Ball 6 on EDD

Six Articles on Electronic Data Discovery

Hitting the High Points of the New e-Discovery Rules
Discovery of Electronic Mail: The Path to Production
Metadata: Beyond Data About Data
The Plaintiffs' Practical Guide to E-Discovery
The Perfect Preservation Letter
"Ball in Your Court" EDD Columns April 2005 - September 2007

Craig Ball

Six on EDD

Six Articles on Electronic Data Discovery

By Craig Ball

Everyone uses computers—at home, at work, on the road, leaving voicemail, opening card key doors--everywhere, every day. Nearly all information is born digitally, and just a fraction of it is printed. Zealous advocates can't walk away from the electronic evidence that never makes its way to paper or turn a blind eye to its metadata. We must master electronic discovery and learn to exploit its power to bring forward the best evidence in cost-effective ways. These six articles offer practical strategies geared to helping you succeed in e-discovery.

Contents:

Hitting the High Points of the New e-Discovery Rules
By Craig Ball

The last time the Federal Rules of Civil Procedure were amended to deal with electronic evidence, eight-track tapes were the hot technology, the Internet and cell phones were the stuff of science fiction and computers were room-sized behemoths owned by corporations, universities and governments. Times have changed, and the Rules have again changed with the times.

For six years, some of the best minds of the bench and bar worked to amend the Rules to address the enormous challenge posed by discovery of electronic evidence. These amendments took effect on December 1, 2006 and, even if you don't regularly appear in Federal court, the new rules merit your attention because they're sure to rapidly impact state court practice, too.

Here's a synopsis of the principal amendments, along with some observations about their operation and impact.

Introducing ESI
There's a new species of evidence in town. It's called **ESI**, for **E**lectronically **S**tored **I**nformation, and it encompasses any potentially relevant data that's stored on computers, disks, tape, gadgets and the Internet.

The amendments don't so much create new rights as compel lawyers and litigants to deal with the central role computers and the Internet play in business and our lives. ESI comprises a startling 95% of all information created nowadays, yet legions of lawyers have been remiss in marshalling this rich evidentiary resource, preferring instead to focus on familiar paper documents. The Rules make clear that discovery of ESI stands on equal footing with discovery of paper documents and require that any request for production of documents be understood to include a request for ESI. Although the Committee that drafted the ESI amendments could have stretched the definition of "document" to include ESI, they wisely recognized that more was needed. After all, so much of the electronic information that impacts our lives—like databases, web content, voice messaging, even spreadsheets—bears little resemblance to conventional documents. Instead, ESI is defined broadly to encompass the forms computer-based information takes today and adapt to whatever tomorrow brings.

The upshot of the new Rules is that:

- ESI is discoverable
- Litigants must preserve and produce ESI
- Lawyers must understand how to request, protect, review and produce ESI
- The courts have the tools to rectify abusive or obstructive electronic discovery

Preservation of ESI
Though they don't detail what litigants must do to meet their obligation, the amended Rules are grounded on the expectation that all parties will preserve potentially relevant ESI. Not only must accessible ESI be preserved, but electronic information that a party deems inaccessible must also be preserved so as not to preempt an opponent's right to compel production.

Preservation of ESI is challenging. Information stored on computers always consists of two or more "chunks" of data, typically a "file" plus information called system "metadata" describing the characteristics of the file and its place within the computing environment: location, size, name, origins and history. Metadata take many forms and are often important evidence in their own right. However, as metadata are designed to change in order to track such things as file access, modification and relocation, metadata can be quite fluid and its preservation demands special handling.

In a nod to preservation pitfalls inherent to a fluid and autonomous environment, the Rules now grant a measure of protection to those who act diligently to preserve data but fail. Absent exceptional circumstances, amended Rule 37(f) prohibits a court from imposing sanctions under the Rules for the failure of a party to provide ESI lost as a result of the "routine, good-faith operation of an electronic information system." But it's a shallow "safe harbor" because, absent reasonable and timely preservation, a court isn't likely to see good faith.

Third Party Preservation and Production

Many custodians store your ESI, and some are third parties subject to your direction or control. In framing your litigation hold, you may need to advise outside counsel, accountants, application service providers and contractors to hang on to what they have and, if you're the requesting party, don't forget to address third parties in preservation letters. Note that Rule 45 governing subpoenas has been amended to support discovery of ESI from third parties, as well as to protect third parties from unduly burdensome requests for ESI.

Stepping Forward

Amended Rule 26(a) requires that a party must, without awaiting a discovery request, promptly identify or produce ESI, documents and tangible things in the party's possession, custody or control that the party may use to support its claims or defenses. ***You've got to step forward without a discovery request.***

The duty to affirmatively disclose material supporting claims and defenses has been in the Rules for some time, but its burdens were constrained by paper's self-limiting nature. Paper is expensive and takes up space, so we tended to organize it or get rid of it. But trillions of bytes weigh nothing at all and occupy little space. Consequently, it's all-too-easy to amass electronic information in unstructured volumes that, if on paper, would have driven us from our homes and offices long ago.

In order to promptly identify and produce ESI, we must first be able to find, preserve, collect, manage, review and duplicate it—a capacity few litigants and fewer lawyers currently possess, but that all must acquire. Until that occurs—and until better electronic records management emerges—the need to promptly step forward with ESI will be a frequent stumbling block.

Meet and Confer

Under the new Rules, parties must not only be prepared to swiftly produce the ESI they expect to use, but they must also, very early on, be fluent and forthcoming about their preservation of ESI and issues relating to its disclosure or discovery. Amended Rule 26(f) requires that parties meet and confer shortly after the response date to address any issues relating to ESI, including its preservation and the form or forms in which it should be produced. The courts expect the

conferees to arrive with answers and exert a genuine, good faith effort to resolve e-discovery questions.

In some jurisdictions, the Rule 26(f) conference has been something of a "drive by" event. Now, counsel must be prepared to field questions about information systems, back up and retention practices, customary formats and applications, data location, volume and composition and a host of other unfamiliar topics. Counsel needs to either know the systems and applications well or bring along someone who does. Early and earnest cooperation with the other side and transparency of process will be essential, and adversarial posturing is best checked at the door.

Inaccessibility

Under the Rules, no party need produce ESI that's "not reasonably accessible," but if an opponent objects, the claim of reasonable inaccessibility must be proven. No one yet knows what is and is not "reasonably accessible" where ESI is concerned, and the definitions extant vary according to whose ox is gored.

When a substantial volume of ESI is implicated, producing parties will claim inaccessibility despite ready access to individual files arguing that burden alone makes such access unreasonable. Information will be claimed to be inaccessible if stored on tape or consisting of data rarely accessed or simply so disorganized or commingled with privileged material that it's costly to review.

Requesting parties will counter that anything's accessible if you devote sufficient money and effort to the task—a contention technically accurate and generally impracticable. Judges—principally federal magistrates—will bridge the chasm as best they can. Plan on the court asking *why* ESI is difficult to access and pressing counsel to articulate *exactly* what they're seeking and *why* they need it. Some judges will say, "Just give it to them" or "Do you want it badly enough to pay for it?" Mostly, we can expect to hear, "Get your technical people down here, go into my jury room and don't leave until you've worked it out." The inevitable "splitting of the baby" necessitates that companies hone their ability to rapidly assess the who, what, when, where and how much for their ESI.

The Next Hurdle

Even if you satisfy the court that your side's ESI isn't reasonably accessible, the war isn't over. A requesting party may compel production of inaccessible ESI by showing good cause. If you're ordered to produce for good cause shown, you can ask the court to tailor the production order to minimize your burden, perhaps in ways such as cost shifting that may persuade the requesting party to narrow or abandon the request. The court may also impose conditions to minimize undue burden by, e.g., granting access to less than the all potentially responsive ESI (*sampling* parts of the data to assess its value to the case) or by requiring the use of data filtering and search mechanisms to narrow the scope of review and production.

Cost Shifting

Cost shifting has a salutary chilling effect on abusive or sloppy discovery, but risks closing the courthouse to meritorious claims against large enterprises and parties with poorly managed information. Here again, magistrates are the ones in the trenches and, in balancing the equities, must be sensitive to the impact of costs and cost shifting, promoting proportionality without erecting barriers to justice. Abusive and overbroad discovery must have consequences. So, too, must the twin goals of expediency and cost-efficiency enunciated by the Rules be furthered.

From either side's vantage point, there's little incentive to be fast, frugal or focused when the opposition's footing the bill.

Forms of Production
Unless the other side expressly agrees or the court orders it, you can't produce paper printouts of documents when the originals you hold are electronically searchable.

Per amended Rule 34(b), your opponent selects the form or forms in which you produce ESI. If you don't produce as designated, you must produce as maintained in the course of business or in a reasonably usable form. Whether you preserved electronic searchability will be a decisive factor in assessing usability.

You can't produce ESI in a form different from that selected by the requesting party unless you advise them of the form or forms you'll supply and afford them an opportunity to object and seek assistance of the court. An unceremonious "here it is" courts trouble. It's unclear how long before the production deadline you must make the alternate format disclosure; that is, it's not specified whether a producing party designating a production format on the thirtieth day may then wait for objection, or must make the designation earlier and produce in the specified format on the thirtieth day.

Clawback Protection
The inadvertent production of privileged information is every lawyer's nightmare. It occurred with regularity in the discovery of paper evidence even when we examined every page before production. But as the volume of ESI has grown, we progressively lost our ability to review everything item-by-item. Plus it's increasingly common for privileged and non-privileged content to insidiously mix, e.g., privileged exchanges embedded in a thread of an apparently benign e-mail or within metadata. Recognizing the growing potential for inadvertent production, amended Rule 26(b)(5)(B) permits a party who has produced privileged or work product data to notify any party receiving the data of the fact of and basis for the privilege claim. After notice, any party receiving the allegedly privileged material must return, sequester or destroy the specified information, retrieve any copies shared with non-parties and may not use or disclose the allegedly privileged ESI until the claim of privilege is resolved.

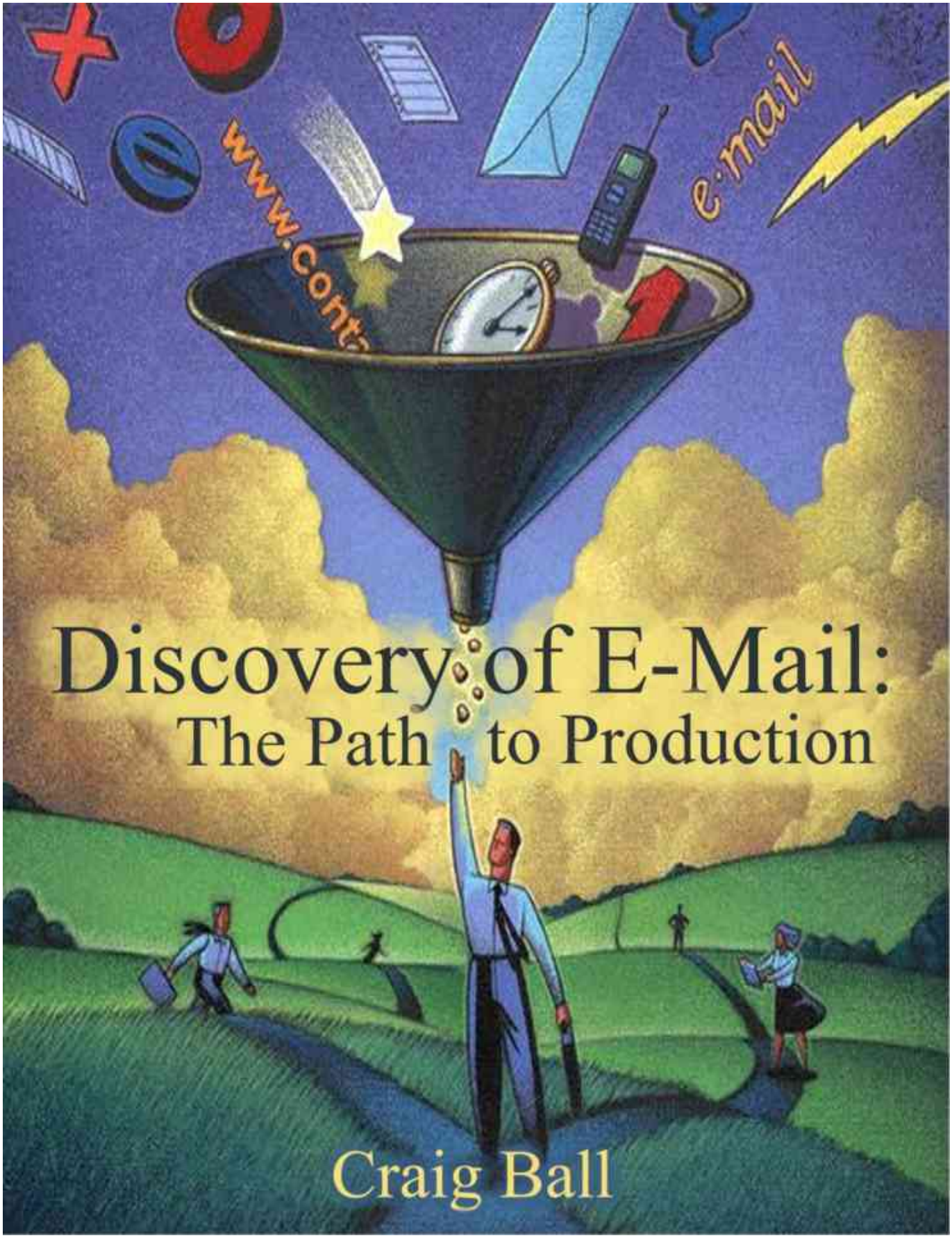
The amendment allows any party challenging the claimed privilege the right to promptly present the disputed ESI to the court under seal, with the producing party obliged to preserve the disputed material until resolution.

The amended Rule *doesn't* affect the substantive law governing privilege; that is, the law of the forum or other applicable law still controls. Instead, the Rule describes a framework for preservation of rights and presentation of claims. The ESI at issue must be logged as other privileged materials but it's unclear how much detail is required. Must privileged ESI be logged item-by-item or may be described more broadly so long as the description is sufficiently clear and complete to permit the parties and the court to understand the basis for the claim and determine whether waiver has occurred?

Fear, Uncertainty and Doubt
Life under the new Rules won't be easy, but overdue changes are often hardest. Requesting parties will curse the delay and unpredictability of the two-tiered inaccessibility analysis. Responding parties will bemoan the necessity and cost to collect, review and produce all the

relevant electronic evidence they've blissfully been ignoring heretofore. First forays into meet-and-confer will consist of two people who don't trust each other negotiating issues neither understands. Judges and magistrates will see dockets swell with e-discovery disputes. All will wish they better understood computers and that, in the rush to embrace automation, we hadn't all been so quick to abandon records management.

Electronic evidence isn't going away. It grows more important—more revealing, more varied, more detailed—each day. Despite the confusion and cost, the Rules amendments insure that electronic discovery receives the overdue focus it warrants, and as we learn more about digital evidence and become adept at seeking, identifying, preserving, searching and producing ESI, the fear, uncertainty and doubt will go the way of eight-track tapes.


www.cont
e-mail
Discovery of E-Mail:
The Path to Production
Craig Ball

Discovery of Electronic Mail: The Path to Production

Asked, “Is sex dirty,” Woody Allen quipped, “Only if it’s done right.” That’s electronic discovery: if it’s ridiculously expensive, enormously complicated and everyone‘s lost sight of the merits of the case, you can be pretty sure you’re doing it right.

But it doesn’t *have* to be that way.

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. It offers no short cuts, but that’s not the point. The goal is to keep you from stepping off a cliff. Not every point outlined here is suited to every production effort, but all deserve *consideration* every time.

Think Ahead

False starts and missteps in electronic discovery are painfully expensive, or even unredeemable if data has been lost. One way to avoid re-treading ground is to question expectations from the outset.

Will the data produced:

- *Integrate paper and electronic evidence?*
- *Be electronically searchable?*
- *Preserve all relevant metadata from the host environment?*
- *Be viewable and searchable using a single application?*
- *Be Bates numbered, and by what method?*
- *Be easily authenticable for admission into evidence?*

After attorney review, data harvest is byte-for-byte the costliest phase of electronic discovery. Understandably, producing parties want to search once and be done with it and confine the requesting party to a single list of keywords. From the requesting party’s perspective, it’s often impossible to frame effective keyword searches absent familiarity with the argot used to describe the events and objects central to the case, resulting in keyword searching missing what well-trained reviewers would find.

Producing parties are often forced to return to the well. Where you anticipate that new keywords will emerge or different search techniques will be used, securing the least costly outcome warrants the most expensive beginning: compiling a comprehensive review set of all potentially relevant e-mail. This entails *identification, preservation, harvest* and *population.*

Identification

“Where’s the e-mail?” It’s a simple question, but one answered too simply—and erroneously--by, “It’s on the e-mail server” or “The last sixty days of mail is on the server and the rest is purged.” Certainly some of the e-mail will reside on the server, but just as certainly more, even *most,* e-mail is elsewhere, and it’s *never all gone* notwithstanding retention policies dictating it disappear. The true location and extent of the e-mail depends on systems configuration, user habits, back up procedures and other hardware, software and behavioral factors. This is true for mom-and-pop shops, for large enterprises and for everything in-between.

How thorough is your effort to identify e-mail? E-mail resides in some or all of the following venues, grouped according to relative accessibility:

Easily Accessible:

- Online e-mail residing in active files on enterprise servers
  - *MS Exchange e.g., (.EDB, .STM, .LOG files)*
  - *Lotus Notes (.NSF files)*
  - *Novell GroupWise (.DB files)*
- E-mail stored in active files on local or external hard drives and network shares
  - *User workstation hard drives (e.g., .PST, .OST files for Outlook and .NSF for Lotus Notes)*
  - *Laptops (same as above)*
  - *"Local" e-mail data files stored on networked file servers ("network shares")*
  - *Mobile devices (PDA, "smart" phones, Blackberry)*
  - *Home systems, particularly those with remote access to office networks*
- Nearline e-mail
  - *Optical "juke box" devices*
  - *Back ups of individual users' e-mail folders (i.e., "brick-level" back ups)*
- Offline e-mail stored in networked repositories
  - *e.g., Zantaz EAS®, EMC EmailXtender®,* Waterford MailMeter Forensic®

Accessible, but Often Overlooked:

- E-mail residing on remote servers
  - *ISPs (IMAP, POP, HTTP servers), Gmail, Yahoo Mail, Hotmail, etc.*
- E-mail forwarded and carbon copied to third-party systems
  - *Employee forwards e-mail to self at personal email account*
- E-mail threaded behind subsequent exchanges
  - *Subject and latest contents diverge from earlier exchanges lodged in body of email*
- Offline local e-mail stored on removable media
  - *External hard drives, thumb drives and memory cards*
  - *Optical media: CD-R/RW, DVD-R/RW*
  - *Floppy Drives, Zip Drives*
- Archived e-mail
  - *Auto-archived to additional .PST by Outlook or saved under user-selected filename*
- Common user "flubs"
  - *Users experimenting with export features unwittingly create e-mail archives*
- Legacy e-mail
  - *Users migrate from e-mail clients "abandoning" former e-mail stores*
- E-mail saved to other formats
  - *.pdf, .tiff, .txt, .eml, etc.*
- E-mail contained in review sets assembled for other litigation/compliance purposes
- E-mail retained by vendors or third-parties (e.g., former service provider)
- Print outs to paper

More Difficult to Access:

- Offline e-mail on server back up media
  - *Back up tapes (e.g., DLT, AIT)*

- E-mail in forensically accessible areas of local hard drives
  Deleted e-mail
  Internet cache
  Unallocated clusters

The issues in the case, key players, relevant times, agreements between the parties and orders of the court determine the extent to which locations must be examined; however, the failure to *identify* all relevant e-mail carries such peril that caution should be the watchword. Isn't it wiser to invest more to know *exactly* what the client has than concede at the sanctions hearing the client failed to preserve and produce evidence it didn't know it had because *no one bothered to look for it*?

What's more, thorough identification of both accessible and inaccessible electronically stored information isn't just sound practice; it's a tenet of federal procedure. Effective December 1, 2006, the Federal Rules of Civil Procedure require a party to provide, *without a discovery request*, "a copy of, or a description by category and location of, all documents, *electronically stored information*, and tangible things that are in the possession, custody, or control of the party and that the disclosing party may use to support its claims or defenses, unless solely for impeachment." FRCP Amended Rule 26(a)(1)(B) [emphasis added]. Amended Rule 26(b)(2)(B) requires a party to identify electronically stored information [ESI] it deems "not reasonably accessible because of undue burden or cost." The commentary to the new Rules makes clear that the duty to identify attaches to every item of potentially responsive ESI:

> "The responding party must also identify, by category or type, the sources containing potentially responsive information that it is neither searching nor producing. The identification should, to the extent possible, 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 identified sources."
>
> *FRCP, Comments to Amended Rule 26, subdivision (b)(2).*

Preservation

The duty to preserve potentially relevant evidence is generally triggered by the anticipation of a claim. Fulfilling a preservation duty with respect to e-mail is made harder by the control reposed in individual users, who establish quirky folder structures, commingle personal and business communications and—most dangerous of all—control deletion and retention of their messages. Individual users who may hold responsive ESI must be directed to retain all potentially relevant messages and be furnished *sufficient* information at appropriate intervals to assess relevance *consistently.* You've got to guard against human frailty in your preservation strategy, so *don't leave the fox guarding the henhouse.* Act promptly to protect data from spoliation at the hands of users most inclined to sweep it under the rug.

Consider the following as parts of an effective e-mail preservation effort:

- Litigation hold notices to users, including clear, practical and specific retention directives
  Notices should remind users of relevant places where their email may reside
  Be sure to provide for notification to new hires and collection from departing employees
- Suspension of "retention" policies that call for purging email

- Suspension of re-use ("rotation") of back up media containing email
- Suspension of hardware and software changes which make email inaccessible
  Replacing back up systems without retaining the means to read older media
  Re-tasking or re-imaging systems for new users
  Selling, giving away or otherwise disposing of systems and media
- Preventing users from deleting/altering/corrupting email
  Immediate and periodic "snapshots" of relevant user email accounts
  Modifying user privileges settings on local systems and networks
  Archival by auto-forwarding selected e-mail traffic to protected storage
- Restricting activity—like moving or copying files—tending to irreparably alter file metadata
- Packet capture of Instant Messaging (IM) traffic or *effective* enforcement of IM prohibition
- Preserve potential for forensic recovery
  Imaging of key hard drives or sequestering systems
  Suspension of defragmentation
  Barring use of wiping software and encryption, with audit and enforcement

A threshold issue is whether there exists a duty of preservation going forward, e.g., with respect to information created during the pendency of the action. If not, timely harvest of data, imaging of drives and culling of relevant back ups from rotation (to name a few) may sufficiently satisfy the preservation duty so as to allow machines to be re-tasked, systems upgraded and back up tape rotation re-initiated. Seeking guidance from the court and working with opposing counsel to craft a preservation order help to insulate a producing party acting in good faith from subsequent claims of spoliation.

Harvest

Knowing what e-mail exists and where, and having taken proper steps to preserve it, it's time to gather potentially relevant messages and attachments into a **comprehensive review** set or select and assemble responsive items into a **preliminary production** set. The difference between the two is that a comprehensive review set is compiled largely without regard to what information will be selected for production. It's a "kitchen sink" assemblage, though ultimately its scope is constrained by the business units, facilities, machines and media selected for examination. By contrast, a preliminary production set is comprised of only those e-mails and attachments that the persons collecting the data from the various files and machines deem responsive to the production requests. When a corporate defendant relies upon each employee to locate and segregate responsive e-mails or when a legal assistant goes from office-to-office selecting e-mails, the resulting collection is a preliminary production set.

The principal advantage of selective harvest is that it cuts the number of messages and attachments subject to attorney review, reducing short run cost. These savings come with attendant risks, among them the need to return to every machine if the initial harvest proves insufficient, the much greater potential for loss or corruption of overlooked evidence and inconsistencies between reviewer judgments. Also, if keyword or concept searches are employed to select e-mail for harvest, be sure to weigh the concerns about such techniques that are discussed later in this article.

The advantage of a comprehensive review set is that despite a larger initial outlay, as new requests and issues arise, the comprehensive collection can be culled again-and-again at little incremental expense. Moreover, by broadly preserving e-mail, a comprehensive review set is a valuable hedge against spoliation claims. For entities subject to ongoing litigation and compliance production, such a comprehensive collection may also be availing in multiple matters.

Whichever method is used, special care must be taken during data harvest to preserve the integrity of the evidence. It's essential to maintain a sound c*hain of custody* for harvested data and be able to establish the *origin* of the e-mail (e.g., system, user account, folder and file from which it was collected) as well as the *custodian* of the e-mail. It's critical to understand that there is more to an e-mail than what a client application like Microsoft Outlook or Lotus Notes displays onscreen. When authenticity is challenged, the unseen header information or encoded attachment data is needed. Accordingly, select a harvest method that preserves *all of the data* in the e-mail.

Another chain of custody requirement is the ability to demonstrate that no one tampered with the data *between* the time of harvest and its use in court. Testimony of the custodians about handling and storage is one solution. Better still, cryptographic hashing, a form of digital "fingerprinting" applied to sections of each e-mail and attachments, generates a alphanumeric value that can be archived with the evidence and used to conclusively establish data integrity, if challenged.

Finally, there is also even more to an e-mail than its contents because, as is true of every file stored on a computer, there is associated *metadata* (data *about* data). Each email must be tracked and indexed by the e-mail client application ("application metadata") and every file containing the e-mail must be tracked and indexed by the file system of the computer storing the data ("system metadata"). E-mail metadata can be important evidence in its own right, helping to establish, e.g., whether or when a message was received, read, forwarded, changed or deleted. System metadata is particularly fragile since most computer users think themselves fully capable of copying a file from one medium to another and fail to appreciate that simply copying a file from, *e.g.,* one hard drive to another *changes the file's metadata* and potentially destroys critical evidence. Select your methods carefully to insure that the act of harvesting data as evidence doesn't alter the evidence or its metadata. If method chosen alters metadata, *archive the correct metadata before it changes*. Though cumbersome, a spreadsheet reflecting the original metadata is preferable to spoliation. Electronic discovery and computer forensics experts can recommend approaches to resolve these and other data harvest issues.

Population

Your scrupulous e-mail harvest is complete, but what you've reaped is no more ready to be searched for evidence than wheat is fit to be a sandwich. Harvested data arrives in varying incompatible formats on different media. Expect massive database files pulled from Microsoft Exchange and Lotus Domino Servers, .PST and .NSF files copied from local hard drives, HTML pages of browser-based e-mail, paper printouts, .PDF and .TIFF images (some searchable, some not) and all manner of forms and formats described in the Identification section, above. Were you to dump it all on a big hard drive and try to view it or run keyword searches, you'd quickly discover it yields up little information. That's because most of the data isn't stored as text. Some of it is locked up (password protected), some encrypted (e.g., Lotus Notes files) and some compressed, which frustrates text searching as effectively as encryption. The scanned

data is a picture, not text, and the e-mail attachments are encoded in a hieroglyphic called "Base 64." Other information lies buried in so-called "compound files" housing "nested" data, created when attachments themselves contain attachments.

Before search tools and reviewers can do their jobs, the harvested data must be deciphered and reconstituted to be accessible and re-appear as the *words* we see when using e-mail clients and word processors. This is accomplished by, for example,

- Opening password protected files
- Decrypting container files and items (e.g., Lotus Notes .NSF)
- Decompressing email container files (e.g., Outlook .PST, .EBD, .OST)
- Converting attachments to compatible formats (e.g., Base64, MIME)
- Decompressing and decrypting attachments (e.g., .ZIP, .XLS,)
- Optical character recognition of document image attachments (e.g., .TIFF)
- Identifying Unicode-formatted and foreign language attachments and documents (e.g., .DOC)
- Accessing files in obscure or proprietary formats
- Repairing corrupted files

By this point, decisions must be made as to what media and methods will be used to host and review the data. Will counsel for the producing party pore over CDs, DVDs or portable hard drives or wade through network attached storage or online repositories? The assembled data should be organized to make it possible to pair the e-mail with its metadata and to trace messages and attachments back to their origins, by, e.g., custodian, interval, location, business unit or other taxonomy.

