v. UBS Warburg* and *Coleman (Parent) Holdings v. Morgan Stanley*.

Why are backup tapes such troublemakers?

Ideally, the contents of a backup system would be entirely cumulative of the active "online" data on the servers, workstations and laptops that make up a network. But because businesses entrust the power to destroy data to every computer user--including some moved to make evidence disappear—and because companies configure systems to purge electronically stored information as part of records retention plans, backup tapes may be the only evidence containers beyond the reach of those who've failed to preserve evidence or inclined to destroy or fabricate it. Going back as far as Col. Oliver North's deletion of e-mail subject to subpoena in the Iran-Contra affair, it's long been the backup systems that ride to truth's rescue with the "smoking gun" evidence.

But, the unique, "last resort" information on backup tapes may be drops in an ocean of irrelevant, duplicative or privileged data on the tapes. Tapes can also be the target of aimless fishing expeditions mounted without regard for the cost and burden of restoring tapes, or tape may be targeted prematurely, before more accessible data sources have been exhausted.

All of this becomes the court's problem when discoverable information may reside on backup tape and:

1. A party obliged to preserve ESI overwrites or discards tapes;
2. A party seeks to be released from a litigation hold on backup tapes;
3. Restoration and review of backup tape will engender significant delays; or,
4. The burden or cost of tape restoration is unreasonable given the circumstances of the case.

Grappling with Backup Tapes

Backup tapes are copies of data on a computer obtained for *disaster recovery*, i.e., picking up the pieces of a damaged or corrupted system. Some call backups "snapshots" of data and, like a photo, backup tapes capture only what's in focus. Backup tapes store data in significantly ways from the computer systems they protect. To save time and space, backups typically store less information about individual files and ignore commercial software programs that can be reinstalled in the event of disaster, so *full backups* tend to focus on all *user created* data and, *differential backups* hold all files created or changed since the last full backup and *incremental backups* grab just what's been created or changed since the last incremental backup. Together, they put Humpty-Dumpty back together again in a process called *tape restoration*.

Jargon Watch
disaster recovery
full backup
differential backup
incremental backup
tape restoration
tape rotation
legacy tapes
serial access
vertical deduplication

Tape is cheap, durable and portable, the last important because backups need to be stored away from the systems at risk. But, tape is also slow and cumbersome--foibles forgiven because it's so rarely needed for restoration. Back up systems have but one legitimate purpose, being the retention of data required to get a business

information system "back up" on its feet in the event of disaster. A business only needs disaster recovery data for a brief interval since no business wants to replicate its systems as they were six months or even six weeks before a crash. As the only backup tapes that matter are the last complete set before the river rose, *in theory*, older tapes are supposed to be recycled by overwriting them in a practice called "*tape rotation*."

But, as theory and practice are rarely on speaking terms, companies may keep backup tapes long past their usefulness for disaster recovery--often *years* past, and even beyond the companies' ability to access tapes created with obsolete software or hardware. These *legacy tapes* are business records—sometimes the last surviving copy—but afforded little in the way of *records management*. Even businesses that overwrite tapes every two weeks replace their tape sets from time to time as faster, bigger options hit the market. The old tapes are often set aside and forgotten in offsite storage or a box in the corner of the computer room.

Like the Delorean in "Back to the Future," legacy tapes allow you to travel back in time. It doesn't take 1.2 million gigawatts of electricity, just lots of cabbage.

Why is Tape So Slow?
Actually, tape is pretty remarkable technology that's seen great leaps in speed and capacity.

Still, there are those pesky laws of physics.

Tape is *serially accessed media,* meaning you must plow through its contents to get to what you're seeking. It's like trying to find the start of a show on a VCR tape: you fast forward or rewind to get there. What's more, information on tape may be recorded in a serpentine fashion, like a mountain switchback, so the tape drive must shuttle back and forth through the entire tape repeatedly to get to the data. As this is a mechanical process, it's occurs at a glacial pace relative to the speed with which computer circuits or even hard drives move data.

Although newer backup tape technologies build in some indexing features, older systems are limited in their ability to find and extract particular files. Further, recalling that backup is an incremental process, reconstructing reliable data sets may require data from multiple tapes to be combined. Add to the mix the fact that as hard drive capacities have exploded, tape must store more and more information to keep pace. Gains in performance are offset by growth in volume.

But remember, tape is cheap, durable and portable. As the cost of hard drives and other media plummets, tape will go the way of the floppy disk; but it's with us for some years yet.

Restoration Environment
The restored contents of back up tapes have to *go* somewhere. It is, after all, a snapshot of *all* user-generated data at one or more points in time; so, the volume of restored information can be considerable. Small and mid-size companies typically lack the idle capacity to effect restoration without a significant investment in equipment and storage. Larger enterprises devote more stand-by resources to disaster recovery and may have alternate environments ready to receive restored data, but those resources must be at the ready in the event of emergency. It may be unacceptably risky to dedicate them, even briefly, to electronic discovery.

In assessing the burden of in-house tape restoration, courts should weigh the cost and effort needed to set up a system to receive the restored data. It frequently proves to be less costly and disruptive to

engage a vendor specializing in tape restoration for electronic discovery than to undertake the task in-house; consequently, courts should expect to receive testimony about the relative costs and burdens of each option.

Deduplication

Companies that archive backup tapes may retain years of tapes, numbering in the hundreds or thousands. Because each full backup is a snapshot of a computer system at the time it's created, there is a substantial overlap between backups. An e-mail in a user's Sent Items mailbox may be there for months or years, so every backup replicates that e-mail, and restoration of every backup adds an identical copy to the material to be reviewed. Restoration of a year of monthly backups would generate 12 copies of the same message, thereby wasting reviewers' time, increasing cost and posing a risk of inconsistent treatment of identical evidence (as occurs when one reviewer flags a message as privileged but another decides it's not).

Consider, too, how many messages and attachments are dispatched to all employees or members of a product team. Across an enterprise, there's a staggering level of repetition.

Accordingly, an essential element of backup tape restoration is deduplication; that is, using computers to identify and cull identical electronically stored information before review. Deduplicating within a single custodian's mailboxes and documents is called "*vertical deduplication*," and it's a straightforward process. However, corporate backup tapes aren't geared to single users. Instead, business' backup tapes hold messages and documents for multiple custodians storing identical messages and documents. Restoration of backup tapes generates duplicates within individual accounts (vertically) and across multiple users (horizontally). Deduplication of messages and documents across multiple custodians is called (not surprisingly) "*horizontal deduplication*."

Horizontal deduplication significantly reduces the volume of information to be reviewed and minimizes the potential for inconsistent characterization of identical items; however, it can make it impossible to get an accurate picture of an individual custodian's data collection, since many constituent items may be absent, eliminated after being identified as identical to another user's items.

Tips on Tapes

Tape ≠ Inaccessible

Tape, like paper, is just a medium to store information. Knowing that information is on paper tells you nothing about the complexity of the content. Is it, "See Spot run" or particle physics? Is it in English or Swahili?

Litigants may mistakenly speak of backup tape as "inaccessible," but unless physically damaged or written in some obscure format, tape is not inaccessible. What they *mean* to argue is that the expenditure of resources required to access the contents is unreasonable.

Accessing some backup systems requires no more than popping in a tape and searching its contents. But, backup systems of large companies with multiple business units are enormously complex and may lack straightforward correspondence between individuals and data. Every system is different.

Consequently, there is no rule-of-thumb for accessibility of backup tapes. When challenged, it's the responsibility of the party claiming inaccessibility to bring forward proof of unreasonable burden or cost.

Proof

Proof of undue burden or cost must consist of more than lawyer gesticulation and stentorian protest. It should come from testimony of IT personnel, outside vendors or qualified experts recounting the actual cost and time to complete a restoration of information on tape. Such effort should segregate the cost of search, filtering or attorney review, these being costs ordinarily borne by a responding party searching accessible ESI.

The court may want to inquire:

- Does the information on the tape exist in any more accessible forms?
- Can the responding party offer a reasonable alternative to restoration of backup tapes that will afford the requesting party comparable access to the information sought?
- Does the responding party routinely restore backup tapes in the ordinary course to, e.g., insure the system is functioning properly or as a service to those who have mistakenly deleted files?
- Have any of the backup tapes at issue been restored in other circumstances and thus exist as accessible information in other cases or held by third parties?
- Does the responding party have the system capacity and in house expertise to restore the data? *Not everyone has the idle system resources or personnel required to temporarily restore a prior version of the data alongside the current version.*
- Can the contents be searched and selectively extracted without wholesale restoration of the tapes? *Emerging software and tape technologies sometimes make this feasible.*
- Has the responding party compared its in-house restoration cost against the services of so-called "tape houses" equipped to process large numbers of tapes at competitive prices? *"Do it yourself" is not always cheaper.*

Threshold Question

The oft-overlooked threshold question is, "What is the likelihood the tapes contain relevant evidence?"

If backup tapes hold potentially responsive information, but a responding party declines to search or produce the contents as being not reasonably accessible, the responding party must identify the tapes and, as feasible, "provide enough detail to enable the requesting party to evaluate the burdens and costs of providing the discovery and the likelihood of finding responsive information" on the tape. *Committee Notes to FRCP Rule 26(b)(2)(B).*

How can a party gauge the likelihood of finding responsive information on tapes without searching them?

The answer lies in recognizing that backup tapes don't exist in a vacuum but as part of an information *system*. A properly managed system incorporates labeling, logging and tracking of tapes, permitting reliable judgments to be made about what's on particular tapes insofar as tying contents to business units, custodians, machines, data sets and intervals. If the responding party has so poorly managed backup tapes that nothing can be apprehended about their contents, the court must decide whether it's fair to deny access to the evidence or shift costs. In such case, the Court may opt to instruct the

responding party to generate an index of contents but defer requiring a full restoration and review pending examination of the index.

First, Pick the Low Hanging Fruit

Too often, parties battle about backup tapes before they've looked to see what's in easily accessible sources and material already produced. Be sure that the parties have exhausted accessible sources for the information. Has the producing party searched the contents of servers, desktops, laptops, external hard drives, handheld devices and removable media like CDs and floppies that may hold the information also stored on tape? Restoring a few tapes can start to look attractive to a party who hasn't thoroughly explored active data repositories. Searching for information costs money, and it may be preferable to look in one harder-to-access place than a hundred easy ones.

Imposing Conditions on Discovery from Tapes

If the court determines that backup tapes are reasonably accessible, then the responding party is obliged to search them and produce responsive information at its own cost, subject to the FRCP Rule 26(b)(2)(C) limitations that apply to all discovery. FRCP Rule 26(b)(2)(B). But, if the court determines that the tapes are not reasonably accessible based on cost or burden, then the requesting party must satisfy the court that there is good cause to compel production from the tapes. Id. If the court then opts to order discovery from the tapes, the court may wish to specify conditions for the discovery to ameliorate hardship and guard against overreaching. Such conditions might include:

1. Sampling to explore relevance

Sampling backup tapes is like drilling for oil: You identify the best prospects, drill exploratory wells and if you hit dry holes, you pack up and move on. But, if a well starts producing, you keep on developing the field.

Sampling backup tapes entails selecting parts of the tape collection deemed most likely to yield responsive information and restoring and searching only those selections before deciding whether to restore more tapes. The size and distribution of the sample hinges on many variables, among them the breadth and organization of the tape collection, relevant dates, fact issues, business units and custodians, resources of the parties and the amount in controversy. Ideally, the parties can agree on a sample size or they can be encouraged to arrive at an agreement through a mediated process. If the parties cannot agree, jurists typically apply their own well-honed sense of fairness. However, the court may wish to seek guidance from a knowledgeable expert as sampling is a well-defined statistical discipline and modest adjustments to sample size and selection can have a substantial impact on reliability.

Recognizing that a backup snapshot often consists of more than a single tape, the court should take pains to insure that each sample is complete for a selected date or interval; that is, the number of tapes shouldn't be arbitrary but should fairly account for the totality of information captured in a single backup event. To better understand this undertaking, look at the example of a backup report and tape inventory included as Appendix 1 of this article. Making the leap between a single custodian's e-mail messages and the tapes which house them is not apparent and may require the court to compel the producing party to furnish some technical assistance to facilitate an effective sampling scheme.

As important as the sample size and distribution is the question of who gets to select the backup sets for particular dates, machines, business units or custodians that will comprise the sample? Here, the best approach is to either allow the requesting party to specify the goals of the examination and require the producing party to tailor the selection to best meet those goals or, better still, require the

producing party to share sufficient information and technical assistance to allow the requesting party to reliably gauge what might be on the tape (by date, business unit, users or whatever information can be ascertained), then let the requesting party choose the "slices" of data to restore and search.

For further discussion of backup tape sampling, see *Hagemeyer v. Gateway*, 222 F.R.D. 594 (E.D.Wis. 2004), *Zubulake v. UBS Warburg*, 217 F.R.D. 309 (S.D.N.Y. 2003) and McPeek v. Ashcroft, 202 F.R.D. 31 (D.D.C. 2001)

2. Shifting the cost of making information on tape accessible for search and review

Cost shifting is a magical tool. It has the power to transform parties into more reasonable, efficient and cooperative creatures or to slam the courthouse door to persons of modest means with meritorious claims and defenses. If the court determines that backup tapes may contain responsive information, the court may order the requesting party to bear the reasonable cost of converting the contents of those backup tapes to more accessible forms, e.g., the cost to have a tape conversion service provider extract the compressed data to accessible formats on external hard drives. Once the information is made reasonably accessible, the cost to review and produce from the accessible sources remains the responsibility of the producing party.

This is frequently an equitable approach, but it should be used with care. A party's willingness to bear the cost of restoration is insufficient to justify discovery from sources unlikely to contain responsive information, and it may nonetheless impose an unreasonable burden in terms of the added volume of data for review. The well-heeled shouldn't get broader access to an opponent's ESI or unfairly increase an opponent's review burden just because they can afford to be curious.

The decisions in *Wiginton v. CB Richard Ellis, Inc*., 229 F.R.D. 568 (N.D.Ill. 2004) offers a helpful analysis for courts to undertake when weighing cost shifting in connection with backup tape restoration.

3. Permitting an inspection of contents

Whatever happened to making information available for inspection as it was kept in the usual course of business? Backup tapes are the modern counterpart to banker's boxes in the warehouse. The court may wish to explore the feasibility of affording a requesting party access to duplicates of backup tapes in order to permit them to inspect contents, with appropriate safeguards for privilege, privacy and trade secret considerations. The producing party may need to offer technical assistance to insure that the requesting party can access the information on tape. This may entail, *inter alia,* sharing passwords required to decrypt locked data and details concerning the computing environment and backup software and hardware. It may be necessary for the producing party to afford access to technical manuals or versions of backup software no longer commercially available. *Compel cooperation.*

Restoration of Backup Tapes in Mitigation of Spoliation

A circumstance where production of responsive information from back up tapes may be unconditionally ordered notwithstanding cost or burden is where the responding party failed in its duty to preserve information in accessible sources. This can occur when, by guile or negligence, a party deletes, discards or corrupts information that should have been subject to a litigation hold directive or when a party fails to discontinue routine migration of information from sources easily accessed to less accessible formats. Certainly, an organization need not discontinue its routine archival of information, but neither should that migration operate to deny an opponent access to information that the producing party was bound to preserve for use in an anticipated suit or claim.

Conclusion

Backup tapes epitomize the cross purposes of information technology and litigation. Compelling a large organization to interrupt its tape rotation, set aside backup tapes and purchase a fresh set can carry a princely price tag; but if the tapes aren't preserved, deleted data may be gone forever. Must a litigant forego essential evidence or pay more to secure it than the amount in controversy? *In a world where virtually no electronic evidence disappears, but only gets harder to reach, how much is enough?* These are a few of the Hobson's choices of e-discovery, but they are made easier when the court understands the technical challenges and can help fashion practical, equitable solutions.

Appendix 1: Example of Backup Tape Backup Report and Tape Inventory

2/19/2008 4:51:52 PM HQEXCH002 Root [ID /var/log/notice] Solstice Backup Release 5.1.1
Windows Storage Server

Start Time: 2/18/2008 00:03:35
End Time: 2/19/2008 4:51:52
--Successful Saved Sets--

HQEXCH002_A	/	LEVEL = FULL
HQEXCH002_A	/BIN	LEVEL = FULL
HQEXCH002_A	/DEV	LEVEL = FULL
HQEXCH002_A	/HQ	LEVEL = FULL
HQEXCH002_A	/HQ/MS	LEVEL = FULL
HQEXCH002_A	/HQ/MS/EXCH	LEVEL = FULL
HQEXCH002_A	/HQ/MS/EXCH/CORP	LEVEL = FULL
HQEXCH002_B	/HQ/MS/EXCH/LAW	LEVEL = FULL

HQ_Exchange Server: Backup Tape Inventory (Format: LTO Ultrium 1)

Tape Number	*Date Recorded*	*Volume*		*% Used*	*Pool*
BU123DLT	**2/12/2008**	***135***	***GB***	***full***	***DAILY***
BU124DLT	**2/12/2008**	***154***	***GB***	***full***	***DAILY***
BU125DLT	**2/12/2008**	***130***	***GB***	***full***	***DAILY***
BU126DLT	**2/12/2008**	***168***	***GB***	***full***	***DAILY***
BU127DLT	**2/12/2008**	***116***	***GB***	***full***	***DAILY***
BU244DLT	**2/12/2008**	***128***	***GB***	***full***	***DAILY***
BU245DLT	**2/12/2008**	***131***	***GB***	***full***	***DAILY***
BU303DLT	**2/12/2008**	***145***	***GB***	***full***	***DAILY***
BU304DLT	**2/12/2008**	***157***	***GB***	***full***	***DAILY***
BU306DLT	**2/12/2008**	***135***	***GB***	***full***	***DAILY***
BU115DLT	**2/13/2008**	***130***	***GB***	***full***	***DAILY***
BU118DLT	**2/13/2008**	***118***	***GB***	***full***	***DAILY***
BU128DLT	**2/13/2008**	***111***	***GB***	***full***	***DAILY***
BU164DLT	**2/13/2008**	***117***	***GB***	***full***	***DAILY***
BU301DLT	**2/13/2008**	***120***	***GB***	***full***	***DAILY***
BU302DLT	**2/13/2008**	***116***	***GB***	***full***	***DAILY***
BU120DLT	***2/14/2008***	***45***	***GB***	***45%***	***DAILY***
BU091DLT	***2/15/2008***	***4434***	***MB***	***4%***	***DAILY***
BU228DLT	***2/15/2008***	***5359***	***MB***	***5%***	***DAILY***

BU553DLT	***2/15/2008***	***114***	***GB***	***full***	***DAILY***
BU225DLT	**2/16/2008**	***8391***	***MB***	***8%***	***DAILY***
BU093DLT	***2/17/2008***	***123***	***GB***	***100%***	***DAILY***
BU095DLT	***2/17/2008***	***40***	***GB***	***41%***	***DAILY***
BU096DLT	***2/18/2008***	***60***	***GB***	***60%***	***DAILY***
BU098DLT	**2/18/2008**	***32***	***GB***	***32%***	***DAILY***
BU554DLT	***2/18/2008***	***116***	***GB***	***full***	***DAILY***
BU555DLT	***2/19/2008***	***167***	***GB***	***full***	***YEARLY***
BU556DLT	***2/19/2008***	***149***	***GB***	***full***	***YEARLY***
BU557DLT	***2/19/2008***	***166***	***GB***	***full***	***YEARLY***
BU558DLT	***2/19/2008***	***138***	***GB***	***full***	***YEARLY***
BU559DLT	***2/19/2008***	***128***	***GB***	***full***	***YEARLY***

Musings on Electronic Discovery

"Ball in Your Court" Columns Selected for Judges August 2005 – April 2008

© Craig Ball

The *Law Technology News* column "Ball in Your Court" is the 2008 honoree as "Best How-To Article" by the American Society of Business Publication Editors (ASBE). It was also the 2007 Gold Medal honoree as "Best Regular Column" as awarded by Trade Association Business Publications International and the 2007 ASBE Silver Medalist honoree as "Best Contributed Column" and their 2006 Silver Medalist honoree as "Best Feature Series" and "Best Contributed Column."