De-duplication

You *finally* made it. The e-mails are assembled, accessible and intelligible. You *could* begin your review right away, but unless your client has money to burn, there's one more thing to do before diving in: *de-duplication*. If Jane e-mails Tom, with copies to Dick and Harry and Tom responds with an attachment by clicking "Reply to All," Tom's response is in *both* Tom's Sent Items folder *and* his Inbox, as well as in Jane, Dick and Harry's Inboxes. Save for variations in time of receipt, the messages are functionally identical. Absent de-duplication, Tom's response will be reviewed five times. Not only is this a costly waste of time, it creates the potential for conflicting decisions respecting relevance and privilege issues. The better course would be to use specialized software to remove all but a single instance of Tom's response from the review set.

De-duplication is typically achieved using metadata, cryptographic hashing or a mix of the two. It may be implemented *vertically*, within a single mailbox, folder or custodian, or *horizontally* (also called *globally*) across multiple mailboxes and multiple custodians. It's essential to *track and log all de-duplication* to permit re-population of duplicated items to be produced.

Be careful with horizontal de-duplication as discovery strategies change. An e-mail sent to dozens of recipients may have been de-duplicated from all but one custodian's mailbox in the expectation that the message would be reviewed and a production decision made on review of that single mailbox. If that custodian's e-mail is excluded from review, the de-duplicated e-mail is *never* reviewed, even if all other custodian's mailboxes are examined. Here, de-duplication could result in the failure to produce a discoverable document.

Review
At last, you and your staff are looking at the e-mail, ready to flag:

- Relevant, discoverable and non-privileged items
- Items responsive to particular requests
- Privileged communications (attorney-client, doctor-patient, work product)
- Confidential communications (trade secrets, proprietary data, personal and private)

If the review set is large, counsel may employ keyword or concept search tools to identify privileged or responsive items. Though a cost effective approach and useful when responding to objective requests (e.g., "produce all e-mail between Jane and Tom"), the value of automated search tools is considerably less clear when used to process subjective requests (e.g., "produce all e-mail expressing product safety concerns."). As previously noted, it's often impossible to frame effective keyword searches absent familiarity with the lingo used to describe the events and objects central to the case. Even then, the crucial communiqué, "*Say nothing*" or "*Dump her*" may be overlooked.

Properly used by those who understand their strengths and recognize their limitations, text and concept search tools are an important adjunct to—but an inadequate substitute for—the judgment of a diligent, well-trained reviewer. If you use automated search tools, be prepared to demonstrate to the court and opposing counsel how such tools compare with the efficacy of human reviewers and the basis for such comparison. Know that in the only litigation study comparing the two this author has found, keyword searching fared poorly, finding only about one-fifth of the relevant items identified by human reviewers. The safest approach is to work cooperatively with opposing counsel to select the keywords and frame the searches to be run against the review set. Mailboxes of key witnesses *always* merit careful message-by-message review for relevant intervals.

Re-population
Once it's been decided what to produce and withhold, the production set should be re-populated with all relevant and discoverable non-privileged messages and attachments that were de-duplicated for review. Alternatively, discuss the issue with opposing counsel and determine counsel's preference. Counsel for the requesting party may be satisfied with a log detailing other recipients, if it serves to simplify his review without causing undue confusion. Don't produce de-duplicated e-mail without establishing and memorializing that opposing counsel knows of the de-duplication and waives re-population.

Redaction
When a paper record held discoverable and privileged content, the time-honored solution was to conceal the privileged text with heavy black marking pen and produce a photocopy of the redacted original. Shortsighted efforts to carry that practice into the realm of electronic discovery proved embarrassing when it was discovered that simply obscuring text on the image layer of, e.g., a document file in Adobe Portable Document Format (.PDF) did nothing to conceal the same text in the file's data layer. Electronic evidence demands different methods to remove privileged and confidential information from discoverable items. Any method employed must eradicate redacted data from all source data including:

- MIME/UU/BASE64 encoded attachments

All e-mails are plain text file, yet we use them to transport photos, music, programs and all manner of binary files as "attachments". In truth, non-text data aren't "attached" at all. Thanks to an encoding scheme called Base64, binary data hitch a ride, embedded within the body of the e-mail, masquerading as text. If an attachment contains privileged content, know that producing the complete contents of the e-mail (that is, not just the message but the file's headers and footers, too) enables the privileged content to be decoded. Accordingly, Base64 encoded attachments must be redacted from MIME e-mails before their production

- Data layer of document image files (.tiff, .pdf)
- All copied and forwarded counterparts, including :bcc transmittals

Production

Decisions about the medium and format of production, as well as the handling of exceptional attachments, must be made before production of the e-mail can proceed:

- Medium for production: What container will be used for delivery?
  Electronic transmittal (e-mail attachment, FTP transfer)
  External hard drive
  Optical disks
  Online repository
  Hard copies
- Format of Production: In what form will the data files be delivered?
  Native *(.PST, .NSF)*
  Discrete files (.eml)
  Text files (.txt, .rtf)
  Load files (Concordance, Summation)
  Image files without data layer ("naked" .tiff)
  Image files with data layer (.pdf)
  Delimited files
- Protocol for production of exceptional files, for example:
  Databases that must be queried to deliver relevant information
  Spreadsheets and tables containing Z-axis data and embedded formulae
  Voice mail messages and associated metadata
  Data requiring proprietary software
  Data that could not be opened or decrypted; corrupted data
  Other data not lending itself to presentation in a letter size, paper-like format
  Scanned data with handwritten entries and marginalia missed by OCR
- What information will be included in privilege logs?
- What information will be furnished respecting de-duplicated items?

Documentation

Inevitably, something will be overlooked or lost, but sanctions need not follow every failure. Avoid sanctions by documenting diligence at every stage of the discovery effort, to be able to demonstrate why the decision that proved improvident was sound *at the time and place it was made*. Keep a record of where the client looked and what was found, how much time and money was expended and what was sidelined and why.

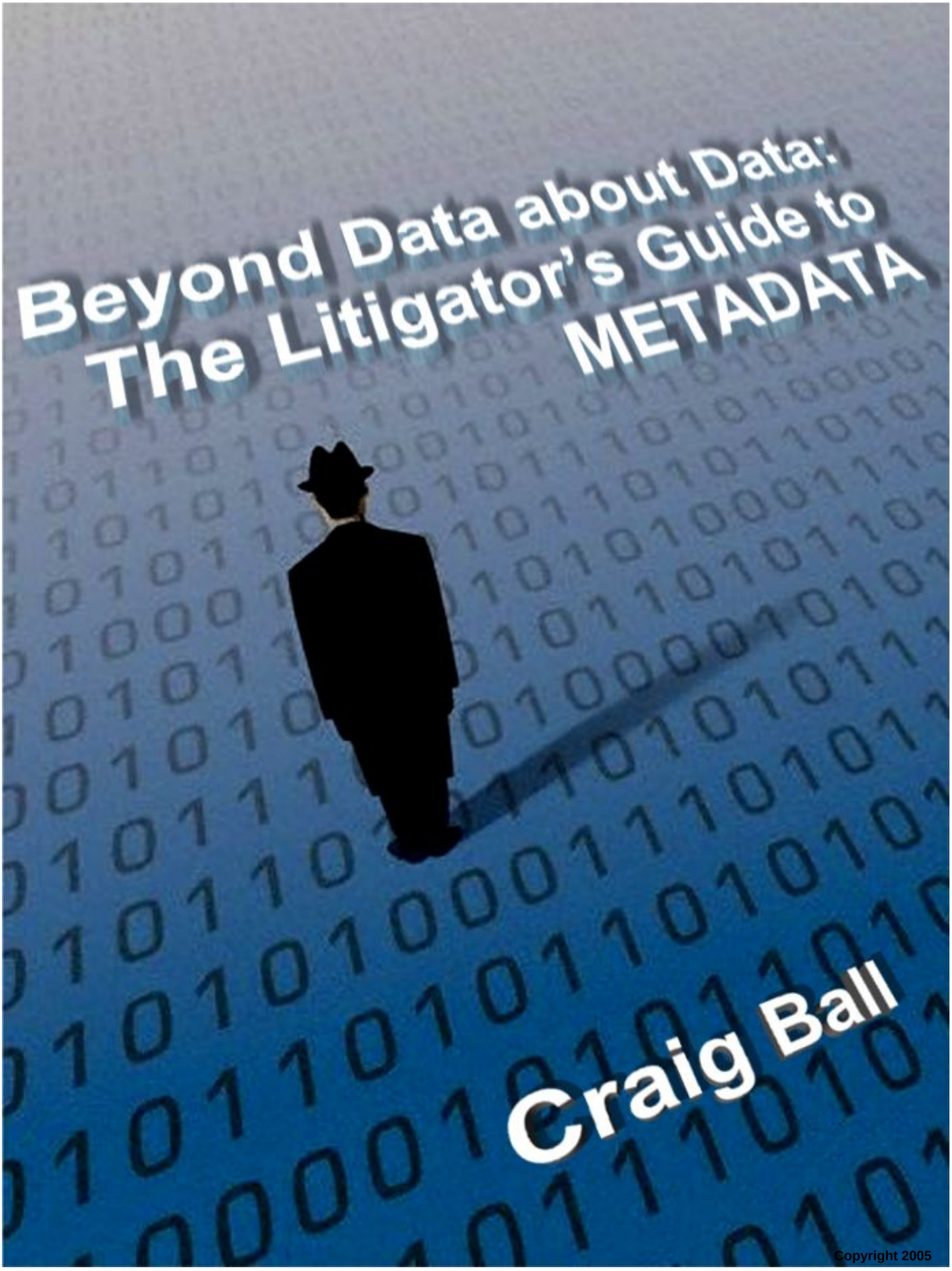
Conclusion

Responding to electronic discovery is a complex and challenging task--all the more so as we venture beyond the familiar confines of e-mail to the vast and varied sweep of all digital evidence. In the rush to embrace personal computing, businesses got ahead of sensible records management. Empowering individuals with networked PCs delegated responsibility for evidence preservation without adequate guidance or oversight. In short, businesses—and all of us—reaped the benefits of computers at the cost of discovery becoming harder and more expensive.

Some argue that we must make it easier and cheaper to litigate by deeming electronic evidence "out of bounds." Others respond that neither difficulty nor cost can justify curtailing full and fair access to evidence. One fact remains: ***most evidence is electronic***. If we want cases decided on the evidence, *discovery means electronic discovery*, and identifying, preserving, harvesting, managing and presenting digital evidence must be as vital and as accepted as cross-examination or trial by jury.

Electronic discovery is discovery in unfamiliar territory. When you figure out the steps and uncover the traps, it's like any other journey. Here's hoping this article helps you navigate the e-mail trail.

If you feel this outline omits a step or offers incorrect information, please share proposed additions or corrections with me at craig@ball.net. For further information about discovery of electronic mail, please read, "Meeting the Challenge: E-Mail in Civil Discovery" (http://www.ballpoint.org/emailpaper.pdf).


Beyond Data about Data:
The Litigator's Guide to
METADATA
Craig Ball
Copyright 2005

Beyond Data about Data: The Litigator's Guide to Metadata

By Craig Ball

In the old joke, a balloonist descends to ask directions, calling out, "Where am I?" A man on the ground yells back, "You're in a hot air balloon about a hundred feet above the ground." When the frustrated balloonist replies, "Thanks for nothing, Counselor," the man on the ground says, "Hey, how did you know I'm a lawyer?" "Simple," says the balloonist, "your answer was 100% accurate and totally useless."

It's time to get beyond defining metadata as "data about data."

Ask an electronic evidence expert, "What's metadata?" and there's a good chance you'll hear, "Metadata is data about data"--another answer that's 100% accurate, and totally useless!

Perhaps it's more helpful to say that, "metadata is evidence, typically stored electronically, that describes the characteristics, origins, usage and validity of other electronic evidence." There are all kinds of metadata found in various places in different forms. Some is supplied by the user and some created by the system. Some is crucial evidence and some just digital clutter. Understanding the difference--knowing what metadata exists and what evidentiary significance it holds--is an essential skill for attorneys dealing with electronic discovery.

It's time to move beyond defining metadata as "data about data" and establish labels and classifications that better describe and distinguish metadata in ways that allow lawyers to better assess relevance and accessibility.

Why Should You Care About Metadata?

In ***Williams v. Sprint/United Mgmt Co.*****,** 230 F.R.D. 640 (D. Kan. 2005), the Federal court ruled that "when a party is ordered to produce electronic documents as they are maintained in the ordinary course of business, the producing party should produce the electronic documents with their metadata intact, unless that party timely objects to production of metadata, the parties agree that the metadata should not be produced, or the producing party requests a protective order."

Our ability to advise a client about how to find, preserve and produce metadata, or to object to production, discuss or forge agreements about metadata, hinges upon how well we understand metadata.

Metadata is discoverable evidence that our clients are obliged to preserve and produce. Metadata sheds light on the origins, context, authenticity, reliability and distribution of electronic evidence, as well as providing clues to human behavior. It's the electronic equivalent of DNA, ballistics and fingerprint evidence, with a comparable power to exonerate and incriminate.

Finally, we need to care about metadata because it's some of the most fragile electronic evidence around. It's a short trip from mishandled metadata to spoliation sanctions.

Misunderstood Metadata

To the extent lawyers have heard of metadata at all, it's likely in the context of its potential to reveal confidential or privileged information hidden within electronic documents. The oft-cited culprit is Microsoft Word, and a cottage industry has grown up offering utilities to strip embedded information from Office applications. Because of the potential to embarrass lawyers—or worse—metadata has acquired an unsavory reputation amongst the bar. But metadata is much more than simply the ***application* metadata** that affords those who know how to find it to dredge up a document's secrets. That's just one species of metadata.

Application metadata is embedded in the file it describes and moves with the file when you copy it. However, not all metadata is embedded for the same reason that cards in a library card catalog aren't stored between the pages of the books. You have to know where the information resides to reach it. Contrast application metadata with ***system* metadata**, which is *not* embedded within the file it describes but stored externally and used by the computer's file system to track file locations and store demographics about each file's name, size, creation, modification and usage. Having both embedded application metadata and external system metadata is advantageous because, when metadata is stored both within and without a file, *discrepancies* between the metadata can expose data tampering.

Like all data, embedded application metadata is just a sequence of ones and zeroes and, in that respect, no less "accessible" than any other data. Accessibility is a measure of an application's ability to convert those ones and zeroes into intelligible information. A programmer configures applications to display selected information—but not necessarily *all* information—by default. Information not displayed by default may be accessible by reconfiguring the program's default settings (such as when a user sets a spreadsheet program to display formulae instead of calculated values). Viewing other embedded data may require drilling down through application menus, such as when a user explores file properties for a Microsoft Office document. These properties are at hand and comprehensible, but tend not to lend themselves to easy printing. None of this is surreptitious data—it's there if the user elects to review it. In fact, despite the common practice to call metadata "hidden," the only application metadata to warrant that description is the information the program employs internally to track, replicate or manage its actions. This data is, indeed, not readily accessible to the user via the program's menus and user-configurable settings, instead requiring specialized computer forensic tools and expertise to extract and interpret.

Every active file stored on a computer has at least one corresponding external block of system metadata—every one, no exceptions. Files may also have multiple associated metadata blocks as well as embedded metadata fields. You will never face the question of *whether* a file has metadata—all active files do—instead, the issues are *what kinds* of metadata exist, *where* it resides and whether it's potentially *relevant* such that it must be preserved and produced.

Every active file stored on a computer has at least one corresponding external block of system metadata—every one, *no exceptions*.

Is Metadata Just the Unprintable Information?

Some commentators mistakenly characterize metadata as the part of a file that "doesn't print out." While that flawed definition might have some merit if all files were Microsoft Word documents, it's far afield as applied it to other applications and formats. Consider a spreadsheet. The user enters formulae in various cells to produce calculated values. The

information keyed in by the user is certainly not metadata. It's the data. Yet, the formulae typically do not "print out." Or consider voicemail, which exists as both the recorded digitized sound of the message and the textual or encoded information describing the time and date of the call, mailbox identifier, etc. Though the sound doesn't print out, it's the *data*, whereas the dates and times may print but are the *metadata*.

The ability to print metadata varies within applications. It's often a simple task to display metadata onscreen—such as by a review of a document's "properties"—and not too difficult to print out. Though printable, it's metadata. As graphical user interfaces proliferate and applications become multimedia, computer data stray farther and farther from printable information. Sound and animation don't print at all; consequently, the animated PowerPoint, MPEG training video, even the Flash web page, fail "the data is what prints out" test. Moreover, data is also increasingly three-dimensional. Excel spreadsheets and database files, as well as hyperlinked documents, structure data not just across the X- and Y-axes of the printed page but "into" the page as well, layering "invisible" or unprintable linked values and sites, formulae, pivot tables and sounds on a Z-axis beneath onscreen information.

Metadata from the Ground Up

The best way to understand metadata from the ground up is to start with the fundamental building block of computerized information: the binary digit or "bit." Perhaps you already know that all computer data exists as a series of ones and zeroes, but have you stopped to consider what that really *means*? What is the consequence of constructing *everything* stored on digital devices from just two signals, on and off? Consider that written English conveys all information using fifty-two upper- and lowercase letters of the alphabet, ten numerical digits (0-9), some punctuation marks and a few formatting conventions, like spaces, line feeds, pages, etc. You can think of these collectively as a seventy or eighty signal "code." In turn, much of the same information could be communicated or stored in Morse code, where a three-signal code composed of dot, dash and pause serves as the entire "alphabet."

We've all seen movies where a tapping sound is heard and someone says, "Listen! It's Morse code!" Suddenly the tapping is a *message* because someone has furnished metadata ("It's Morse code!") *about* the data (tap, tap, pause, tap). Likewise, all those ones and zeroes on a computer only make sense when other ones and zeroes—the metadata—communicate a framework for parsing and interpreting the data stream.

All those ones and zeroes on a computer only make *sense* when other ones and zeroes—the metadata—communicate a framework for parsing and interpreting the data stream.

Sometimes metadata is elemental, like the contents of a computer's master file table detailing where the sequences of one and zeroes for particular files begin and end. This is metadata altogether invisible to a user without special tools called hex editors capable of peering through the walls of the Windows interface into the utilitarian plumbing of the operating system. Without file location metadata, each time a user sought to access a file or program, the operating system would be either wholly unable to find it or required to examine every stored byte. It'd be like looking for someone by knocking on every door in town!

At other times, metadata supports enhanced functionality not essential to the operation of the system. The metadata that tracks the date a file is created, last accessed or last modified might be expendable, but makes it much easier to manage important functions like system back ups.

Likewise, metadata indicating who has access privileges to particular files is unimportant to the user, but a network administrator would be hard-pressed to run a secure network without it.

As we move up the evolutionary ladder for system metadata, some metadata is recorded just in case it's needed to support a specialized task for the operating system or an application. Standard system metadata fields like "Camera Model" or "Copyright" may seem an utter backwater to a lawyer concerned with spreadsheets and word processed documents, but if the issue is the authenticity of a photograph or pirated music, these fields can make or break the case. ***It's all about relevance.***

Metadata is like the weather reports from distant cities which run in the daily paper. Though only occasionally relevant, you want the information available when you need it. Likewise, you should consider metadata preservation in every case involving digital evidence, even when you're uncertain you'll need it.

Much More Metadata

Modern operating systems record a ream of data detailing the creation, use and status of files as well as the use and configuration of associated applications. Windows users see a few these characteristics tracked in the Details view of a folder. By default, only a file's name, size, type and date modified are displayed; however, right click on the column titles in Windows XP and another thirty-four-odd metadata fields can be displayed, including creation date, author and comments. But even this broad swath of metadata is just *part* of the information about the file recorded by the operating system.

Within the Master File Table and index records used by Windows XP to track all files, still more attributes are encoded in hexadecimal notation. In fact, an ironic aspect of Windows is that the record used to track information about a file may be larger than the file itself! Stored within the hives of the System Registry—the "Big Brother" database that tracks attributes covering almost any aspect of the system—are thousands upon thousands of attribute values called "registry keys." Other records and logs track network activity and journal virtually every action. Within this maelstrom of metadata, some information is readily accessible and comprehensible while other data is so Byzantine and cryptic as to cause even highly skilled computer forensic examiners to scratch their heads.

An ironic aspect of Windows is that the record used to track information about a file may be larger than the file itself!

Relevance

How much of this metadata is relevant and discoverable? Would I be any kind of lawyer if I didn't answer, "It depends?" In truth, it *does* depend upon what issues the data bears upon. If the origin, use, distribution, destruction or integrity of electronic evidence is at issue, the "digital DNA" of metadata is essential evidence that needs to be preserved and produced.

Does this then mean that every computer system and data device in every case must be forensically imaged and analyzed by experts? Certainly not! *Once we understand what metadata exists and what it signifies, a continuum of reasonableness will inform our actions*. A competent police officer making a traffic stop collects relevant information, such as, e.g., the driver's name, address, vehicle license number, driver's license number and date, time and location of offense. We wouldn't expect the traffic cop to collect a bite mark impression, cheek swab or shoe print from the driver; but make the matter a murder investigation, and the

investigator is far more interested in a DNA sample than a driver's license number. The crucial factor isn't burden. It's *relevance*, assessed by those with the knowledge and experience to recognize and gauge relevance

There are easily accessible, frequently valuable metadata that, like the information collected by the traffic cop, we should expect to preserve routinely. Examples of these might be originating path and filename, and origination MAC (Modified-Accessed-Created) dates for each file. For e-mail, the obligatory metadata might include complete header data, not just the To/From/Date/Subject items culled from the header by the e-mail program. Proper evidence handling entails a sound chain-of-custody, even in civil matters. Metadata functions as the tag attached to evidence in a police property room. The preservation of a file's external system metadata, in particular its name, system origins and dates of creation, last access and modification, is as fundamental to meeting chain-of-custody obligations as preserving certified mail receipts or tracking the elements of the business records hearsay exception, perhaps more so because metadata is so fluid. Fail to preserve metadata at the earliest opportunity and you may never be able to replicate what was lost.

Fail to preserve metadata at the earliest opportunity and you may never be able to replicate what was lost.

So where do we draw the relevance line? Begin by recognizing that the advent of electronic evidence hasn't changed the fundamental dynamics of discovery: Litigants are entitled to discover relevant, non-privileged information, and determination of relevance hinges on the issues before the court. Relevance assessments aren't static, but change as evidence emerges and issues arise. Metadata irrelevant at the start of a case may be decisive when allegations of data tampering or spoliation enter the fray. A producing party must periodically re-assess the adequacy of preservation and production and act to meet changed circumstances.

Periodically re-assess the adequacy of preservation and production and act to meet changed circumstances.

The Path to Production of Metadata

The balance of this paper discusses steps typically taken in shepherding a metadata production effort. Don't look for a recitation of established best practices—those rules are very much in flux—or expect a comprehensive checklist of "Do's" and "Don'ts." Instead, the goal is to introduce challenges unique to discovery of metadata and explore the "Why" behind them. These steps include:

- Gauge spoliation risks before you begin: *Don't peek!*
- Identify potential forms of metadata
- Assess relevance
- Consider Authentication and Admissibility
- Evaluate Need and Methods for Preservation
- Collect Metadata
- Plan for Privilege and Production Review
- Resolve Production Issues

Gauge spoliation risks before you begin: *Don't peek!*

German scientist Werner Heisenberg thrilled physicists and philosophy majors alike when he posited that the very act of observing alters the reality observed. Heisenberg's Uncertainty Principal speaks to the world of subatomic particles, but it aptly describes a daunting challenge

to lawyers dealing with metadata: *When you open any document in Windows without first employing specialized hardware or software, metadata changes and prior metadata values are lost.* Altered metadata implicates not only claims of spoliation, but also severely hampers the ability to filter data chronologically. How, then, can a lawyer evaluate documents for production without *reading* them?

One solution is to preserve original metadata values *before* examination. This can be achieved using software that archives the source metadata to a table or spreadsheet. Then, if an examination results in a corruption of metadata, the original values can be ascertained. Another approach is to conduct the examination using only a forensically qualified duplicate of the data. The techniques used to create a forensically qualified image guard against alteration of the original evidence, which remains available and unaltered should original metadata values be needed. A third approach is to use write blocking hardware or software to intercept all changes to the evidence media. Finally--and most commonly—an electronic discovery vendor can harvest and preserve all metadata on read-only media (e.g., a CD-R or DVD-R) or in a hosted environment, permitting examination without metadata corruption.

Identify potential forms of metadata

To preserve metadata and assess its relevance, you have to know it exists. So, for each category of data subject to discovery, assemble a list of associated metadata. You'll likely need to work with an expert the first time or two, but once you have a current and complete list, it will serve you in future matters. You'll want to know not only what the metadata field contains, but also its location and its significance.

There are at least eighty accessible application and system metadata fields tracked for a Microsoft Word

The numbers may surprise you. There are at least ***eighty*** easily accessible application and system metadata fields tracked for each Microsoft Word, PowerPoint and Excel document, *excluding* tracked changes, comments and Registry data (though a few are redundant and the majority of them rarely used). For unfamiliar or proprietary applications and environments, enlist help identifying metadata from the client's IT personnel. Most importantly, *seek your opponent's input, too*. Your job is simpler when the other side is conversant in metadata and can expressly identify fields of interest. The parties may not always agree, but at least you'll know what's in dispute.

Assess relevance

Are you going to preserve and produce dozens and dozens of metadata values for every document and e-mail in the case? Probably not, although you may find it easier to preserve all than selectively cull out just those values you deem relevant.

Relevance is always subjective and is as fluid as the issues in the case. For example, two seemingly innocuous metadata fields common to Adobe Portable Document Format (PDF) files are "PDF Producer" and "PDF Version." These are listed as "Document Properties" under the "File" menu in any copy of Adobe Acrobat. Because various programs can link to Acrobat to create PDF files, the PDF Producer field stores information concerning the source application, while the PDF Version field tracks what release of Acrobat software was used to create the PDF document. These metadata values may seem irrelevant, but consider how that perception changes if the dispute turns on a five-year-old PDF contract claimed to have been recently forged. If the metadata reveals the PDF was created using a scanner introduced to market last

year and the latest release of Acrobat, that metadata supports a claim of recent fabrication. In turn, if the metadata reflects use of a very old scanner and an early release of Acrobat, the evidence bolsters the claim that the document was scanned years ago. Neither is conclusive on the issue, but both are relevant evidence needing to be preserved and produced.

Consider Authentication and Admissibility

Absent indicia of authenticity like signature, handwriting and physical watermarks, how do we establish that electronic evidence is genuine or tie an individual to the creation of an electronic document? Computers may be shared or unsecured and passwords lost or stolen. Software permits alteration of documents sans the telltale signs that expose paper forgeries. Once, we relied upon dates in correspondence to establish temporal relevance, but now documents may generate a new date each time they are opened, inserted by a word processor macro as a "convenience" to the user. Without metadata, it's often impossible to establish authenticity or establish relevance. Where the origins and authenticity of evidence are in issue, preservation of original date and system user metadata is essential. When deciding what metadata to preserve or request, consider, *inter alia,* network access logs and journaling, evidence of other simultaneous user activity and version control data.

Without metadata, it's often impossible to establish authenticity or establish relevance.

In framing a preservation strategy, balance the burden of preservation against the likelihood of a future need for the metadata, but remember, if you act to preserve metadata for documents supporting your case, it's hard to defend a failure to preserve metadata for items bolstering the opposition's case. Failing to preserve metadata could deprive you of the ability to challenge the relevance or authenticity of material you produce.

Evaluate Need and Methods for Preservation

Not every item of metadata is important in every case, so what factors should drive preservation? The case law, rulings of the presiding judge and regulatory obligations are paramount concerns, along with obvious issues of authenticity and relevance; but another aspect to consider is the *stability* of particular metadata. As discussed, some metadata fields, like Last Access Date, change the instant the file is opened for review, copied or even checked for viruses. If you don't preserve these fragile fields, you lose the ability to go back to the source data and extract metadata when needed. Where a preservation duty has attached, by, e.g., issuance of a preservation order or operation of law, the loss of metadata may constitute spoliation subject to sanction.

How, then, do you avoid spoliation occasioned by review and collection? What methods will preserve the integrity and intelligibility of metadata? Often, collection activities required by litigation hold notices themselves serve to corrupt metadata. When, for example, a custodian or reviewer copies responsive files to new media, prints documents or forwards e-mail, metadata is altered or lost. Consequently, metadata preservation must be addressed *before* a preservation protocol is implemented. Be certain to document what was done and why. Advising your opponents of the proposed protocol in sufficient time to allow them to make objection, seek court intervention or propose an alternate protocol helps to protect against belated claims of spoliation.

Often, collection activities required by litigation hold notices themselves serve to corrupt metadata.

Collect Metadata

Because metadata is stored both within and without files, simply duplicating a file without capturing its system metadata may be insufficient. However, not all metadata preservation efforts demand complex and costly solutions. It's possible to tailor the method to the case in a proportional way. For example, if only a handful of files are implicated, the simplest and most expedient way to preserve and produce metadata might be to simply record the relevant metadata values by hand. As the number of files increase, you might create a file listing or spreadsheet detailing the original metadata. Even just archiving the files ("zipping") may be a sufficient method to preserve associated metadata. In other cases, you'll need to employ forensic imaging or use vendors specializing in electronic discovery.

Whatever the method chosen, be careful to preserve the association between the data and its metadata. For example, if the data is the audio component of a voice mail message, it may be of little use unless correlated with the metadata detailing the date and time of the call and the identity of the voice mailbox user.

When copying file metadata, know the limitations of the environment and medium in which you're working. I learned this lesson the hard way several years ago while experimenting with recordable CDs as a means to harvest files and their metadata. Each time I tried to store a file and its MAC dates (modified/accessed/created) on a CD, I found that the three *different* MAC dates derived from the hard drive would always emerge as three *identical* MAC dates when read from the CD! I learned that CD-Rs aren't formatted in the same manner as magnetic media. Whereas the operating system formats a hard drive to store three distinct dates, CD-R media stores just one. In a sense, a CD hasn't the "slots" to store all three dates. When the CD's contents are copied back to magnetic media, the operating system re-populates the slots for the three dates with the single date found on the optical media. Thus, *using a CD in this manner serves to both corrupt and misrepresent the metadata.* Similarly, different operating systems maintain different metadata fields, so, e.g., moving data from a Windows XP environment to a Windows 98 environment results in truncation or loss of metadata.

Plan for Privilege and Production Review

The notion of reviewing metadata for privileged communications may seem odd unless you consider that application metadata potentially contains deleted content and commentary. When the time comes to review metadata for production and privilege, the risks of spoliation faced in harvest may re-appear during review. Consider:

- How will you efficiently access metadata?
- Will the metadata exist in a form you can interpret?
- Will your examination alter the metadata?
- How will you flag particular metadata for production?
- How can you redact privileged or confidential metadata?

If a vendor or in-house discovery team has extracted the metadata to a slip-sheet in an image format like TIFF or PDF, review is as simple as reading the data. However, if review will take place in native format, some metadata fields may be inaccessible, encoded or easily corrupted. If the review set is hosted online, be certain you understand which metadata fields are accessible and intelligible via the review tool and which are not.

In ***Williams v. Sprint/United Mgmt Co.*****,** 230 F.R.D. 640 (D. Kan. 2005), concerns about privileged metadata prompted the defendant to strip out *all* metadata from the native-format spreadsheet files it produced in discovery. The court responded by ordering production of all metadata as maintained in the ordinary course of business, save only privileged and expressly protected metadata, but offered no guidance as to how one might *effect* selective redaction of application metadata.

The court was right to recognize that privileged information need not be produced, wisely distinguishing between surgical redaction and blanket excision. One is redaction following examination of content and a reasoned judgment that particular matters are privileged. The other excises data in an overbroad and haphazard fashion, grounded only on an often-unwarranted concern that the data pared away ***might*** contain privileged information. The baby goes out with the bathwater. Moreover, blanket redaction based on privilege concerns doesn't relieve a party of the obligation to log and disclose such redaction. The defendant in *Williams* not only failed to examine or log items redacted, it left it to the plaintiff to figure out that something was missing.

None of this obliges a party to use applications that generate substantial metadata or refrain from steps taken to minimize metadata in the creation of electronically stored information. The underlying principle is that the requesting party is entitled to the metadata benefits available to the producing party. That is, the producing party may not vandalize or hobble electronic evidence for production without adhering to the same rules attendant to redaction of privileged and confidential information from paper documents.

The requesting party is entitled to the metadata benefits available to the producing party.

Resolve Production Issues

Like other forms of electronic evidence, metadata may be produced in its native file format, as a database load file, exported to a compatible delimited dataset, in an image format, hosted in an online database or even as a paper printout. However, metadata presents more daunting production challenges than other electronic evidence. One hurdle is that metadata is often unintelligible outside its native environment without processing and labeling. How can you tell if an encoded value describes the date of creation, modification or last access without both decoding the value *and* preserving its significance with labels? Another issue is that metadata isn't always textual. It may consist of no more than a flag in an index entry—just a one or zero—wholly without meaning unless you know what it denotes. A third challenge producing metadata lies in finding ways to preserve the relationship between metadata and the data it describes and, when obliged to do so, present both the data and metadata so as to be electronically searchable.

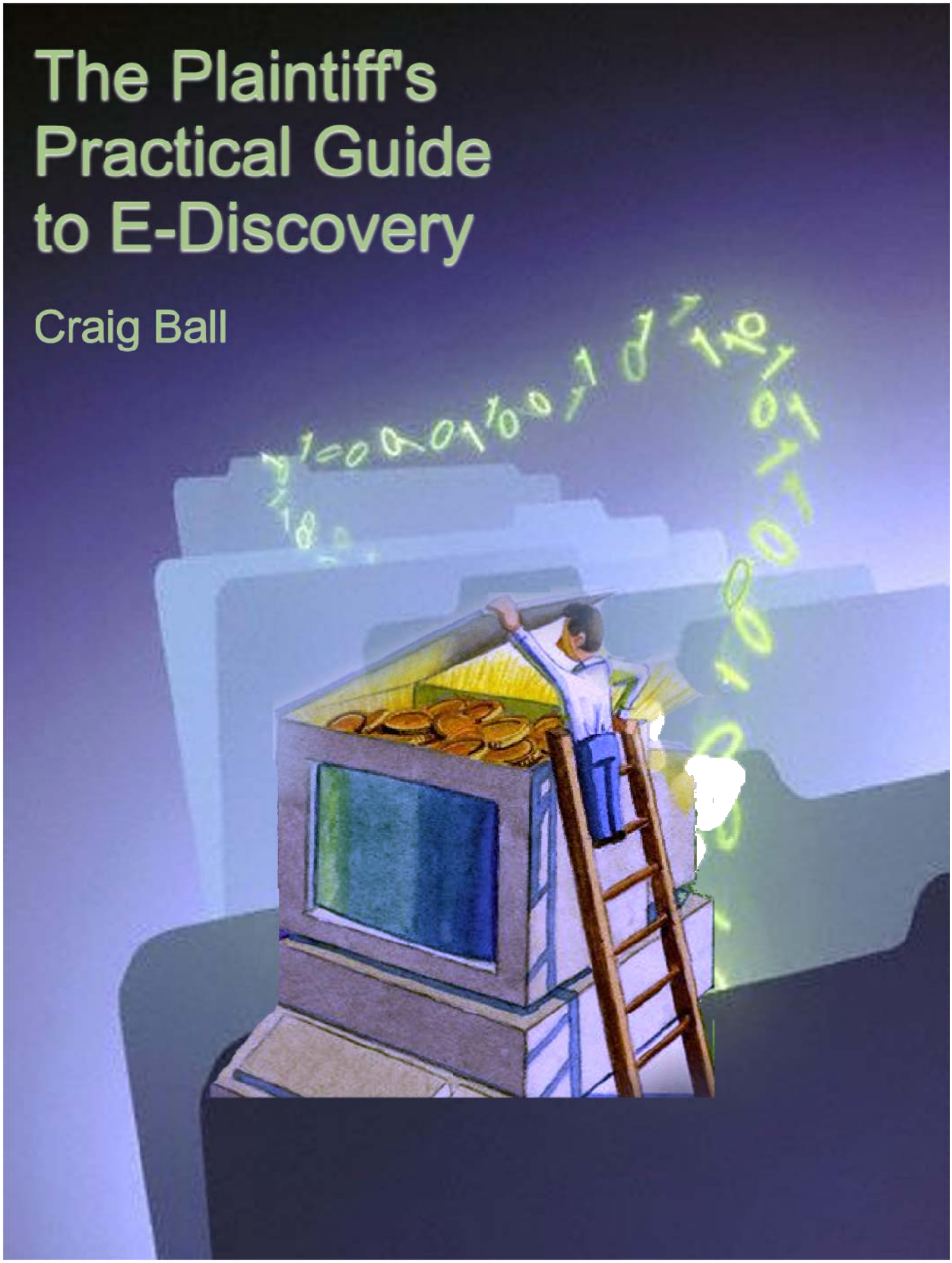
When files are separated from their metadata, we lose much of the ability to sort, manage and authenticate them. Returning to the voice mail example, unless the sound component of the message (e.g., the WAV file) is paired with its metadata, a reviewer must listen to the message in real time, hoping to identify the voice and deduce the date of the call from the message. It's a Herculean task without metadata, but a task made much simpler had the producing party, *e.g.,* dropped the WAV file into an Adobe PDF file as an embedded sound file then inserted the metadata in the image layer. Now, a reviewer can both listen to the message and search and sort by the metadata.

Sometimes, simply producing a table or spreadsheet detailing originating metadata values will suffice. On other occasions, only native production or forensically qualified imaging will suffice to carry forward relevant metadata. Determining the method of metadata production best suited to the case demands planning, guidance from experts and cooperation with the other side.

Beyond Data about Data

The evidentiary value of metadata will only increase as the world moves inexorably to digitization and electronic communications. Already, some 95% of all information is born electronically, the data bound to and defined by its metadata as we are by our DNA. Metadata grows ever more vital in discovery; dictating that we move beyond unhelpful definitions like "data about data," toward an effective vocabulary to describe metadata in its many forms, and toward sensible standards governing its preservation, relevance and production. We also must foster improved electronic records management practices and systems, all the while encouraging those who design and build operating systems and software to offer products better suited, not just to litigation—at best, a tertiary concern for them—but to the goals of data verifiability, portability, integrity, stability and eradication shared by businesses and litigators alike.

Already, some 95% of all information is born electronically, the data bound to and defined by its metadata as we are by our DNA.


The Plaintiff's
Practical Guide
to E-Discovery
Craig Ball

The Plaintiff's Practical Guide to E-Discovery, Part I
By Craig Ball

It's challenging. It's expensive. But it's the latest--and maybe the greatest—tool at the disposal of plaintiffs' counsel willing to learn the ropes and aggressively assert their clients' rights. It's electronic data discovery (EDD).

The world has changed, and the traditional approach to discovery--casting the widest net and poring over bankers' boxes of documents--is history. Most of the evidence in your case isn't on paper and never will be. The volume of discoverable electronic information is exploding, growing at a rate that makes paper review unimaginable. Although data volume depends on the case--with a car wreck case generating less data than a pharmaceutical products liability class action--even the lowest-end personal computer stores *millions* of pages of information. Hence, volume concerns impact every case.

Getting discoverable data and making sense of it requires plaintiffs' counsel to gain an understanding of where digital evidence lives, the forms it takes and how it's preserved, altered and destroyed. It also entails learning as much as you can about the architecture and operations of the defendant's systems and networks, and how the key players—such as those whose conduct forms the basis of the contemplated claim or suit—interact with a fast-growing universe of digital devices and data repositories.

Challenges Unique to EDD

Plaintiffs pursuing electronic discovery face entrenched ignorance and outright obstinacy. Evidence on paper was never destroyed and suppressed with the ease and frequency seen with electronic evidence. Folks who wouldn't dream of shredding paper files hit the "delete" button without a moment's hesitation.

One EDD challenge is that foxes guard the henhouse. The common practice in large organizations is to issue a "litigation hold" notice to employees instructing them to retain and segregate relevant electronic evidence. This is the starter pistol for the race by those with something to hide to delete it as quickly as possible. Though motivated to hide workplace pornography or e-Bay shopping, the clumsy delete-o-thons that follow sweep away relevant and discoverable electronic evidence, too.

Another vexing problem is defense counsels' cluelessness about electronic discovery. Your opponent may be a courtroom whiz, but if he or she has a tenuous grasp of computer systems or doesn't understand his or her client's devices and data, defense counsel can't give sound guidance about preserving digital evidence or pose the right questions to knowledgeable IT personnel. It's astounding how often attorneys *brag* about how *little* they know about computers! Whether due to poor judgment or a client focused on shortsighted cost savings, computer-illiterate counsel don't always seek out the expert help they need. Is this malpractice? Probably, but the lack of expertise is so pervasive, it may take years and several more high-profile e-discovery debacles before lawyers fully appreciate how much their lack of knowledge hurts their clients.

This leaves you on the horns of a dilemma. Do you lay low while evidence is overlooked or lost, banking on spoliation sanctions to protect your client, or do you seek to educate your opponent and improve the odds that you'll get the electronic evidence? The latter is clearly the better

approach, and your ability to succeed in securing sanctions for discovery abuse is enhanced when you can show how hard you tried to help the defendant "get it."

Failure: The Dirty Little Secret of E-Discovery

The reality of electronic discovery is that the responding party will fail, partly due to the near-absence of prudent electronic records management. Once upon a time, a discovery request sent a file clerk scurrying to a file room set aside for orderly information storage. There, the clerk sought a labeled drawer or box and the labeled folders within. He didn't search *every* drawer, box or folder, but went only to the place where the company kept items responsive to the request. From cradle to grave, paper had its *place,* tracked by standardized, compulsory practices. Correspondence was dated and its contents or relevance described below the date. Documents and files reposed, sorted and aggregated, within a clear and consistent information framework. Responding parties could affirm that discovery was complete on the strength that they'd looked in the places where responsive items customarily resided. This was the *power of place,* and it freed respondents from the obligation to look elsewhere.

Today, evidence splays willy-nilly across servers, back up tapes, workstation hard drives, laptops, home systems, personal digital assistants, online storage and thumb drives. Collection efforts overlook many e-evidence venues. What records management exists takes a thousand quirky forms as individual employees adopt their own peculiar folder structures and retention practices. Those fearing the consequences of their digital evidence are empowered by the delete key to attempt its destruction. The power of place is gone, leaving high cost and failure in its stead.

Recognizing that your opponents will fail to preserve and produce all discoverable evidence and may even seek to destroy it, what can you do to forestall that failure or ameliorate the harm to your client?

Elements of Successful Electronic Discovery

A requesting party's successful e-discovery effort entails most or all of the following elements:

- Identifying relevant systems and data
- Compelling preservation of potentially relevant digital evidence
- Seeking production of digital evidence in manageable formats
- Honing preservation and production through "meet and confer" sessions
- Memorializing preservation and production duties as court orders
- Assimilating, analyzing and using the electronic data produced
- Identifying discovery abuses and seeking the Court's intervention

A successful effort also requires us to rethink our traditional approach to discovery. Pursuing paper discovery, we wove expansive nets of the finest mesh to seine anything and everything. Electronic discovery demands the narrow aim of a harpoon. No opponent can satisfy nor court enforce a demand for "any and all electronic communications and records." That's a fishing expedition. The watchword for e-discovery is, ***be careful what you wish for lest you get it, have to manage it and pay for its production.***

Relevant Systems and Data

Successful cost-effective electronic discovery is far easier when you know what your case is about and what you need to prove it. Trials still come down to a few key people and documents. You don't win by forcing production of the most data. You win by securing the *right* data.

Start by learning all you can about the defendant's systems. Examine what you have for clues about what else is out there (e.g., by checking path statements on documents or e-mail circulation lists). Check the defendant's website, press releases and public filings for descriptions of IT assets. Talk to former employees about system architecture and data retention practices. Take 30(b)(6) depositions of designated IT personnel—but be sure you also get past the managers and talk to those in the trenches who know about that box of old back up tapes in the storeroom. Get copies of document retention policies and ask witnesses about compliance. Demand that the defendant furnish network topology diagrams and asset tracking inventories for computer hardware, but take these with a grain of salt, as they're often outdated or fail to reflect real-world configurations. Finally, remember that you're not alone. Network with other lawyers who've sought e-discovery against the defendant to learn the digital landscape and hone in on helpful witness and experts.

Compelling Preservation

Effective e-discovery begins before suit, with your preservation letter to the defendant. The goal of the preservation letter is to remind opponents to preserve evidence and reduce the chance that evidence will disappear. But, the preservation letter also serves as the linchpin of a subsequent claim for spoliation, helping to establish bad faith and conscious disregard of the duty to preserve relevant evidence. It is today's clarion call that underpins tomorrow's, "I told you so." The more effectively you give notice and convey what must be retained—including methodologies for preservation and consequences of failure—the greater the likelihood your opponent will be punished for destruction of evidence.

Wouldn't it be sufficient to remind your opponent that the laws and rules prohibiting destruction of evidence apply to electronically stored information in the same manner that they apply to other evidence and that they must take every reasonable step to preserve this information until the final resolution of the case? Perhaps in a decade or so it will be enough; but today, we face an explosion of electronic evidence untamed by sound records management and marshaled by litigators and in-house counsel who don't understand information systems. *The reality of electronic discovery is that it starts off as the responsibility of those who don't understand the technology and ends up as the responsibility of those who don't understand the law*. A well-drafted preservation letter helps bridge this knowledge gap.

> **For a comprehensive discussion of the elements of an effective preservation letter, *see* "The Perfect Preservation Letter," at page 40, *infra*.**

The Sharper Tools in the Shed: Meet and Confer

Proper preservation of electronic data can be costly and complicated. If your opponent is exceptionally well versed in electronic discovery, he or she should seek to meet and confer with you to arrive at an agreed preservation plan. The defendant benefits by reining in the high cost of preservation while minimizing the threat of sanctions. What's in it for your side?

You should welcome a meet and confer opportunity if the defendant furnishes complete and accurate information about the nature and location of their electronic evidence (including back

up procedures and retention schemes) and the identity of, and devices used by, key players. Absent such disclosure, it's very risky to agree that the defendant may, e.g., rotate server back up tapes, replace systems or delete older e-mail. Don't concede that the defendant can destroy any information item in any form until the defendant demonstrates that the information lost is not likely to be relevant to any issue in the case or lead to the discovery of admissible evidence. Such representation need to be unambiguously confirmed in writing and supported by a showing that a person with the requisite knowledge and skill actually knows the contents of the media to be overwritten, discarded or destroyed. For voluminous or costly-to-restore data sets, sampling contents is prudent if there is any question as to what information resides on media slated for destruction.

Your agreement may also serve to blunt the defendant's ability to seek cost sharing for the expense of electronic production and open doors to broader or easier access. To gain your client's consent to a narrower preservation obligation, the defendant may be willing to undertake and bear the cost of, e.g., producing digital evidence in your preferred format or computer forensic analysis of key player's office and home computers.

Your opponent isn't the only one who needs to prepare for a meet and confer session. Though it may be difficult early in the case, you must be able to put forward what you want and why you're entitled to it. You don't get the e-mail just because it's there. Be prepared to articulate the relevance and the appropriate interval for the digital evidence you seek.

Finally, it's essential to reduce all preservation and production agreements to writing. Where possible, submit agreed orders to the court, remembering that judges are much more willing to impose sanctions for violations of their orders than for breach of an agreement between counsel.

Part II

Having decided what you need and advised the defendant to preserve it, it's time to seek production. In Part II, we look at the pros and cons of production formats and explore common e-mail systems, concluding with tips for getting the most out of your e-discovery efforts and budget.

The Plaintiff's Practical Guide to E-Discovery, Part II

By Craig Ball

It's challenging. It's expensive. But, it's the single greatest litigation advantage for plaintiffs' counsel willing to learn the ropes and aggressively assert their clients' rights. It's electronic data discovery (EDD). In Part I, we addressed challenges unique to EDD, elements of a successful e-discovery effort and steps to compel preservation of digital evidence. In Part II, we look at the pros and cons of production formats and explore common e-mail systems, concluding with tips for getting the most out of your e-discovery efforts and budget.

Formats

One of the biggest mistakes a requesting party makes is requesting or accepting production of electronic evidence in a format ill suited to their needs. Electronic evidence can be produced in five principal formats:

- Paper print outs,
- Paper-like image of the data in TIFF or PDF,
- Export of the native data to a common electronic format, e.g., Access database or load file,
- Native data, and
- Hosted data.

Sometimes you'll have the chance to designate the format for production of electronic evidence. In Federal court, litigants will have this right as of December 1, 2006. Your choice of format should factor in both the type of data being produced as well as the way in which you and your staff are capable of managing the evidence. In a perfect world, you'd want everything in its native electronic format, but in the real world, you may lack the systems and software to deal with and preserve the evidentiary integrity of all native formats. Plus, redaction issues and other fears (some rational, some not) mean that your opponents may be unwilling to offer native data.

Paper Production: I've heard defense counsel deride as "dinosaurs" the plaintiffs' lawyers who ask that electronic evidence be "blown back" to paper. True, converting searchable electronic data to costly and cumbersome paper is usually a step backwards, but not always. Paper still has its place. For example, in a case where the entire production consists of a few hundred e-mails and several thousand e-documents, searching and volume aren't a problem and paper remains about as good a medium as any. But once the volume or complexity increases much beyond that which you can easily manage by memory, you're better off insisting on production in electronically searchable formats.

Image Production: Here, production consists of files that are digital "pictures" of the documents, e-mails and other electronic records. These images are typically furnished in accessible file formats like Adobe's Portable Document Format (PDF) or in one of the Tagged Image File Formats (TIFF). It's not the evidence. It's a facsimile of certain aspects of the evidence. But as long as the information lends itself to a printed format and is electronically searchable, image formats work reasonably well, though they can be the most costly approach. But for embedded information (such as the formulae in spreadsheets) or when the evidence moves beyond the confines of printable information (as voice mail, databases or video), image production breaks down. Additionally, the requesting party must insure that the images are accompanied by electronically searchable data layers and relevant metadata. Beware the defendant who tries to pawn off "naked" TIFF images (devoid of searchable information and metadata) as responsive.

> **For further discussion of the pros and cons of image production, *see* "E-Discovery Report Card: How Well Do TIFF Images Perform," *infra*.**

Exported Formats: Some electronic evidence easily adapts to any of several production formats. For example, e-mail may be readable in multiple e-mail client programs (Outlook, Eudora, Lotus Notes) or in generic e-mail formats (e.g., .EML, MSG). The contents of simple databases can be exported to generic formats (e.g., comma or tab delimited output) and then imported into compatible applications (e.g., Excel spreadsheets to Access databases). When discoverable data works as exported formats, you may prefer to obtain exported, delimited data in order to review the data in the compatible application of your choice. The key is to be sure you don't corrupt or lose any important data--or surrender needed abilities to manipulate it--in

the export/import process. Consider exported data for production of e-mail and simple databases (like contact lists). However, as data structures grow more complex, it's much harder--or impossible--to present exported data in a way that accurately reflects the native environment, forcing you to seek native production.

Native Production: In native production, the defendant produces duplicates of the actual data files containing responsive information. That is, the producing party produces the actual evidence, not an extraction or facsimile of it. The benefit is that, if you have copies of software programs used to create and manipulate the data or a compatible viewer utility, you have the ability to see the evidence more-or-less exactly as it appears to the producing party. A major benefit is that, apart from issues of privilege and redaction, native data requires no expensive processing for production. Thus, it may be a far less costly form of production. Sounds great, but native production is not without its drawbacks. The native applications required to view the data in its native format may be prohibitively expensive or difficult to operate without extensive training (e.g., an Oracle or SAP database). Additionally, reviewers must take care not to change the native data while viewing it. The rule of thumb is that native production is preferable, but only when the requesting party has the experience, expertise and resources to manage native data.

Defendants often resist production of native data because of the difficulty they face in redacting privileged information. An Outlook post office (.PST) file can hold both discoverable e-mail and privileged attorney-client communications, but as it's a unified and complex database file, it's very difficult to produce the former without also producing the latter. Another risk to defendants is that native data (like Word .DOC files) can contain embedded, revealing metadata. Native production is also less amenable to Bates numbering

Hosted Data: This is production without possession, in that the information produced resides on a controlled-access website. The requesting party reviews the data through an online application (similar to a web browser) capable of displaying information from a variety of electronic formats. More commonly, hosted data and online review tools are used by counsel for the producing party to search the production set for privileged and responsive items rather than as a means to afford access to the requesting party. The items identified as responsive are then burned to CD or DVD and produced, often in image formats as discussed above.

Rules of Thumb for Formats

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) of the Federal Rules of Civil Procedure contemplates that producing parties not remove or significantly degrade the searchability of electronically stored information, 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. Read on for further discussion of e-mail systems and formats.

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. Some companies take the head-in-the-sand posture that none of their employees use IM and, pretending IM doesn't exist, take no steps to preserve IM transactions. But 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." Because individual users control whether or note to log IM exchanges, a responding party can't make 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 backup 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 backup 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 backup 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.

Discovery of E-Mail

Futurist Arthur C. Clarke said, "Any sufficiently advanced technology is indistinguishable from magic." E-mail is one of those magical technologies most of use every day without really understanding how it works. Though a discovery request for "the e-mail" may secure adequate production, understanding e-mail systems helps you to see if something is missing and gauge whether the methods used to assemble responsive e-mail were calculated to locate *all* responsive messages. More to the point, e-discovery is increasingly a two-way street, and plaintiff's counsel needs to prepare for electronic discovery of client e-mail. Can you instruct your client where to find and how to produce e-mail?

Get the e-mail! It's the watchword in discovery today. Some label the press for production of electronic mail a feeding frenzy, but it's really just an inevitable recognition of how central to our lives e-mail has become. More than fifty billion e-mails traverse the Internet daily, far more than telephone and postal traffic combined, and the average businessperson sends and receives between 50 and 150 e-mails every business day. At that rate, a company employing 100,000 people could find itself storing *3 billion* e-mails annually. E-mail contributes 500 times greater volume to the Internet than web page content. Trial lawyers go after e-mail because it accounts for the majority of business communications, and e-mail users tend to let their guard down and share things online that they'd never dare put in a memo.

Did You Say *Billion*?

Aggregate volume is only part of the challenge for discovery and production of e-mail. Unlike paper records, e-mail tends to be natively stored in massive data blobs. The single file containing my Outlook e-mail is over three gigabytes in size and holds some 35,000 messages, many with multiple attachments, covering virtually every aspect of my life, and many other people's lives, too. In thousands of those e-mails, the subject line bears only a passing connection to the contents as "Reply to" threads stray further and further from the original topic. E-mails meander through disparate topics or, by absent-minded clicks of the "Forward" button, lodge in my inbox dragging with them, like toilet paper on a wet shoe, the unsolicited detritus of other people's business. To respond to a discovery request for e-mail on a particular topic, I'd either need to skim/read all 35,000+ messages or I'd have to have a very high degree of confidence that a keyword search would flush out all responsive material. If the request for production implicated material I no longer kept on my current computer, I'd be forced to root around through a motley array of old systems, obsolete disks, outgrown hard drives, ancient backup tapes (for which I have no tape reader) and unlabeled CDs, uncertain whether I've lost the information or just overlooked it somewhere along the way. The situation isn't much different in corporate America.