You can download the complete collection of "Ball in Your Court" from http://www.craigball.com/BIYC.pdf. New columns run each month at www.lawtechnews.com.

Contents

Don't Try This at Home

by Craig Ball

[Originally published in Law Technology News, August 2005]

The legal assistant on the phone asked, "Can you send us copies of their hard drives?"

As court-appointed Special Master, I'd imaged the contents of the defendant's computers and served as custodian of the data for several months. The plaintiff's lawyer had been wise to lock down the data before it disappeared, but like the dog that caught the car, he didn't know what to do next. Now, with trial a month away, it was time to start looking at the evidence.

"Not unless the judge orders me to give them to you," I replied.

The court had me act as custodian because the discoverable evidence on a hard drive lives cheek by jowl with all manner of sensitive stuff, such as attorney-client communications, financial records and pictures of naked folks engaged in recreational activity. In suits between competitors, intellectual property and trade secrets such as pricing and customer contact lists need protection from disclosure when not evidence. As does all that full-of-surprises deleted data accessible by forensic examination.

"Even if the court directs me to turn over the drive images, you probably won't be able to access the data without expert assistance."

I explained that, like most computer forensic specialists, I store the contents of hard drives as a series of compressed image files, not as bootable hardware that can be attached to a computer and examined. Doing so is advantageous because the data is easier to access, store and authenticate, as well as far less prone to corruption by the operating system or through examination. Specialized software enables me to assemble the image files as a single virtual hard drive, identical in every way to the original. On those rare occasions when a physical duplicate is needed, I reconstitute those image files to a forensically sterile hard drive and use cryptographic algorithms to demonstrate that the restored drive is a faithful counterpart of the original. Of course, putting the digital toothpaste back in the tube that way takes time and costs money.

"Do we ask the court for a restored drive?"

"You could," I said, "and you might get it if the other side doesn't object."

Incredibly, lawyers who'd never permit the opposition to fish about in their client's home or office blithely give the green light when it comes to trolling client hard drives. No matter how much you want to demonstrate good faith or that your client has "nothing to hide," be wary of allowing the other side to look at the drives.

Even when you've checked the contents, you can't see all that a forensic exam can turn up, and your client may not tell you about all those files she deleted last night.

"But," I warned, "as soon as you attach the drive to your computer and start poking around, you'll alter the evidence."

Microsoft Windows acts like a dog marking territory. As soon as you connect a hard drive to Windows, the operating system writes changes to the drive. Forensic examiners either employ devices called "write blockers" to intercept these alterations or perform their examination using operating systems less inclined to leave their mark all over the evidence. Without similar precautions, opening files, reading e-mail or copying data irretrievably alters file metadata, the data-about-data revealing, inter alia, when a file was last modified, accessed or created. You may find the smoking gun, but good luck getting it into evidence when it emerges you unwittingly altered the data! This is why smart lawyers never "sneak a peek" at digital evidence.

"It'd be a violation of the software licensing to use the programs on the duplicate so you'll need to have the right software to read the e-mail and other documents and to crack any passwords you run into. However, you can't load your software on the duplicate drive because that will overwrite recoverable deleted files. Don't forget to take steps to isolate the system you'll use for examination from your office network and the internet as well as to…."

She stopped me. "We shouldn't be doing this ourselves, should we?"

"Not unless you know what you're doing. Anyway, I doubt the court will allow it without a showing of good cause and some provision to protect privileged and non-discoverable confidential data."

Now I got the question I was waiting for: "What should we do?"

"As the court's neutral," I answered, "I'm not in a position to answer that question, but before I'd burn a lot of time and money pursuing electronic discovery of particular media, I'd work out the answers to, 'What's this case about, and what am I really looking for?'"

What I wanted to add is that electronic discovery is no more about hard drives than traditional discovery was "about" paper. The hard drive is just a gigantic file cabinet, locked up like some Houdini vanishing act and packed with contents penned in Sanskrit. We don't gear discovery to metal boxes, big or small.

Sure, it's smart to focus on specific media and systems when you seek preservation, but when your goal is discovery, media ceases to be an end in itself. Then, the objectives are the e-mail, documents and other digital evidence relating to the issues in the case, narrowly targeted by time, topic, and custodian. Sorry Marshall McLuhan, it's not the medium. It's the message.

A Golden Rule for E-Discovery

by Craig Ball

[Originally published in Law Technology News, March 2006]

Albert Einstein said, "In the middle of every difficulty lies opportunity." Electronic data discovery is certainly one of the greatest difficulties facing litigants today. So wouldn't you know some genius would seize upon it as an opportunity for abuse? Perhaps Einstein meant to say, “In the middle of every difficulty is an opportunity for lies.”

I'm not talking about the pyrotechnic failures to produce email or account for back up tapes that brought low the mighty in such cases as *Zubulake v. UBS Warburg* and *Coleman (Parent) Holdings v. Morgan Stanley*. Stonewalling in discovery predated electronic discovery and will likely plague our progeny’s progeny when they grapple with photonic or neuronal discovery. But while an opponent’s “No, we won’t give it to you,” may be frustrating, it’s at least sufficiently straightforward to join the issue and promote resolution. The abuses lately seen make stonewalling seem like fair play.

Playing the Telephone Game

I’m talking sneaky stuff, like printing electronic information to paper, then scanning and running it through optical character recognition (OCR), or "printing" electronic information to a TIFF image format then OCR’ing the TIFF.

If you’ve played the parlor game, “Telephone,” you’ve seen how transmitting messages introduces errors. The first listener interprets the message, as does the next listener and the next. Each mangles the message and the errors compound hilariously. “Send reinforcements--we’re going to advance” emerges as, “Send three and four pence--we’re going to a dance.”

When you print electronic evidence, part of the message (*e.g.,* its metadata) is lost in the printing. When you scan the printout, more distortion occurs, and then optical character recognition further corrupts the message, especially if the scanned image was askew, poorly resolved or included odd typefaces. Page layouts and formatting suffer in the translation process, too. If you’re lucky, what emerges will bear a resemblance to the original evidence. If not, the output will be as distorted as the Telephone game message, but no laughing matter. Much of its electronic searchability is gone.

Speaking on a panel at New York LegalTech 2006, I groused, "Imaging data to TIFF and then OCR'ing it ought to be a crime in all 50 states." Was I surprised when that drew applause from the EDD-savvy audience! Their enthusiastic response confirmed that others are fighting TIFF/OCR abuse, too.

There's always been gamesmanship in discovery, but it wasn't hard to detect dirty pool with paper. Bad copies *looked* bad. Redaction stood out. Page numbers and dates exposed omission. But e-discovery creates fresh-and-furtive opportunities for shenanigans, and they're harder to detect and prove.

Bad OCR

Take OCR. We tend to think of optical character recognition as a process that magically transforms pictures of words into searchable text. OCR is OCR, right? In fact, error rates for OCR applications

vary widely. Some programs are superb, correctly interpreting better than 99% of the words on most pages, even when the page is askew, the fonts obscure and the scan a mess. Other applications are the Mr. Magoos of the OCR world, misinterpreting so many words that you might as well retype the document. In between are OCR apps that do well with some typefaces and formatting and poorly with others.

The OCR application or service provider that processes electronic evidence has an enormous impact on the usability of the production. Bad OCR insures that text searches will come up short and spreadsheet data will be worthless. But how do you know when a producing party furnishes bad OCR, and how do you know if it's an effort to hamper your investigation? Start by checking whether the other side depends on the same bad data or if they are relying on the pristine originals.

"Even a dog," observed Justice Oliver Wendell Holmes, "knows the difference between being tripped over and being kicked." True, but e-discovery can leave you feeling dumber than a dog when you can't tell if the opposition's messing with you or just plain incompetent. One day, it will be a distinction without a difference for purposes of enforcement--sloppy and slick will both draw sanctions. Until then, courts need to explore whether the data produced is hobbled compared with that used by the producing party and its counsel.

Level the Playing Field

So how do you deal with opponents who convert native data to naked TIF formats and deliver bad OCR? The answer is to insist that the source data stay in its native digital format. That doesn't necessarily mean native file production, but be sure that the text and the relevant metadata are ported directly to the production format *without* intervening OCR. It's cheaper, faster and much more accurate.

A level playing field means that the form in which information's produced to me isn't more cumbersome or obscure than what's available to you. The elements needed to sort, read, classify, search, evaluate and authenticate electronic evidence—elements like accurate text and relevant metadata—should be in my hands, too.

In short, *it shouldn't be much harder to use or understand the information you've produced when it's on my system than when it's on yours.* This digital Golden Rule has yet to find its full expression in the Sedona Guidelines or the new Federal e-discovery rules, but it's a tenet of fairness that should guide the hand of every Solomon grappling with e-discovery.

Rules of Thumb for Forms of ESI Production

by Craig Ball

[Originally published in Law Technology News, July 2006]

Come December 2006, amended Rule 34(b) of the Federal Rules of Civil Procedure has a gift for requesting parties both naughty and nice. It accords them the right to specify the form or forms of production for electronically stored information (ESI) sought in discovery. Though December may seem remote in these dog days of July, litigators better start making their lists and checking them twice to insure that, come December, they'll know what forms are best suited to the most common types of ESI.

Last month, I covered the five principal forms ESI can take:

1. Hard copies;
2. Paper-like images of data in, e.g., TIFF or PDF;
3. Data exported to "reasonably usable" electronic formats like Access databases or load files;
4. Native data; and
5. Hosted data.

This month, we'll look at considerations in selecting a form of production for the kinds of data most often seen in e-discovery.

Word Processed Documents

In small productions (e.g., less than 5,000 pages), paper and paper-like forms (.PDF and .TIFF) remain viable. However, because amended Rule 34(b) contemplates that producing parties not remove or significantly degrade the searchability of ESI, both parties must agree to use printouts and "naked" image files in lieu of electronically searchable forms. When the volume dictates the need for electronic searchability, image formats are inadequate unless they include a searchable data layer or load file; otherwise, hosted or native production (e.g., .DOC, .WPD, .RTF) are the best approaches. Pitfalls in native production include embedded macros and auto date features that alter the document when opened in its native application. Moreover, word processor files can change their appearance and pagination depending upon the fonts installed on, or the printer attached to, the computer used to view the file. Be careful referring to particular pages or paragraphs because the version you see may format differently from the original.

Consider whether system and file metadata are important to the issues in your case. If so, require that original metadata be preserved and a spreadsheet or other log of the original system metadata be produced along with the files.

E-Mail

Again, very small productions may be managed using paper or images if the parties agree on those forms, but as volume grows, only electronically searchable formats suffice. These can take the form of individual e-mails exported to a generic e-mail format (.EML or .MSG files), image files (i.e., .PDF or TIFF) coupled with a data layer or load file, hosted production or native production in one of the major e-mail storage formats (.PST for Outlook, .NSF for Lotus Notes, .DBX for Outlook Express). While native formats provide greatest flexibility and the potential to see far more information than hard copies or images, don't seek native production if you lack the tools and skill to access the native format without corrupting its contents or commingling evidence with other files.

All e-mail includes extensive metadata rarely seen by sender or recipient. This header data contains information about the routing and timing of the e-mail's transmission. Require preservation and production of e-mail metadata when it may impact issues in the case, particularly where there are questions concerning origin, fabrication or alteration of e-mail.

Spreadsheets
Even when spreadsheets fit on standard paper, printed spreadsheets aren't electronically searchable and lack the very thing that separates a spreadsheet from a table: the formulae beneath the cells. If the spreadsheet is just a convenient way to present tabular data, a print out or image may suffice, but if you need to examine the methodology behind calculations or test different theories by changing variables and assumptions, you'll need native file production. Hosted production that allows virtual operation may also suffice. When working with native spreadsheets, be mindful that embedded variables, such as the current date, may update automatically upon opening the file, changing the data you see from that previously seen by others. Also, metadata about use of the spreadsheet may change each time it is loaded into its native application. Once again, decide if metadata is important and require its preservation when appropriate.

PowerPoint Presentations:
You can produce a simple PowerPoint presentation as an electronically searchable image file in PDF or TIFF, but if the presentation is animated, it's a poor candidate for production as an image because animated objects may be invisible or displayed as incomprehensible layers. Instead, native or hosted production is appropriate. Like spreadsheets, native production necessitates preservation of original metadata, which may change by viewing the presentation.

Voice Mail
Often overlooked in e-discovery, voice mail messages and taped conversations (such as recorded broker-client transactions) may be vitally important evidence. As voice mail converges with e-mail in so-called integrated messaging systems, it's increasingly common to see voice mail messages in e-mail boxes. Seek production of voice mail in common sound formats such as .WAV or .MP3, and be certain to obtain voice mail metadata correlated with the audio because information about, e.g., the intended recipient of the voice message or time of its receipt, is typically not a part of the voice message.

Instant Messaging
Instant messaging or IM is similar to e-mail except that exchanges are in real-time and messages generally aren't stored unless the user activates logging or the network captures traffic. IM use in business is growing explosively despite corporate policies discouraging it. In certain regulated environments, notably securities brokerage, the law requires preservation of IM traffic. Still, requests for discovery of IM exchanges are commonly met with the response, "We don't have any;" but because individual users control whether or note to log IM exchanges, a responding party can make no global assertions about the existence of IM threads without examining each user's local machine. Although IM applications use proprietary formats and protocols, most IM traffic easily converts to plain text and can be produced as an ASCII- or word processor-compatible files.

Databases
Enterprises increasingly rely on databases to manage business processes. Responsive evidence may exist only as answers obtained by querying a database. Databases present enormous e-discovery challenges. Specify production of the underlying dataset and application and you'll likely face objections that the request for production is overbroad or intrudes into trade secrets or the privacy

rights of third parties. Producing parties may refuse to furnish copies of database applications arguing that doing so violates user licenses. But getting your own license for applications like Oracle or SAP and assembling the hardware needed to run them can be prohibitive.

If you seek the dataset, specify in your request for production the appropriate back up procedure for the database application geared to capture all of the data libraries, templates and configuration files required to load and run the database. If you simply request the data without securing a back up of the entire database environment, you may find yourself missing an essential component. By demanding that data be backed up according to the publisher's recommended methodology, you'll have an easier time restoring that data, but be sure the back up medium you specify is available to the producing party (i.e., don't ask for back up to tape if they don't maintain a tape back up system).

An approach that sometimes works for simpler databases is to request export of records and fields for import to off-the-shelf applications like Microsoft Access or Excel. One common export format is the Comma Separated Variable or CSV file, also called a Comma Delimited File. In a CSV file, each record is a single line and a comma separates each field. Not all databases lend themselves to the use of exported records for analysis, and even those that do may oblige you to jump through hoops or engage an expert.

If you aren't confident the producing party's interrogation of the database, will disgorge responsive data, consider formulating your own queries using the application's query language and structure. For that, you'll need to understand the application or get expert help, e.g., from a former employee of the responding party or by deposing a knowledgeable employee of your opponent to learn the ins-and-outs of structuring a query.

Summer Reading

ESI. CSV. WAV. It's a new language for lawyers, but one in which we must be fluent if we're to comply with amended Rule 26(f)(3) and its requirement that parties discuss forms of production in the pre-discovery meet-and-confer. So, this summer, lay down that Grisham novel in favor of a work that has us all in suspense: *The Rules.*

Copy That?

by Craig Ball

[Originally published in Law Technology News, October 2006]

One of the frustrating things about e-discovery is that two lawyers discussing preservation will use the same words but mean entirely different things. Take "copying." When a producing party agrees to copy a paper document, there's rarely a need to ask, "What method will you use," or "Will you copy the entire page?" It's understood they'll capture all data on both sides of the page and produce a duplicate as nearly equivalent as possible to the original.

But when data is stored electronically, "making a copy" is susceptible to meanings ranging from, "We'll create a forensically sound, authenticated image of the evidence media, identical in the smallest detail," to "We'll duplicate some parts of the evidence and change other parts to substitute misleading information while we irreparably alter the original." Of course, nobody defines "making a copy" the latter way, but it's an apt description of most data copying efforts.

Unlike paper, electronically stored information (ESI) always consists of at least two components: a block of data called a file and at least one other block of data containing, inter alia, the file's name, location and its last modified, accessed, and created dates (MAC dates) of the file. This second block, called system metadata, is often the only place from which the file name, location and dates can be gleaned. Anyone working with more than a handful of files appreciates the ability to sort and search by MAC dates. Take away or corrupt system metadata and you've made ESI harder to use.

So, copying a file means more than just duplicating the data in the file. It also means picking up the system metadata for the file stored in the disk's "Master File Table" or "File Allocation Table."

The good news is that Microsoft Windows automatically retrieves both the file and its system metadata when copying a file to another disk. The bad news is that Windows automatically changes the creation date of the duplicate and the last access date of the original to the date of copying. The creation date changes because Microsoft doesn't use it to store the date a user authored the contents of the file. Instead, Creation Date denotes the date on which the file was created on the particular medium or system housing it. Copying a file re-creates it. Spoliation *and* misrepresentation in a click!

But wait! It gets *worse*.

Floppy disks, thumb drives, CDs, and DVDs don't use the same file systems as hard drives running Windows. They don't record the same system metadata in the same way. If a Windows computer is an old roll-top desk with many small drawers and pigeonholes to hold file metadata, then a thumb drive or recordable CD is a modern desk with just a few. If you try to shift the contents of the roll-top to the modern desk, there aren't as many places to stash stuff. Likewise, file systems for floppy disks, thumb drives, CDs, and DVDs aren't built to store the same or as many metadata values for a file as Windows. So, when a file is copied from a hard drive to a thumb drive, floppy disk or optical media, some of its system metadata gets jettisoned and only the last modified value stays aboard. That's bad.

Now, copy the data from the thumb drive, floppy or optical media back to a Windows machine and the operating system has a bunch of empty metadata slots and pigeonholes to fill. Not receiving a value for the jettisoned system metadata, it simply makes something up! That is, it takes the last modified

date and uses it to fill both the slot for last modified date and the slot for last accessed date. That's worse. So, if we can't copy a file by…copying it, what do we do?