A Snippet about Protocols

Computer network specialists are always talking about this "protocol" and that "protocol." Don't let the geek-speak get in the way. An application protocol is a bit of computer code that facilitates communication between applications, i.e., your e-mail client, and a network like the Internet. When you send a snail mail letter, the U.S. Postal Service's "protocol" dictates that you place the contents of your message in an envelope of certain dimensions, seal it, add a defined complement of address information and affix postage to the upper right hand corner of the envelope adjacent to the addressee information. Only then can you transmit the letter through the Postal Service's network of post offices, delivery vehicles and postal carriers. Omit the address, the envelope or the postage--or just fail to drop it in the mail--and Grandma gets no Hallmark this year! Likewise, computer networks rely upon protocols to facilitate the

transmission of information. You invoke a protocol—Hyper Text Transfer Protocol—every time you type http:// at the start of a web page address.

E-Mail Systems: POP, IMAP, MAPI and HTTP

Although Microsoft Exchange Server rules the roost in enterprise e-mail, with Lotus Notes coming in a distant second, these are by no means the most common e-mail system for the individual and small business user. When you access your personal e-mail from your own Internet Service Provider (ISP), chances are your e-mail comes to you from your ISP's e-mail server in one of three ways, POP, IMAP or HTTP, the last commonly called web- or browser-based e-mail. Understanding how these three protocols work—and differ—helps in identifying where e-mail can or cannot be found.

POP (for Post Office Protocol) is the oldest and most common of the three approaches and the one most familiar to users of the Outlook Express, Netscape and Eudora e-mail clients. Using POP, you connect to a mail server, download copies of all messages and, unless you have configured your e-mail client to leave copies on the server, the e-mail is deleted on the server and now resides on the hard drive of the computer you used to pick up mail. Leaving copies of your e-mail on the server seems like a great idea, since you have a back up if disaster strikes and can access all your e-mail using different computers. However, few ISPs afford unlimited storage space on their servers for users' e-mail, so mailboxes quickly become "clogged" with old e-mails and the servers start bouncing new messages. As a result, POP e-mail typically resides only on the local hard drive of the computer used to read the mail and on the back up system for the servers which transmitted, transported and delivered the messages. In short, POP is locally-stored e-mail that supports some server storage.

IMAP (Internet Mail Access Protocol) is the e-mail protocol used by Lotus Notes and, since 2004, supported by America Online (though AOL continues to furnish a proprietary e-mail client to its subscribers). IMAP differs from POP in that, when you check your e-mail, your e-mail client downloads just the headers (To, From, Date, Subject, etc.) of e-mail it finds on the server and only retrieves the body of a message when you open it for reading. Else, the entire message stays in your account on the server. Unlike POP, where e-mail is searched and organized into folders locally, IMAP e-mail is organized and searched on the server. Consequently, the server (and its back up tapes) retains not only the messages but also the way the user structured those messages for archival. Since IMAP e-mail "lives" on the server, how does a user read and answer it without staying connected all the time? The answer is that IMAP e-mail clients afford users the ability to synchronize the server files with a local copy of the e-mail and folders. When an IMAP user reconnects to the server, local e-mail stores are updated (synchronized) and messages drafted offline are transmitted. So, to summarize, IMAP is server-stored e-mail, with support for synchronized local storage.

MAPI (Messaging Application Programming Interface) is the e-mail protocol at the heart of Microsoft's Exchange Server application. Like IMAP, MAPI e-mail is typically stored on the server, not the client machine. Likewise, the local machine may be configured to synchronize with the server mail stores and keep a copy of mail on the local hard drive, but this is user- and client application-dependent. If the user hasn't taken steps to keep a local copy of e-mail, e-mail is not likely to be found on the local hard drive, except to the extent fragments may turn up through computer forensic examination.

HTTP (Hyper Text Transfer Protocol) mail, or web-based/browser-based e-mail, dispenses with the local e-mail client and handles all activities on the server, with users managing their e-mail using their Internet browser to view an interactive web page. Although some browser-based e-mail services support local (POP) synchronization with an e-mail client, typically users do not have any local record of their browser-based e-mail transactions except for messages they've affirmatively saved to disk or portions of e-mail web pages which happen to reside in the browser's cache (e.g., Internet Explorer's Temporary Internet Files folder). Gmail, Hotmail and Yahoo Mail are well-known examples of browser-based e-mail services, although many ISPs (including all the national providers) offer browser-based e-mail access in addition to POP and IMAP connections.

The protocol used to carry e-mail is not especially important in electronic discovery except to the extent that it signals the most likely place where archived e-mail can be found. Companies choose server-based e-mail systems (e.g., IMAP and MAPI) for two principal reasons. First, such systems make it easier to access e-mail from different locations and machines. Second, it's easier to back up e-mail from a central location. Because IMAP and MAPI systems store all e-mail on the server, the back up system used to protect server data can yield a mother lode of server e-mail. Depending upon the back up procedures used, access to archived e-mail can prove a costly and time-consuming task or a relatively easy one. The enormous volume of e-mail residing on back up tapes and the often-high cost to locate and restore that e-mail makes discovery of archived e-mail from back up tapes a big bone of contention between litigants. In fact, most reported cases addressing cost-allocation in e-discovery seem to have been spawned by disputes over e-mail on server back up tapes.

Local E-Mail Storage Formats and Locations

Faced with a discovery request for e-mail, where does one look to find stored e-mail, and what form will that storage take? Because individual e-mails are just text files, they could be stored as discrete text files. However, that's not an efficient or speedy way to manage a large number of messages, so e-mail client software doesn't do it that way. Instead, e-mail clients employ proprietary database files housing e-mail messages, and each of the major e-mail clients uses its own unique format for its database. Some programs encrypt the message stores. Some applications merely display e-mail housed on a remote server and do not store messages locally (or only in fragmentary way). The only way to know with certainty if e-mail is stored on a local hard drive is to look for it. Merely checking the e-mail client's settings is insufficient because settings can be changed. Someone not storing server e-mail today might have been storing it a month ago. Additionally, users may create new identities on their systems, install different client software, migrate from other hardware or take various actions resulting in a cache of e-mail residing on their systems without their knowledge.

For many, computer use is something of an unfolding adventure. One may have first dipped her toes in the online ocean using browser-based e-mail or an AOL account. Gaining computer-savvy, she may have signed up for broadband access or with a local ISP, downloading e-mail with Netscape Messenger or Microsoft Outlook Express. With growing sophistication, a job change or new technology at work, the user may have migrated to Microsoft Outlook or Lotus Notes as an e-mail client. Each of these steps can orphan a large cache of e-mail, possibly unbeknownst to the user but still fair game for discovery. Accordingly, a producing party shouldn't assert that a user has no e-mail unless the user's local machine(s) and server storage areas have both been thoroughly searched by someone who knows where active and orphaned e-mail reside.

Finding E-Mail on Exchange Servers

150 million people get their e-mail via a Microsoft product called Exchange Server. Though the preceding paragraphs dealt with finding e-mail stores on local hard drives, in disputes involving medium- to large-sized businesses, the e-mail server is likely to be the principal focus of electronic discovery efforts. The server is a productive venue in electronic discovery for many reasons, among them:

- Periodic back up procedures, routine parts of prudent server operation, tend to shield e-mail stores from those who, by error or guile, might delete or falsify data on local hard drives.
- The ability to recover deleted mail from archival server back ups may obviate the need for costly and sometimes fruitless forensic efforts to restore lost messages.
- Data stored on a server is often less prone to tampering by virtue of the additional physical and system security measures typically dedicated to centralized computer facilities as well as the inability of the uninitiated to manipulate data in the more-complex server environment.
- The centralized nature of an e-mail server affords access to many users' e-mail and may lessen the need for access to workstations at multiple business locations or to laptops and home computers.
- Unlike e-mail client applications, which store e-mail in varying formats and folders, e-mail stored on a server can usually be located with ease and adheres to a common file format.
- The server is the crossroads of corporate electronic communications and the most effective chokepoint to grab the biggest "slice" of relevant information in the shortest time, for the least cost.

Of course, the big advantage of focusing discovery efforts on the mail server (i.e., it can deliver up thousands or millions of messages) is also its biggest disadvantage (someone has to extract and review thousands or millions of messages). Absent a carefully-crafted and, ideally, agreed-upon plan for discovery of server e-mail, both requesting and responding parties run the risk of runaway costs, missed data and wasted time.

Server-based e-mail data is generally going to fall into two realms, being online "live" data, which is easily accessible, and offline "archival" data, which may be fairly inaccessible. Absent a change in procedure, "chunks" of data shift from the online to the offline realm on a regular basis--daily, weekly or monthly—as selected information on the server is duplicated onto backup media and deleted from the server's hard drives. The most common back up mechanism is a tape drive, really just a specialized version of a cassette tape recorder or VCR. These back up drives store data on magnetic tape cartridges not unlike a VHS tape As time elapses, the back up media may deteriorate, be discarded or re-used, such that older offline archival data entirely disappears (except, of course, from the many different places it may exist, in bits and pieces, on other servers and local systems).

When e-mail is online, it's an easy and inexpensive task to duplicate the messages and their attachments in their native form to a discrete file or files and burn those to CD or otherwise transmit the e-mail for review and production. When e-mail is offline, it can be no mean feat to get to it, and the reason why it's challenging and costly has to do with the way computers are backed up. The customary practice for backing up a server is to make a copy of specified files

and folders containing data. Sometimes a backup will copy everything, including the operating system software and the date; but, more often, time and cost constraints mean that only the stuff that can't be reloaded from other sources gets copied. Another common practice is to copy all the data every once and a while (e.g., monthly) and record only changes to the data at more frequent intervals. This is *incremental* backup.

A daunting challenge to using backup tapes to restore e-mail is the very large volume of duplicate e-mail found on successive backups. If a user tends to keep e-mail on the server, a backup of the user's Inbox in one month will look very much like a backup the next. Perhaps 90% of the messages will be identical, month-to-month. Unless steps are taken to filter out these identical e-mails (to "de-duplicate" the data), both the producing party and the requesting party may have to plow through a bloated, redundant production.

Another pitfall of backup tapes is that any e-mail received and deleted between backups isn't always saved. For example, if the defendant's e-mail server is backed up nightly, an e-mail received in the morning and immediately deleted likely won't be backed up because the system no longer "sees" the deleted e-mail in the user's Inbox.

Tips for Seeking E-Mail

As the volume of e-mail mounts, producing parties are increasingly turning to search tools to identify relevant messages. Be wary of searches based upon subject lines alone. As an e-mail "conversation" threads from one message to the next, aided by the Reply button, contents can veer far afield of the stated subject. Keyword searches alone also tend to overlook a significant percentage of responsive items, particularly when users employ unfamiliar terms-of-art, shorthand references and nicknames in their exchanges. If the producing party employs search tools in lieu of human judgment when responding to discovery, you have a right to know about it and to challenge the methodology if it proves inadequate to the task.

If relevant, be sure to seek production of BCC fields for responsive e-mail, which fields only exist on the sender's copy of the e-mail. Understand that what a user sees in their e-mail client program (e.g., Outlook or Eudora) is just part of the data that accompanies every e-mail. The data fields seen by the user may be sufficient for your purposes, but sometimes you'll need more extensive information, such as e-mail header data and routing information.

Don't give up. There is always at least one "e-mail pack rat" who keeps a copy of everything, notwithstanding policies to the contrary. The more Draconian the policy compelling e-mail destruction, the greater the likelihood employees have invented ways to circumvent the system (such as by forwarding old e-mail to himself or herself, burning it to disc or shipping it off to a free Gmail account). When opposing counsel says, "There isn't anything else," they may or may not be leveling with you, but they are almost certainly wrong. It's the rare case where a truly exhaustive e-discovery search is even begun prior to the first motion to compel and for sanctions. It's the rarer case where that subsequent search fails to turn up items that should have been produced in the first place.

Six Tips for Getting the Most from E-Discovery

First (and chanting this like a mantra is a good idea), *compel broad e-retention but seek narrow e-production.* This is the approach the courts are most likely to sustain and has the incidental strategic value of being most challenging to your opponent. Narrow requests aimed at clearly relevant evidence necessitate careful qualitative review by the producing party.

Second, get your preservation letter and your e-discovery out fast. Data will disappear over time, and you'll be in a poor position to complain about it if you didn't ask for the evidence while it was still around.

Third, don't expect to get it all in a single set of discovery. Digital evidence is everywhere, and you may have to go back to the well many times as you learn more about your opponent's operations and systems. Don't allow the fear of missing something to cause you to cast your net too wide. Keep the focus on what you need to prove your case, and be tenacious.

Fourth, remember that electronic evidence is easy to corrupt but hard to eradicate. It's probably not gone. Few private entities take effective steps to obliterate electronic information, and those that do tend to implement such practices only after the duty to preserve evidence arises. *There is no more compelling evidence than the void left by those who've deleted that which they were obliged to preserve.*

Fifth, get help. You may have to hire an EDD expert to guide you through the first time or two and anytime you're in over your head. Ask the court for help, too. Seek a TRO or protective order and move for appointment of a special master skilled in electronic discovery.

Finally, don't forget: ***what goes around comes around***. Plaintiffs are increasingly vulnerable to e-discovery, so anticipate boomerang discovery, meet and confer early and often, share expectations, seek solutions and don't promise what you may not be able to deliver.

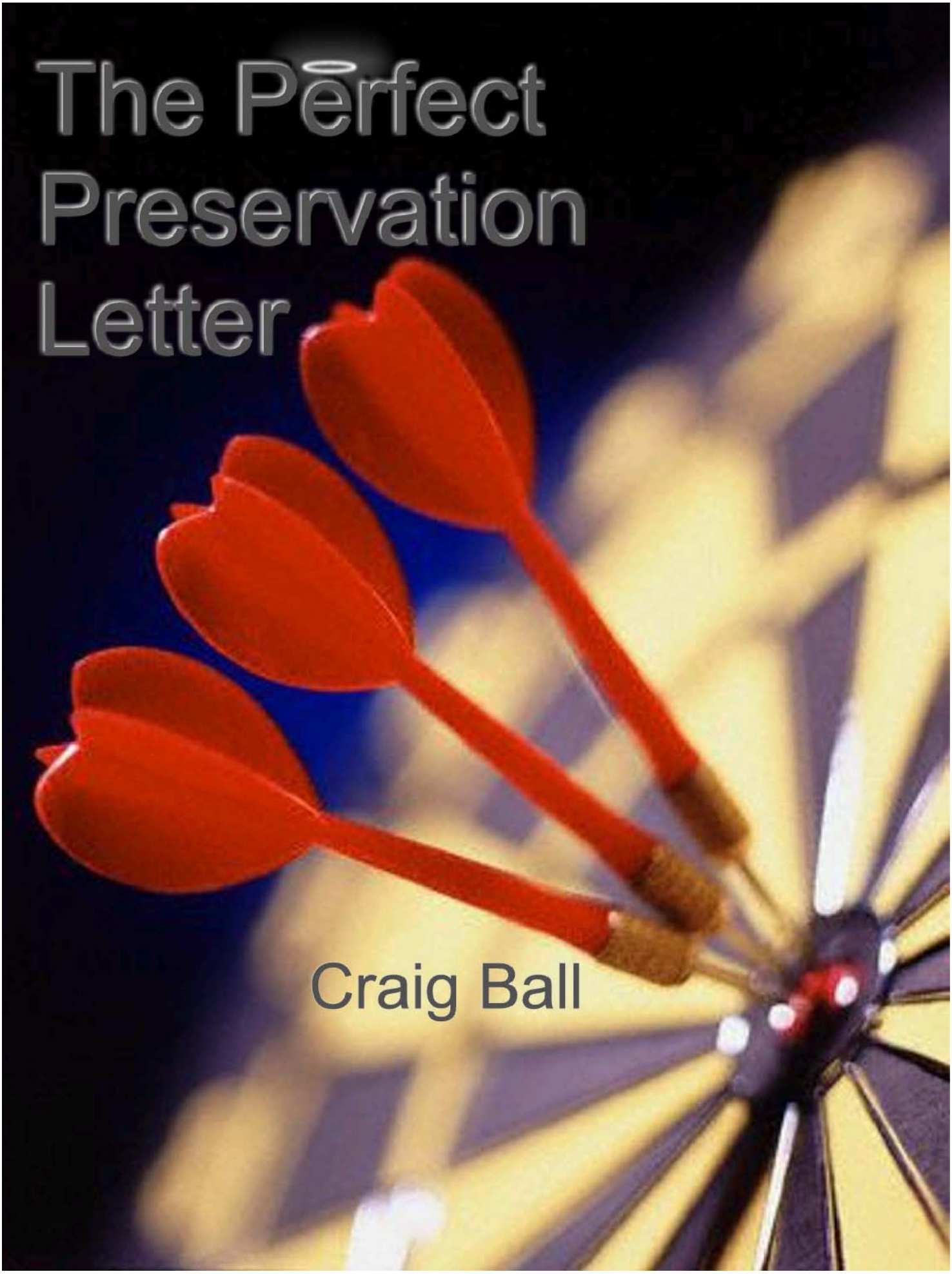
Twenty Tips for Counsel Seeking Discovery

1. Get your preservation letter out early and be *both* specific and general. Assume that the recipients don't know their own systems and don't understand computer forensics. Use the letter to educate them so they can't use ignorance as an excuse.
2. Do your homework: use the Net and ask around to learn about the nature and extent of your opponent's systems and practices. You're probably not the first person to pursue e-discovery against the defendant. Others may know where the bodies are buried.
3. Get your e-discovery out swiftly. Data is going to disappear. You're in a poor position to complain about it if you didn't seek the missing evidence while it was still around.
4. Force broad retention, but pursue narrow discovery.
5. What they must keep and what they must give you are different obligations. Keeping the first broad protects your client's interests and exposes their negligence and perfidy. Keeping requests for production narrow and carefully crafted makes it hard for your opponent to buy delays through objection. Laser-like requests mean that your opponents must search with a spoon instead of a backhoe. Tactically, ten single, surgical requests spread over 20 days are more effective than 20 requests in one.
6. Be aware that opposing counsel may not understand the systems as well as you do, and won't want anyone—especially his client--to know. Help him "get it," so he can pose the right questions to his client.
7. Question the IT people and focus on the grunts. They've spent less time in the woodshed than the managers, and they know the *real* retention practices.
8. Get their document retention policies, network topology and inventory of computing resources (including laptops, home systems, PDAs and removable media).
9. Invoke the court's injunctive power early to force preservation. The agreement reached to avoid a court order may be better than the relief you'll get from the judge.
10. If you can't get make any headway, seek appointment of a neutral expert or special master.
11. Ask all opponent employee witnesses what they were told to do in the way of e-document retention, then find out what they actually did.
12. Digital data is easily forged. Know how and when to check for authenticity of data produced.
13. Be sure to get metadata whenever it may be relevant.
14. Don't accept image data (TIFF or PDF) when you need native data.
15. Have the principal cases on e-discovery and cost shifting at hand. Tailor your requests to the language of the cases.
16. Set objections for hearing immediately. Require assertions of burden and cost to be supported by evidence.
17. Analyze what you get promptly after you get it, and pin down that it's being tendered as "everything" that's responsive. Follow up with additional requests based upon your analysis.
18. Don't let yourself be railroaded into cost sharing, but if it happens, be sure you're protected from waste and excess by the other side, and leverage your role as underwriter to gain greater access.
19. Be prepared to propose a "claw back" production, if advantageous.
20. Don't accept assertions about cost or complexity unless you know them to be accurate. Independently evaluate all such claims and be prepared to propose alternatives.

E-Discovery Report Card: How Well Do TIFF Images Perform?
Craig Ball

Mark Twain observed, "To a man with a hammer, everything looks like a nail." To e-discovery vendors—paid by the TIFF image and heavily invested in systems to convert digital evidence to TIFF images--*everything* looks like a candidate for production in TIFF. But how does TIFF *really* stack up against native production for the most common file types in electronic discovery? Here's TIFF's report card:

Source Evidence	Form or Forms of TIFF Production	Grade	Explanation
Paper Documents	TIFF images with a load file containing OCR data	A	TIFF images paired with good OCR are ideally suited to electronic production when the source evidence exists *only* as paper documents
Electronic Mail	TIFF images with a load file containing full message text and To, From, CC/BCC, Date and Subject Fields	B	TIFF is suited to production of plain text e-mail when paired with production of searchable text and header fields (sometimes mistakenly called "metadata"). TIFF's won't always suffice for HTML-formatted e-mail or for some common e-mail attachments.
Word Processed Documents	TIFF images with a load file containing full text and relevant system and application metadata fields *(e.g.,* author, path, filename, last modified date, last accessed date, last printed date and *both* stored creation dates).	B	Though the printed version of a document can be faithfully reproduced and electronically searchable using a load file, application metadata (comments, tracked changes, circulation lists, edit times, etc.) will be lost unless extracted and produced. Formatting won't fairly represent the original unless the TIFF is virtually "printed" using the same formatting settings.
Spreadsheets	TIFF images with a load file containing full text and relevant system and application metadata	D	Spreadsheets contain three-dimensional data (e.g., a grid of computed values derived from underlying formulae and/or multiple stacked/linked worksheets) that TIFF can't handle even with load files. But TIFF might suffice for spreadsheets without formulae (e.g., tables) that fit on a single page or two.
Databases	TIFF images with a load file containing full text and relevant system and application metadata fields	F	Enterprises use databases throughout their operations, and even everyday programs like Microsoft's Outlook and Intuit's Quicken store their data in *database* formats. TIFF images are incapable of presenting database information in a useful or intelligible manner, with or without load files.
Digital Photographs	TIFF images	B	TIFF does a good job on photographic information when resolution is adequate and TIFF formats supporting color are used. TIFF productions often miss the non-pictorial data found in common digital file formats, such as camera identification and EXIF and RAW data, and alteration ("Photoshopping") is harder to detect using TIFF.
Audio Files	Not feasible	F	Voice mail and integrated messaging use sound formats to store data. TIFF can't store audio.
Video Files	Not Feasible	F	TIFF can't store video or animation.
Redacted Data	TIFF images with an OCR load file	A	Ironically, TIFF's limitations make it a superior format for producing redacted data because it's incapable of preserving parts of the evidence not seen on a printed page. Hence, it's closest to paper in its ability to support obscured (i.e., "blacked out") text.


The Perfect Preservation Letter
Craig Ball

The Perfect Preservation Letter

By Craig Ball

Well, I was drunk the day my Mom got outta prison,
And I went to pick her up in the rain;
But before I could get to the station in my pickup truck,
She got runned over by a damned old train.

From "You Never Even Called Me By My Name"
(a/k/a "The Perfect Country and Western Song")
By Steve Goodman, performed by David Allan Coe

Outlaw musician David Allan Coe sings of how no country and western song can be "perfect" unless it talks of Mama, trains, trucks, prison and getting drunk. Likewise, no digital evidence preservation letter can be "perfect" unless it clearly identifies the materials requiring protection, educates your opponent about preservation options and lays out the consequences of failing to preserve the evidence*. You won't find the perfect preservation letter in any formbook.* You have to build it, custom-crafted from a judicious mix of technical boilerplate and *fact-specific* direction. It compels broad retention while appearing to ask for no more than the bare essentials. It rings with reasonableness. It keeps the focus of e-discovery where it belongs: relevance. This article discusses features of the perfect preservation letter and offers suggestions as to how it can be effectively drafted and deployed.

Contents

The Role of the Preservation Letter

> "The reality of electronic discovery is it starts off as the responsibility of those who don't understand the technology and ends up as the responsibility of those who don't understand the law."

You can read the Federal Rules of Civil Procedure from cover to cover and not see a reference to preservation letters. So why invest a lot of effort creating the perfect preservation letter? Wouldn't it be sufficient to remind your opponent that the laws and rules prohibiting destruction of evidence apply to electronically stored information in the same manner that they apply to other evidence and that they must take every reasonable step to preserve this information until the final resolution of the case? Perhaps in a decade or so it will be enough; but, today we face an explosion of electronic evidence untamed by sound records management. Too many litigators and in-house counsel are clueless about information systems. The reality of electronic discovery is it starts off as the responsibility of those who don't understand the technology and ends up as the responsibility of those who don't understand the law. A well-drafted preservation letter helps bridge this knowledge gap.

The goal of the preservation letter is, of course, to remind opponents to preserve evidence, to be sure the evidence doesn't disappear. But, the preservation letter also serves as the linchpin of a subsequent claim for spoliation, helping to establish bad faith and conscious disregard of the duty to preserve relevant evidence. It is today's clarion call that underpins tomorrow's, "I told you so." The more effectively you give notice and convey what must be retained--including methodologies for preservation and consequences of failure--the greater the likelihood your opponent will be punished for destruction of evidence.

The Proposed Amendments to the Rules of Civil Procedure

Though serving a preservation letter isn't a formal component of civil discovery procedures, it's likely to be a *de facto* practice as federal and local rules of civil procedure impose express e-discovery "meet and confer" obligations upon litigants. For example, effective December 1, 2006, Rule 26 of the Federal Rules of Civil Procedure will require litigants to "discuss any issues relating to preserving discoverable information,"[1] as well as "any issues relating to disclosure or discovery of electronically stored information, including the form or forms in which it should be produced."[2] The preservation letter is sure to frame the agenda for such discussions.

The preservation letter will may play an important role in a court's consideration of whether a party acted in good faith in connection with information lost to routine operations of an electronic information system.[3] Assessment of good faith turns on the subjective awareness of the party failing to preserve evidence. The preservation letter can establish such awareness, bolstering a claim that the party destroying evidence knew of its discoverability and recklessly or intentionally disregarded it. Per commentary to the proposed rule, "Good faith in the routine operation of an

[1] Proposed Amendment to Rule 26(f) of the Federal Rules of Civil Procedure. All proposed Amendments and commentary cited are available at http://www.uscourts.gov/rules/EDiscovery_w_Notes.pdf.

[2] Id. Rule 26(f)(3)

[3] Id. Proposed Rule 37(f), entitled "Electronically Stored Information" provides, "Absent exceptional circumstances, a court may not impose sanctions under these rules on a party for failing to provide electronically stored information lost as a result of the routine, good-faith operation of an electronic information system."

information system may involve a party's intervention to modify or suspend certain features of that routine operation to prevent the loss of information, if that information is subject to a preservation obligation. [4] A clear and instructive preservation letter that serves to educate your opponent isn't just a professional courtesy; it also compels recognition of a duty to intervene to prevent data loss and deprives an opponent of a sanctions "safe harbor."

"A clear and instructive preservation letter that serves to educate your opponent isn't just a professional courtesy…."

What is Electronic Evidence Preservation?

When evidence is a paper document, preserving it is simple: We set the original or a copy aside, confident that it will come out of storage exactly as it went in. Absent destructive forces or tampering, paper stays pretty much the same. But despite lawyers' ardor for paper, 95% of information is born *digitally*, and the overwhelming majority is never printed.

By contrast, preserving electronic data poses unique challenges because:

- "Touching" data changes it
- Digital evidence is increasingly ill-suited to printing
- Data must be interpreted to be used
- Storage media are fragile and changing all the time
- Digital storage media are dynamic and recyclable

Touching Data Changes It

Route a document through a dozen hands and, aside from a little finger grime or odd coffee stain, the document won't spontaneously change just by moving, copying or reading it. But open that same document in Microsoft Word, or copy it to a CD, and you've irretrievably changed that document's *system metadata*, the data-about-data items, like the document's creation or last access dates that may themselves be evidence.