The answer is that you have to use tools and techniques designed to preserve system metadata or you must record the metadata values before you alter them by copying. Various tools and techniques exist to duplicate files on Windows systems without corrupting metadata. One that Windows users already own is Microsoft Windows Backup. If you have Windows XP Pro installed, you'll probably find Windows Backup in Accessories>System Tools. If you use Windows XP Home Edition, Windows Backup wasn't automatically installed, but you can install it from valueadd/MSFT/ntbackup on your system CD.

So far, we've talked only about copying a file and its system metadata. But each file comes from a complex environment containing lots of data illuminating the origins, usage, manipulation and even destruction of files. Some of this information is readily accessible to a user, some is locked by the operating system and much more is inaccessible to the operating system, lurking in obscure areas such as "unallocated clusters" and "slack space." When you copy a file and its metadata, all of this information is left behind. Even if you copy all the active files on the hard drive, you won't preserve the revealing latent data. To do that, you have to go deeper than the operating system and create a forensically sound copy.

The classic definition of a forensically sound copy is that it's an authenticable duplicate of a storage medium by a method that doesn't alter the source and reflects or can reliably reconstruct every readable byte and sector of the source with nothing added, altered or omitted. It's a physical, rather than a logical duplicate of the original.

A forensically sound copy may be termed a clone, drive image, bit stream duplicate, snapshot or mirror. As long as the copy is created in a way that preserves latent information and can be reliably authenticated, the name doesn't matter, though drive image denotes a duplicate where the contents of the drive are stored or compressed in one or more files which can be reconstituted as a forensically sound copy, and some use snapshot to mean a full system backup of a server that doesn't preserve latent data.

Beware the misguided use of the Symantec Corp.'s Ghost or other off-the-shelf duplication programs. Though it's possible to create a forensically sound drive clone with Ghost, I've never seen it done correctly in the wild. Instead, IT personnel invariably use Ghost in ways that don't preserve latent data and alter the original. Usually this flows from ignorance; occasionally, it's an intentional effort to frustrate forensic examination.

There is no single approved way to create a forensically sound copy of a drive. Several hardware and software tools are well suited to the task, each with strengths and weaknesses. Notables include Guidance Software Inc.'s EnCase, the no-cost Linux "dd" (data dump) function, AccessData Corp.'s Forensic Toolkit, X-Ways Software Technology AG's X-Ways Forensics, Paraben Corp.'s Replicator and drive duplication devices from Intelligent Computer Solutions Inc. and Logicube Inc. There are many different types of digital media out there, and a tool appropriate to one may be incapable of duplicating another. You have to know what you're doing and select the correct application for the job.

And there's the takeaway: Not all copies are created equal. Successful preservation of ESI hinges not only on selecting the tools, but also on your planning and process, e.g., defining your goals, protecting

the chain of custody, authenticating the duplicate, documenting the effort and understanding the consequences of your chosen method. Copy that?

In Praise of Hash

by Craig Ball

[Originally published in Law Technology News, November 2006]

I love a good hash. Not the homey mix of minced meat and potato Mom used to make. I mean *hash values*, the results of mathematical calculations that serve as reliable digital "fingerprints" of electronically stored information. If you haven't come to love hash values, you will, because they're making electronic discovery easier and less costly.

Using hash algorithms, any amount of data—from a tiny file to the contents of entire hard drives and beyond—can be uniquely expressed as an alphanumeric sequence of fixed length.

The most common forms of hashing are MD5 and SHA-1. The MD5 hash value of Lincoln's Gettysburg Address is E7753A4E97B962B36F0B2A7C0D0DB8E8. Anyone, anywhere performing the same calculation on the same data will get the same unique value in a fraction of a second. But change "Four score and seven" to "Five score" and the hash becomes 8A5EF7E9186DCD9CF618343ECF7BD00A. However subtle the alteration—an omitted period or extra space—the hash value changes markedly. The chance of an altered electronic document having the same MD5 hash—a "collision" in cryptographic parlance--is one in 340 *trillion, trillion, trillion*. Though supercomputers have fabricated collisions, it's still a level of reliability far exceeding that of fingerprint and DNA evidence.

Hashing sounds like rocket science—and it's a miraculous achievement—but it's very much a routine operation, and the programs used to generate digital fingerprints are freely available and easy to use. Hashing lies invisibly at the heart of everyone's computer and Internet activities and supports processes vitally important to electronic discovery, including identification, filtering, Bates numbering, authentication and de-duplication.

Identification

Knowing a file's hash value enables you to find its identical counterpart within a large volume of data without examining the contents of each file. The government uses this capability to ferret out child pornography, but you might use it to track down company secrets that flew the coop when an employee joined the competition.

Hash algorithms are one-way calculations, meaning that although the hash value identifies just one sequence of data, it reveals nothing *about* the data; much as a fingerprint uniquely identifies an individual but reveals nothing about their appearance or personality. Thus, hashing helps resolve how to search for stolen data on a competitor's systems without either side revealing trade secrets. It's done by comparing hash values of their files against hash values of your proprietary data. The hash values reveal nothing about the contents of the files except whether they match. It's not a foolproof solution because altered data present different hash values, but it's sometimes a sufficient and minimally intrusive method. A match conclusively establishes that purloined data resides on the competitor's system.

Filtering

Matching to known hash values simplifies e-discovery and holds down costs by quick and reliable exclusion of irrelevant data from processing and search. Matching out-of-the-box values for entire operating systems and common applications like Microsoft Windows or Intuit's Quicken, culls huge chunks of patently irrelevant files from consideration without risk of overlooking relevant information

excluded based on location or file extension. Hashing thwarts efforts to hide files by name change or relocation because hash-matching flushes out a file's true nature--so long, that is, as the contents of the file haven't changed.

Bates Numbering

Hashing's ability to uniquely identify e-documents makes it a candidate to replace traditional Bates numbering in electronic production. Though hash values don't fulfill the sequencing function of Bates numbering, they're excellent unique identifiers and enjoy an advantage over Bates numbers because they eliminate the possibility that the same number might attach to different documents. An electronic document's hash value derives from its contents, so will never conflict with that of another document unless the two are identical.

Authentication

I regularly use hashing to establish that a forensically sound duplicate of a hard drive faithfully reflects every byte of the source and to prove that my work hasn't altered the original evidence.

As e-discovery gravitates to native production, concern about intentional or inadvertent alteration requires lawyers to have a fast, reliable method to authenticate electronic documents. Hashing neatly fills this bill. In practice, a producing party simply calculates and records the hash values for the items produced in native format. Once these hash values are established, the slightest alteration of the data would be immediately apparent when hashed.

De-duplication

In e-discovery, vast volumes of identical data are burdensome and pose a significant risk of conflicting relevance and privilege assessments. Hashing flags identical documents, permitting one review of an item that might otherwise have cropped up hundreds of times. This is de-duplication, and it drastically cuts review costs.

But because even the slightest difference triggers different hash values, insignificant variations between files (e.g., different Internet paths taken by otherwise identical e-mail) may frustrate de-duplication when hashing an entire e-document. An alternative is to hash relevant *segments* of e-documents to assess their relative identicality, a practice called "near de-duplication."

Here's to You, Math Geeks

So this Thanksgiving, raise a glass to the brilliant mathematicians who dreamed up hash algorithms. They're making electronic discovery and computer forensics a whole lot easier and less expensive.

Unlocking Keywords

by Craig Ball

[Originally published in Law Technology News, January 2007]

The notion that words hold mythic power has been with us as long as language.

We know we don't need to ward off evil spirits, but we still say, "Gesundheit!" when someone sneezes. Can't hurt.

But misplaced confidence in the power of word searches can seriously hamper electronic data discovery. Perhaps because keyword searching works so well in the regimented realm of automated legal research, lawyers and judges embrace it in EDD with little thought given to its effectiveness as a tool for exploring less structured information. Too bad, because the difference between keyword searches that get the goods and those that fail hinges on thoughtful preparation and precaution.

Text Translation

Framing effective searches starts with understanding that most of what we think of as textual information isn't stored as text. Brilliant keywords won't turn up anything if the data searched isn't properly processed.

Take Microsoft Outlook e-mail. The message we see isn't a discrete document so much as a report assembled on-the- fly from a database. As with any database, the way information is stored little resembles the way we see it onscreen after our e-mail program works its magic by decompressing, decoding and decrypting messages.

Lots of evidence we think of as textual isn't stored as text, including fax transmissions, .tiff or PDF documents, PowerPoint word art, CAD/CAM blueprints, and zip archives. For each, the search software must process the data to insure content is accessible as searchable text.

Be certain the search tool you or your vendor employ can access and interpret all of the data that should be seen as text.

Recursion

Reviewing a box of documents that contains envelopes within folders, you'd open everything to ensure you saw everything.

Computers store data within data such that an Outlook file can hold an e-mail transmitting a zip archive containing a PowerPoint with an embedded .tiff image.

It's the electronic equivalent of Russian nesting dolls. If the text you seek is inside that .tiff, the search tool must drill down through each nested item, opening each with appropriate software to ensure all content is searched. This is called recursion, and it's an essential feature of competent search. Be sure your search tool can dig down as deep as the evidence.

Exceptions

Even when search software opens wide and digs deep, it will encounter items it can't read: password protected files, proprietary formats, and poor optical character recognition. When that happens, it's important the search software generates an exceptions log flagging failures for follow up.

Know how the search tool tracks and reports items not searched or incompletely searched.

Search Term Tips
So far, I've talked only about search tools; but search terms matter, too.

You'll get better results when you frame searches to account for computer rigidity and human frailty. Some tips:

Stemming: Computers are exasperatingly literal when searching. Though mechanized searches usually overlook differences in capitalization, they're easily confounded by variances in prefixes or suffixes of the sort that human reviewers easily assimilate (e.g., flammable and inflammable or exploded and exploding).

You'll miss fewer variations using stemmed searches targeting common roots of keywords; e.g., using "explod" to catch both exploded and exploding.

But use stemming judiciously as the more inclusive your search, the more challenging and costly the review. Be sure to include the correct stemming operator for the search tool.

Boolean Search: Just as with legal research, pinpoint responsive items and prioritize review using Boolean operators to find items containing both of two keywords, or keywords within a specified proximity.

Misspelling: It's scary how many people can't spell. Even the rare good speller may hit the wrong key or resort to the peculiar shorthand of instant messaging.

Sometimes you can be confident a particular term appears just one way in the target documents—e-mail addresses are prime examples—but a thorough search factors in common misspellings, acronyms, abbreviations and IM-speak.

Synonyms: Your search for "plane" won't get off the ground if you don't also look for "jet," "bird," "aircraft, "airliner" and "crate."

A comprehensive search incorporates synonyms as well as lingo peculiar to those whose data is searched.

Noise words: Some words occur with such regularity it's pointless to look for them. They're "noise words," the static on your ESI radio dial.

I recently encountered a situation where counsel chose terms like "law" and "legal" to cull data deemed privileged. Predictably, the results were disastrously overinclusive.

I recommend testing keywords to flush out noise words. There's irrelevant text all over a computer—in spelling dictionaries, web cache, help pages, and user license agreements. Moreover, industries have

their own parlance and noise words, so it's important to assess noisiness against a representative sample of the environment you're searching.

Noise words are particularly nettlesome in computer forensic examinations, where searches extend beyond the boundaries of active files to the wilds of deleted and fragmented data. Out there, just about everything has to be treated as a potential hiding place for revealing text.

Because computers use alphabetic characters to store non-textual information, billions or trillions of characters randomly form words in the same way a million typing monkeys will eventually produce a Shakespearean sonnet. The difference is that the monkeys are theoretical while there really are legions of happenstance words on every computer. Consequently, searching three- and four-letter terms in forensic examinations—e.g., "IBM" or "Dell"—can be a fool's errand requiring an examiner to plow through thousands of false hits. If you must use noisy terms, it's best to frame them as discrete occurrences (flanked by spaces) and in a case-specific way (IBM but not iBm).

Striking a Balance

Effective keyword searching demands more than many imagine. You don't have to put every synonym and aberrant spelling on your keyword list, but you need to appreciate the limits of text search and balance the risk of missing the mark against the burden of grabbing everything and the kitchen sink. The very best results emerge from an iterative process: revisiting potentially responsive data using refined and expanded search terms.

Getting to the Drive

by Craig Ball

[Originally published in Law Technology News, April 2007]

Traditionally, we've relied on producing parties to, well, *produce*. Requesting parties weren't entitled to rifle file cabinets or search briefcases. When evidence meant paper documents, relying on the other side's diligence and good faith made sense. Anyone could read paper records, and when paper was "deleted," it was gone.

But, as paper's given way to electronically stored information (ESI), producing parties lacking computer expertise must blunder through or depend upon experts to access and interpret the evidence. Lawyers get disconnected from the evidence. When discoverable ESI resides in places the opposition can't or won't look, how can we accept a representation that "discovery responses are complete?" When there's a gaping hole in the evidence, sure, you can do discovery about discovery, but sometimes, you've just got to "get to the drive."

“Getting to the drive” means securing forensically qualified duplicates of relevant computer disk drives used by the other side, and having them examined by a qualified expert. Often lumped together, it's important to consider these tasks independently because each implicates different concerns.

When not writing or teaching, I examine computer hard drives voluntarily surrendered by litigants or pried from their fingers by court order. Serving as neutral or court-appointed special master, my task is to unearth ESI bound up with privileged or confidential content, protecting the competing interests of the parties. The parties can separate wheat from chaff for conventional, accessible data, but when the data's cryptic, deleted or inaccessible, I'm brought in to split the baby.

Increasingly, I see lawyers awakening to the power of computer forensics and wanting access to the other side's drives, but unsure when it's allowed or how to proceed. Some get carried away.

In a recent Federal District Court decision, *Hedenburg v. Aramark American Food Services*, 2007 WL 162716 (W.D. Wash.), the defendant in a discrimination and wrongful termination case suspected the plaintiff's e-mail or internet messaging might be useful for impeachment concerning her mental state. Apparently, Aramark didn't articulate more than a vague hunch, and Hedenburg dubbed it a "fishing expedition."

Judge Ronald Leighton denied access, analogizing that, "If the issue related instead to a lost paper diary, the court would not permit the defendant to search the plaintiff's property to ensure that her search was complete."

True enough, and the right outcome here, but what if a credible witness attested to having seen the diary on the premises, or the plaintiff had a history of disappearing diaries? What if injury or infirmity rendered the plaintiff incapable of searching? On such facts, the court might well order a search.

In weighing requests to access hard drives, judges should distinguish between the broad duty of preservation and the narrower one of production. It's not expensive to preserve the contents of a drive by forensic imaging (comparable in cost to a half-day deposition transcript), and it permits a computer to remain in service absent concerns that data will be lost to ongoing usage.

A drive can be forensically imaged without the necessity of anyone viewing its contents; so, assuming the integrity of the technician, no privacy, confidentiality or privilege issues are at stake. Once a drive image is "fingerprinted" by calculating its hash value (See, LTN Nov. 2005), that value can be furnished to the court and the other side, eliminating potential for undetected alteration.

Considering the volatility of data on hard drives and the fact that imaging isn't particularly burdensome or costly, courts shouldn't hesitate to order forensically-qualified preservation when forensic examination is foreseeable. In contrast, such forensic examination and production is an expensive, intrusive, exceptional situation.

Hard drives are like diaries in how they're laced with intimate and embarrassing content alongside discoverable information. Drives hold privileged spousal, attorney and health care communications, not to mention a mind-boggling incidence of sexually-explicit content (even on "work" computers). Trade secrets, customer data, salary schedules, passwords abound.

So how does a court afford access to the non-privileged evidence without inviting abuse or exploitation of the rest? An in-camera inspection might suffice for a diary, but what judge has the expertise, tools, and time to conduct an in-camera computer forensic examination?

With so much at stake, courts need to approach forensic examination cautiously. Granting access should hinge on demonstrated need and a showing of relevance, balanced against burden, cost or harm. It warrants proof that the opponent is either incapable of, or untrustworthy in, preserving and producing responsive information, or that the party seeking access has some proprietary right with respect to the drive or its contents. Showing that a party lost or destroyed ESI is a common basis for access, as are situations like sexual harassment or data theft where the computer was instrumental to the alleged misconduct.

Of course, parties often consent. Seeking to prove your client has "nothing to hide" by granting the other side unfettered access to computers is playing Russian roulette with a loaded gun. You won't know what's there, and if it's sufficiently embarrassing, your client won't tell you. Instead, the cornered client may wipe information and the case will turn on spoliation and sanctions.

Orders granting examination of an opponent's drive should provide for handling of confidential and privileged data and narrow the scope of examination by targeting specific objectives. The examiner needs clear direction in terms of relevant keywords and documents, as well as pertinent events, topics, persons and time intervals. A common mistake is to agree upon a search protocol or secure an order without consulting an expert to determine feasibility, complexity or cost. The court should encourage the parties to jointly select a qualified neutral examiner as this will not only keep costs down but will also help ensure that the agreed-upon search protocol is respected.

Getting to the drive isn't easy, nor should it be. When forensics may come into play, e.g., cases of data theft, spoliation and computer misuse, demand prompt, forensically-sound preservation. When you want to look, be ready to show good cause and offer appropriate safeguards.

Who Let the Dogs Out?

by Craig Ball

[Originally published in Law Technology News, May 2007]

What is evidence? I won't quote *Black's Law Dictionary* or *McCormick on Evidence*, partly because I boxed mine when online legal research made my library obsolete, and because my well-thumbed copies inhabited a time when evidence was largely a thing or statement. We examined things. Witnesses made statements.

After law school and apart from the occasional trial, lawyers rarely reflect on the nature of evidence. Like pornography, we know it when we see it. But with electronic evidence, we hardly see it anymore. No longer can we open a file drawer and wade in.

Now, we rely on experts and technicians using searches and filters to troll roiling oceans of data and process the catch of the day. By the time lawyers "see" electronic evidence, it's frozen fish sticks and canned tuna. Sorry, Charlie McCormick, 21st century lawyers don't go near the water.

Rethinking Assumptions

Fundamentals of evidence mastered in law school are still helpful, but some electronically stored evidence is so foreign to traditional assumptions that we need to rethink them. Who is charged with its content and custody? What's an original? How do we authenticate it? When/how do we allow its use?

We still expect lawyers to know the evidence in their cases and produce it, but electronic evidence forces counsel to rely on crude tools and methodologies and work through technical intermediaries of uneven ability who speak in acronyms and jargon. Lawyers are increasingly so disconnected from the evidence that when we search for evidence, we tend to find only what we seek instead of what's there to be found.

I see this glaringly manifested by colleagues who regard a text search for a handful of keywords as a sufficient effort. Just because Lexis or Westlaw make you feel like the Amazing Kreskin, a seat-of-the-pants keyword search in unstructured data is a whole different kettle of fish.

Ever run a pack of bloodhounds to find a fugitive? Me neither, but we've *seen* it a million times in old movies. Outskirts of city at night. Hardboiled detective hands tattered shirt sleeve to dog wrangler. Ol' Blue sniffs the rag. "Go git 'em, boy." Cut to thick forest. Baleful "roof, roof, a-roof" signals auspicious time to wade down fortuitously encountered stream and throw off scent. Segue to confused hound. Fade to shot of grinning anti-hero sipping Mojitos with Brazilian beauty on Ipanema Beach. Roll credits.