It's common to use recordable CDs to transfer data between systems or as a medium of e-production. But how many lawyers are aware that you *can't easily* copy all of a file's metadata when it's moved from hard drive to a recordable CD? The two media use different file systems such that the CD-R doesn't offer a structure capable of storing all of a file's Windows metadata. That is, where the Windows file system offers three slots for storing file dates (i.e., Modified, Accessed and Created), the CD-R file structure has but one. With no place to go, metadata is jettisoned in the CD recording process, and the missing metadata may be misreported on the destination system.

Digital Evidence Is Increasingly Ill-Suited to Printing

Much modern evidence doesn't lend itself to paper. For example, a spreadsheet displays values derived from embedded formulae, but you can't embed those formulae in paper. In large databases, information occupies expansive grids that wouldn't fit on a printed page or make much sense if it could. And, of course, sound and video evidence can't make the leap to paper. So preserving on paper isn't always an option, and it's rarely an inexpensive or efficient proposition.

[4] *Id.* Committee Note to Proposed Rule 37(f).

Data Must Be Interpreted To Be Used

If legible and in a familiar language, a paper document conveys information directly to the reader. A literate person can interpret an alphabet, aided by blank space and a few punctuation marks. It's a part of our grade school "programming." But ***all*** digital data are just streams of ones and zeroes. For those streams of data to convey anything intelligible to people, the data must be interpreted by a computer using specialized programming called "applications." Without the right application—sometimes even without the correct *version* of an application—data is wholly inaccessible or may be inaccurately presented. Successfully preserving data also entails preserving legacy *applications* capable of correctly interpreting the data as well as legacy computing environments—hardware and software—capable of running these applications.

Storage Media Are Fragile and Changing

If your great grandfather put a letter in a folder a century ago, chances are good that notwithstanding minor signs of age, you could pull it out today and read it. But changes in storage technology and rapid obsolescence have already rendered fifteen-year-old digital media largely inaccessible absent considerable effort and expense. How many of us still have a computer capable of reading a 5.25" floppy? The common 3.5" floppy disk is disappearing, too, with even CD-ROMs fast on its heels to oblivion. Data stored on back up tapes and other magnetic and even optical media fades and disappears over time like the contents of once-common thermal fax paper. Disks expected to last a century are turning up illegible in a decade. Back up tapes may stretch a bit each time they are used and are especially sensitive to poor storage conditions. Long-term data preservation will entail either the emergence of a more durable medium or a relentless effort to migrate and re-migrate legacy data to new media.

Digital Storage Media Are Dynamic and Recyclable

By and large, paper is not erased and reused for information storage; at least not in a way we'd confuse its prior use as someone's Last Will & Testament with its reincarnation as a cardboard carton. By contrast, a hard drive is constantly changing and recycling its contents. A later version of a document may overwrite—and by so doing, destroy—an earlier draft, and storage space released by the deletion of one file may well be re-used for storage of another. This is in sharp contrast to paper preservation, where you can save a revised printout of a document without affecting—and certainly not obliterating-- a prior printed version.

Clearly, successful preservation of digital data isn't always as simple as copying something and sticking it in a folder; but your opponent may not appreciate the planning and effort digital preservation requires. When that's the case, the requesting party is at a crossroads: Do you seek to educate the producing party or its counsel about how and why to properly preserve digital evidence, or do you keep mum in hopes that an advantage will flow from your opponent's ineptitude? That is, do you want the evidence or the sanction? Most of the time, you'll want the evidence.

The Duty to Preserve

At what point does the duty to preserve evidence arise? When the lawsuit is filed? Upon receipt of a preservation letter? When served with a request for production?

The duty to preserve evidence may arise before—and certainly arises without—a preservation letter. In fact, the duty can arise long before. A party's obligation to preserve evidence is generally held to arise when the party knows or has reason to know that evidence may be

relevant to future litigation. This "reasonable anticipation of litigation" standard means that any person or company who should see a claim or lawsuit on the horizon must act, *even before a preservation letter or lawsuit has materialized*, to cease activities likely to destroy electronic or tangible evidence and must take affirmative steps to preserve such evidence.

Thus, the preservation letter may be only one of a number of events—albeit a decisive one--sufficient to trigger a duty to preserve evidence. Arrival of the preservation letter is often the first time responding parties focus on what evidence exists and what they will elect to save.

Balance and Reasonableness

The problem with preservation letters is that they often must be sent when you know little-to-nothing about your opponent's information systems; consequently, they tend to be everything-but-the-kitchen-sink requests, created without much thought given to the "how" and "how much" issues faced by the other side.

A preservation letter that demands the moon and paralyzes your opponent's operations won't see compliance or enforcement. Absent evidence of misconduct (such as shredding or other overt destruction of evidence), a court isn't likely to sanction a party for failing to comply with a preservation letter so onerous that no one dare turn on their computer for fear of spoliation! For a preservation letter to work, it must be reasonable on its face. Remember: all you're trying to do is keep the other side from destroying relevant evidence, and just about any judge will support you in that effort *if your demands aren't cryptic, overbroad or unduly burdensome*.

If it could be accomplished with paper evidence, judges expect a corollary accomplishment with electronic evidence. Still, digital is different, and some of the ways we approach paper discovery just won't fly for electronic evidence. For example, using the term "any and all" in a request for digital evidence is a red flag for potential over breadth. Demanding that an opponent retain "any and all electronic communications" is nonsense. After all, phone conversations are electronic communications, and it's unlikely that a court would require a litigant to tape all phone calls, though a judge shouldn't hesitate to compel *retention* of the tapes *when phone calls are already recorded and relevant.* If what you want preserved is e-mail, or instant messaging or voice mail, *spell it out.* Your opponent may squawk, but at least the battle lines will be drawn on specific evidentiary items your opponent may destroy instead of fighting about what constitutes a "communication?" The risk to this approach is that your opponent may fail to preserve what you haven't specified. Still, to the extent the evidence destroyed was relevant and material, that risk may be adequately addressed by a demand to retain all information items bearing on the claims made the basis of the claim. Further, the preservation letter neither creates the duty to preserve nor constrains it. If the evidence was relevant and discoverable, then destroyed at a time when your opponent should have known to keep it, it's still spoliation.

Preservation Essentials

A perfect preservation letter must, first and foremost, seek to halt routine business practices geared to the destruction of potential evidence. It might call for an end to server back up tape rotation (as appropriate), electronic data shredding, scheduled destruction of back up media, re-imaging of drives, drive hardware exchanges, sale, gift or destruction of computer systems and, (especially if you know computer forensics may come into play) disk defragmentation and maintenance routines. A lot of digital evidence disappears because of a lack of communication ("legal forgot to tell IT") or of individual initiative ("this is MY e-mail and I can delete it if I want

to"). So, be sure to highlight the need to effectively communicate retention obligations to those with hands-on access to systems, and suggest steps to forestall personal delete-o-thons. **Remember:** When you insist that communications about preservation obligations ***reach*** every custodian of discoverable data and that such communications stress the importance of the duty to preserve, you are demanding no more than the developing law suggests is warranted. See, e.g., *Zubulake v. UBS Warburg LLC, No. 02 Civ. 1243 (S.D.N.Y. July 20, 2004) ("Zubulake V").*

Next, get fact specific! Focus on items specifically bearing on the claim or suit, like relevant business units, activities, practices, procedures, time intervals, jurisdictions, facilities and key players. Here, follow the "who, what, when, where and how" credo of good journalism. Preservation letters are more than a boilerplate form into which you pack every synonym for document and computer in the thesaurus. If your preservation letter boils down to "save everything about anything by everyone, everywhere at any time," it's time to re-draft it because not only will no trial court enforce it, many will see it as discovery abuse.

The preservation letter's leading role is to educate your opponent about the many forms of relevant electronic evidence and the importance of taking prompt, affirmative steps to be sure that evidence remains accessible. Educating the other side isn't a noble undertaking—it's sound strategy. Spoliation is frequently defended on the basis of ignorance; *e.g.,* "Your honor, we had no idea that we needed to do that," and your goal is to slam the door on the "it was an oversight" excuse. Doing so entails more than just reciting a litany of storage media to be preserved--you've got to educate, clearly and concisely.

Finally, don't be so focused on electronic evidence that you fail to direct your opponent to retain the old-fashioned paper variety. And remember that turnabout is fair play. Don't compel your opponent to preserve data to an extent much greater than *your* client could sustain. Doing so could hurt your credibility with the court right out of the gate.

The Nature of the Case

As formal discovery requests come after service of a complaint, the parties know what the dispute is about by the time the discovery requests arrive. But a pre-suit preservation letter may be your opponent's first inkling that they face litigation. Don't just assume that those receiving your preservation letter know what the dispute is about: *spell it out for them*. Though you may be unprepared to draft your formal complaint, furnish sufficient information about the nature of the case to sustain a later claim that a reasonable person reading the preservation letter would have known to preserve particular evidence. Names of key players, dates, business units, office locations and events will all be weighed in deciding what's relevant and must be retained. The more you can offer, the less likely you are to someday hear, "If you wanted Bob's e-mail, why didn't you mention Bob in the preservation letter?"

When to Send a Preservation Letter

The conventional wisdom is that preservation letters should go out as soon as you can identify potential defendants. But there may be compelling reasons to delay sending a preservation letter. For example, when you face opponents who won't hesitate to destroy evidence, a preservation letter is just the starting gun and blueprint for a delete-o-thon. Instead, consider seeking a temporary restraining order or the appointment of a special master (but recognize that the Comments to the proposed Rules amendments strongly discourage entry of *ex parte* preservation orders). Delay in sending the letter may be wise when your investigation is

ongoing and the service of a preservation letter will cause the other side to hire a lawyer or trigger privileges running from the anticipation of litigation. There may even be circumstances where you ***want*** your opponent's routine, good faith destruction of information to continue, such as where information unfavorable to your position will be lost in the usual course of business.

> "There may be circumstances where you ***want*** your opponent's routine destruction of information to continue…."

Who Gets the Letter?

If counsel hasn't appeared for your opponent, to whom should you direct your perfect preservation letter? Here, the best advice is err on the side of as many appropriate persons as possible. Certainly, if an individual will be the target of the action, he or she should receive the preservation letter. However, if you know of others who may hold potential evidence (such as the defendant's spouse, accountant, employer, banker, customers and business associates), it's smart to serve a preservation demand upon them, making clear that you are also seeking preservation of physical and electronic records in their possession pertaining to the matters made the basis of the contemplated action. Some litigants use the preservation letter as a means to put pressure upon customers lost to or solicited by a competitor-defendant. ***Beware***…as the preservation letter isn't a discovery mechanism expressly countenanced by the rules of procedure, its use as an instrument of intimidation may not be privileged and could provoke a counterclaim based on libel or tortuous interference.

If the potential defendant is a corporation, a presentation addressed to the wrong person may be ignored or late in reaching those able to place litigation holds on records. Consequently, it's wise to direct preservation letters to several within the organization, including, *inter alia*, the Chief Executive Officer, General Counsel, Director of Information Technologies and perhaps even the Head of Corporate Security and registered agent for service of process. You may want to copy other departments, facilities or business units.

You naturally want to be sure that as many who hold evidence as possible are put on notice, but you also want to disseminate the preservation duty widely to foment uncertainty in those who might destroy evidence but for the possibility that others in the organization will retain copies. Of course, if counsel has entered an appearance, weigh whether you are constrained from communicating directly with represented parties.

If possible, consider who is most likely to *unwittingly* destroy evidence and be certain that person receives a preservation letter. Sending a preservation letter to a person likely to destroy evidence *intentionally* is a different story. The letter may operate as the triggering event to spoliation, so you may need to balance the desire to give notice against the potential for irretrievable destruction.

Of course, preservation letters, like any important notice, should be dispatched in a way enabling you to prove receipt, like certified mail, return receipt requested.

How *Many* Preservation Letters?

Turning to the obligatory litigation-as-war metaphor, is a preservation letter best delivered as a single giant salvo across the opponent's bow, or might it instead be more effectively launched as several carefully-aimed shots? It's common to dispatch a single, comprehensive request, but

might it instead be wiser to present your demands in a *series* of focused requests, broken out by, *e.g.*, type of digital medium, issues, business units, or key players? Your preservation letter may be destined to be an exhibit to a motion, so when the time comes to seek sanctions for a failure to preserve evidence, wouldn't it be more compelling to direct the court to a lean, specific preservation notice than a bloated beast? Also, consider supplementing a "master" preservation notice with specific notices directed at key players as the matter proceeds. It's difficult to claim, "We didn't realize you wanted ***Bob's*** e-mail" when Bob got his very own, custom-tailored preservation letter.

Specifying Form of Preservation

The proposed amendments to the Federal Rules of Civil Procedure permit a requesting party to specify the form or forms in which the requesting party wants electronic evidence produced. Some states, notably Texas, already permit such a designation in their rules governing discovery. Often, there's no additional trouble or expense for the producing party to generate one format over another and occasionally a non-native production format proves easier or cheaper to manage. But, should the *preservation letter* specify the form in which the data should be preserved? Generally, the answer is "No," because you don't want your preservation letter to appear to demand anything more onerous than maintaining evidence in the way it's kept in the ordinary course of business. On the other hand, when your specification operates to ***ease*** the cost or burden to the producing party or otherwise ***helps*** the producing party fulfill its preservation obligation, a format might be *suggested* (although, be clear that the specified form is just one acceptable format).

Special Cases: Back Up Tapes, Computer Forensics and Metadata

The e-discovery wars rage in the mountains of e-mail and flatlands of Excel spreadsheets, but nowhere is the battle so pitched as at the front lines and flanks called *back up tapes, computer forensics and metadata*. These account for much of the bloodshed and so deserve special consideration in a preservation letter.

Back Up Tapes

In the "capture the flag" e-discovery conflicts now waged, the primary objective is often your opponent's server back up tapes or, more particularly, forcing their retention and 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. To this end, a business need retain disaster recovery data for a brief interval since there are few instances where a business would wish to re-populate its information systems with stale data. Because only the latest data has much utility in a properly designed back up system, the tapes containing the oldest backed-up information are typically recycled over time. This practice is "tape rotation," and the interval between use and reuse of a particular tape or set of tapes is the "rotation cycle" or "rotation interval."

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 those motivated to make evidence disappear--backup tapes are often the only evidence containers beyond the reach of those with the incentive to destroy or fabricate evidence. Going back to 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 "smoking gun" evidence.

Another reason back up tapes lie at the epicenter of e-discovery disputes is that many organizations elect to retain back up tapes for archival purposes (or in response to litigation hold instructions) long after they've lost their usefulness for disaster recovery. Here again, when data has been deleted from the active systems, the stale back up tapes are a means (joined by, *inter alia*, computer forensics and discovery from local hard drives) by which the missing pieces of the evidentiary puzzle can be restored.

In organizations with many servers, back up systems are complex, hydra-headed colossi. There may be no simple one-to-one correspondence between a server and a particular user, and most tape back up systems structure stored data differently from active data on the server, complicating restoration and exploration. Volume, complexity and the greater time it takes to access tape compared to disk all contribute to the potentially high cost of targeting back up tapes in discovery. Compelling a large organization to interrupt its tape rotation, set aside back up 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. This is the Hobson's choice[5] of e-discovery.

A preservation letter should target just the backup tapes likely to contain deleted data relevant to the issues in the case—a feat easier said than done. Whether by Internet research, contact with former employees or consultation with other lawyers who've plowed the same ground, seek to learn all you can about the architecture of the active and backup systems. The insight gleaned from such an effort may allow for a more narrowly tailored preservation request or justify a much broader one.

The responding party need not retain purely cumulative evidence, so once established that data has not been deleted and all relevant information still exists on the servers, the back up tapes should be released to the rotation. Again, this is a goal more easily stated than achieved because it requires three elements too often absent from the adversarial process: ***communication, cooperation and trust*.** Hopefully, the advent of compulsory meet-and-confer sessions will force litigants to focus on e-discovery issues sufficiently early to stem unnecessary costs by narrowing the breadth of preservation efforts to just those actions or items most likely to yield discoverable data.

Drive Cloning and Imaging

Data deleted from a personal computer isn't gone. The operating system simply releases the space the data occupies for reuse and treats the space as available for reuse. Deletion rarely erases data. In fact, there are three and *only* three ways that information's destroyed on personal computer:

1. Completely overwriting the deleted data on magnetic media (*e.g.,* floppy disks, tapes or hard drives) with new information;
2. Strongly encrypting the data and then "losing" the encryption key; or,
3. Physically damaging the media to such an extent that it

There are three and *only* three ways that information's destroyed on a personal computer

[5] Thomas Hobson was a British stable keeper in the mid-1600s whose policy was that you either took the horse nearest the stable door or he wouldn't rent you a horse. "Hobson's choice" has come to mean an illusory alternative. Back up tapes are problematic, but the unacceptable alternative is letting evidence disappear.

cannot be read.

Computer forensics is the name of the science that, *inter alia,* resurrects deleted data. Because operating systems turn a blind eye to deleted data (or at least that which has gone beyond the realm of the Recycle Bin), a copy of a drive made by ordinary processes won't retrieve the sources of deleted data. Computer forensic scientists use specialized tools and techniques to copy every sector on a drive, including those holding deleted information. When the stream of data containing each bit on the media (the so-called "bitstream") is duplicated to another drive, the resulting forensically qualified duplicate is called a "clone." When the bitstream's stored in files, it's called a "drive image." Computer forensic tools analyze and extract data from both clones and images.

In routine computer operation, deleted data is overwritten by random re-use of the space it occupies or by system maintenance activities; consequently, the ability to resurrect deleted data with computer forensics erodes over time. When the potential for discovery from deleted files on personal computers is an issue, a preservation letter may specify that the computers on which the deleted data reside should be removed from service and shut down or imaged in a forensically sound manner. Such a directive might read:

> You are obliged to take affirmative steps to prevent anyone with access to your data, systems and archives from seeking to modify, destroy or hide electronic evidence on network or local hard drives (such as by deleting or overwriting files, using data shredding and overwriting applications, defragmentation, re-imaging or replacing drives, encryption, compression, steganography or the like). With respect to workstation and laptop hard drives, one way to protect existing data on is the creation and authentication of a forensically-qualified image of all sectors of the drive. Such a forensically-qualified duplicate may also be called a bitstream image or clone of the drive. Be advised that a conventional back up or "Ghosting" of a hard drive are not forensically-qualified procedures because they capture only active data files and fail to preserve forensically-significant data that may exist in such areas as unallocated space, slack spaces and the swap file.
>
> For the hard drives and other digital storage devices of each person named below, and of each person acting in the capacity or holding the job title named below, as well as each other person likely to have information pertaining to the instant action on their computer hard drive(s) or other electronic storage media, demand is made that you immediately obtain, authenticate and preserve forensically-qualified images of the hard drives and other storage media in (or used in conjunction with) any computer system (including portable and home computers) used by that person during the period from ____ to _____, as well as recording and preserving the system time and date of each such computer.
>
> [Insert names, job descriptions and titles here].
>
> Once obtained, each such forensically-qualified image should be labeled to identify the date of acquisition, the person or entity creating the image, the deviation (if any) of the system time and date and the system from which it was obtained. Each such image should be preserved without alteration.

Be advised that booting a drive or other electronic storage media, examining its contents or running any application will irretrievably alter the evidence it contains and may constitute unlawful spoliation of evidence.

Metadata

Metadata, the "data about data" created by computer operating systems and applications, may be critical evidence in your case, and its preservation requires prompt and definite action. Information stored and transmitted electronically must be tracked by the system which stores it and often by the application that creates it.

For example, a Microsoft Word document is comprised of information you can see (*e.g.,* the text of the document and the data revealed when you look at the document's "Properties" in the File menu), as well as information you can't see (*e.g.*, tracked changes, revision histories and other data the program uses internally). This *application* metadata is stored as part of the document file and moves with the file when it is copied or transmitted. In contrast, the computer system on which the document resides keeps a record of when the file was created, accessed and modified, as well as the size, name and location of the file. This *system* metadata is typically not stored within the document, at least not completely. So when a file is copied or transmitted—as when it's burned to disk for production—potentially relevant and discoverable system metadata is not preserved or produced. Worse, looking at the document or copying it may irreparably alter the metadata. Absent proper steps to protect metadata, even a virus scan can corrupt metadata evidence.

Metadata is not a critical element in all disputes, but in some the issue of ***when*** a document or record was created, altered or copied lies at the very heart of the matter. If you reasonably anticipate that metadata will be important, be sure to direct the producing party to preserve relevant metadata evidence and warn of the risks threatening corruption. Because many aren't aware of metadata—and even those who are may think of it just in the context of application metadata—the preservation letter needs to define metadata and educate your opponent about where to find it, the common operations that damage it and, if possible, the means by which it's preserved.

For further information about metadata, see "*Beyond Data about Data: the Litigators Guide to Metadata*" at http://www.craigball.com/metadata.pdf.

End Game

Are you prepared to let relevant evidence disappear without a fight? ***No!***
Can the perfect preservation letter really make *that* much difference? ***Yes!***

The preservation letter demands your best effort for a host of reasons. It's the basis of your opponent's first impression of you and your case. A well-drafted preservation letter speaks volumes about your savvy, focus and preparation. An ill-drafted, scattergun missive suggests a formbook attorney who's given little thought to where the case is going. A letter that demonstrates close attention to detail and preemptively slams the door on cost-shifting and "innocent" spoliation bespeaks a force to be reckoned with and signals a case that deserves to be a settlement priority. The carefully-crafted preservation letter serves as a blueprint for meet and confer sessions and a touchstone for efforts to remedy destruction of evidence.

Strategically, the preservation letter forces your opponent to weigh potential costs and business disruption at the outset, often before a lawsuit is filed. If it triggers a litigation hold, everyone from the board room to the mail room may learn of the claim and be obliged to take immediate action. It may serve as the starting gun for a reckless delete-o-thon or trigger a move toward amicable resolution. But done right, ***the one thing it won't be is ignored*****.**

APPENDIX: Exemplar Preservation Demand to Opponent

What follows *isn't* the perfect preservation letter *for your case*, so I don't recommend adopting it as a form. I include it here as a drafting aid and to flag issues unique to EDD. You should tailor *your* electronic discovery efforts to the issues, parties and systems in your case. Be thorough insofar as data may be relevant, but eschew the "everything and the kitchen sink" approach. Use common sense. If your preservation demand effectively requires your opponent to pull the plug on every computer, what good is it? If you can't articulate *why* particular ESI is potentially relevant, perhaps you shouldn't demand its preservation. CDB

Demand for Preservation of Electronically Stored Information

Plaintiffs demand that you preserve all documents, tangible things and electronically stored information potentially relevant to the issues in this cause. As used in this document, "you" and "your" refers to [**NAME OF DEFENDANT**], and its predecessors, successors, parents, subsidiaries, divisions or affiliates, and their respective officers, directors, agents, attorneys, accountants, employees, partners or other persons occupying similar positions or performing similar functions.

You should anticipate that much of the information subject to disclosure or responsive to discovery in this matter is stored on your current and former computer systems and other media and devices (including personal digital assistants, voice-messaging systems, online repositories and cell phones).

Electronically stored information (hereinafter "ESI") should be afforded the broadest possible definition and includes (by way of example and not as an exclusive list) potentially relevant information electronically, magnetically or optically stored as:

- Digital communications (e.g., e-mail, voice mail, instant messaging);
- Word processed documents (e.g., Word or WordPerfect documents and drafts);
- Spreadsheets and tables (e.g., Excel or Lotus 123 worksheets);
- Accounting Application Data (e.g., QuickBooks, Money, Peachtree data files);
- Image and Facsimile Files (e.g., .PDF, .TIFF, .JPG, .GIF images);
- Sound Recordings (e.g., .WAV and .MP3 files);
- Video and Animation (e.g., .AVI and .MOV files);
- Databases (e.g., Access, Oracle, SQL Server data, SAP);
- Contact and Relationship Management Data (e.g., Outlook, ACT!);
- Calendar and Diary Application Data (e.g., Outlook PST, Yahoo, blog tools);
- Online Access Data (e.g., Temporary Internet Files, History, Cookies);
- Presentations (e.g., PowerPoint, Corel Presentations)
- Network Access and Server Activity Logs;
- Project Management Application Data;
- Computer Aided Design/Drawing Files; and,
- Back Up and Archival Files (e.g., Zip, .GHO)

ESI resides not only in areas of electronic, magnetic and optical storage media reasonably accessible to you, but also in areas you may deem *not* reasonably accessible. You are obliged to *preserve* potentially relevant evidence from *both* these sources of ESI, even if you do not anticipate *producing* such ESI.

The demand that you preserve both accessible and inaccessible ESI is reasonable and necessary. Pursuant to amendments to the Federal Rules of Civil Procedure that have been approved by the United States Supreme Court (eff. 12/1/06), you must identify all sources of ESI you decline to produce and demonstrate to the court why such sources are not reasonably accessible. For good cause shown, the court may then order production of the ESI, even if it finds that it is not reasonably accessible. Accordingly, even ESI that you deem reasonably inaccessible *must be preserved in the interim* so as not to deprive the plaintiffs of their right to secure the evidence or the Court of its right to adjudicate the issue.

Preservation Requires Immediate Intervention
You must act immediately to preserve potentially relevant ESI including, without limitation, information with the *earlier* of a Created or Last Modified date on or after [**DATE**] through the date of this demand and concerning:

1. The events and causes of action described in [**Plaintiffs' Complaint**];
2. ESI you may use to support claims or defenses in this case;
3. ….
4. ….

Adequate preservation of ESI requires more than simply refraining from efforts to destroy or dispose of such evidence. You must also intervene to prevent loss due to routine operations and employ proper techniques and protocols suited to protection of ESI. *Be advised that sources of ESI are altered and erased by continued use of your computers and other devices.* Booting a drive, examining its contents or running any application will irretrievably alter the evidence it contains and may constitute unlawful spoliation of evidence. Consequently, alteration and erasure may result from your failure to act diligently and responsibly to prevent loss or corruption of ESI.

Nothing in this demand for preservation of ESI should be understood to diminish your concurrent obligation to preserve document, tangible things and other potentially relevant evidence.

Suspension of Routine Destruction
You are directed to immediately initiate a litigation hold for potentially relevant ESI, documents and tangible things, and to act diligently and in good faith to secure and audit compliance with such litigation hold. You are further directed to immediately identify and modify or suspend features of your information systems and devices that, in routine operation, operate to cause the loss of potentially relevant ESI. Examples of such features and operations include:

- Purging the contents of e-mail repositories by age, capacity or other criteria;
- Using data or media wiping, disposal, erasure or encryption utilities or devices;
- Overwriting, erasing, destroying or discarding back up media;
- Re-assigning, re-imaging or disposing of systems, servers, devices or media;
- Running antivirus or other programs effecting wholesale metadata alteration;

- Releasing or purging online storage repositories;
- Using metadata stripper utilities;
- Disabling server or IM logging; and,
- Executing drive or file defragmentation or compression programs.

Guard Against Deletion

You should anticipate that your employees, officers or others may seek to hide, destroy or alter ESI and act to prevent or guard against such actions. Especially where company machines have been used for Internet access or personal communications, you should anticipate that users may seek to delete or destroy information they regard as personal, confidential or embarrassing and, in so doing, may also delete or destroy potentially relevant ESI. This concern is not one unique to you or your employees and officers. It's simply an event that occurs with such regularity in electronic discovery efforts that any custodian of ESI and their counsel are obliged to anticipate and guard against its occurrence.