We didn't see Blue bounding by his quarry's e-ticket confirmation to Rio and the thumb drive storing offshore account numbers. It wasn't a bad search, it was just too single-minded.

Form Above Substance

Processing volume in this narrow way without assimilating it is emblematic of the lengths we go to elevate form above substance. Hacking through terabytes of data, we've become the child squinting at the scary parts of the movie through hands over our eyes, looking as narrowly as possible at the content.

Too cavalier about locating responsive evidence, we are disproportionately obsessed with inadvertent production of privileged information—to the point that much of the time and cost of e-discovery is consumed by the effort.

Are confidential attorney-client communications really so much a part of every custodian's data that e-discovery must slow to a costly crawl? If so, we need to encapsulate and tag these privileged items at the time they're created to isolate them from mainstream electronically stored information. Better to treat lawyers like vestal virgins than let the taint of their work bloat the cost and complexity of review.

When will we see that clients self-immolate far more often through incomplete production than inadvertent production?

We need to devote more time to thinking about what the evidence is instead of where it lodges. Too often, we fixate on the containers—the e-mail, spreadsheets and databases—with insufficient regard for the content. This isn't just a rant against producing parties. I see the failure as well in requesting parties determined to get to the other side's tapes and hard drives, but unable to articulate what they're seeking.

Saying, "I want the e-mail" is as meaningless as saying, "I want the paper." E-mail, voicemail, ledgers or lipstick on the mirror are just media used to hold and convey information. It's the transaction and the content that make them evidence.

The form matters, but only for reasons of accessibility (Can I view or hear it?), preservation (How do I protect it?), utility (Can I search and sort it?), completeness (Is something added or absent?) and authentication (Can I rely on it?).

Pondering the essential nature of evidence can't remain the exclusive province of law review commentators and law school professors. As never before, trial lawyers in the trenches must think hard about just what is the evidence? What are we really looking for? What gets us closer to the truth?

Page Equivalency and Other Fables

by Craig Ball

[Originally published in Law Technology News, August 2007]

When the parties to a big lawsuit couldn't agree on a vendor to host an electronic document repository, the court appointed me to help. Poring over multimillion dollar bids, I saw the vendors were told to assume that a gigabyte of data equals 22,500 pages. If the dozens of entities involved produced their documents in a mix of .tiff images and native formats—spreadsheets, word processed documents, e-mail, compressed archives, maps, photos, engineering drawings and more—how sensible, I wondered, was it to assume 22,500 pages per gig?

It's comforting to quantify electronically stored information as some number of pieces of paper or bankers' boxes. Paper and lawyers are old friends. But you can't reliably equate a volume of data with a number of pages unless you know the composition of the data. Even then, it's a leap of faith.

I've been railing against page equivalency claims for years because they're so elusive and often abused to misstate the burden and cost of electronic data discovery.

"Your Honor, Megacorp's employees each have 80 GB laptops. That means we will have to review 40 million pages per machine. Converting those pages to .tiff images will cost Megacorp 4 million dollars per laptop."

Nonsense!

If you troll the internet for page equivalency claims, you'll be astounded by how widely they vary, though each is offered with utter certitude. A GB of data is variously equated to an absurd 500 million typewritten pages, a naively accepted 500,000 pages, the popularly cited 75,000 pages and a laggardly 15,000 pages. The other striking aspect of page equivalency claims is that they're blithely accepted by lawyers and judges who wouldn't concede the sky is blue without a supporting string citation.

In testimony before the committee drafting the federal e-discovery rules, ExxonMobil representatives twice asserted that one GB yields 500,000 typewritten pages. The National Conference of Commissioners on Uniform State Laws proposes to include that value in its Uniform Rules Relating to Discovery of Electronically Stored Information. The Conference of Chief Justices cites the same equivalency in its "Guidelines for State Trial Courts Regarding Discovery of Electronically-Stored Information." Scholarly articles and reported decisions pass around the 500,000 pages per GB value like a bad cold.

Yet, 500,000 pages per GB isn't right. It's not even particularly close to right.

Several years ago, my friend Kenneth Withers, now with The Sedona Conference and then e-discovery guru for the Federal Judicial Center, wrote a section of the fourth edition of the Manual on Complex Litigation that equated a terabyte of data to 500 billion typewritten pages. It was supposed to say million, not billion. Withers, who owned up to the error with his customary grace and candor, has contributed so much wisdom to the bench and bar that he can't be faulted. But the echoes of that innocent thousand-fold miscalculation still reverberate today. Anointed by the prestige of the manual, the 500 billion-page equivalency was embraced as gospel. Even when the value was "corrected" to 500 million pages per terabyte—equal to 500,000 pages per GB—we're still talking an equivalency

with all the credibility of an Elvis sighting.

Now, with more e-discovery miles in the rear view mirror, it's clear we've got to look at individual file types and quantities to gauge page equivalency, and there is no reliable rule of thumb geared to how many files of each type a typical user stores. It varies by industry, by user and even by the lifespan of the media and the evolution of particular applications. A reliable page equivalency must be expressed with reference to both the quantity and form of the data, e.g., "a gigabyte of single page .tiff images of 8½"x11" documents scanned at 300 dpi equals approximately 18,000 pages."

Consider the column you're reading. In plain text, it's a file just 5 kilobytes in size and prints as one to two typewritten pages. As a rich text format document, the file quadruples to 20 KB. The same text as a Microsoft Word document is 25 KB. Converted to a .tiff image, it's 123 KB without an accompanying load file. Applying a page equivalency of 500,000 pages per GB, a vendor using per page pricing may quote this column as being anything from one page to as many as 61 pages. Billed by the GB, you'll pay almost five times more for the article as two .tiff pages than as a native Word document. A flawed page equivalency hits the bottom line...hard.

So how many pages are in a gigabyte of data? Lawyers know this answer: *it depends.* To know, perform a data biopsy of representative custodians' collections and *gauge*—don't guess— page volume.

Well Begun is Half Done

by Craig Ball

[Originally published in Law Technology News, November 2007]

It's easy to feel overwhelmed by the daunting complexity of electronic discovery. There's so much to do in an arena where lawyers feel distinctly disadvantaged. We know we've got to hit the ground running, but so often we're paralyzed instead of galvanized. If only lawyers knew what to do first, certain of making the right choice.

Take heart. There is a reliably correct first step, and it's the identification of sources of electronic evidence. Do it well, and much of the fog hiding the hazards of e-discovery lifts. Pitfalls remain, but you're less likely to stumble into them.

Identification of electronically stored information (ESI) involves more than just a head count of machines, backup tapes, custodians, network storage areas, and thumb drives. Certainly, it's important to have a current inventory, but identification of potentially responsive sources of ESI goes deeper. You've got to know what you've got, who's got it, how much they have, where it is, and when it's going away.

Identification anticipates obligations imposed by the Federal Rules of Civil Procedure, such as Rule 26(a)(1)(B)'s requirement that litigants describe and supply the location of ESI going to claims or defenses and Rule 26(f)'s dictate that litigants discuss the forms of ESI. Then there's the duty to identify ESI claimed not reasonably accessible pursuant to Rule 26(b)(2)(B) or as privileged under Rule 26(b)(5)(A). Both must be identified with sufficient particularity to enable your opponent to gauge the merits of the objection.

If you can't properly identify the sources of ESI, you may be compelled to overproduce at enormous cost or run the risk of sanctions for failure to do so. That's not a Catch-22. It's an avoidable consequence of failing to do what the law requires.

Jump start the identification process by obtaining IT asset inventories and system diagrams. Most medium-size to large businesses track the acquisition, deployment, and disposal of computer systems. These assets tend to be depreciated for tax purposes, so the bean counters have to know when they come and go. Follow the money trail.

Similarly, IT departments often track deployment of systems and software for warranty, support and licensure, and they certainly track intranet connections and user privileges, if only to know where the wires from the patch panel lead! Check to see if the IT staff has a network map laying out the relationship between servers, users, business units and backup systems. Even an out-of-date network diagram is a leg up. Now, you're on the hardware and software trail.

Identify potentially responsive ESI along the people trail. Who are the persons most knowledgeable about the matters in contention? Pin down the principal software applications, data storage practices, devices, and media used by these key custodians. A phone call or e-mail may suffice to gather what you need, but better results flow from visits to the custodians' workplace and face-to-face interviews. Using a checklist tailored to the issues and computing environment is desirable, but don't let it get in the way of listening and observing.

It helps to lay eyes on the external hard drive or the mothballed system on the floor beside the desk. Ask about that stack of CDs on the shelf. Probe to find the pack rats. Remember: Even benign ESI hurts if you've sworn it doesn't exist.

Collect machine service tags and serial numbers, e-mail addresses, and user logon IDs. Record the overall capacity of hard drives along with their active data volume. Determine if there are local e-mail stores and archives on the machine, their file types, and sizes. Be sure to inquire about former

machines, applications, and e-mail systems and to what extent legacy data migrated to current systems. Meet representations of, "That's gone," with, "How can you be certain?"

While identifying ESI, you're also collecting information about foreseeable threats to its integrity and existence. For backup media, you want to know the rotation cycle and anticipated changes to hardware and software. Explore whether desktop systems, laptops, or portable data storage devices are slated for replacement or modification. For e-mail servers and voicemail systems, pin down purge settings that dictate when and how deleted messages become unrecoverable.

Of course, it's not enough to identify when potentially relevant ESI will disappear. You've got to be poised to preserve it. Ensure that those identifying spoliation hazards are trained to react to them.

The goal of all this is to generate a spreadsheet or database allowing an evolving view of the lay of your client's data landscape by custodian, volume, location, and other criteria. Thus equipped, you can more reliably gauge the cost and complexity of e-discovery and implement right-sized preservation. Plus, you'll be better able to fulfill your "meet and confer" obligations and build trust with the other side.

So have no fear. Identification of ESI is always the right thing to do; and done well, it greases the wheels for the labors to follow.

Ask the Right Questions

by Craig Ball

[Originally published in Law Technology News, December 2007]

Sometimes it's more important to ask the right questions than to know the right answers, especially when it comes to nailing down sources of electronically stored information, preservation efforts and plans for production in the FRCP Rule 26(f) conference, the so-called "meet and confer."

The federal bench is deadly serious about meet and confers, and heavy boots have begun to meet recalcitrant behinds when Rule 26(f) encounters are perfunctory, drive-by events. Enlightened judges see that meet and confers must evolve into candid, constructive mind melds if we are to take some of the sting and "gotcha" out of e-discovery. Meet and confer requires intense preparation built on a broad and deep gathering of detailed information about systems, applications, users, issues and actions. An hour or two of hard work should lay behind every minute of a Rule 26(f) conference. Forget "winging it" on charm or bluster, and forget, "We'll get back to you on that."

Here are 50 questions of the sort I think should be hashed out in a Rule 26(f) conference. If you think asking them is challenging, think about what's required to deliver answers you can certify in court. It's going to take considerable arm-twisting by the courts to get lawyers and clients to do this much homework and master a new vocabulary, but, there is no other way.

These 50 aren't all the right questions for you to pose to your opponent, but there's a good chance many of them are . . . and a likelihood you'll be in the hot seat facing them, too.

1. What are the issues in the case?
2. Who are the key players in the case?
3. Who are the persons most knowledgeable about ESI systems?
4. What events and intervals are relevant?
5. When did preservation duties and privileges attach?
6. What data are at greatest risk of alteration or destruction?
7. Are systems slated for replacement or disposal?
8. What steps have been or will be taken to preserve ESI?
9. What third parties hold information that must be preserved, and who will notify them?
10. What data require forensically sound preservation?
11. Are there unique chain-of-custody needs to be met?
12. What metadata are relevant, and how will it be preserved, extracted and produced?
13. What are the data retention policies and practices?
14. What are the backup practices, and what tape archives exist?
15. Are there legacy systems to be addressed?
16. How will the parties handle voice mail, instant messaging and other challenging ESI?
17. Is there a preservation duty going forward, and how will it be met?
18. Is a preservation or protective order needed?
19. What e-mail applications are used currently and in the relevant past?
20. Are personal e-mail accounts and computer systems involved?

21. What principal applications are used in the business, now and in the past?
22. What electronic formats are common, and in what anticipated volumes?
23. Is there a document or messaging archival system?
24. What relevant databases exist?
25. Will paper documents be scanned, at what resolution and with what OCR and metadata?
26. What search techniques will be used to identify responsive or privileged ESI?
27. If keyword searching is contemplated, can the parties agree on keywords?
28. Can supplementary keyword searches be pursued?
29. How will the contents of databases be discovered? Queries? Export? Copies? Access?
30. How will de-duplication be handled, and will data be re-populated for production?
31. What forms of production are offered or sought?
32. Will single- or multi-page .tiffs, PDFs or other image formats be produced?
33. Will load files accompany document images, and how will they be populated?
34. How will the parties approach file naming, unique identification and Bates numbering?
35. Will there be a need for native file production? Quasi-native production?
36. On what media will ESI be delivered? Optical disks? External drives? FTP?
37. How will we handle inadvertent production of privileged ESI?
38. How will we protect trade secrets and other confidential information in the ESI?
39. Do regulatory prohibitions on disclosure, foreign privacy laws or export restrictions apply?
40. How do we resolve questions about printouts before their use in deposition or at trial?
41. How will we handle authentication of native ESI used in deposition or trial?
42. What ESI will be claimed as not reasonably accessible, and on what bases?
43. Who will serve as liaisons or coordinators for each side on ESI issues?
44. Will technical assistants be permitted to communicate directly?
45. Is there a need for an e-discovery special master?
46. Can any costs be shared or shifted by agreement?
47. Can cost savings be realized using shared vendors, repositories or neutral experts?
48. How much time is required to identify, collect, process, review, redact and produce ESI?
49. How can production be structured to accommodate depositions and deadlines?
50. When is the next Rule 26(f) conference (because we need to do this more than once)?

For alternate views on the EDD topics to be addressed at a Rule 26(f) conference, Magistrate Judge Paul Grimm's committee's "Suggested Protocol for Discovery of ESI," (www.mdd.uscourts.gov/news/news/ESIProtocol.pdf), and the U.S.D.C. for the District of Kansas'"Guidelines for Discovery of Electronically Stored Information" (www.ksd.uscourts.gov/guidelines/electronicdiscoveryguidelines.pdf).

Redaction Redux

by Craig Ball

[Originally published in Law Technology News, February 2008]

"The forceps of our minds are clumsy forceps," observed H. G. Wells, "and crush the truth a little in taking hold of it." Clumsier still is a method commonly used to redact information from electronically stored information—one that so crushes truth, it's alarming *anyone* defends it, let alone promotes it as a "standard."

I speak of redacting electronic documents by converting them to .tiff images, blacking out privileged and confidential content, then clumsily attempting to recreate electronic searchability by optical character recognition (OCR). When applied to spreadsheets and databases, it simply doesn't work. Why, then, are we content to spin invisible cloth rather than acknowledge the emperor's privates are on parade?

Good sense and fair play dictate that redaction methods preserve the integrity of unredacted content and the searchability and usability of the document. Instead, expediency and anxiety drive use of .tiff and OCR for redaction, enabling counsel to cling to familiar, if shopworn, "black line" redaction methods out of fear that privileged contents lurk in some dark digital recess.

To appreciate the problem, consider a complex spreadsheet like those routinely encountered in e-discovery. Spreadsheets are data grids made up of "cells" formed at the intersection of rows and columns. Cells contain hidden formulae entered by the user that generate calculated values seen as numbers in the cell. Formulae are what distinguish a spreadsheet from a word-processed table and may be important evidence in that they establish the origins, dependency and sensitivity of the calculated values. Put differently, *formulae make the numbers dance*. Without them, cell values are runes bereft of rhyme or reason.

With its embedded content, page-defying proportions and dynamic functionality, the exemplar spreadsheet fairly cries out for native production. Alas, it also harbors privileged or confidential content that must be excised.

If the requesting party isn't vigilant, here's how redaction goes wrong:

First, the producing party images the spreadsheet in .tiff format. It sprawls beyond the bounds of an 8½ x 11-inch page, so the data spills confusingly across multiple pages of .tiff images, obscuring column and row relationships. It's a mess.

Second, converting the spreadsheet to .tiff strips away all the underlying formulae, destroying spreadsheet function and undermining a key advantage of native production.

Finally, converting to .tiff means the data is no longer intelligible as data—i.e., it's not electronically searchable. A .tiff is just a picture—static ink on a virtual page—and no more electronically searchable than a Gutenberg Bible.

But it gets worse. To this point, the spreadsheet has been folded across unnatural dimensions, stripped of its usability and rendered electronically unsearchable. Now, the producing party redacts

objectionable information like it was any 2D paper document—by using a drawing utility to black it out or printing it to paper for obliteration by a trusty felt-tip marker!

The spreadsheet's on life support. Seeking to resuscitate its electronic searchability, the producing party administers OCR.

OCR is inherently error-prone, but when the optically recognized data is text, spell checking corrects egregious recognition errors and restores some of the electronic searchability the federal rules require. When the data is numeric, however, there are no means to spell-check the inevitably myopic OCR. Wrong numbers replace right ones, and the data becomes wholly untrustworthy. By the time the spreadsheet reaches the requesting party, it's a goner:

• Usability: **gone**.

• Searchability: **crippled**.

• Integrity: **destroyed**.

• Content: **affirmatively misrepresented**.

The operation was a success, but the patient died.

If this is an "industry standard" practice, then we must recall that an entire industry can be negligent. As Judge Learned Hand wrote, "Courts must in the end say what is required; there are precautions so imperative that even their universal disregard will not excuse their omission." *The T.J. Hooper*, 60 F.2d 737 (2d Cir. 1932).

Preemptively, requesting parties should hone in on how ESI will be redacted, and if flawed redaction techniques will materially impair usability or searchability, they must act swiftly to combat their use and promote alternatives.

Redaction of ESI should be tailored to the nature of the data, using the right tool for the task. Where once native redaction was daunting, now there are reliable, cost-effective techniques for Adobe Systems Inc. PDF and Microsoft Corp. Office documents, including spreadsheets. For example, Adobe Acrobat 8.0 supports data layer redaction, and the latest release of Microsoft's Office productivity suite stores documents in readily redactable XML formats.

In sum, .tiff-OCR has its place, but when it's the *wrong* approach, don't use it. Opt instead for techniques that preserve the intelligibility and integrity of the unredacted content.

The Science of Search

by Craig Ball

[Originally published in Law Technology News, April 2008]

Federal Magistrate Judge John Facciola is a remarkable fellow. He hails from Brooklyn, wears bow ties, knows the Bruce Springsteen songbook by heart and doesn't hesitate to bring the White House to heel when the administration gets sloppy in its electronic evidence preservation. But his most heretical act may be his observation in *United States v. O'Keefe*, No. 06-249 (D.D.C. Feb. 18, 2008), that keyword search of electronically stored information is a topic "clearly beyond the ken of a layman." By a layman, he means any lawyer or judge who isn't an expert in computer technology, statistics and linguistics.

Facciola adds that, given the complexity of the science of search, "[F]or lawyers and judges to dare opine that a certain search term or terms would be more likely to produce information than [other] terms . . . is truly to go where angels fear to tread."

Heeding the call, the crack team of Forensically-trained Offerers of Legal Services (FOOLS) at Ball Labs have rushed in to formulate 36 search terms guaranteed to grab the smoking gun in any English-language ESI collection. The 36 terms are the letters of the alphabet and the numbers 0-9.

Ridiculous? Sure! But in a case where I serve as special master for ESI, a party proposed that the letter "S" be used as a search term. In another appointment, the plaintiff wanted to search for the number 64.

These earnest requests came from good lawyers offering credible rationales. They saw only that the term would be found within the evidence they sought, not appreciating that it would also appear in just about everything else, too. In the parlance of information retrieval, the terms scored high on recall but failed miserably in precision.

The parties advocating their use failed to appreciate that keyword search in e-discovery is less a means to find information than it is a method to filter it—and a pretty poor one at that. Keyword search of ESI is a sampling strategy—a way to look at less than everything with some assurance that you're examining the parts most likely to hold responsive data.

The notion that lawyers are unqualified per se to concoct keyword searches is likely to shake some sensibilities. Lawyers believe themselves adept at keyword search in e-discovery because they've mastered keyword search in online legal research. The correlation is superficial at best. Unlike the crazy quilt of ESI, the language of reported cases is precise, consistent and structured. Misspellings are rare. Legal research is Disneyland. E-discovery is Baghdad.

Judge Facciola is right to point to lawyers' misplaced reliance on keyword search and lack of expertise. *Search is a science*, yet we approach it on faith, gambling that intuition and luck are enough. Still, noting the profession's lack of expertise doesn't address the knottier problem of *where* to secure the expertise we now must bring to court to establish or challenge the efficacy of search.

The answer isn't to spawn a new breed of self-anointed cyberlinguistics experts for hire. Neither will a wholesale move to concept search tools suffice. Smarter search tools employing algebraic and

probabilistic analysis are unquestionably an improvement on the crude tools we employ, but hardly dispense with the need for experts to explain their operation and defend their performance.

The answer is that lawyers need to learn more about the science of search as part of our legal and continuing education. We need to become skilled at tools and methods that help us refine searches and routinely test them against representative data so we can distinguish noisy terms from effective ones and learn to zero in on relevant ESI.

Law schools teach the science and art of legal research when modern methods have all but eliminated the need to navigate the reporter system. Instead, students and lawyers must be afforded the means to master the art and science of digital information. We must dare to tread in these areas, not as fools but as professionals skilled in eliciting, testing and marshaling evidence wherever it may be found.

"The right to practice law is not one of the inherent rights of every citizen . . . [but] is a peculiar privilege granted and continued only to those who demonstrate special fitness in intellectual attainment and in moral character." *Matter of Keenan*, 314 Mass. 544, 546 (1943). So it has been, and so it must remain as evidence takes new forms, if we are to be afforded that peculiar privilege.

Grimm Prognosis

by Craig Ball

[Originally published in Law Technology News, July 2008]

There's a double standard in e-discovery. Keyword search is deemed "good enough" for identifying responsive electronically stored information; yet when privilege is on the line, lawyers insist on page-by-page review. It's a tacit recognition that keyword search is a blunt instrument—a point artfully made twice this year by Magistrate Judge John Facciola in *U.S v. O'Keefe*, 537 F. Supp. 2d 14, 24 (D.D.C. 2008), and *Equity Analytics v. Lundin*, 248 F.R.D. 331, 333 (D.D.C. 2008), and emphatically underscored lately by Magistrate Judge Paul Grimm in *Victor Stanley, Inc. v. Creative Pipe, Inc.*, Civil Action No. MJG-06-2662 (D. Md. May 29, 2008)

It's assumed that lawyers are qualified to review documents for relevance, responsiveness and privilege character. But are we qualified to craft *proxies* for our judgment in the form of keyword searches?

In *Victor Stanley*, 165 documents slipped by a privilege review employing keyword search and a cursory-sounding "title page" analysis for non-searchable items. Defendants had unwisely abandoned efforts to secure a clawback agreement (a nonwaiver agreement providing that inadvertently produced privileged materials may not be used).

Plaintiff's counsel spotted the documents and dutifully reported their potentially privileged character, but argued defendants waived privilege by using a faulty review process. The court agreed, pointing to defendants' failure to provide information regarding keywords used, how they were selected, steps taken to assess the reliability of the outcome and the qualifications of the attorneys to design an effective and reliable search.

Thus another jurist dismisses the legal profession's ability to search ESI without demonstrated expertise. It's enough to give Perry Mason an inferiority complex!

Do lawyers have so insightful a grasp of the words and semantic relationships behind our relevance and privilege decisions that we can distill the *je ne sais quoi* of our well-honed legal minds into quotidian keyword spotting?

We'd like to think we do, despite studies showing we possess little ability to frame effective keyword searches. We're shocked when our magic words catch barely 20% of responsive documents.

We shouldn't be.

Language is deceptively complex, and meaning is an elusive, protean quarry. We depend upon context for meaning, but keyword search ignores context entirely. Boolean search is only marginally better at gleaning context.

That leaves lawyers in a tough spot. Mushrooming volumes of ESI require us to rely more on automated search tools at the same time courts and opposing counsel are less willing to indulge the fiction that these tools perform in unskilled hands. The jig is up, and lawyers are now obliged to *prove* these proxies really work.

How do we meet that burden of proof? Judge Facciola deems both lawyers and judges keyword naifs, instead summoning a phalanx of linguists, statisticians and computer experts. Though expecting searches to be designed by qualified persons, Judge Grimm leaves the door open to lawyer-initiated keyword search when counsel can demonstrate adequate quality assurance and quality control.

This is a subtle but important distinction. Lawyers can become "qualified persons," though they may never be linguists, statisticians or computer experts. Still, Judge Grimm sets the bar high:

> "Use of search and information retrieval methodology, for the purpose of identifying and withholding privileged or work product protected information from production, requires the utmost care in selecting methodology that is appropriate for the task … [and] careful advance planning by persons qualified to design effective search methodology. The implementation of the methodology selected should be tested for quality assurance; and the party selecting the methodology must be prepared to explain the rationale for the method chosen to the court, demonstrate that it is appropriate for the task, and show that it was properly implemented."

Victor Stanley departs from *O'Keefe* in another subtle way. By emphasizing collaboration, Judge Grimm preserves counsel's ability to negotiate and agree upon search methods. Judge Facciola is no less a proponent of collaboration and transparency in e-discovery, but declaring both counsel and courts unequipped to oversee keyword search without expert assistance imperils the parties' freedom to agree on search methods and the court's authority to ratify such agreements.

What court, deeming itself unqualified to weigh such matters, could endorse a search protocol framed by those equally unequal to the task? Thus *Victor Stanley* preserves the litigants' inalienable right to be wrong, so long as everyone agrees that wrong is right. It's a Faustian bargain, but one permitting cases to move forward by simply ignoring pesky questions concerning the integrity and completeness of electronic discovery.

The Victor Stanley decision gives teeth to the duty to use better search techniques. Avoiding privilege waiver is a powerful incentive to:

- **Get expert help,**
- **Collaborate on search methods,**
- **Test your searches,**
- **Check the discard pile, and**
- **Get that clawback agreement.**

About the Author

Craig Ball, of Austin is a Board Certified Texas trial lawyer and accredited computer forensics expert, who's dedicated his career to teaching the bench and bar about forensic technology and the art and science of persuasion. Craig hung up his trial lawyer spurs to serve as a court-appointed special master and consultant in electronic evidence, as well as publishing and lecturing on computer forensics, emerging technologies, digital presentation and electronic discovery. Fortunate to supervise, consult on or serve as Special Master in connection with some of the world's largest electronic discovery projects and most challenging cases, Craig also greatly values his role as an instructor in computer forensics and electronic evidence to the Department of Justice and other law enforcement and security agencies.

Craig Ball is a prolific contributor to continuing legal and professional education programs throughout the United States, having delivered over 600 presentations and papers. Craig's articles on forensic technology and electronic discovery regularly appear in the national media, including in American Bar Association, ATLA and American Lawyer Media print and online publications. He also writes a multi-award winning monthly column on computer forensics and e-discovery for Law Technology News and Law.com called "Ball in your Court." Craig is a recipient of the Presidents' Award, the State Bar of Texas' most esteemed recognition of service to the profession.

Craig's been married to a trial lawyer for 21 years. He and Diana have two delightful teenagers and share a passion for world travel, cruising and computing.

EDUCATION
Rice University (B.A., triple major, English, Managerial Studies, Political Science, 1979); University of Texas (J.D., with honors, 1982); Oregon State University (Computer Forensics certification, 2003); EnCase Intermediate Reporting and Analysis Course (Guidance Software 2004); WinHex Forensics Certification Course (X-Ways Software Technology AG 2005); numerous other classes on computer forensics and electronic discovery.

SELECTED PROFESSIONAL ACTIVITIES
Law Offices of Craig D. Ball, P.C.; Licensed in Texas since 1982.
Board Certified in Personal Injury Trial Law by the Texas Board of Legal Specialization
Certified Computer Forensic Examiner, Oregon State University and NTI
Certified Computer Examiner (CCE), International Society of Forensic Computer Examiners
Admitted to practice U.S. Court of Appeals, Fifth Circuit; U.S.D.C., Southern, Northern and Western Districts of Texas.
Member, Editorial Advisory Boards, Law Technology News and Law.com (American Lawyer Media)
Board Member, Georgetown University Law School Advanced E-Discovery Institute
Member, Sedona Conference WG1 on Electronic Document Retention and Production
Experienced Special Master in Electronic Discovery, Federal and State District Courts
Instructor in Computer Forensics and Electronic Discovery, United States Department of Justice
Instructor, Cybercrime Summit, 2006, 2007
President, Houston Trial Lawyers Association (2000-01); President, Houston Trial Lawyers Foundation (2001-02)
Director, Texas Trial Lawyers Association (1995-2003); Chairman, Technology Task Force (1995-97)

Member, High Technology Crime Investigation Association and International Information Systems Forensics Assn.

ACADEMIC APPOINTMENTS AND HONORS

2008 recipient of the Texas Bar CLE "Standing Ovation" award for exceptional contributions
2006 recipient of "Lifetime Achievement Award for Promoting Technology in the Law"
Selected by peers as one of "Best Lawyers in America" and a "Texas Super Lawyer"
Faculty, Texas College of Trial Advocacy, 1992 and 1993
Adjunct Professor, South Texas College of Law, 1983-88
Rated AV by Martindale-Hubbell

Craig Ball's *Law Technology News* column "Ball in Your Court" is the 2008 honoree as "Best How-To Article" by the American Society of Business Publication Editors (ASBE). It was also the 2007 Gold Medal honoree as "Best Regular Column" as awarded by Trade Association Business Publications International and the 2007 ASBE Silver Medalist honoree as "Best Contributed Column" and their 2006 Silver Medalist honoree as "Best Feature Series" and "Best Contributed Column."

Goals for E-Discovery

Transparency in Rule 26(f) process

- **Get the techies talking**
- **Database disclosures: schemas & docs**
- **Share search, sampling and selection facts**

Goals for E-Discovery

Transparency in Rule 26(f) process

Preservation: Lawyer “boots on ground”

- **Talking to the grunts**
- **Involving key custodians**
- **Guarding against human frailty**

Goals for E-Discovery

Transparency in Rule 26(f) process

Preservation: Lawyer "boots on ground"

Preserve broadly. Produce sensibly.

- **Preserve forensically when forensics implicated**
- **Metadata means many things. Ask why sought?**

Goals for E-Discovery

Transparency in Rule 26(f) process

Preservation: Lawyer "boots on ground"

Preserve broadly. Produce sensibly.

Protect usability and searchability of ESI

- **Maintain a level ESI playing field**

Goals for E-Discovery

- **Just because two lawyers agree they can fly, doesn't mean they should head to the roof**

Require expertise where needed

Demand solid proof of burden and cost

Goals for E-Discovery

Consider authentication & presentation

Require expertise where needed

Demand solid proof of burden and cost

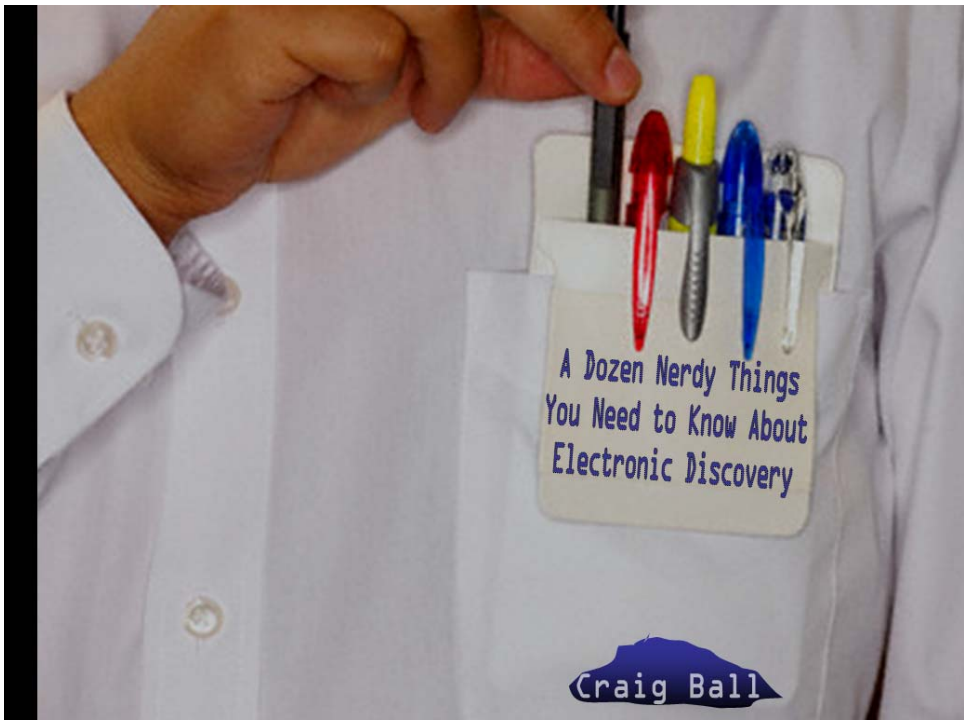
Goals for E-Discovery

Rule 1 Proportionality: *a firm hand early*

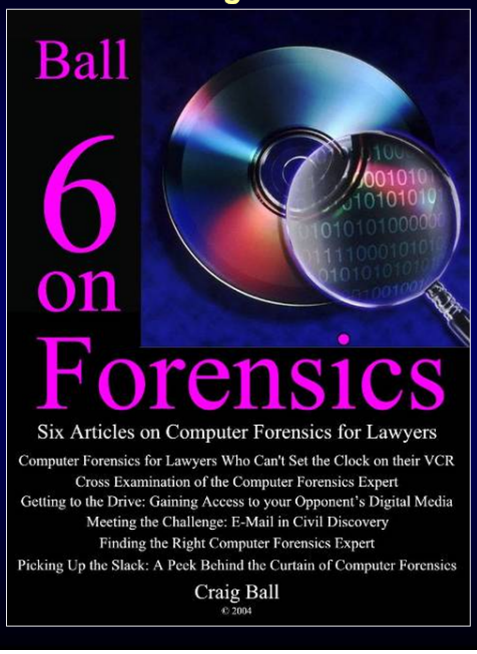
Consider authentication & presentation

Require expertise where needed

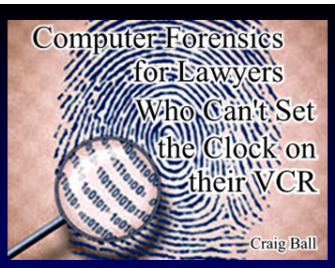
Demand solid proof of burden and cost

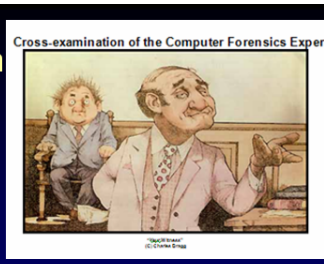


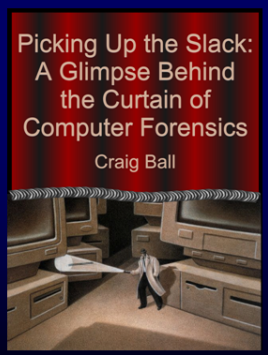
Papers
www.craigball.com

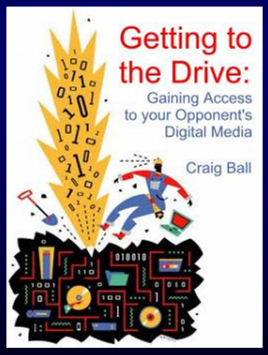


Papers
www.craigball.com



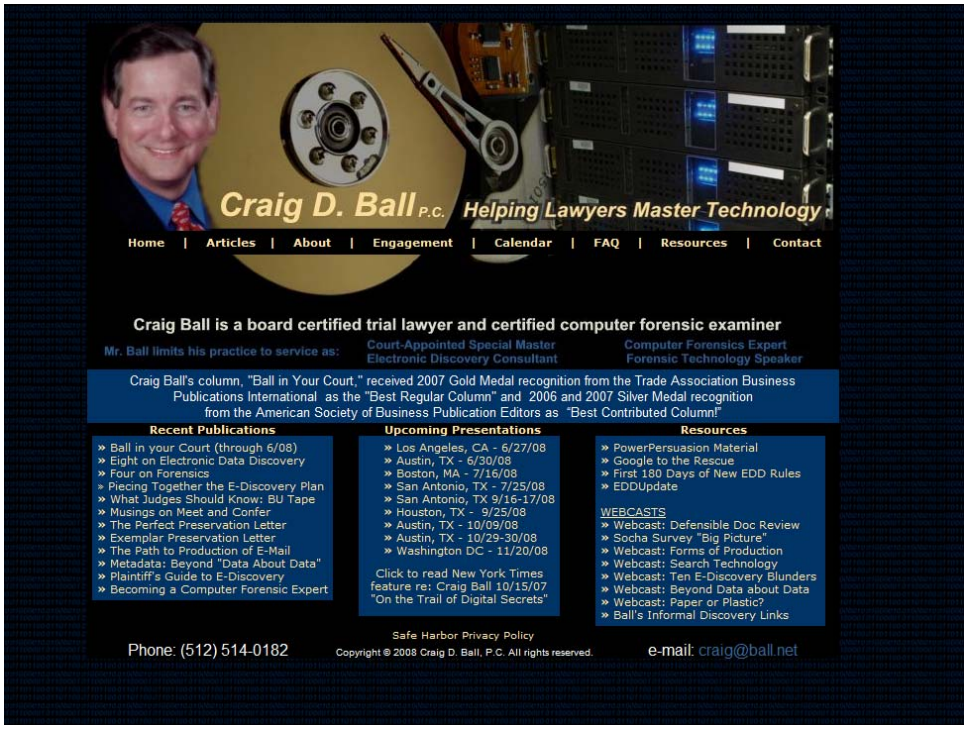







Papers
www.craigball.com
Ball
8
on
EDD
Eight Articles on Electronic Data Discovery
Musings on E-Discovery: Ball in Your Court
Hitting the High Points of the New e-Discovery Rules
What Judges Should Know About Discovery from Backup Tapes
The Plaintiffs' Guide to Meet and Confer
Metadata: Beyond Data About Data
The Perfect Preservation Letter
The Plaintiffs' Practical Guide to E-Discovery
Discovery of Electronic Mail: The Path to Production
Craig Ball