[OPTIONAL PARAGRAPHS]

Preservation by Imaging

You should take affirmative steps to prevent anyone with access to your data, systems and archives from seeking to modify, destroy or hide electronic evidence on network or local hard drives (such as by deleting or overwriting files, using data shredding and overwriting applications, defragmentation, re-imaging or replacing drives, encryption, compression, steganography or the like). With respect to local hard drives, one way to protect existing data on local hard drives is by the creation and authentication of a forensically qualified image of all sectors of the drive. Such a forensically qualified duplicate may also be called a bitstream image or clone of the drive. Be advised that a conventional back up of a hard drive is not a forensically qualified image because it only captures active, unlocked data files and fails to preserve forensically significant data that may exist in such areas as unallocated space, slack space and the swap file.

With respect to the hard drives and storage devices of each of the persons named below and of each person acting in the capacity or holding the job title named below, as well as each other person likely to have information pertaining to the instant action on their computer hard drive(s), demand is made that you immediately obtain, authenticate and preserve forensically qualified images of the hard drives in any computer system (including portable and home computers) used by that person during the period from ____ to _____, as well as recording and preserving the system time and date of each such computer.

[Insert names, job descriptions and titles here].

Once obtained, each such forensically qualified image should be labeled to identify the date of acquisition, the person or entity acquiring the image and the system and medium from which it was obtained. Each such image should be preserved without alteration.

Preservation in Native Form
You should anticipate that certain ESI, including but not limited to spreadsheets and databases, will be sought in the form or forms in which it is ordinarily maintained. Accordingly, you should preserve ESI in such native forms, and you should not select methods to preserve ESI that remove or degrade the ability to search your ESI by electronic means or make it difficult or burdensome to access or use the information efficiently in the litigation.

You should additionally refrain from actions that shift ESI from reasonably accessible media and forms to less accessible media and forms if the effect of such actions is to make such ESI not reasonably accessible

Metadata
You should further anticipate the need to disclose and produce system and application metadata and act to preserve it. System metadata is information describing the history and characteristics of other ESI. This information is typically associated with tracking or managing an electronic file and often includes data reflecting a file's name, size, custodian, location and dates of creation and last modification or access. Application metadata is information automatically included or embedded in electronic files but which may not be apparent to a user, including deleted content, draft language, commentary, collaboration and distribution data and dates of creation and printing. Be advised that metadata may be overwritten or corrupted by careless handling or improper steps to preserve ESI. For electronic mail, metadata includes all header routing data and Base 64 encoded attachment data, in addition to the To, From, Subject, Received Date, CC and BCC fields.

Servers
With respect to servers like those used to manage electronic mail (e.g., Microsoft Exchange, Lotus Domino) or network storage (often called a user's "network share"), the complete contents of each user's network share and e-mail account should be preserved. There are several ways to preserve the contents of a server depending upon, e.g., its RAID configuration and whether it can be downed or must be online 24/7. If you question whether the preservation method you pursue is one that we will accept as sufficient, please call to discuss it.

Home Systems, Laptops, Online Accounts and Other ESI Venues
Though we expect that you will act swiftly to preserve data on office workstations and servers, you should also determine if any home or portable systems may contain potentially relevant data. To the extent that officers, board members or employees have sent or received potentially relevant e-mails or created or reviewed potentially relevant documents away from the office, you must preserve the contents of systems, devices and media used for these purposes (including not only potentially relevant data from portable and home computers, but also from portable thumb drives, CD-R disks and the user's PDA, smart phone, voice mailbox or other forms of ESI storage.). Similarly, if employees, officers or board members used online or browser-based e-mail accounts or services (such as AOL, Gmail, Yahoo Mail or the like) to send or receive potentially relevant messages and attachments, the contents of these account mailboxes (including Sent, Deleted and Archived Message folders) should be preserved.

Ancillary Preservation
You must preserve documents and other tangible items that may be required to access, interpret or search potentially relevant ESI, including logs, control sheets, specifications, indices, naming protocols, file lists, network diagrams, flow charts, instruction sheets, data entry forms, abbreviation keys, user ID and password rosters or the like.

You must preserve any passwords, keys or other authenticators required to access encrypted files or run applications, along with the installation disks, user manuals and license keys for applications required to access the ESI.

You must preserve any cabling, drivers and hardware, other than a standard 3.5" floppy disk drive or standard CD or DVD optical disk drive, if needed to access or interpret media on which ESI is stored. This includes tape drives, bar code readers, Zip drives and other legacy or proprietary devices.

Paper Preservation of ESI is Inadequate
As hard copies do not preserve electronic searchability or metadata, they are not an adequate substitute for, or cumulative of, electronically stored versions. If information exists in both electronic and paper forms, you should preserve both forms.

Agents, Attorneys and Third Parties
Your preservation obligation extends beyond ESI in your care, possession or custody and includes ESI in the custody of others that is subject to your direction or control. Accordingly, you must notify any current or former agent, attorney, employee, custodian or contractor in possession of potentially relevant ESI to preserve such ESI to the full extent of your obligation to do so, and you must take reasonable steps to secure their compliance.

System Sequestration or Forensically Sound Imaging
We suggest that, with respect to [**NAME KEY PLAYERS**] removing their ESI systems, media and devices from service and properly sequestering and protecting them may be an appropriate and cost-effective preservation step.

In the event you deem it impractical to sequester systems, media and devices, we believe that the breadth of preservation required, coupled with the modest number of systems implicated, dictates that forensically sound imaging of the systems, media and devices is expedient and cost effective. As we anticipate the need for forensic examination of one or more of the systems and the presence of relevant evidence in forensically accessible areas of the drives, we demand that you employ forensically sound ESI preservation methods. Failure to use such methods poses a significant threat of spoliation and data loss.

By "forensically sound," we mean duplication, for purposes of preservation, of all data stored on the evidence media while employing a proper chain of custody and using tools and methods that make no changes to the evidence and support authentication of the duplicate as a true and complete bit-for-bit image of the original. A forensically sound preservation method guards against changes to metadata evidence and preserves all parts of the electronic evidence, including in the so-called "unallocated clusters," holding deleted files.

Preservation Protocols
We are desirous of working with you to agree upon an acceptable protocol for forensically sound preservation and can supply a suitable protocol, if you will furnish an inventory of the systems and media to be preserved. Else, if you will promptly disclose the preservation protocol you intend to employ, perhaps we can identify any points of disagreement and resolve them. A successful and compliant ESI preservation effort requires expertise. If you do not currently have such expertise at your disposal, we urge you to engage the services of an expert in electronic evidence and computer forensics. Perhaps our respective experts can work cooperatively to secure a balance between evidence preservation and burden that's fair to both sides and acceptable to the Court.

Do Not Delay Preservation
I'm available to discuss reasonable preservation steps; however, *you should not defer preservation steps pending such discussions if ESI may be lost or corrupted as a consequence of delay.* Should your failure to preserve potentially relevant evidence result in the corruption, loss or delay in production of evidence to which we are entitled, such failure would constitute spoliation of evidence, and we will not hesitate to seek sanctions.

Confirmation of Compliance
Please confirm by [**DATE**], that you have taken the steps outlined in this letter to preserve ESI and tangible documents potentially relevant to this action. If you have not undertaken the steps outlined above, or have taken other actions, please describe what you have done to preserve potentially relevant evidence.

Respectfully,

Musings on Electronic Discovery

"Ball in Your Court"
April 2005 – September 2007

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

© Craig Ball

Craig Ball is a Board Certified trial lawyer, certified computer forensic examiner and electronic evidence expert He has dedicated his career to teaching the bench and bar about forensic technology and trial tactics. After decades trying lawsuits, Craig now limits his practice solely to serving as a court-appointed special master and consultant in computer forensics and electronic discovery, and to publishing and lecturing on computer forensics, emerging technologies, digital persuasion and electronic discovery. Craig's award-winning e-discovery column, "Ball in Your Court," appears in Law Technology News. He has consulted or served as a testifying expert in computer forensics and electronic discovery in some of the most challenging and well-known cases in the U.S. Named as one of the Best Lawyers in America and a Texas Superlawyer, Craig is a recipient of the Presidents' Award, the State Bar of Texas' most esteemed recognition of service to the profession and of the Bar's Lifetime Achievement Award in Law and Technology.

A prolific contributor to continuing legal and professional education programs throughout the United States, Craig Ball has delivered over 450 presentations and papers. His articles on forensic technology and electronic discovery frequently appear in the national media, including in American Bar Association, ATLA and American Lawyer Media print and online publications. The presentation, "Craig Ball on PowerPoint" (a/k/a "PowerPersuasion") is consistently one of the top rated continuing legal education programs in the nation.

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 Board, 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
Special Master, Electronic Discovery, Federal and Harris County (Texas) District Courts
Instructor in Computer Forensics, 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

Faculty, Texas College of Trial Advocacy, 1992 and 1993
Adjunct Professor, South Texas College of Law, 1983-88
Listed in "Best Lawyers in America" and Selected as a "Texas Super Lawyer," 2003-2007
Rated AV by Martindale-Hubbell

The DNA of Data

by Craig Ball

[Originally published in Law Technology News, April 2005]

Discovery of electronic data compilations has been part of American litigation for two generations, during which time we've seen nearly all forms of information migrate to the digital realm. Statisticians posit that only five to seven percent of all information is "born" outside of a computer, and very little of the digitized information ever finds its way to paper. Yet, despite the central role of electronic information in our lives, electronic data discovery (EDD) efforts are either overlooked altogether or pursued in such epic proportions that discovery dethrones the merits as the focal point of the case. At each extreme, lawyers must bear some responsibility for the failure. Few of us have devoted sufficient effort to learning the technology, instead deluding ourselves that we can serve our clients by continuing to focus on the smallest, stalest fraction of the evidence: paper documents. When we do garner a little knowledge, we abuse it like the Sorcerer's Apprentice, by demanding production of "any and all" electronic data and insisting on preservation efforts sustainable only through operational paralysis. We didn't know how good we had it when discovery meant only paper.

However, electronic evidence isn't going away. It's growing…exponentially, and some electronic evidence items, like databases, spreadsheets, voice mail and video, bear increasingly less resemblance to paper documents. Proposed changes in the rules of procedure wending their way through the system require lawyers to discuss ways to preserve electronic evidence, select formats in which to produce it and manage volumes of information dwarfing the Library of Congress. Litigators must learn it or find a new line of work.

My goal for this column is to help make electronic discovery and computer forensics a little easier to understand, never forgetting that this is exciting, challenging—and very cool—stuff.

Accessible versus Inaccessible

You can't talk about EDD today without using the "Z" word: Zubulake (pronounced "zoo-boo-lake"). Judge Shira Scheindlin's opinions in *Zubulake v. UBS Warburg, L.L.C.,* 217 F.R.D. 309 (S.D.N.Y. 2003) triggered a whirlwind of discussion about EDD. Judge Scheindlin cited the "accessibility" of data as the threshold for determining issues of what must be produced and who must bear the cost of production. Accessible data must be preserved, processed and produced at the producing party's cost, while inaccessible data is available for good cause and may trigger cost shifting.

But what makes data "inaccessible?" Is it a function of the effort and cost required to make sense of the data? If so, do the boundaries shift with the skill and resources of the producing party such that ignorance is rewarded and knowledge penalized? To understand when data is truly inaccessible requires a brief look at the DNA of data.

Everything's Accessible

Computer data is simply a sequence of ones and zeroes. Data is only truly inaccessible when you can't read the ones and zeroes or figure out where the sequence starts. To better grasp this, imagine you had the unenviable responsibility of typing the complete works of Shakespeare on a machine with only two keys, "A" and "B," and if you fail, all the great works of the Bard

would be lost forever. As you ponder this seemingly impossible task, you'd figure out that you could encode the alphabet using sequences of As and Bs to represent each of the twenty-six capital letters, their lower case counterparts, punctuation and spaces. The uppercase "W" might be "ABABABBB" and the uppercase "S," "ABABAABB." Cumbersome, but feasible. Armed with the code and knowing where the sequence begins, a reader can painstakingly reconstruct every lovely foot of iambic pentameter.

This is just what a computer does when it stores data in ones and zeroes, except computers encode many "alphabets" and work with sequences billions of characters long. Computer data is only "gone" when the media that stores it is obliterated, overwritten or strongly encrypted without a key. This is true for all digital media, including back up tapes and hard drives. But, inaccessibility due to damage, overwriting or encryption is rarely raised as grounds for limiting e-discovery or shifting costs.

Just Another Word for Burdensome?

Frequently, lawyers will couch a claim of undue burden in terms of inaccessibility, arguing that it's too time-consuming or costly to restore the data. But, burden and inaccessibility are opposite sides of the same coin, and "inaccessibility" adds nothing to the mix but confusion. Arguing *both* burden and inaccessibility is two bites at the apple.

Worse, there is a risk in branding particular media as "inaccessible." Parties resisting discovery shouldn't be relieved of the obligation to demonstrate undue burden simply because evidence resides on a back up tape. We must be vigilant to avoid a reflexive calculus like:

All back up tapes are inaccessible

⇩

Inaccessible means undue burden presumed

⇩

Good cause showing required for production

⇩

Requesting party pays cost of conversion to "accessible" form.

Zubulake put EDD on every litigator's and corporate counsel's radar screen and proved invaluable as a provocateur of long-overdue debate about electronic discovery. Still, its accessibility analysis is not a helpful touchstone, especially in a fast-moving field like computing. Codifying it in proposed amendments to F.R.C.P. Rule 26(b)(2) would perpetuate a flawed standard. Even if that occurs, don't be cowed by the label, "inaccessible," and don't shy away from seeking discovery of relevant media just because it's cited as an example of something inaccessible. Instead, require the producing party to either show that the ones and zeroes can't be accessed or demonstrate that production entails an undue burden.

Unclear on the Concept

by Craig Ball

[Originally published in Law Technology News, May 2005]

A colleague buttonholed me at the American Bar Association's recent TechShow and asked if I'd visit with a company selling concept search software to electronic discovery vendors. Concept searching allows electronic documents to be found based on the ideas they contain instead of particular words. A concept search for "exploding gas tank" should also flag documents that address fuel-fed fires, defective filler tubes and the Ford Pinto. An effective concept search engine "learns" from the data it analyzes and applies its own language intelligence, allowing it to, e.g., recognize misspelled words and explore synonymous keywords. I said, "Sure," and was delivered into the hands of an earnest salesperson who explained that she was having trouble persuading courts and litigators that the company's concept search engine worked. How could they reach them and establish credibility? She extolled the virtues of their better mousetrap, including its ability to catch common errors, like typing "manger" when you mean "manager."

But when we tested the product against its own 100,000 document demo dataset, it didn't catch misspelled terms or search for synonyms. It couldn't tell "manger" from "manager." Phrases were hopeless. Worse, it didn't reveal its befuddlement. The program neither solicited clarification of the query nor offered any feedback revealing that it was clueless on the concept.

The chagrined company rep turned to her boss, who offered, "100,000 documents are not enough for it to really learn. The program only knows a word is misspelled when it sees it spelled both ways in the data it's examining and makes the connection."

The power of knowledge lies in using what's known to make sense of the unknown. If the software only learns what each dataset teaches it, it brings nothing to the party. Absent from the application was a basic lexicon of English usage, nothing as fundamental as Webster's Dictionary or Roget's Thesaurus. There was no vetting for common errors, no "fuzzy" searching or any reference foundation. The application was the digital equivalent of an idiot savant (and I'm taking the savant on faith because this application is the plumbing behind some major vendors' products).

Taking the Fifth?

In the Enron/Andersen litigation, I was fortunate to play a minor role for lead plaintiff's counsel as an expert monitoring the defendant's harvesting and preservation of electronic evidence. The digital evidence alone quickly topped 200 terabytes, far more information than if you digitized all the books in the Library of Congress. Printed out, the paper would reach from sea-to-shining sea several times. These gargantuan volumes — and increasingly those seen in routine matters — can't be examined without automated tools. There just aren't enough associates, contract lawyers and paralegals in the world to mount a manual review, nor the money to pay for it. Of necessity, lawyers are turning to software to divine relevancy and privilege.

But as the need for automated e-discovery tools grows, the risks in using them mount. It's been 20 years since the only study I've seen pitting human reviewers against search tools. Looking at

a (paltry by current standards) 350,000 page litigation database, the computerized searches turned up just 20 percent of the relevant documents found by the flesh-and-bone reviewers.

The needle-finding tools have improved, but the haystacks are much, much larger now. Are automated search tools performing well enough for us to use them as primary evidence harvesting tools?

Metrics for a Daubert World

Ask an e-discovery vendor about performance metrics and you're likely to draw either a blank look or trigger a tap dance that would make the late Ann Miller proud. How many e-discovery products have come to market without any objective testing demonstrating their efficacy? Where is the empirical data about how concept searching stacks up against human reviewers? How has each retrieval system performed against the National Institute of Standards and Technology text retrieval test collections?

If the vendor response is, "We've never tested our products against real people or government benchmarks," how are users going to persuade a judge it was a sound approach come the sanctions hearing?

We need to apply the same Daubert-style standards [*Daubert v. Merrell Dow Pharmaceuticals* (92-102) 509 U.S. 579 (1993)] to these systems that we would bring to bear against any other vector for junk science: Has it been rigorously tested? Peer-reviewed? What are the established error rates?

Calibration and Feedback

Like the airport security staff periodically passing contraband through the x-ray machines and metal detectors to check the personnel and equipment, automated search systems must be periodically tested against an evolving sample of evidence scrutinized by human intelligence. Without this ongoing calibration, the requesting party may persuade the court that your net's so full of holes, only a manual search will suffice. If that happens, what can you do but settle?

Thanks to two excellent teachers, I read Solzhenitsyn in seventh grade and Joyce Carol Oates in the ninth. I imagine that if I re-read those authors today, I'd get more from them than my adolescent sensibilities allowed. Likewise, if software gets smarter as it looks at greater and greater volumes of information, is there a mechanism to revisit data processed *before* the software acquired its "wisdom" lest it derive no more than my 11-year-old brain gleaned from One Day in the Life of Ivan Denisovitch? What is the feedback loop that ensures the connections forged by progress through the dataset apply to the entire dataset?

For example, in litigation about a failed software development project, the project team got into the habit of referring to the project amongst themselves as the "abyss" and the "tar baby." Searches for the insider lingo, as concepts or keywords, are likely to turn up e-mails confirming that the project team knowingly poured client monies into a dead end.

If the software doesn't make this connection until it processes the third wave of data, what about what it missed in waves one and two? Clearly, the way the data is harvested and staged impacts what is located and produced. Of course, this epiphany risk—not realizing what you saw until after you've reviewed a lot of stuff—afflicts human examiners too, along with fatigue, inattentiveness and sloth to which machines are immune.

But, we trust that a diligent human examiner will sense when a newly forged connection should prompt re-examination of material previously reviewed.

Will the software know to ask, "Hey, will you re-attach those hard drives you showed me yesterday? I've figured something out."

Concept Search Tools

Though judges and requesting parties must be wary of concept search tools absent proof of their reliability, even flawed search tools have their place in the trial lawyer's toolbox.

Concept searching helps overcome limitations of optical character recognition, where seeking a match to particular text may be frustrated by OCR's inability to read some fonts and formats. It also works as a lens through which to view the evidence in unfamiliar ways, see relationships that escaped notice and better understand your client's data universe while framing filtering strategies.

I admire the way EDD-savvy Laura Kibbe, in-house counsel for pharmaceutical giant Pfizer, Inc., uses concept searching. She understands the peril of using it to filter data and won't risk having to explain to the court how concept searching works and why it might overlook discoverable documents. Instead, Laura uses concept searching to brainstorm keywords for traditional word searches and then uses it again as a way to prioritize her review of harvested information.

For producing parties inclined to risk use of concept searching as a filtering tool, inviting the requesting party to contribute keywords and concepts for searching is an effective strategy to forestall finger pointing about non-production. The overwhelming volume and the limitations of the tools compel transformation of electronic discovery to a collaborative process. Working together, both sides can move the spotlight away from the process and back onto the merits of the case.

Cowboys and Cannibals

by Craig Ball

[Originally published in Law Technology News, June 2005]

With its quick-draw replies, flame wars, porn and spam, e-mail is the Wild West boom town on the frontier of electronic discovery--all barroom brawls, shoot-outs, bawdy houses and snake oil salesman. It's a lawless, anyone-can-strike-it-rich sort of place, but it's taking more-and-more digging and panning to get to the gold.

Folks, we need a new sheriff in town.

A Modest Proposal

E-mail distills most of the ills of e-discovery, among them massive unstructured volume, mixing of personal and business usage, wide-ranging attachment formats and commingled privileged and proprietary content. E-mail epitomizes "everywhere" evidence. It's on the desktop hard drive, the server, back up tapes, home computer, laptop on the road, Internet service provider, cell phone and personal digital assistant. Stampede!

There's more to electronic data discovery than e-mail, but were we to figure out how to simply and cost-effectively round up, review and produce all that maverick e-mail, wouldn't we lick EDD's biggest problem?

The e-mail sheriff I envision is a box that pops up when you hit send and requires designation of the e-mail as personal or business-related. If personal, it's sent and a copy is immediately forwarded to your personal e-mail account. The personal message is then purged from the enterprise system. If business related, you must assign the message to its proper place within the organization's data structure. If you don't put it where it belongs, the system won't send it. Tough love for a wired world. On the receiving end, when you seek to close an e-mail you've read, you're likewise prompted to file it within your organization's data structure, deciding if it's personal or business and where it belongs.

When I first broached this idea to my e-discovery colleagues, the response was uniformly dismissive: "Our people wouldn't do it" being the common reply. Hogwash! They'll do it if they have to do it. They'll do it if there's a carrot and a stick. They'll do it if the management system is designed well and implemented aggressively. I ask them, "Why do you make employees punch in a code to use the photocopier, but require no accountability for e-mail that may sink the company?"

Some claim, "Our people will just call everything personal or file all business correspondence as 'office general.'" Possibly, but that means that business data will be notable by its absence from its proper place. Eventually, the boss will say, "Dammit Dusty, why can't you keep up with your e-filing?" In addition, Dusty won't want the system to report that he characterizes 95% of the at-work electronic communications he handles each day as personal in nature. Certainly, there needs to be audit and oversight, and the harder you make it to for a user to punt or evade the system, the better the outcome. This model worked for paper. It can work for e-mail.

Once, a discovery request sent a file clerk scurrying to a file room set aside for orderly information storage. There, the clerk sought a labeled drawer or box and the labeled folders within. He didn't search every drawer, box or folder, but went only to the places where the company kept items responsive to the request. From cradle to grave, paper had its place, tracked by standardized, compulsory practices. Correspondence was dated and its contents or relevance described just below the date. Files bore labels and were sorted and aggregated within a structure that generally made sense to all who accessed them. These practices enabled a responding party to affirm that discovery was complete on the strength of the fact that they'd looked in all the places where responsive items were kept.

By contrast, the subject lines of e-mails may bear no relation to the contents or be omitted altogether. There is no taxonomy for data. Folder structures are absent, ignored or unique to each user. Most users' e-mail management is tantamount to dumping all their business, personal and junk correspondence into a wagon hoping the Google cavalry will ride to the rescue. The notion "keep everything and technology will help you find it" is as seductive as a dance hall floozy…and just as treacherous.

E-discovery is not more difficult and costly than paper discovery simply because of the sheer volume of data or even the variety of formats and repositories. Those concerns are secondary to the burdens occasioned by the lack of electronic records management. We could cope with the volume if it were structured because we could rely on that structure to limit our examination to manageable chunks. Satirist Jonathan Swift was deadly humorous when, in his 1729 essay, "A Modest Proposal," he suggested the Irish eat their children to solve a host of societal ills, but I'm deadly serious when I modestly propose we swallow our reluctance and impose order on enterprise e-mail. The payback is genuine and immediate. Tame the e-mail bronco and the rest of the herd will fall in line.

Does imposing structure on electronic information erase the advantages of information technology? Is it horse-and-buggy thinking in a jet age? No, but it's has its costs. One is speed. If the sender or recipient of an e-mail is obliged to think about where any communication fits within their information hierarchy and designate a "location," that means the user has to pause, think and act. They can't just expectorate a message and hit send. Dare we re-introduce deliberation to communication? The gun-slinging plaintiff's lawyer in me will miss the unvarnished, *res gestae* character of unstructured e-mail, but in the end, we can do with a little law west of the Pecos.

Give Away Your Computer

by Craig Ball

[Originally published in Law Technology News, July 2005]

With the price of powerful computer systems at historic lows, who isn't tempted to upgrade? But, what do you do with a system you've been using if it's less than four or five-years old and still has some life left in it? Pass it on to a friend or family member or donate it to a school or civic organization and you're ethically obliged to safeguard client data on the hard drive. Plus, you'll want to protect your personal data from identity thieves and snoopers. Hopefully you already know that deleting confidential files and even formatting the drive does little to erase your private information—it's like tearing out the table of contents but leaving the rest of the book. How do you be a Good Samaritan without jeopardizing client confidences and personal privacy?

Options

One answer: replace the hard drive with a new one before you donate the old machine. Hard drives have never been cheaper, and adding the old hard drive as extra storage in your new machine ensures easy access to your legacy data. But, it also means going out-of-pocket and some surgery inside both machines—not everyone's cup of tea.

Alternatively, you could remove or destroy the old hard drive, but those accepting older computers rarely have the budget to buy hard drives, let alone the technician time to get donated machines running. Donated systems need to be largely complete and ready to roll.

Probably the best compromise is to wipe the hard drive completely and donate the system recovery disk along with the system. Notwithstanding some largely theoretical notions, once you overwrite every sector of your hard drive with zeros or random characters, your data is gone forever. The Department of Defense recommends several passes of different characters, but just a single pass of zeros is enough to frustrate all computer forensic data recovery techniques in common use.

Free is Good

You can *buy* programs to overwrite your hard drive, but why do so? Effective erasure tools are available as free downloads from the major hard drive manufacturers, and most work on other manufacturers' drives. Western Digital offers its Data Lifeguard Diagnostic Tool at http://support.wdc.com/download. Seagate's DiscWizard Starter Edition is found at www.seagate.com/support/disc/drivers/discwiz.html and Maxtor's PowerMax utilities is found by drilling down from www.maxtor.com/support. DBAN (for Darik's Boot and Nuke), a free Linux program, will also obliterate all data on a Windows system and is available at http://dban.sourceforge.net/. Each application offers bells-and-whistles, but all you're seeking is the ability to create a boot floppy that can write zeroes to the hard drive. If your system has no floppy drive, each site also offers a boot CD image download.

Why a boot floppy or CD? Because no wiping program running under Windows can erase all of the data on a Windows drive. Running under DOS (or, in the case of DBAN, Linux) insures that no file is locked out to the wiping utility while it does its job. To this end, check to be sure that whatever wiping application you select "sees" the entire hard drive. If it only recognizes, say,

the first 32 GB of a 40 GB drive, check your settings or use a different utility. Fortunately, these utilities are user-friendly and report what they see and do.

Careful!
Wiping every sector on a hard drive is a time consuming process. Allow hours of (largely) unattended operation to get the job done, and if it's an option, be sure to select a full overwrite (or "low level format") and not a quick version. There are no shortcuts to overwriting every sector to sterilize a drive. Check to be sure there is only one hard drive in the system. If multiple drives are present, wipe each of them. Above all, understand that *there is no turning back from this kind of data erasure*. No Recycle Bins. No Undo command. No clean room magic. Be absolutely certain you have another working copy of anything you mean to keep.

An Important Courtesy
When you sterilize a drive, your privileged data obliterated along with the operating system and all applications. A wiped drive can't boot a computer, but can return to service if you remember to donate the system restore disk with the hardware. For computers lacking restore disks, supply the operating system installation disk and any application disks you wish to donate. As long as you're not continuing to use the same applications loaded from the same disks (or copies) on your new machine, your end user license is likely to be freely transferable. If the donated system came without disks, you or your recipient will need to contact the manufacturer and request a restore disk. If, as is often the case in larger firms, the operating systems are site licensed, it may be a violation of that license to share them. Your recipient will then need to purchase their own license or seek out someone who'll donate an operating system. School districts typically have their own site licenses.