The Perfect Preservation Letter
Craig Ball
Papers
www.craigball.com
The Plaintiff's Practical Guide to E-Discovery
Craig Ball
Discovery of E-Mail:
The Path to Production
Craig Ball


Craig D. Ball P.C.
Helping Lawyers Master Technology
Home | Articles | About | Engagement | Calendar | FAQ | Resources | Contact
Craig Ball is a board certified trial lawyer and certified computer forensic examiner
Mr. Ball limits his practice to service as:
Court-Appointed Special Master
Electronic Discovery Consultant
Computer Forensics Expert
Forensic Technology Speaker
Craig Ball's column, "Ball in Your Court," received 2007 Gold Medal recognition from the Trade Association Business Publications International as the "Best Regular Column" and 2006 and 2007 Silver Medal recognition from the American Society of Business Publication Editors as "Best Contributed Column!"
Recent Publications
» Ball in your Court (through 6/08)
» Eight on Electronic Data Discovery
» Four on Forensics
» Piecing Together the E-Discovery Plan
» What Judges Should Know: BU Tape
» Musings on Meet and Confer
» The Perfect Preservation Letter
» Exemplar Preservation Letter
» The Path to Production of E-Mail
» Metadata: Beyond "Data About Data"
» Plaintiff's Guide to E-Discovery
» Becoming a Computer Forensic Expert
Upcoming Presentations
» Los Angeles, CA - 6/27/08
» Austin, TX - 6/30/08
» Boston, MA - 7/16/08
» San Antonio, TX - 7/25/08
» San Antonio, TX 9/16-17/08
» Houston, TX - 9/25/08
» Austin, TX - 10/09/08
» Austin, TX - 10/29-30/08
» Washington DC - 11/20/08
Click to read New York Times feature re: Craig Ball 10/15/07 "On the Trail of Digital Secrets"
Resources
» PowerPersuasion Material
» Google to the Rescue
» First 180 Days of New EDD Rules
» EDDUpdate
WEBCASTS
» Webcast: Defensible Doc Review
» Socha Survey "Big Picture"
» Webcast: Forms of Production
» Webcast: Search Technology
» Webcast: Ten E-Discovery Blunders
» Webcast: Beyond Data about Data
» Webcast: Paper or Plastic?
» Ball's Informal Discovery Links
Safe Harbor Privacy Policy
Phone: (512) 514-0182
Copyright © 2008 Craig D. Ball, P.C. All rights reserved.
e-mail: craig@ball.net


1
Courts to specify
conditions for
production of
inaccessible ESI

FRCP 26(b)(2)(B) Two-Tiered Analysis

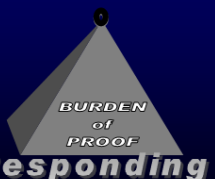
- Requesting Party serves discovery request
  - "Produce e-mail from Bob to Mary in June 2003"
- Responding Party objects "Not Reasonably Accessible"
- Requesting Pty moves to compel

Responding Party

Requesting Party

FRCP 26(b)(2)(B) Two-Tiered Analysis

- Requesting Party serves discovery request
  - "Produce e-mail from Bob to Mary in June 2003"
- Responding Party objects "Not Reasonably Accessible"
- Requesting Pty moves to compel
- Responding Pty must identify nature/location of ESI and prove not reasonably accessible—some discovery allowed
  - "We have fourteen 70GB tapes in legacy DLT format for June 2003 but no tape drive or software to read them."



Responding Party

Requesting Party

FRCP 26(b)(2)(B) Two-Tiered Analysis

- Requesting Party serves discovery request
  - "Produce e-mail from Bob to Mary in June 2003"
- Responding Party objects "Not Reasonably Accessible"
- Requesting Pty moves to compel
- Responding Pty must identify nature/location of ESI and prove not reasonably accessible—some discovery allowed
  - "We have fourteen 70GB tapes in legacy DLT format for June 2003 but no tape drive or software to read them."
- If you find inaccessible, Req Pty must show good cause
  - "It's the only source for this relevant, important information, Judge"

FRCP 26(b)(2)(B) Two-Tiered Analysis

- Requesting Party serves discovery request
  - "Produce e-mail from Bob to Mary in June 2003"
- Responding Party objects "Not Reasonably Accessible"
- Requesting Pty moves to compel
- Responding Pty must identify nature/location of ESI and prove not reasonably accessible—some discovery allowed
  - "We have fourteen 70GB tapes in legacy DLT format for June 2003 but no tape drive or software to read them."
- If you find inaccessible, Req Pty must show good cause
  - "It's the only source for this relevant, important information, Judge"
- You may require production under specified conditions
  - Court orders restoration of four dates as sample.
    Shifts cost to convert from DLT tape to hard drive.

sampling cost shifting searching filtering

2

Computers and other gadgets tell compelling stories

"[D]iscovery of electronically stored information stands on equal footing with discovery of paper documents."

ESI

Electronically Stored Information

"[A] request for production of "documents" should be understood to encompass, and the response should include, electronically stored information …."

Comments to Amended FRCP Rule 34

Broad enough to cover all current types of computer-based information
Flexible enough to encompass future changes and developments

ESI

Electronically Stored Information

Comments to Amended FRCP Rule 34

ESI

Electronically Stored Information

- ESI is as discoverable as paper.
- Parties must preserve and produce it.
- Counsel must understand and review it.

The Universe of Electronic Evidence

Beyond the PC and Server

- Thumb drives and memory cards
- Cell phones and Blackberries
- Access Control and Timekeeping
- MP3 Players
- Digital cameras and surveillance
- External drives
- Copiers & fax machines
- The Internet
- GPS and air bag systems
- Toll Tag data

Many Ways Secrets Hide in Microsoft Windows

1. Unallocated space
2. Slack space
3. Swap/page file
4. Prefetch
5. Windows logs
6. Registry
   - MRUs
   - User Assist Keys
   - USBStor Records
7. Internet activity: INDEX.DAT
8. Recycler INFO2
9. Temp LNK files
10. BAK files

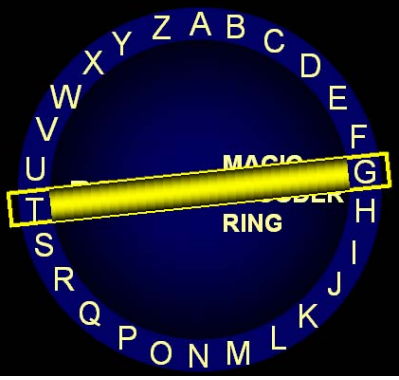
Windows Registry

User Assist Keys

Windows tracks web surfing and file access
in a ROT-13 encrypted Registry key called:

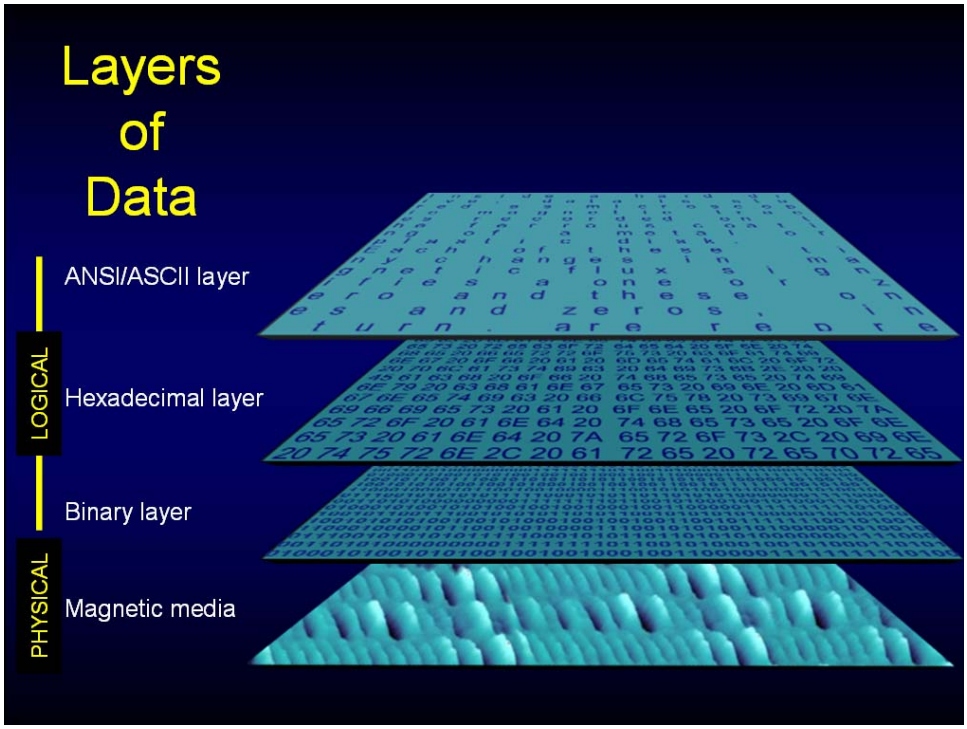
HKEY_USERS\[SID]\Software\Microsoft\Windows\CurrentVersion\Explorer
\UserAssist\{75048700-EF1F-11D0-9888-006097DEACF9}\Count

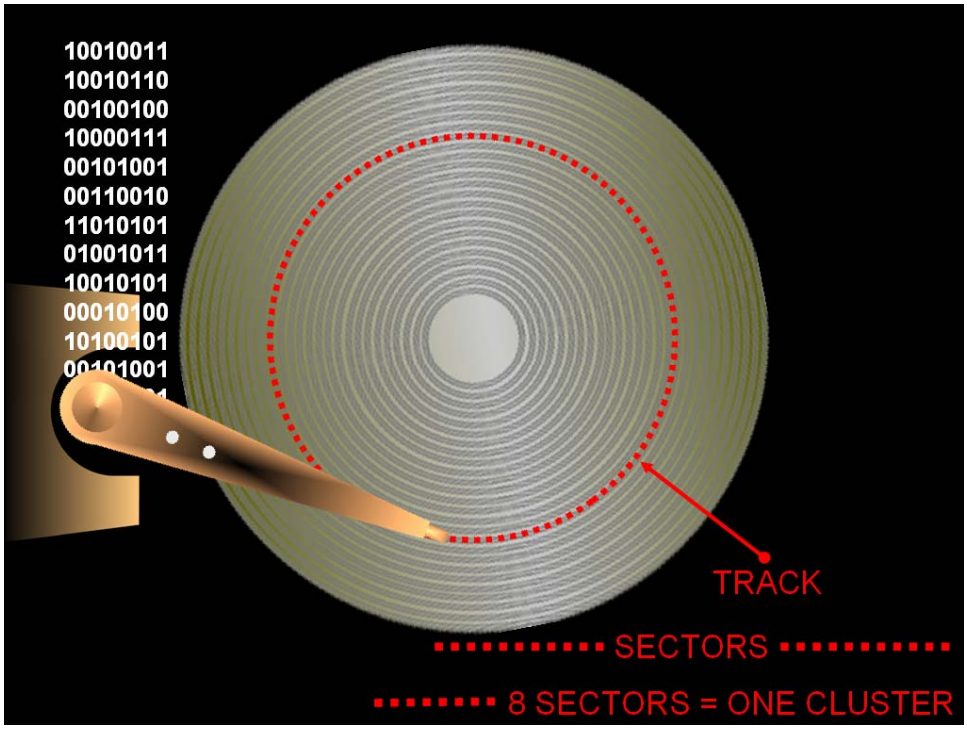
DOG = QBT

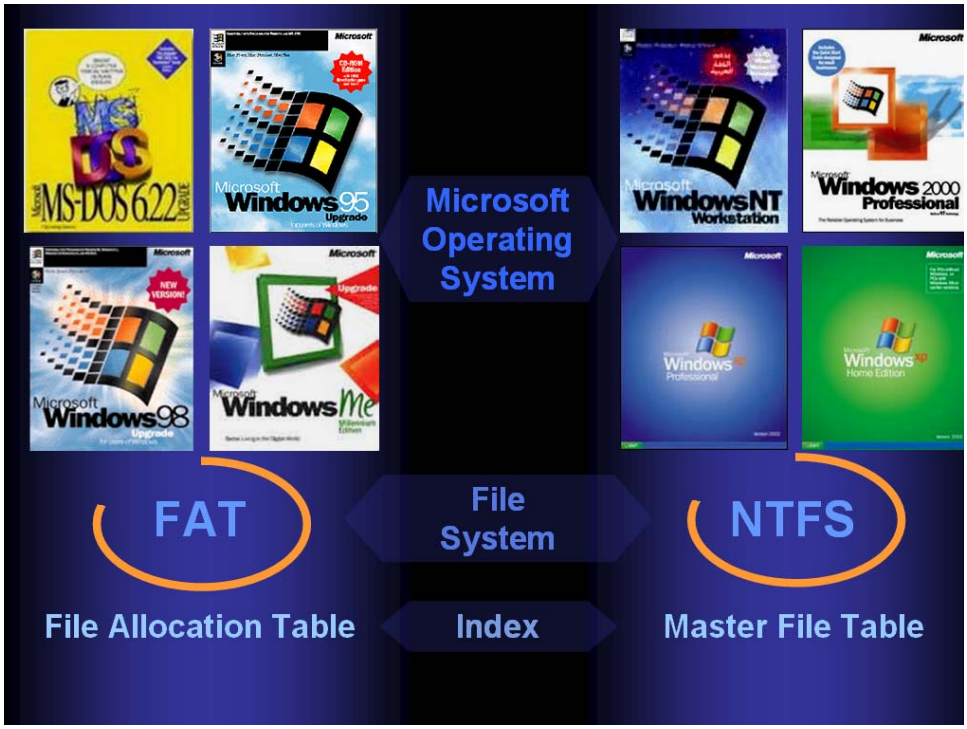


3

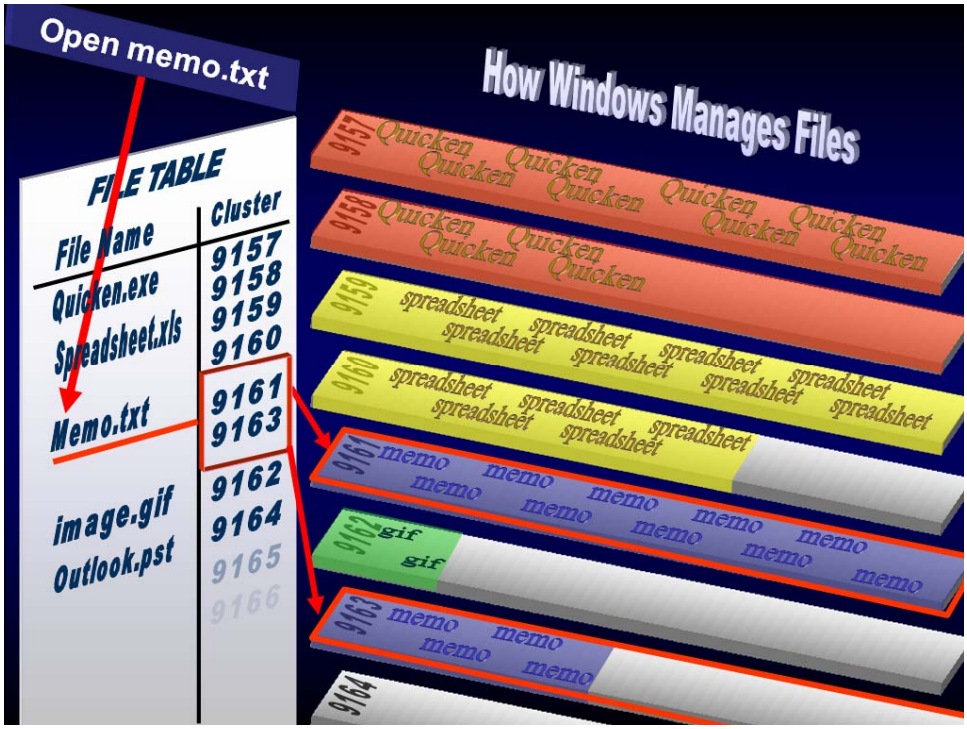
Deleted data's
not gone…
Why?