Dodging Blasts from The Past
Be sure to caution your recipient that it's very important to promptly download critical security patches and service packs for the restored operating system and applications. A restored machine is like a step back in time to when many now-closed security holes were wide open, so the recipient needs to slam these vulnerabilities shut at the very first connection to the Internet.

Help for the Helper
Worries about data security needn't keep you from helping others by donating your used computer. For additional guidance, contact TechSoup (www.techsoup.org) or the National Cristina Foundation (www.cristina.org), and seek out — or organize — the computer donation program in your community.

Breaking News:
Clearing your donated, sold or discarded hard drives of sensitive information isn't just good practice, it's now also required by law. Effective June 1, 2005, the Federal Trade Commission's Disposal Rule 16 CFR Part 682, requires businesses—including lawyers and law firms—to take reasonable measures to dispose of sensitive information derived from credit reports and background checks so that the information cannot practicably be read or reconstructed. The Rule, which applies to both paper and digital media, requires implementing and monitoring compliance with disposal policies and procedures for this information. Comments to the rule suggest using disc wiping utilities, but also offer that electronic media may be economically disposed of by "simply smashing the material with a hammer." Sounds like a great stress reliever, but don't forget your safety goggles!

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.

Yours, Mine and Ouch!

by Craig Ball

[Originally published in Law Technology News, September 2005]

When star performer Sarit Shmueli was fired from her real estate agent job with The Corcoran Group, she returned to her desk to find that she'd been locked out of the company computer system. Shmueli was barred from retrieving virtual property, in particular, her client list. When demands for her data were unavailing, Shmueli sued Corcoran for three million dollars.

Though opinions vary on whether to drop the axe on Friday or Monday, human resource experts agree that immediate steps must be taken to block the fired individual's access to company computers. That's wise, considering more than a few heading to the door have trashed important files or attempted to sabotage entire networks. But, what about personal computer data? Fired workers are routinely furnished a cardboard box and the supervised opportunity to collect personal belongings before being escorted to the door. Is denying access to personal data stored on a company computer a violation of the discharged party's property rights?

Overruling a motion for summary judgment, a recent decision in Sarit Shmueli's case holds that when The Corcoran Group blocked Shmueli from accessing her records on the company computer system, Corcoran may have been guilty of conversion. "There should be no reason why [a] practical view should not apply equally to the present generation of documents—electronic documents—which are just as vulnerable to theft and wrongful transfer as paper documents, if not more so," reasoned New York Supreme Court Justice Herman Cahn. *Shmueli v. The Corcoran Group*, 104824/03 (N.Y.S.Ct. July 25, 2005) online at http://decisions.courts.state.ny.us/fcas/FCAS_docs/2005JUL/30010482420033SCIV.PDF.

Of course, Justice Cahn is right to ascribe value to electronic documents, but what are the ramifications of affording a conversion cause of action to employees and contractors against companies that refuse to recognize personal property interests in data stored on their systems?

The bulk of my work as a special master in computer forensics revolves around employees who've allegedly purloined company data to build rival businesses. If the employees haven't already jumped ship, they're canned as soon as the boss realizes they're playing for the other team. Then, much time and money goes into assessing what information was taken and what tracks were covered. Initial reports of the Shmueli decision made me wonder if workers being shown the door might now be entitled to access the company network for the purpose of copying information they deem to be their own property. Are they perhaps at liberty to delete such "personal" items, too? What of the longstanding notion that anything stored on the company computer belongs to the company?

The Order makes clear that Shmueli's unique employment status as independent contractor and not an employee played a decisive role in the court's view that Corcoran may be liable for converting its former agent's digital property just as if the data had been printed to paper. Though Corcoran asserted its ownership of the computer trumped the plaintiff's rights, the court countered that because Shmueli worked with Corcoran and not as an employee of Corcoran, the computer was "licensed" for plaintiff's use to facilitate the independent contract.

The problem with the court's distinction is that employers enjoy no privilege to confiscate and convert personal property. Just because you have your spouse's photo on your desk at work doesn't mean the boss can keep it when you're fired. Instead, the employer's right to retain the information on the computer must be grounded on a presumption that information stored on an employer's computer is either company property in the first instance or acquires that character by virtue of being the fruit of an employee's labors.

The commingling of personal and business property on company computers is a growing concern. From personal e-mail to screenplays written on the lunch hour, employers should anticipate the obligation to identify, segregate and return personal data "belonging" to fired employees. Likewise, employees need to tailor their personal use of company systems to the possibility of lock out.

This is going to be harder than it sounds because employers aren't disposed to grant "computer visitation rights" after a firing to let former employees—particularly ones now working for the competition—clean out their digital lockers. But without such access, how well could any of us identify from memory all personal items on our office computers, and what fired employee wants a former boss or co-worker sifting through their personal information?

A well-drafted acceptable use policy signed by the employee helps by defining rights and responsibilities in the use of business computer systems during employment. However, if employers are lax in their enforcement of the policy such that employees harbor a reasonable expectation of privacy in their use of company systems, terminations may entail the added ugliness of a custody battle over data and potential liability for conversion. Even where a violation of an AUP is clear, will courts decide that personal data escheats to a former employer simply because it's found on their virtual premises? That is, which schoolyard canon will prevail: the altruistic "If it's not yours, give it back," or the Draconian "You shouldn't have brought it here, so now it's mine?"

The Path to E-Mail Production

(Part I of IV)

by Craig Ball

[Originally published in Law Technology News, October 2005]

Asked, "Is sex dirty," Woody Allen quipped, "Only if it's done right." That's electronic discovery: if it's ridiculously expensive, enormously complicated and everyone's lost sight of the merits of the case, you're probably doing it right.

But it doesn't have to be that way. Over the next few issues, we'll walk a path to production of e-mail — perhaps the trickiest undertaking in EDD. The course we take may not be the shortest or easiest, but that's not the point. We're trying to avoid stepping off a cliff. Not every point is suited to every production effort, but all deserve consideration.

Think Ahead

EDD missteps are painfully expensive, or even unredeemable, if data is lost. Establish expectations at the outset.

Will the data produced:

- Integrate paper and electronic evidence?
- Be electronically searchable?
- Preserve all relevant metadata from the host environment?
- Be viewable and searchable using a single application, such as a web browser?
- Lend itself to Bates numbering?
- Be easily authenticable for admission into evidence?

Meeting these expectations hinges on what you collect along the way through identification, preservation, harvest and population.

Identification

"Where's the e-mail?" It's a simple question, but one answered too simply—and erroneously—by, "It's on the e-mail server" or "The last 60 days of mail is on the server and the rest is purged." Certainly, some e-mail will reside on the server, but most e-mail is elsewhere, and it's never all gone, notwithstanding retention policies. The true location and extent of e-mail depends on systems configuration, user habits, back up procedures and other hardware, software and behavioral factors. This is true for mom-and-pop shops, for large enterprises and for everything in-between.

Consider a recent case where I was asked to assess whether a departing associate stole files and diverted cases. The firm used a Microsoft Exchange e-mail server, so I could have collected or searched the associate's e-mail there. Had I looked only at the server, I would've missed the Hotmail traffic in the temporary internet files folder and the short message service (SMS) exchanges in the PDA synchronization files. Or the Microsoft Outlook archive file (.pst) and offline synchronization file (.ost), both stored on a laptop hard drive, and holding thousands more e-mails.

Just looking at the server wouldn't have revealed the stolen data or the diverted business — searching elsewhere uncovered a treasure trove of damning evidence.

E-mail resides in some or all of the following venues, grouped according to relative accessibility:

Easily Accessible:
• Online e-mail residing in active files on enterprise servers: MS Exchange e.g., (.edb, .stm, .log files), Lotus Notes (.nsf files), Novell GroupWise (.db files)
• E-mail stored in active files on local or external hard drives and network shares: User workstation hard drives (e.g., .pst, .ost files for Outlook and .nsf for Lotus Notes), laptops, "local" e-mail data files stored on networked file servers, mobile devices, and home systems, particularly those with remote access to networks.
• Nearline e-mail: Optical "juke box" devices, backups of user e-mail folders.
• Offline e-mail stored in networked repositories: e.g., Zantaz EAS, EMC EmailXtender, Waterford MailMeter Forensic, etc.

Accessible, but Often Overlooked:
• E-mail residing on remote servers: ISPs (IMAP, POP, HTTP servers), Gmail, Yahoo Mail, Hotmail, etc.
• E-mail forwarded and cc'd to third-party systems: Employee forwards e-mail to self at personal e-mail account.
• E-mail threaded behind subsequent exchanges: Contents diverge from earlier exchanges lodged in body of e-mail.
• Offline local e-mail stored on removable media: External hard drives, thumb drives and memory cards, optical media: CD-R/RW, DVD-R/RW, floppy drives, zip drives.
• Archived e-mail: Auto-archived or saved under user-selected filename.
• Common user "flubs": Users experimenting with export features unwittingly create e-mail archives.
• Legacy e-mail: Users migrate from e-mail clients "abandoning" former e-mail stores.
• E-mail saved to other formats: PDF, .tiff, .txt, .eml, etc.
• E-mail contained in review sets assembled for other litigation/compliance purposes.
• E-mail retained by vendors or third- parties (e.g., former service provider.)
• Print outs to paper.

More Difficult to Access:
• Offline e-mail on server back up media: Back up tapes (e.g., DLT, AIT)
• E-mail in forensically accessible areas of local hard drives: Deleted e-mail, internet cache, unallocated clusters.

The issues in the case, key players, relevant times, agreements between the parties and orders of the court determine the extent to which locations must be examined; however, the failure to identify all relevant e-mail carries such peril that caution should be the watchword. Isn't it wiser to invest more to know exactly what the client has than concede at the sanctions hearing the client failed to preserve and produce evidence it didn't know it had because no one bothered to look for it?

Electronic evidence is fragile and ever changing, so once you've found the e-mail evidence, you must guard against its loss or corruption.

Next month, we'll walk through preservation thicket.

The Path to Production: Retention Policies That Work

(Part II of IV)

by Craig Ball

[Originally published in Law Technology News, November 2005]

In this second in a series, we continue down the path to production of electronic mail. Last month, I reminded you to look beyond the e-mail server to the many other places e-mail hides. Now, having identified the evidence, we're obliged to protect it from deletion, alteration and corruption.

Preservation

Anticipation of a claim is all that's required to trigger a duty to preserve potentially relevant evidence, including fragile, ever-changing electronic data. Preservation allows backtracking on the path to production, but fail to preserve evidence and you've burned your bridges.

Complicating our preservation effort is the autonomy afforded e-mail users. They create quirky folder structures, commingle personal and business communications and — most dangerous of all — control deletion and retention of messages.

Best practices dictate that we instruct e-mail custodians to retain potentially relevant messages and that we regularly convey to them sufficient information to assess relevance in a consistent manner. In real life, hold directives alone are insufficient. Users find it irresistibly easy to delete data, so anticipate human frailty and act to protect evidence from spoliation at the hands of those inclined to destroy it. Don't leave the fox guarding the henhouse.

Consider the following as parts of an effective e-mail preservation effort:

- Litigation hold notices to custodians, including clear, practical and specific retention directives. Notices should remind custodians of relevant places where e-mail resides, but not serve as a blueprint for destruction. Be sure to provide for notification to new hires and collection from departing employees.
- Suspension of retention policies that call for purging e-mail.
- Suspension of re-use (rotation) of back up media containing e-mail.
- Suspension of hardware and software changes which make e-mail inaccessible.
- Replacing back up systems without retaining the means to read older media.

- Re-tasking or re-imaging systems for new users.
- Selling, giving away or otherwise disposing of systems and media.
- Preventing custodians from deleting/ altering/corrupting e-mail.
- Immediate and periodic "snapshots" of relevant e-mail accounts.
- Modifying user privileges settings on local systems and networks.
- Archival by auto-forwarding selected e-mail traffic to protected storage.
- Restricting activity like moving or copying files tending to irreparably alter file metadata.
- Packet capture of Instant Messaging (traffic or effective enforcement of IM prohibition.

- Preserve potential for forensic recovery.
- Imaging of key hard drives or sequestering systems.
- Suspension of defragmentation.
- Barring wiping software and encryption, with audit and enforcement.

Threshold issue
A threshold preservation issue is whether there is a duty of preservation going forward, e.g., with respect to information created during the pendency of the action. If not, timely harvest of data, imaging of drives and culling of relevant back ups from rotation may sufficiently meet the preservation duty so as to allow machines to be re-tasked, systems upgraded and back up tape rotation re-initiated. Securing guidance from the court and cooperating with opposing counsel to fashion practical preservation orders help insulate a producing party from subsequent claims of spoliation.

The Knowledge Hurdle
Thanks to a string of recent, high profile decisions, litigants are gradually awakening to their obligation to preserve electronic evidence. Still, attitudes often range from insufficient ("We'll just stop rotating back up tapes") to incredulous ("Why would we need to preserve voice mail?").

One hurdle is the lack of knowledge on the part of those charged with the responsibility to design and direct preservation efforts: too many don't understand what and how data change or what triggers those changes. They fail to appreciate how the pieces fit together.

For example, in a lawsuit concerning a plant explosion, the defendant, a major oil company, preserved monthly "full" back ups of its e-mail server but failed to hang on to four weeks of incremental back ups immediately preceding the blast.

A full back up is a snapshot of the e-mail system at a single point in time. An incremental back up records changes to the e-mail system between snapshots. Did someone think that full back ups were cumulative of the incremental sessions? If so, they missed the fact that any e-mail received and deleted between snapshots might exist on the incremental back ups but be absent from the monthly tapes. They didn't consider how the pieces fit together.

If you've done a good job identifying where e-mail lives, preservation is largely a matter of duplicating the e-mail without metadata corruption or shielding it from subsequent loss or alteration. Both demand technical competence, so you'll need expert help the first time or two. If you ask questions and seek out reasons behind actions, knowledge gained from one effort will guide you through the next.

Minimize Burden and Cost
With digital storage costs at all time lows, it's tempting to minimize spoliation risks by simply keeping everything. ***Don't***. Keeping everything merely postpones and magnifies the cost and complexity of production. Yet, you can suspend document retention and tape rotation without triggering a costly data logjam, if you adapt your preservation from reflexive to responsive.

Reflexive preservation describes steps you take while figuring out what's relevant and what's not. It's immediate and encompassing action to preserve the status quo while you sift the facts,

forge agreements with opponents or seek guidance from the court. Calling a halt to back up tape rotation or suspending retention policies is reflexive preservation.

Reflexive preservation is a triage mechanism and a proper first response; but it's too expensive and disruptive for the long haul. Instead, convert reflexive preservation to responsive preservation by continually tweaking your preservation effort to retain only what's relevant to claims or necessary to meet business and regulatory obligations. Narrow the scope of preservation by agreement, motion practice and sound, defensible judgment.

Having identified the e-mail evidence and preserved it, we need to collect it and make it accessible for review and searching. Next month, we hike up harvest hill and perambulate population pass. Wear sensible shoes!

The Path to Production: Harvest and Population

(Part III of IV)
by Craig Ball

[Originally published in Law Technology News, December 2005]

On the path to production, we've explored e-mail's back alleys and trod the mean streets of the data preservation warehouse district. Now, let's head to the heartland for harvest time. It's data harvest time.

After attorney review, data harvest is byte-for-byte the costliest phase of electronic data discovery. Scouring servers, local hard drives and portable media to gather files and metadata is an undertaking no company wants to repeat because of poor planning.

The Harvest

Harvesting data demands a threshold decision: Do you collect all potentially relevant files, then sift for responsive material, or do you separate the wheat from the chaff in the field, collecting only what reviewers deem responsive? When a corporate defendant asks employees to segregate responsive e-mail, (or a paralegal goes from machine-to-machine or account-to-account selecting messages), the result's are "field filtered."

Field filtering holds down cost by reducing the volume for attorney review, but it increases the risk of repeating the collection effort, loss or corruption of evidence and inconsistent selections. If keyword or concept searches alone are used to field filter data, the risk of under-inclusive production skyrockets.

Initially more expensive, comprehensive harvesting (unfiltered but defined by business unit, locale, custodian, system or medium), saves money when new requests and issues arise. A comprehensive collection can be searched repeatedly at little incremental expense, and broad preservation serves as a hedge against spoliation sanctions. Companies embroiled in serial litigation or compliance production benefit most from comprehensive collection strategies.

A trained reviewer "picks up the lingo" as review proceeds, but a requesting party can't frame effective keyword searches without knowing the argot of the opposition. Strategically, a producing party requires an opponent to furnish a list of search terms for field filtering and seeks to impose a "one list, one search" restriction. The party seeking discovery must either accept inadequate production or force the producing party back to the well, possibly at the requesting party's cost.

Chain of Custody

Any harvest method must protect evidentiary integrity. A competent chain of custody tracks the origins of e-evidence by, e.g., system, custodian, folder, file and dates. There's more to e-mail than what you see on screen, so it's wise to preempt attacks on authenticity by preserving complete headers and encoded attachments.

Be prepared to demonstrate that no one tampered with the data between the time of harvest and its use in court. Custodial testimony concerning handling and storage may suffice, but better approaches employ cryptographic hashing of data — "digital fingerprinting" — to prove nothing has changed.

Metadata

There's more to an e-mail than its contents: there's metadata, too. Each e-mail is tracked and indexed by the e-mail client ("application metadata") and every file holding e-mail is tracked and indexed by the computer's file system ("system metadata"). E-mail metadata is important evidence in its own right, helping to establish whether and when a message was received, read, forwarded, changed or deleted. Metadata's evidentiary significance garnered scant attention until Williams v. Sprint, 2005 W.L. 2401626 (D. Kan. Sept. 29, 2005), where in a dispute over production of spreadsheets, the court held that a party required to produce electronic documents as kept in the ordinary course of business must produce metadata absent objection, agreement or protective order.

System metadata is particularly fragile. Just copying a file from one location to another alters the file's metadata, potentially destroying critical evidence. Ideally, your data harvest shouldn't corrupt metadata, but if it may, archive the metadata beforehand. Though unwieldy, a spreadsheet reflecting original metadata is preferable to spoliation. EDD and computer forensics experts can recommend approaches to resolve these and other data harvest issues.

Processing and Population

However scrupulous your e-mail harvest, what you've reaped isn't ready to be text searched. It's a mish-mash of incompatible formats on different media: database files from Microsoft Exchange or Lotus Domino Servers, .PST and .NSF files copied from local hard drives, HTML fragments of browser-based e-mail and .PDF or .tiff images. Locked, encrypted and compressed, it's not text, so keyword searches fail.

Before search tools or reviewers can do their jobs, harvested data must be processed to populate the review set, i.e., deciphered and reconstituted as words by opening password-protected items, decrypting and decompressing container files and running optical character recognition on image files. Searching now will work, but it'll be slow going thanks to the large volume of duplicate items. Fortunately, there's a fix for that, too.

Next month: de-duplication, deliverables, documentation and the destination on the path to production.

The Path to Production: Are We There Yet?

(Part IV of IV)

by Craig Ball

[Originally published in Law Technology News, January 2006]

The e-mail's assembled and accessible. You could begin review immediately, but unless your client has money to burn, there's more to do before diving in: de-duplication. When Marge e-mails Homer, Bart and Lisa, Homer's "Reply to All" goes in both Homer's Sent Items and Inbox folders, and in Marge's, Bart's and Lisa's Inboxes. Reviewing Homer's response five times is wasteful and sets the stage for conflicting relevance and privilege decisions.

Duplication problems compound when e-mail is restored from backup tape. Each tape is a snapshot of e-mail at a moment in time. Because few users purge mailboxes month-to-month, one month's snapshot holds nearly the same e-mail as the next. Restore a year of e-mail from monthly backups, and identical messages multiply like rabbits.

De-Duplication

De-duplication uses metadata, cryptographic hashing or both to exclude identical messages. De-duplication may be implemented vertically, within a single mailbox or custodian, and horizontally, across multiple mailboxes and custodians. When questioning or prepping a witness, you'll want to see all relevant messages in the witness' mailbox, not just unique messages; so track and log de-duplication to facilitate re-population of duplicated items. De-duplication works best when unique messages and de-duplication logs merge in a database, allowing a reviewer to reconstruct mailboxes.

Be wary of "horizontal" de-duplication when discovery strategies change. An e-mail sent to dozens of recipients de-duplicated from all but one custodian's mailbox may be lost forever if the one custodian's e-mail ends up not being produced.

Review Tools

Rather than plow through zillions of e-mails for responsive and privileged items, reviewers often turn to keyword or concept search tools. Automated search tools make short work of objective requests for "all e-mail between Simpson and Burns," but may choke on "all e-mail concerning plant safety." To frame effective keyword searches, you have to know the lingo describing events and objects central to the case. Even then, crucial communiqués like, "My lips are sealed" or "Excellent" may be missed.

Are tireless black box tools an adequate substitute for human review? The jury's still out. In a seminal study, keyword searching fared poorly, finding only about one-fifth of relevant items identified by human reviewers. However, litigation management consultant Anne Kershaw looked at an advanced search tool and found machines performed almost twice as well as humans. The safest course is to arm conscientious, well-trained reviewers with state-of-the-art search tools and work cooperatively with opposing counsel to frame searches. Even then, examine the mailboxes of key witnesses, message-by-message.

Redaction

Paper redaction was easy: We concealed privileged text using a black marker and photocopied. It's trickier to eradicate privileged and confidential information at the data layer of document image files and within encoded attachments and metadata. Run your approach by an expert.

Re-population
For production, should you re-populate to restore relevant, non-privileged items previously de-duplicated, or will the other side accept a de-duplication log? Never produce de-duplicated e-mail without memorializing that opposing counsel knows of the de-duplication and waives re-population.

Deliverables
There isn't just one "right" media or format for deliverables. Options for production media include network transmittal, external hard drives, optical disks, tape, online repositories and hard copies. Formats range from native (.pst), exported (.eml), text (.txt), load files (Concordance, Summation), image files with or without data layers (.pdf, .tiff) and delimited files. Evidence ill suited to .tiff production (databases, some spreadsheets, etc.), compels native production.

Documentation
Inevitably, something will be overlooked or lost, but sanctions need not follow every failure. Document diligence throughout the discovery effort and be prepared to demonstrate why bad decisions were sound at the time and under the circumstances. Note where the client looked for responsive information, what was found, how much time and money was expended, what was sidelined and why. Avoid sanctions by proving good faith.

Are We There Yet?
The path to production is a long and winding road, but it's heading in the right direction. Knowing how to manage electronic evidence is as vital to trial practice as the ability to draft pleadings or question witnesses. Don't forget what happened on Main Street when they built the Interstate. Paper discovery's the old road. E-discovery's the Interstate.

Locard's Principle

by Craig Ball

[Originally published in Law Technology News, February 2006]

Devoted viewers of the TV show "CSI" know about Locard's Exchange Principle: the theory that anyone entering a crime scene leaves something behind or takes something away. It's called cross-transference, and though it brings to mind fingerprints, fibers and DNA, it applies to electronic evidence, too. The personal computer is Grand Central Station for PDAs, thumb drives, MP3 players, CDs, floppies, printers, scanners and a bevy of other gadgets. Few systems exist in isolation from networks and the Internet. When these connections are used for monkey business like stealing proprietary data, the electronic evidence left behind or carried away can tell a compelling story.

Recently, a colleague owning a very successful business called about an employee who'd quit to start a competing firm. My colleague worried that years of collected forms, research and other proprietary data might have gone out the door, too. The departing employee swore he'd taken nothing, but the unconvinced boss needed reassurance that someone he trusted hadn't betrayed him. He asked me to examine Mr. Not Me's laptop.

Turning to a forensic specialist was a smart move. Had the boss yielded to temptation and poked around the laptop, Locard's Principle dictates he would have irretrievably contaminated the digital crime scene. Last access dates would change. Log entries would be overwritten. Some deleted data might disappear forever. More to the point, an unskilled examiner would have overlooked the wealth of cross-transference evidence painting a vivid picture of theft and duplicity.

Stolen data has to be accessed, copied and then find its way out of the machine. Whether it's sent to a printer, e-mailed, burned to optical disk, written to a floppy or spirited away on a thumb drive, each conduit carries data away and leaves data behind as evidence of the transaction.

Forensic analysis of the employee's laptop turned up many examples of Locard's Principle at work. Windows employs a complex database called the Registry to track preferences and activities of the operating system and installed applications. When a USB storage device like a thumb drive connects, however briefly, to a Windows computer, the operating system interrogates the attachment and dutifully records information about the device and the date in the Registry. A moment-by-moment analysis of every file accessed shortly before the employee's departure and of the Registry revealed attachment of a thumb drive—an event reinforced by the system accessing the sound file played when a device attaches to a USB port. "Bonk-bink." This immediately preceded access to many proprietary files on the network, concluding with the system accessing the sound file signaling removal of the USB device. "Bink-bonk."

Further examination showed access to other proprietary data in conjunction with use of the system driver that writes data to recordable CDs. This evidence, along with an error log file created by a CD burning application detailing the date and time of difficulty encountered trying to burn particular proprietary files to CD-R, left no doubt as to what had transpired.

The coup de grace demonstrating the premeditated nature of the theft emerged from a review of files used to synchronize the laptop with a “smart phone” PDA. These held records of cell phone text messaging between the employee and a confederate in the firm discussing what files needed to be spirited away. Though the messages weren’t created on or sent via the laptop, they transferred to the laptop’s hard drive unbeknownst to the employee when he synched his PDA. Armed with this evidence, the boss confronted the still-employed confederate, who tearfully confessed all to the sadder-but-wiser employer. Case closed, but no happy ending.

Computers, like crime scenes, have stories to tell. Data and metadata in their registries, logs, link files and abandoned storage serve as Greek chorus to the tragedy or comedy of the user’s electronic life. Most cases don’t require the “CSI” treatment, but when the computer takes center stage, don’t overlook the potential for computer forensic analysis—and Dr. Locard’s Exchange Principle--to wring decisive evidence from the machine.

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.

Data Recovery: Lessons from Katrina

by Craig Ball

[Originally published in Law Technology News, April 2006]

When the sea reclaimed New Orleans and much of the Gulf Coast, hundreds of lawyers saw their computers and networks submerged. Rebuilding law practices entailed Herculean efforts to resurrect critical data stored on the hard drives in sodden machines.

Hard drives operate within such close tolerances that a drop of water or particle of silt that works its way inside can cripple them; yet, drives aren't sealed mechanisms. Because we use them from the beach to the mountains, drives must equalize air pressure through filtered vents called "breather holes." Under water, these breather holes are like screen doors on a submarine. When Hurricane Katrina savaged thousand of systems, those with the means and motivation turned to data recovery services for a second chance.

Data recovery, in the words of John Christopher, a veteran data recovery engineer at DriveSavers Inc., (www.drivesavers.com) is "open heart surgery" for hard drives. Companies such as Novato, Calif.-based DriveSavers and Ontrack Data Recovery (a division of Kroll Ontrack Inc., www.ontrack.com) are the courts of last resort for damaged drives. DriveSavers worked on dozens of Katrina-damaged drives, some submerged for weeks. Drive housings were full of crud, and recovery required finding identical drives and sacrificing them for compatible parts. DriveSavers reported that it was able to resurrect data from about two-thirds of the Katrina drives sent in.

Keep Them Wet

Ontrack's vice president of operations Todd Johnson reports that his company recovered useable data from about 70 percent of the 425 Katrina-damaged drives they received. All the drives required clean room treatment, with the best outcomes seen in those kept immersed in water or sealed in airtight plastic bags until delivery.

"Don't dry them out," Johnson warned, because that causes the heads that read data to become affixed to the platters.

Another factor favoring recovery was quick action. Whether you proceed with full-scale data recovery or not, promptly getting a drive cleaned and processed by a professional keeps your options open.

DriveSavers' Christopher echoed the need to move quickly and resist turning on the power to "see what works." He lamented that too many dim their prospects for recovery by letting a tech-savvy relative or electronics superstore take a stab at it.