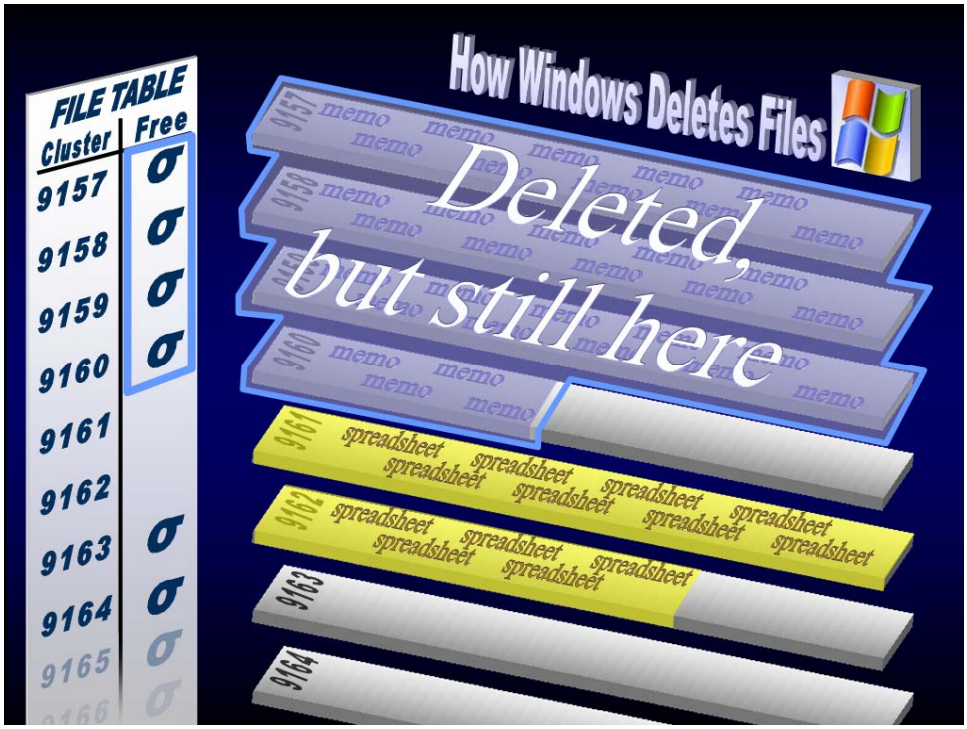

Layers of Data
ANSI/ASCII layer
Hexadecimal layer
Binary layer
Magnetic media
LOGICAL
PHYSICAL


10010011
10010110
00100100
10000111
00101001
00110010
11010101
01001011
10010101
00010100
10100101
TRACK
SECTORS
8 SECTORS = ONE CLUSTER

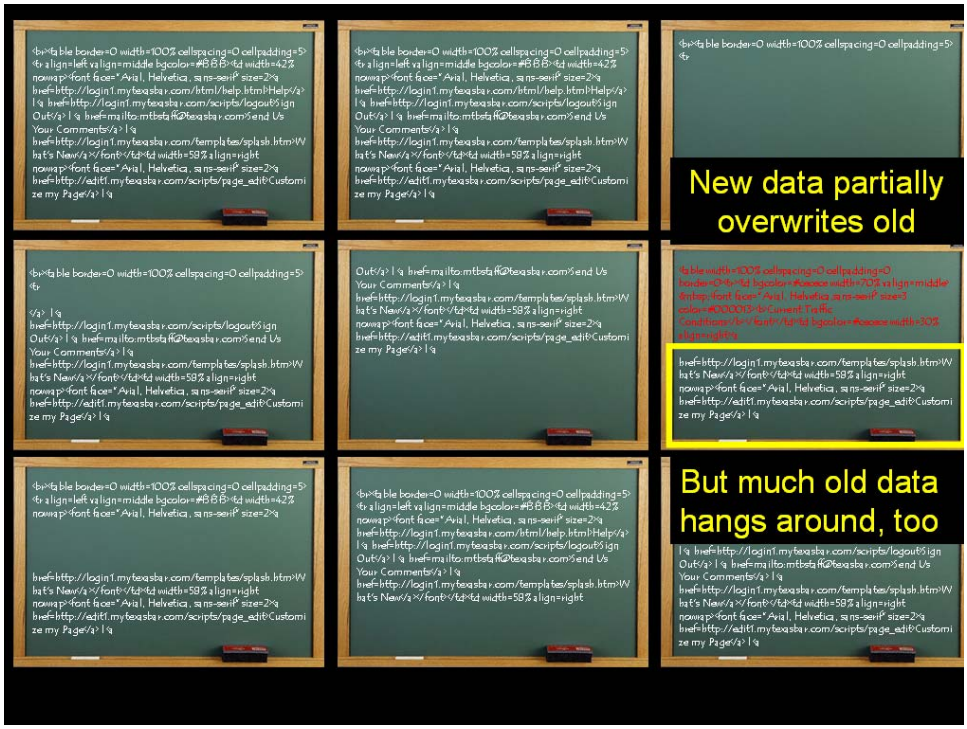

Microsoft Operating System
FAT
File System
NTFS
File Allocation Table
Index
Master File Table


The File System
INDEX
FILES

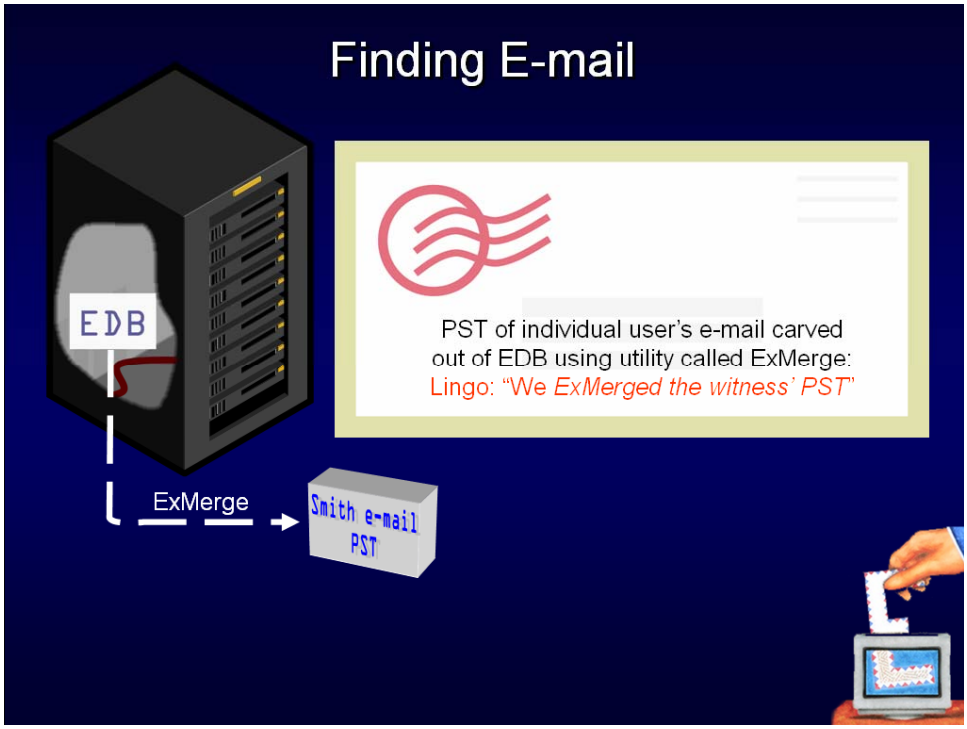

Open memo.txt
How Windows Manages Files
FILE TABLE
File Name
Cluster
Quicken.exe
9157
9158
Spreadsheet.xls
9159
9160
Memo.txt
9161
9163
image.gif
9162
9164
Outlook.pst
9165
9166
Quicken
spreadsheet
memo
gif


How Windows Deletes Files
FILE TABLE
Cluster
Free
9157
9158
9159
9160
9161
9162
9163
9164
9165
9166
Deleted,
but still here
memo
spreadsheet


How Windows Deletes Files
FILE TABLE
Cluster
Free
9157
9158
9159
9160
9161
9162
9163
9164
9165
e-mail
e-mail
e-mail
e-mail
gif
gif
slack space
slack space
unallocated clusters
spreadsheet
spreadsheet
spreadsheet
Most of the deleted memo survives


New data partially overwrites old
But much old data hangs around, too


4
E-mail's rarely gone--
just harder to find


Finding E-mail
EDB
PST of individual user's e-mail carved
out of EDB using utility called ExMerge:
Lingo: "We ExMerged the witness' PST"
ExMerge
Smith e-mail
PST

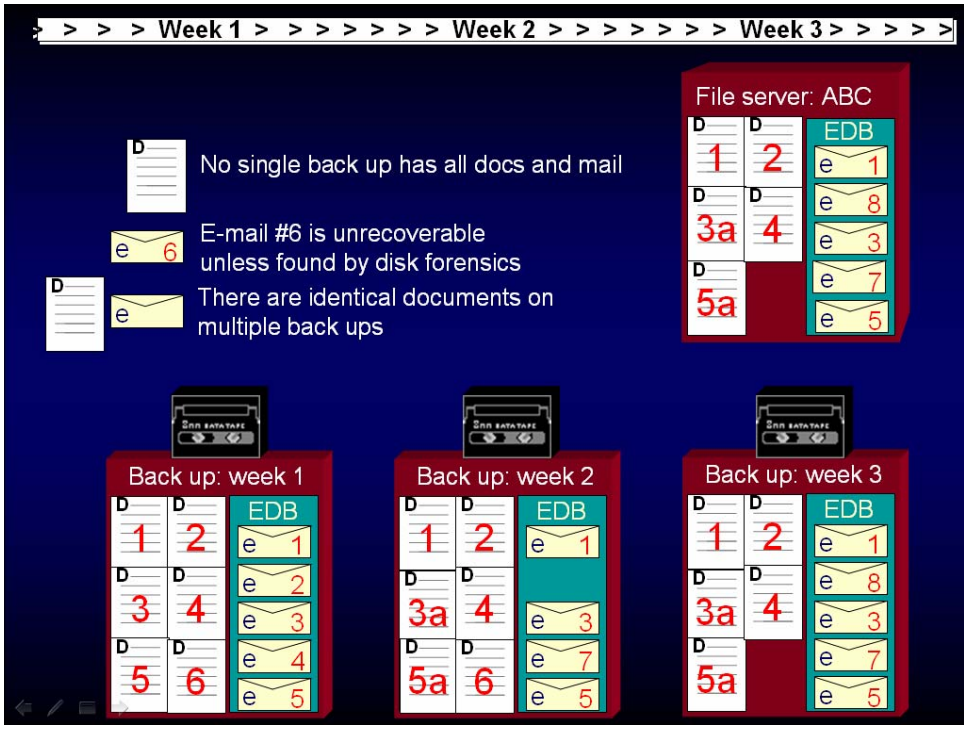

Finding E-mail
Back Up
EDB
EDB files backed up to digital tape
Lingo: "What's the *rotation interval*?"
Smith e-mail PST
Backup Tapes


Finding E-mail
Local e-mail storage:
• Outlook.pst
• Archive.pst
• Outlook.ost
• Local backups
• "Orphaned" mail client collections
• Web cache for web mail
• Forensic recovery from container files
Local Machines
Smith e-mail PST
Local PSTs/OSTs
Backup Tapes


Finding E-mail
Other Sources:
• PDAs and synch files
• Other custodians' collections
• Forwarded to personal accounts
• Legacy machines
• Message threads
• Forensics analysis of media
Smith e-mail PST
Local PSTs/OSTs
Backup Tapes
E-mail is rarely gone.
It just gets harder to find.

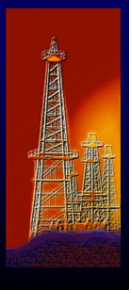

The White House E-Mail Mess
Target Interval: September 30 – October 6, 2003
OVP Exchange Server
Where's the e-mail?
PST Storage Server
Journaling
Journal Server
Backup Tapes
10/21/03
10/21/03
10/21/03


5
Why are backup tapes a headache?


Week 1
Week 2
Week 3
No single back up has all docs and mail
E-mail #6 is unrecoverable unless found by disk forensics
There are identical documents on multiple back ups
File server: ABC
EDB
Back up: week 1
Back up: week 2
Back up: week 3

Tips for Tapes

- Exhaust accessible sources
- Prove cost and burden
  - Was accessible ESI moved to tape?
  - How are tapes used? DR vs. Archive?
- *Sample* likely prospects
  - Furnish info about tapes to Req. Pty
  - Permit Req. Pty to choose samples
  - Tailor to custodians and intervals, not #s
  - Distinguish cost to restore vs. to review



6

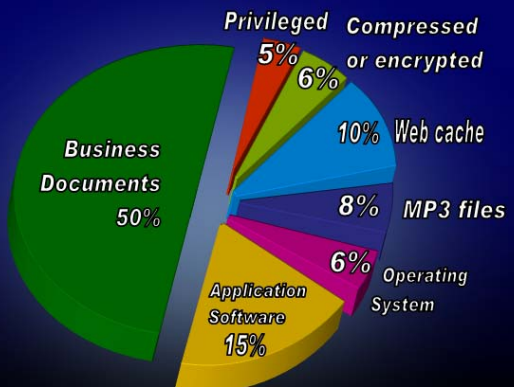
Filtering can be a smart, effective condition to impose

Metrics

Page equivalency absent sampling and metrics is a sham

Best practice: Parties establish metrics using samples

Data Biopsies



Hashing is a need-to-know technology *and* a useful condition

Hashing: 'need to know' technology

Algorithms "fingerprint" data

Different File names

Name	Size
PornDemoPicture1.jpg	7825
WindowsUpdate.dll	7825

Output "Message Digest" value

Identical Hash Values

95168DDADA3F7CE9A7BD98FA67C183BC

95168DDADA3F7CE9A7BD98FA67C183BC

Identical Contents

Identical Contents

PornDemoPicture1.jpg

WindowsUpdate.dll

Most Common: MD5 and SHA-1

Hashing: 'need to know' technology

Algorithms "fingerprint" data

Different File names

Name	Size
PornDemoPicture1.jpg	7825
WindowsUpdate.dll	7825

Identical Contents — Identical Contents

PornDemoPicture1.jpg — WindowsUpdate.dll

Most Common: MD5 and SHA-1

Output "Message Digest" value

Identical Hash Values

95168DDADA3F7CE9A7BD98FA67C183BC

95168DDADA3F7CE9A7BD98FA67C183BC

Single file or entire drive

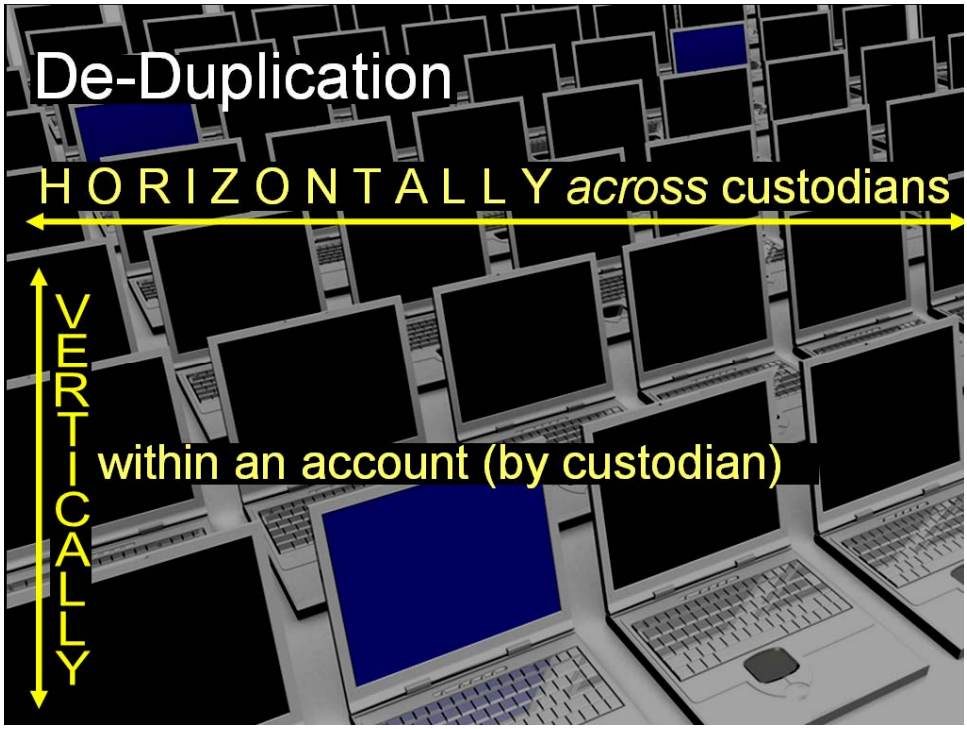
Locate identical files

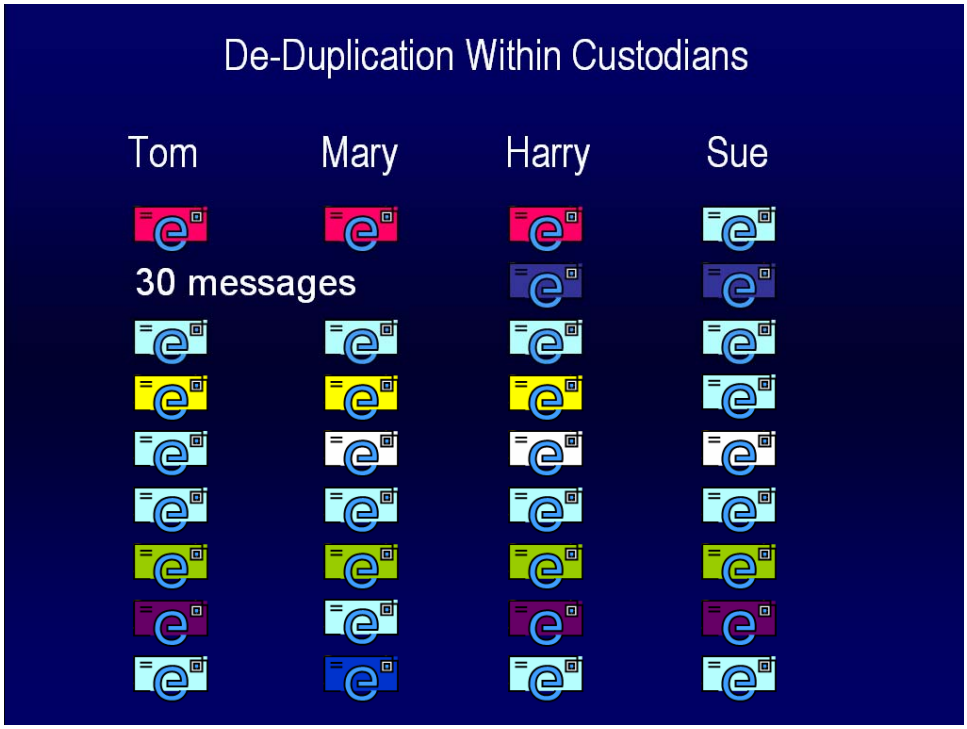
HASH SET

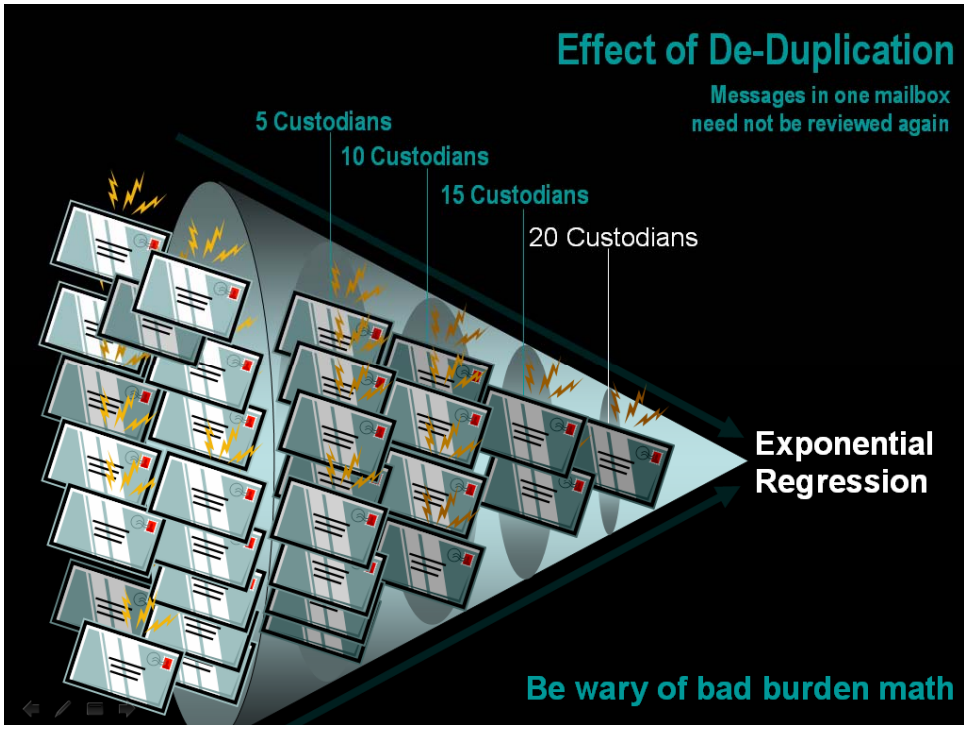
KNOWN HASH VALUES: TRADE SECRETS

FD12F4360A966E52FA988819979402B6
40497C63180EDA91EE580116AC7356B4
1D0B7D58E960AEBE2653181D9FD843AD
95168DDADA3F7CE9A7BD98FA67C183BC
2B8179DB7D7AA3683BF5CAA15FFA595A

Deduplication
is another
can't-do-without-it
technology

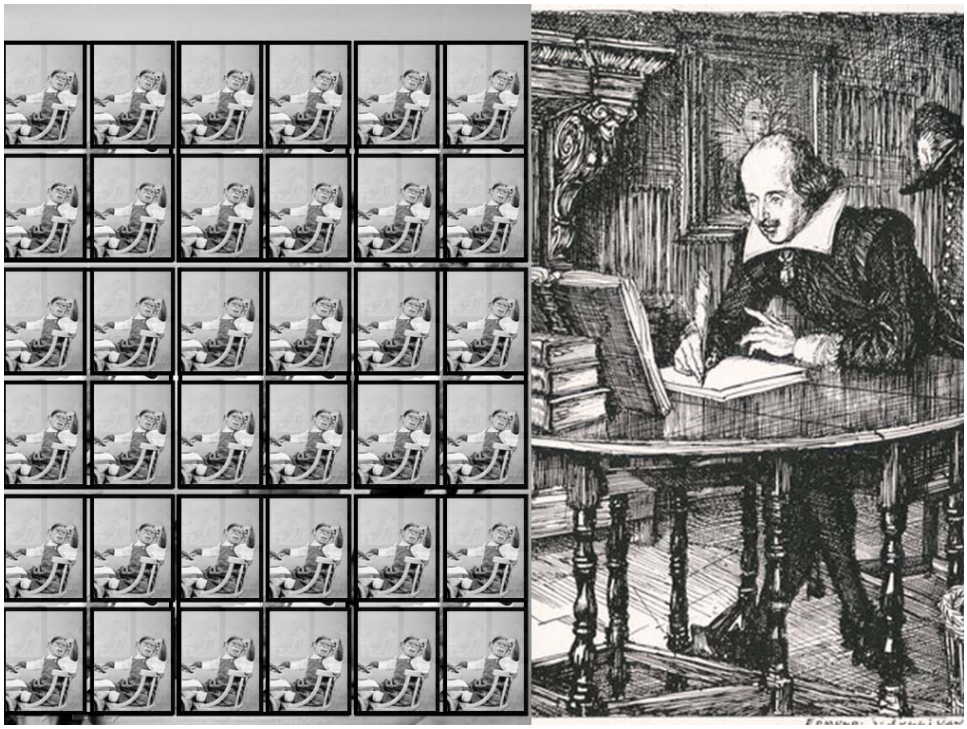

De-Duplication
HORIZONTALLY *across* custodians
VERTICALLY
within an account (by custodian)


De-Duplication Within Custodians
Tom
Mary
Harry
Sue
30 messages



9

Keyword searching is trickier than most people think



Keyword Search Challenges

- Noise Words and Stopwords
  - Too short or too common:
    - Dell" "HP" "JP"

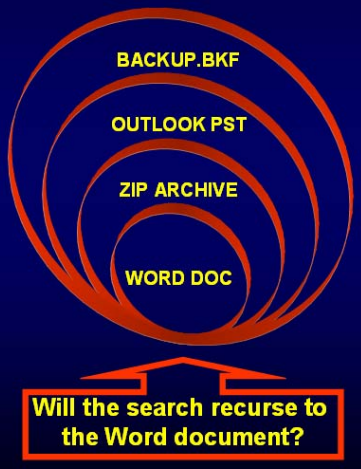
      -Peskoff: 400,000 noise hits in 60GB data
      28,000 in email alone

      39 times in Mary Kelley's e-mail

      -Proposed in Recent Cases: "S" "64"
- Misspellings: Mispellings, Misspelings?
- Acronyms and IM-Speak
- Stemming: Backda*: backdate, backdated, backdating

Keyword Search Challenges

- Recursion
  - Will the search tool drill down as far as needed?

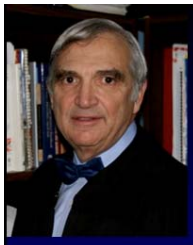


Search is a *science*

A lot rides on it

"Let's try these" isn't good enough

Must *test* searches



United States v. O'Keefe,
No. 06-249 (D.D.C. Feb. 18, 2008)
Hon. John M. Facciola

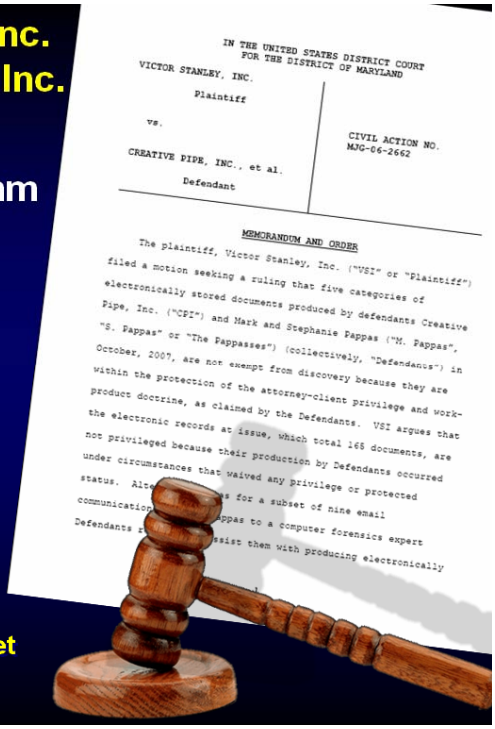
"Whether search terms or "keywords" will yield the information sought is a complicated question involving the interplay, at least, of the sciences of computer technology, statistics and linguistics."

United States v. O'Keefe,
No. 06-249 (D.D.C. Feb. 18, 2008)
Hon. John M. Facciola

"[F]or lawyers and judges to dare opine that a certain search term or terms would be more likely to produce information…is truly to *go where angels fear to tread."*

Victor Stanley, Inc. v. Creative Pipe, Inc.

No. MJG-06-2662
(D.Md 5/29/08)
Hon. Paul Grimm



IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF MARYLAND

VICTOR STANLEY, INC.
Plaintiff
vs.
CREATIVE PIPE, INC., et al.
Defendant

CIVIL ACTION NO. MJG-06-2662

MEMORANDUM AND ORDER

The plaintiff, Victor Stanley, Inc. ("VSI" or "Plaintiff") filed a motion seeking a ruling that five categories of electronically stored documents produced by defendants Creative Pipe, Inc. ("CPI") and Mark and Stephanie Pappas ("M. Pappas", "S. Pappas" or "The Pappasses") (collectively, "Defendants") in October, 2007, are not exempt from discovery because they are within the protection of the attorney-client privilege and work-product doctrine, as claimed by the Defendants. VSI argues that the electronic records at issue, which total 165 documents, are not privileged because their production by Defendants occurred under circumstances that waived any privilege or protected status. Alt... for a subset of hine email communication... ppas to a computer forensics expert Defendants r... ssist them with producing electronically

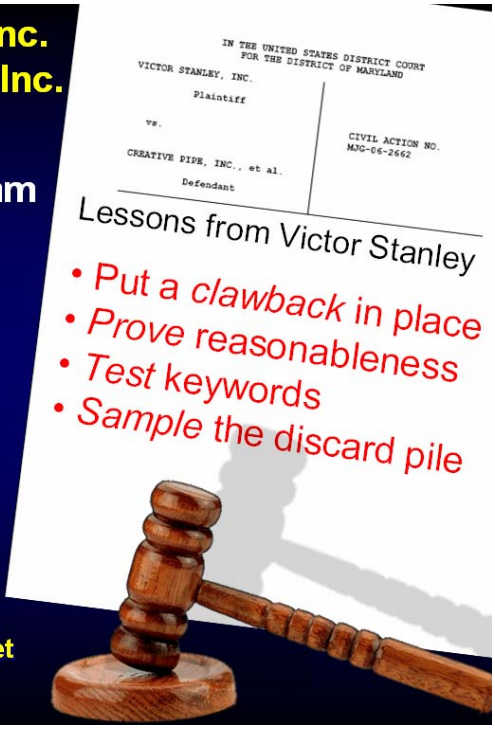
Creative Pipe's lawyers produced 165 documents claimed as privileged.

Ruling: Privilege waived:

- **Abandoned clawback**
- **No search expertise**
- **Failed to test keywords**
- **Didn't sample production set**

Victor Stanley, Inc. v. Creative Pipe, Inc.

No. MJG-06-2662
(D.Md 5/29/08)
Hon. Paul Grimm



IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF MARYLAND

VICTOR STANLEY, INC.
Plaintiff
vs.
CREATIVE PIPE, INC., et al.
Defendant

CIVIL ACTION NO. MJG-06-2662

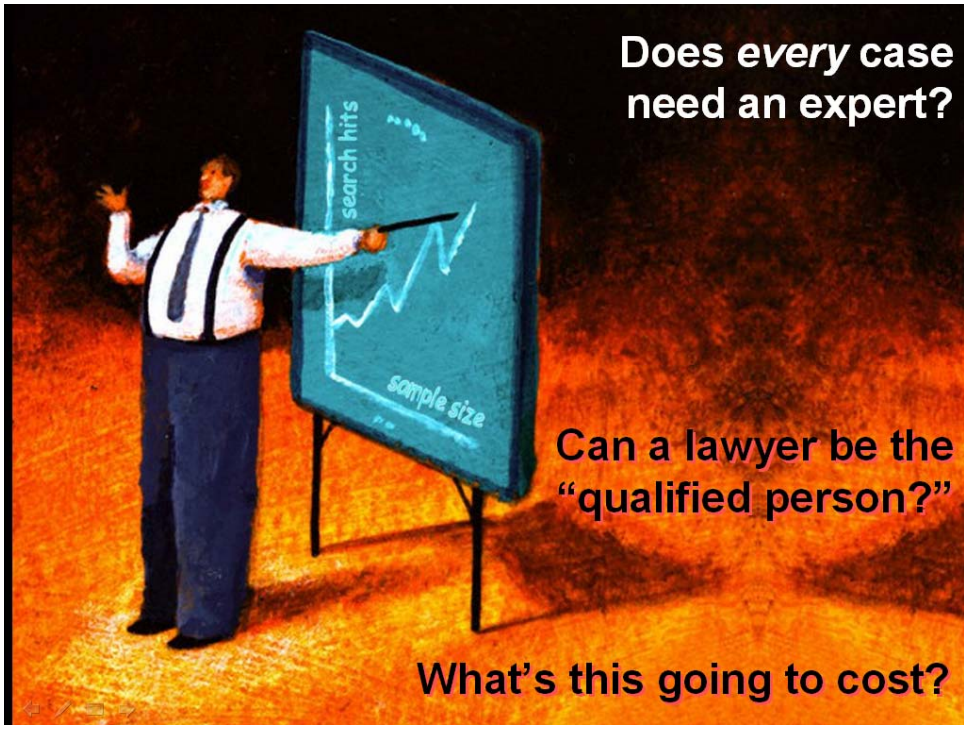
Lessons from Victor Stanley

- *Put a clawback in place*
- *Prove reasonableness*
- *Test keywords*
- *Sample the discard pile*

Creative Pipe's lawyers produced 165 documents claimed as privileged.

Ruling: Privilege waived:

- **Abandoned clawback**
- **No search expertise**
- **Failed to test keywords**
- **Didn't sample production set**


Does *every* case need an expert?
search hits
sample size
Can a lawyer be the "qualified person?"
What's this going to cost?


10
Metadata's the least understood & most abused topic in electronic discovery

How many *accessible* metadata fields for a Word document?

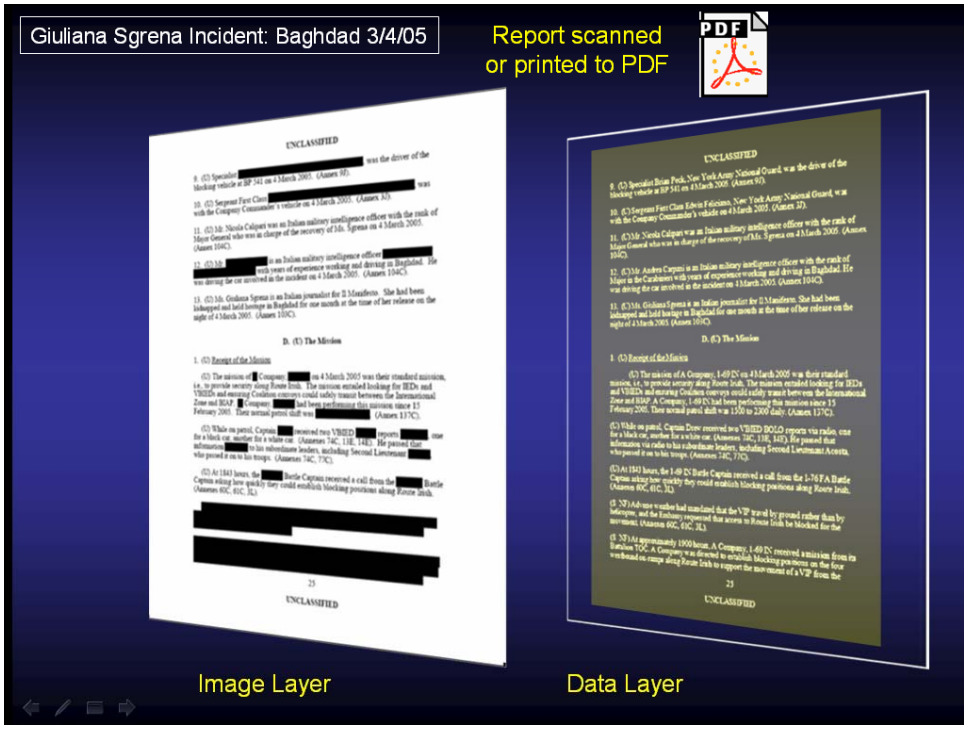
Some relevant

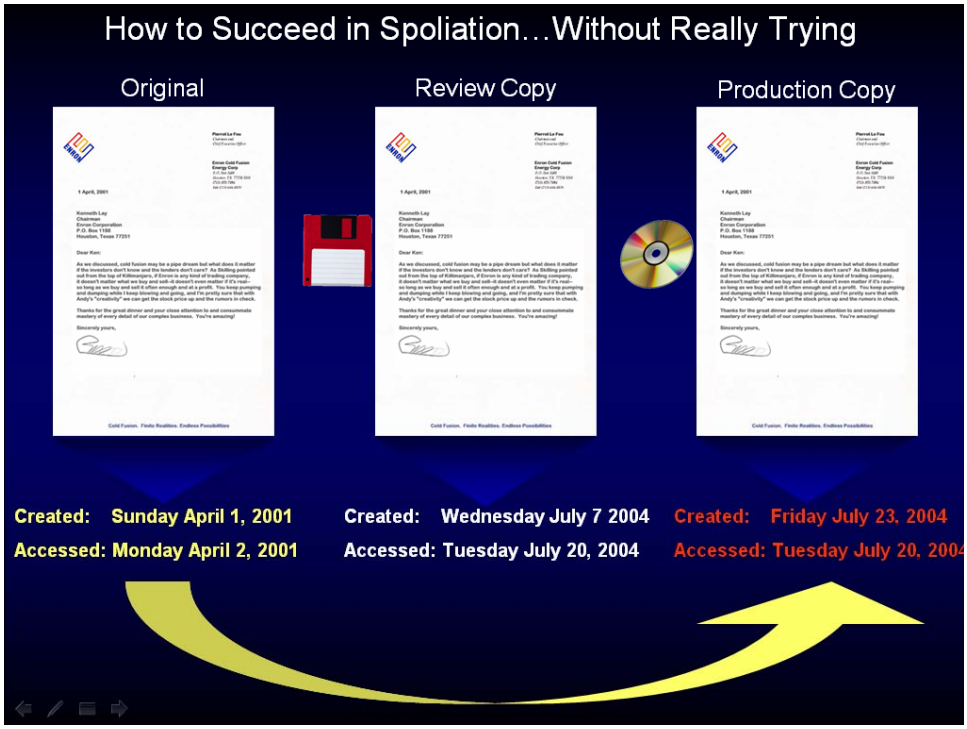
Some less so

Name
Type
Location
Size
MS-DOS name
Date Created
Date Modified
Date Accessed
Date Printed
Attributes
Status
Owner
Author
Title
Subject
Category
Pages
Comments
Copyright
Artist
Album title
Year
Track number
Genre

Dimensions
Episode name
Program description
Audio sample size
Audio sample rate
Channels
Company
Description
File version
Product name
Product version
Keywords
Manager
Hyperlink base
Template
Last saved by
Revision number
Total editing time
Checked by
Client
Date completed
Department
Destination
Duration

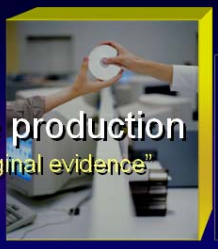
Protected
Camera model
Date picture taken
Disposition
Division
Document number
Editor
Forward to
Group
Language
Mailstop
Matter
Office
Project
Publisher
Purpose
Received from
Recorded by
Recorded date
Reference
Source
Telephone number
Typist
Bit rate




How to Succeed in Spoliation…Without Really Trying
Original
Review Copy
Production Copy
Created: Sunday April 1, 2001
Accessed: Monday April 2, 2001
Created: Wednesday July 7 2004
Accessed: Tuesday July 20, 2004
Created: Friday July 23, 2004
Accessed: Tuesday July 20, 2004


11
Selecting the right
Forms of Production
is important

Forms of production



Native production
"the original evidence"

PROS	CONS
Lowest cost Includes metadata Fully functional "What if" manipulation	Viewer applications Potential for alteration Redaction challenges No Bates numbering

Image production
"a picture of the evidence"

Easy to view Static Easy to Redact Bates numbering	Expensive Strips metadata Not functional No "What if" capability

TIFF Production requires "Load Files"

Other Forms: Joint Hosted
Paper

There's no "one size fits all" form of production.
Different data demands different forms

12

Getting the geeks together really helps

When were the relevant events and intervals?

What's the case about?

Who was involved?

What's being preserved?

How is the data stored?

Asking the right questions

What's relevant?

How will ESI be produced?

Where is the data located?

How is it vulnerable to spoliation?

Who knows the systems, applications and processes?

When will information be produced?

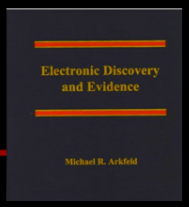
How can it be searched, filtered and reviewed?

What *won't* be considered?

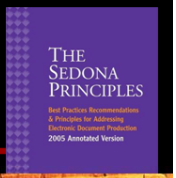
What do we have?

Recommended Reading

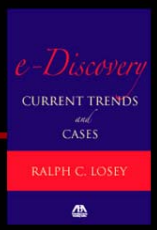
Electronic Discovery and Evidence
by Michael Arkfeld



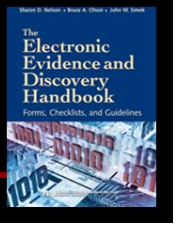
The Sedona Principles: Best Practices Recommendations & Principles for Addressing Electronic Document Production



E-Discovery: Current Trends and Cases
by Ralph C. Losey



The Electronic Evidence and Discovery Handbook: Forms, Checklists and Guidelines
by Sharon D. Nelson, Bruce A. Olson, John W. Simek



E-Discovery's easier when you get in touch with *your* inner nerd

Thank You

craig@ball.net