Back It Up and Lock It Down

Despite the miracles performed by professional disk doctors, data recovery is unpredictable and very expensive. Add the cost of business interruption and frustrated clients, and the IT lesson from Katrina is **back it up and lock it down**. Even when systems survive, they may be inaccessible for prolonged periods due to closed or clogged roadways, hazardous conditions,

areas cordoned off to prevent looting or loss of basic services, like electricity and telecommunications. You've got to have an accessible back up.

Katrina forced firms across the Gulf Coast to come to grips with flawed backup practices. Many had no backup system at all. Others were horrified to discover that never-tested backup tapes were useless. The proliferation of data on desktop drives and laptops off the backup grid meant that even those diligent about backup suffered data loss. Still others found to their dismay that backups stored in the same city were kept too close.

Lessons Learned
Whether the risk is hurricane, earthquake, fire, flood, terrorism, theft or disgruntled IT person, no firm is beyond disaster's reach. Here are steps to help weather the storm:

1. Back up critical data…regularly, thoroughly, obsessively.

2. Do periodic test restores of backed up data.

3. Ensure that key data on laptops and desktops is captured.

4. Mass disasters claim entire regions, so store backed up data out of harm's way. Consider online back- up, which safely ensconces data in distant servers, accessible via high-speed net connection from anywhere.

5. Know the answer to, "What would I grab if I had to leave right now?" Prepare for "grab and go" emergencies by using removable disk drive drawers or external hard drives.

Keep anti-static drive packaging and watertight containers on hand. For desktops, consider simple and inexpensive RAID configurations to make grab and go practical (see "Peace of Mind for a Pittance," in the March issue of Law Technology News, March 2005).

6. Encrypt the back up. Recent high-profile breaches of data security stemmed from poor management of backup media. Be sure your data back- ups are safe from prying eyes and that several in the firm know the encryption key. A backup you can't decrypt might as well have washed away.

Do-It-Yourself Digital Discovery

by Craig Ball

[Originally published in Law Technology News, May 2006]

Recently, a West Texas firm received a dozen Microsoft Outlook PST files from a client. Like the dog that caught the car, they weren't sure what to do next. Even out on the prairie, they'd heard of online hosting and e-mail analytics, but worried about the cost. They wondered: Did they really *need* an e-discovery vendor? Couldn't they just do it themselves?

As a computer forensic examiner, I blanch at the thought of lawyers harvesting data and processing e-mail in native formats. "Guard the chain of custody," I want to warn. "Don't mess up the metadata! Leave this stuff to the experts!" But the trial lawyer in me wonders how a solo/small firm practitioner in a run-of-the-mill case is supposed to tell a client, "Sorry, the courts are closed to you because you can't afford e-discovery experts."

Most evidence today is electronic, so curtailing discovery of electronic evidence isn't an option, and trying to stick with paper is a dead end. We've got to deal with electronic evidence in small cases, too. Sometimes, that means doing it yourself.

The West Texas lawyers sought a way to access and search the Outlook e-mail and attachments in the PSTs. It had to be quick and easy. It had to protect the integrity of the evidence. And it had to be cheap. They wanted what many lawyers will come to see they need: the tools and techniques to stay in touch with the evidence in smaller cases without working through vendors and experts.

What's a PST?

Microsoft Outlook is the most popular business e-mail and calendaring client, but don't confuse Outlook with Outlook Express, a simpler application bundled with Windows. Outlook Express stores messages in plain text, by folder name, in files with the extension .DBX. Outlook stores local message data, attachments, folder structure and other information in an encrypted, often-massive database file with the extension .PST. Because the PST file structure is complex, proprietary and poorly documented, some programs have trouble interpreting PSTs.

What about Outlook?

Couldn't they just load the files in Outlook and search? Many do just that, but there are compelling reasons why Outlook is the wrong choice for an electronic discovery search and review tool, foremost among them being that it doesn't protect the integrity of the evidence. Outlook changes PST files. Further, Outlook searches are slow, don't include attachments and can't be run across multiple mail accounts. I considered Google Desktop--the free, fast and powerful keyword search tool that makes short work of searching files, e-mail and attachments--but it has limited Boolean search capabilities and doesn't limit searches to specific PSTs.

Non-Starters

I also considered several extraction and search tools, trying to keep the cost under $200.00. One, a gem called Paraben E-Mail Examiner ($199.00), sometimes gets indigestion from PST files and won't search attachments. Another favorite, Aid4Mail Professional from Fookes Software ($49.95), quickly extracts e-mail and attachments and outputs them to several

production formats, but Aid4Mail has no search capability. I looked at askSam software ($149.95), but after studying its FAQ and noodling with a demo, askSam proved unable to access any PST except the default profile on the machine—potentially commingling evidence e-mail and the lawyer's own e-mail.

dtSearch

The answer lay with dtSearch Desktop, a $199.00 indexed search application offering a command line tool that extracts the contents of PST files as generic message files (.MSG) indexed by dtSearch. In testing, once I got past the clunky command line syntax, I saved each custodian's mail to separate folders and then had dtSearch index the folders. The interface was wonderfully simple and powerful. Once you select the indices, you can use nearly any combination of Boolean, proximity, fuzzy or synonym searches. Search results are instantaneous and essential metadata for messages and attachments are preserved and presented. It even lets you preview attachments.

dtSearch lacks key features seen in products designed as e-discovery review tools, like the ability to tag hot documents, de-duplicate and redact privileged content. But you can copy selected messages and attachments to folders for production or redaction, preserving folder structures as desired. You can also generate printable search reports showing search results in context. In short, dtSearch works, but as a do-it-yourself e-mail tool, it's best suited to low volume/low budget review efforts.

Wave of the Future?

Any firm handles a fifty-page photocopy job in-house, but a fifty *thousand*-page job is going out to a copy shop. Likewise, e-discovery service providers are essential in bigger cases, but in matters with tight budgets or where the evidence is just e-mail from a handful of custodians, lawyers may need to roll up their sleeves and do it themselves.

Tips for Doing It Yourself

If you'd like to try your hand, dtSearch offers a free 30-day demonstration copy at www.dtsearch.com. Practice on your own e-mail or an old machine before tackling real evidence, and if you anticipate the need for computer forensics, leave the evidence machines alone and bring in an expert.

Whether e-mail is stored locally as a PST, in a similar format called an OST or remotely on an Exchange server depends on the sophistication and configuration of the e-mail system. To find a local PST file on a machine running Windows XP, NT or 2000, look for C:\Documents and Settings*Windows user name*\Local Settings\Application Data\Microsoft\Outlook\Outlook.pst. Archived e-mail resides in another file typically found in the same directory, called Archive.pst. Occasionally, users change default filenames or locations, so you may want to use Windows Search to find all files with a PST extension.

When you locate the PST files, record their metadata; that is, write down the filenames, where you found them, file sizes, and dates they were created, modified and last accessed (right click on the file and select Properties if you don't see this information in the directory). Be sure Outlook's not running and copy the PST files to read-only media like CD-R or DVD-R. Remember that PSTs for *different* custodians tend to have the *same* names (i.e., Outlook.pst and Archive.pst), so use a naming protocol or folder structure to keep track of who's who. When dealing with Outlook Express, search for messages stored in archives with a DBX extension.

Though dtSearch will index DBX files, PSTs must first be converted to individual messages using the included command line tool, mapitool.exe. For DOS veterans, it's old hat, but those new to command line syntax may find it confusing. To use mapitool, you'll need to know the paths to mapitool.exe and to the PSTs you're converting. Then, open a command line window (Start>Run>Command), and follow the instructions included with mapitool.

When mapitool completes the conversion, point the dtSearch Index Manager to the folder holding the extracted messages and index its contents. Name the index to correspond with the custodian and repeat the process for each custodian's PST files.

Function Follows Form

by Craig Ball

[Originally published in Law Technology News, June 2006]

The federal rules amendments governing discovery of electronically stored information have sailed through the U.S. Supreme Court and are now before Congress. Assuming passage, they'll be effective this December.

Though all bolts aren't tight and a few sections of track are missing, we're lining up to board the e-discovery roller coaster. It's going to be a wild ride.

As we countdown to the new rules, we should use the time to explore what powerful tools they'll be and acquire the skills to use them artfully while avoiding the sharp edges.

Rule 34(b): Have It Your Way

My favorite amendment—and let's not tarry over what sort of loon has a "favorite"—is FRCP 34(b), which empowers a requesting party to specify the form or forms in which electronically stored information (ESI) is to be produced.

Form didn't matter for paper production, but it makes all the difference in the ability to manage and search ESI.

The producing party must deliver ESI in the specified form or make an objection stating the reasons it won't and the form or forms it intends to provide. Alternate forms must be either those in which the ESI is ordinarily maintained or that are "reasonably usable." This is a giant leap forward for requesting parties, who get ESI their way or at least in a way that's electronically searchable.

The Committee Notes bear this out. "If the responding party ordinarily maintains the information it is producing in a way that makes it searchable by electronic means, the information should not be produced in a form that removes or significantly degrades this feature."

That means no more "naked" .tif or PDF files stripped of searchable data layers. No more blowbacks to paper. Even printouts of e-mail won't cut it… unless that's what the requesting party wants.

Having the power to specify the forms of production presupposes the ability to make an informed choice; plus changes to Rule 26 (f) (3) require parties to discuss forms of production in the pre-discovery meet-and-confer.

We're all going to have to know this stuff.

Five Forms

One of the biggest mistakes a requesting party makes is requesting or accepting production of electronic evidence in a format ill suited to their needs. ESI production takes five principal forms:

1. Hard copies;

2. Paper-like images of data in, e.g., Adobe's Portable Document Format (PDF) or in one of the Tagged Image File Formats (.tif);

3. Data exported to "reasonably usable" electronic formats like Access databases or load files;

4. Native data; and

5. Hosted data.

Your format specification hinges on both the nature of the data and your in-house capabilities for dealing with it.

In a perfect world, you'd want everything in native electronic format. But in the real world, you may lack the systems, software or expertise to access native data and preserve its evidentiary integrity. Plus, concerns about redaction, alteration and Bates numbering mean your opponents may be unwilling to produce native data.

Hard Copies

Converting searchable electronic data to costly and cumbersome paper is usually a step backwards, but paper still has its place.

In a case where the entire production consists of a few hundred e-mails and several thousand e-documents, searching and volume aren't a problem and paper remains as good a medium as any. But once the volume or complexity increases beyond that which you can easily manage by memory, you're better off insisting on production in electronically searchable forms.

Image Production

Here, production consists of files that are digital "pictures" of the documents, e-mails and other electronic records, typically in accessible file formats (PDF or .tif). As long as the information lends itself to a printed format and is electronically searchable, image formats work reasonably well; but for embedded information (such as the formulae in spreadsheets) or when the evidence moves beyond the confines of printable information (e.g., voicemail, databases or video), image production breaks down.

Requesting parties must ensure that electronically searchable data layers and relevant metadata accompany the images. Beware those who try to pawn off "naked" .tif images (devoid of searchable information and metadata) as responsive.

Exported Formats

Some electronic evidence adapts to multiple production formats, so sometimes you'll want exported, delimited data in order to work with it in the compatible application of your choice.

For example, e-mail may be readable in any of several programs or in generic e-mail formats (e.g., .eml, .msg). The contents of simple databases like contact lists can be exported to generic formats (e.g., comma or tab-delimited output) and imported into compatible applications, such as Microsoft Corp.'s Excel spreadsheets or Access databases.

The key is to be sure that important data or the ability to manipulate it isn't lost in the export/ import process.

Native Production

As data structures grow more complex, it's much harder to present exported data in an accurate or complete way.

In native production, the producing party furnishes duplicates of the actual data files containing responsive information and a requesting party with copies of the software programs used to create and manipulate the data (or compatible viewers) has the ability to see the evidence more-or-less exactly as it appears to the other side.

Sounds great, but native production is not without its problems. The native applications required to view the data in its native format may be prohibitively expensive or difficult to operate without extensive training (e.g., Oracle Corp. or SAP America Inc. databases).

Additionally, care must be taken not to change the native data while viewing it. Native production is best, but only when you have the experience, expertise and resources to manage native data.

Producing parties often fight native production because of difficulty in redacting privileged information. An Outlook post office (.pst) file can hold both discoverable e-mail and privileged attorney-client communications, but as it's a unified and complex database file, it's challenging to separate the two.

Another (largely overblown) risk to defendants is that native data (like Microsoft Office files) can contain embedded, revealing metadata. Where native files are concerned, metadata is evidence, too. See, e.g., Williams v. Sprint/United Management Co., 230 F.R.D. 640 (D. Kan. 2005).

Hosted Data

This is production without production in that the information produced resides on a controlled-access website. The requesting party reviews the data through an online application (similar to a web browser) capable of displaying information from a variety of electronic formats. More commonly, hosted data and online review tools are used by counsel for the producing party to search the production set for privileged and responsive items rather than as a means to afford access to the requesting party. The items identified are then burned to CD or DVD and produced, usually in image formats as discussed above.

Next Month: Specifying the right form of production for the most common ESI.

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.*

Ten Common E-Discovery Blunders

by Craig Ball

[Originally published in Law Technology News, August 2006]

A colleague recently asked me to list 10 electronic data discovery errors lawyers make with distressing regularity. Here's that list, along with suggestions to avoid making them:

1. Committing to EDD efforts without understanding a client's systems or data.
It's Russian roulette to make EDD promises when you haven't a clue how much data your client has, or what and where it is. Instead, map the systems and run digital "biopsies" on representative samples to generate reliable metrics and gain a feel for how much are documents, e-mail, compressed files, photos, spreadsheets, applications and so on.

It matters. A hundred gigabytes of geophysical data or video may be a handful of files and cost next to nothing to produce. The same 100 gigs of compressed e-mail could comprise tens of millions of pages and cost a fortune.

2. Thinking you can just "print it out."
Even if you've the time and personnel to stick with paper, is it ethical to subject your clients to the huge added costs engendered by your unwillingness to adapt?

3. Foolishly believing that enough smart people can take the place of the right technologies or that the right technologies eliminate the need for enough smart people.
No search tool yet invented finds every responsive or privileged e-document, and no law firm can marshal enough qualified people to manually review 100 million pages. The best outcomes in EDD flow from pairing well-trained people with the right tools.

4. Ignoring preservation obligations until the motion to compel.
The duty to preserve evidence doesn't hinge on a preservation notice or lawsuit. You must advise your client to preserve potentially relevant paper and electronic evidence as soon as they reasonably anticipate a suit or claim. Even if they aren't obliged to produce inaccessible electronic evidence, they're probably obliged to preserve it.

5. Thinking that search technology trumps records management.
Sorry, but Google isn't going to save us. Privileged communications once went straight from the printer into a file labeled "Attorney Correspondence." Now, they're jumbled with Viagra ads and notices about donuts in the coffee room. We need to enforce cradle-to-grave management for electronic records and restore the "power of place" that allows us to once more limit where we look for responsive data to just those places "where we keep that stuff." Much of the heavy lifting will be over when users must "file" messages in a virtual "file room" when they're sent or received.

6. Hammering out EDD agreements without consulting an expert.
Just because both sides agree to something doesn't make it feasible, or even a good idea. An agreed order stating that an expert will recover "all deleted files" sounds simple, but it's the sort of muddled directive that needlessly drives up the cost of EDD. The right expert will identify

efficiencies, flag pitfalls and suggest sensible, cost-effective search and sampling strategies from the earliest meet-and-confer session.

If your client can't afford an attending expert — though in the end, amateurs costs much more — at least run proposed agreements by someone in the know before they go to the judge.

7. Taking a "peek" at a computer that may contain critical evidence.
Metadata is the data about data that reveals, inter alia, dates of creation, access and modification. Sometimes it's the "who-knew-what-when" evidence that makes the case. But if you access an electronic document, even for a split second, you irrevocably alter its metadata. So when metadata matters, beware the IT guy who volunteers to "ghost" the drive or run searches. Run—don't walk—to engage a properly trained expert to create a forensically qualified image or clone of the evidence.

8. Failing to share sufficient information or build trust with the other side.
The judges are serious about this meet-and-confer business. You can't complain about the other side's demand to see everything if you're playing hide the ball. EDD-savvy requesting parties appreciate the futility of "any-and-all" requests, but how can they seek less if you keep them in the dark about the who, what and where of your client's electronically stored information? Surviving the mutually assured destruction scenario for EDD means building trust and opening lines of communication. The EDD meet-and-confer isn't the place for posturing and machismo. Save it for court.

9. Letting fear displace reason.
Don't let an irrational fear of sanctions rob you of your good judgment. Clients don't have to keep everything. Judges aren't punishing diligent, good-faith efforts gone awry. Your job is to help manage risk, not eliminate it altogether. Do your homework, talk to the right folks, document your efforts and be forthcoming and cooperative. Then, if it then feels right, it probably is.

10. Kidding ourselves that we don't need to learn this stuff.
O.K., you went to law school because you didn't know enough technology to change the batteries on a remote control. This English major feels your pain. But we can't very well try lawsuits without discovery, and we can't do discovery today without dealing with electronically stored information.

You don't want to work through an expert forever, do you? So, we have to learn enough about EDD to advise clients about preservation duties, production formats, de-duplication, review tools, search methodologies and the other essential elements of e-discovery. Our clients deserve no less.

Ten Tips to Clip the Cost of E-Discovery

by Craig Ball

[Originally published in Law Technology News, September 2006]

E-discovery costs *less* than paper discovery. Honest. *In comparable volumes*, it's cheaper to collect, index, store, copy, transport, search and share electronically stored information (ESI). But we hoard data with an indiscriminate tenacity we'd label "mental illness" if we were piling up paper. It's not just that we keep so *much*; it's that our collections are so *unstructured*. Squirrel away twenty years of National Geographic with an index and you're a "librarian." Without the index, you're that "crazy cat lady."
So the number one way to hold down the cost of e-discovery is:

1. If you don't need to keep it, *get rid of it*
Preservation obligations aside, if you're keeping backup tapes you don't need for disaster recovery or that you can't *read* because you no longer have the hardware or software, *get rid of them*. The same holds for all those old computers, hard drives, floppies, CD-ROMs, Zip disks and former e-mail accounts. Don't stick tapes in a closet intending to someday wipe and sell them on e-Bay*. If they don't hold information you must retain*, wipe them, shred them or pulverize them *now.*

2. Get tough on e-mail
E-mail *should* be easy. It's got those handy subject lines. It's electronically searchable. The circulation list's right up front. It's a cinch to file.

In reality, e-mail conversations (*threads*) veer off topic, search is a hit-or-miss proposition (*CUL8R*), addresses are cryptic (*HotBob37@aol.com*) and only the most organized among us *(anal-retentive*) file e-mail with anything like the effort once accorded paper correspondence. Personal messages rub elbows with privileged communications, spam and key business intelligence.

During WWII, everyone knew, "Loose lips sink ships." But does every employee appreciate the risk and cost of slipshod e-mail? Get tough on e-mail through policy, then train, audit and enforce. Train to manage e-mail, appreciate that *messages never die* and know that hasty words are eaten under oath. Tame the e-mail beast and the rest is easy.

3. Have a data taxonomy and standardize storage
Paper discovery cost less, in part because we generated and retained less paper, but also because we did a better job managing paper. We didn't search everywhere because there was always a file, folder or cabinet where we kept "that stuff." That's the power of place.

Records management isn't a form of personal expression. We must restore the elements of good records management to ESI. Want a desktop background with puppies? *Fine, but you must use the company's folder structure and naming protocols.* Want to send an e-mail? *No problem, but if it's personal, you must designate it as such, and if not, you must assign it a proper place within the company's information management system.*

4. Trim ESI requiring attorney review

The costliest phase of e-discovery is attorney review, so big savings flow from shrinking the volume of ESI reviewed, shifting the review burden to the other side and using cheaper talent.

Pare review volume by filtering and de-duplication to cull non-responsive data *before* attorney review, and work with the other side to identify irrelevant file types and target discovery to specific custodians and date ranges. Discovery rules permit production of ESI as maintained in the usual course of business, so consider leaving review to the opposition, protecting privileged content through claw back agreements. Finally, must a local attorney pore over everything, or can some of the work be done by legal assistants or outsourced to lower-cost lawyers in Indiana or India?

5. Keep responsive ESI on the servers

Between road warriors, at-home workers, local drives and smart phones, ESI has gone off the reservation, straying beyond the confines of the company's servers. Harvesting maverick data is costly, so employ policy and technology to insure that responsive data stays on the servers where it's more efficiently secured, searched and backed up.

6. No new gadgets without an e-discovery plan and budget

Everyone loves new toys, but the price tag on the latest PDA, messaging system or software won't reflect the costs it adds to e-discovery. You don't have to give up gadgets, but factor their impact on e-discovery into the total cost of ownership, and be sure preserving and harvesting their contents is part of your e-discovery plan.

7. Build cross-enterprise search and collection capability

Harvest is e-discovery's second costliest component. Eliminating onsite collection adds up to major savings. Emerging technologies make it possible to remotely search and harvest ESI from all machines on a network. Though still in its infancy, cross-enterprise search and collection makes sense for serial litigants and large workforces.

8. Develop in-house harvest expertise

If you want to destroy evidence, ask the IT guy to preserve it. Forensically sound preservation isn't the same as copying, Ghosting or backing up. It demands special tools and techniques. Oil well firefighter Red Adair put it well: "If you think it's expensive to hire a professional, wait until you hire an amateur!"

Learning to be a computer forensic *examiner* is hard, but learning to do forensically sound *acquisitions* isn't. You'll preserve more data than you'll analyze, so having an IT staffer trained in forensically sound preservation saves money on outside experts…and spoliation sanctions

9. Know the component cost of vendor services

Though e-discovery vendors tout "proprietary technologies," all use pretty much the same prosaic processes. Still, some are especially efficient at particular tasks (like tape restoration or scanning) and price these services competitively. When you understand the pieces of ESI processing and what each adds to the bill, you can match the task to the best-qualified vendor and get the best price.

10. Work cooperatively with the other side

This tip saves more than the others combined. Being forthright about your ESI and transparent in your e-discovery methodology fosters the trust that enables an opponent to say, "You don't have to produce that." The e-discovery horror stories—the ones that end with sanctions—all start with, "Once upon a time, there was a plaintiff and a defendant who couldn't get along."

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.

Santa@NorthPole.com

by Craig Ball

[Originally published in Law Technology News, December 2006]

Dear Santa,

I've been a good boy this year. I spent all my time helping lawyers and judges with electronic discovery and studying really, really hard about ESI, data harvest, spoliation, de-duplication, meet-and-confer, search tools, forms of production and computer forensics. I didn't use the word "solution" in a single column.

Please leave these presents under my tree:

1. I want a container file format for electronically stored information (ESI). We are gathering all this discoverable data but corrupting its metadata in the process. Plus, it's so hard to authenticate and track ESI. The container would safely hold the evidence as we harvest, search and produce it. It would include hash verification of all its parts, a place to store both an image of the document and its native content and even a special pocket to hold an overlay of all that helpful stuff we used to stamp onto paper documents, like Bates numbers and confidentiality warnings. And Santa--this is really important--it needs to be open sourced so no one has to pay to use it and extensible so we can keep using it for a very long time.

2. I want integrally write-protected external hard drives with removable electronic keys. Producing ESI on optical disks is nice because they're read-only media and you can't intentionally or inadvertently corrupt their contents. But nowadays, there's just too much ESI to hand over on optical disks. I want external hard drives designed for e-discovery such that a producing party can fill them with information then remove a USB key or snap off a tab to insure that nothing else can be written to or changed on the drive. If it hashed its contents and burned that hash value to an onboard write-once chip, that would be pretty cool, too.

3. May I have information technology training courses designed expressly for lawyers and litigation support, offering real depth and serious accountability for mastering the subject matter? Lawyers and their staff are waking up to the need to learn this stuff, but the traditional CLE and CPE paths don't offer or demand enough. We don't need another 10,000-foot "certification" course. We need Parris Island.

4. While we're at it big guy, how about making electronic discovery and digital evidence a discrete part of law school curriculum? I understand that teaching the *practice* of law is looked down upon at the best schools, but the assumption that young lawyers who grew up with computers automatically "get it" is misguided.

5. Could there also be licensure for computer forensic examiners geared to insuring genuine expertise and experience? Putting computer forensic examiners under the jurisdiction of the state boards that regulate private investigators and security guards is like putting the football coach in charge of the Physics Department. Weeding out

unqualified computer forensic examiners is a worthwhile goal, but can't legislatures put the task in the hands of those best qualified to judge.

6. Since we're regulating the forensics side of e-discovery, how about a code of ethics for electronic discovery vendors and experts, too? One that's not just lipstick on a pig! All parties need to be confident that information in a vendor's custody, including how that material is reviewed, is secure and that vendors are keeping their software current and adhering to other sound practices.

7. Santa, I'm still hoping to get what I asked for last year, like e-mail clients that compel immediate filing of messages within an information taxonomy and published standards and definitions for metadata fields of common file types. I do hope the elves are working on those, too.

Thanks. Fly carefully. Love, Craig

Thanksgiving

As this goes to press, the e-discovery amendments to the Federal Rules of Civil Procedure finally take effect--a milestone culminating six years of hard work by the Rules committee. We owe them a huge debt of gratitude even as we greet with trepidation the consequences of what they've wrought

Like Y2K, there will be no falling of the sky or trembling of the ground. But like Y2K, the post-FRCP amendments world will never be quite the same. The Y2K apocalypse never materialized, but the world quietly changed as dramatically as if it had. Enormous sums were plowed into computing infrastructure, and the clamor for programming talent threw wide the doors to India and the world, transforming the global economy in the many ways New York Times columnist Thomas Friedman insightfully describes in his bestseller, "The World is Flat."

The changes to the Federal rules are modest. The added language probably amounts to fewer words than this column. But those Amendments are already driving massive investment in infrastructure, training, services and personnel. It's just the tip of the iceberg. Litigants and their lawyers won't feel anything on December 1. Most will escape the maelstrom for another week, month, even a year or two. But it's coming, as inevitably as death and taxes, ready…or not.

Craig Ball, a member of the Editorial Advisory Boards of both LTN and Law.com Legal Technology, is a litigator and computer forensics/EDD special master, based in Austin, Texas. E-mail: craig@ball.net

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.

Climb the Ladder

by Craig Ball

[Originally published in Law Technology News, February 2007]

Though computer forensics is a young discipline, it's not the exclusive province of new graduates of computer forensics degree programs. It's a natural career extension for IT and law enforcement professionals and peripatetic lawyers with a dominant geek gene. Expertise in litigation and computer forensics also opens the door to lucrative opportunities in electronic data discovery consulting. Here are "The Eight Es" to becoming a skilled CF expert:

1. Exploration...The lion's share of CF knowledge is self-taught. The best examiners are insatiably curious and voraciously read about software, hardware, registry keys, root kits, etc. They live for figuring out how it all fits together. Fortunately, there's a wealth of information: in books (search Amazon.com for "computer forensics") and online (www.e-evidence.info) in discussion forums, product FAQs, user groups and confabs.

2. Education...A computer science or law degree is nice, but you can study animal husbandry so long as you go on to study CF in a comprehensive way. Professional certifications that legitimately demonstrate training, testing and practical experience have value in helping courts, clients, and potential employers assess your qualifications. Supplement your college degree with as many courses and certifications as your time and budget allow.

Excellent programs are offered by universities, vendors, professional associations, and the government, such as ***New Technologies Inc.,*** (www.forensics-intl.com), ***Guidance Software*** (www.guidancesoftware.com), ***Access Data*** (www.accessdata.com), the ***International High Technology Crime Investigation Association*** (www.htcia.org), the ***International Association for Computer Information Systems*** (www.iacis.org), and the ***Federal Law Enforcement Training Center*** (www.fletc.gov). But don't fool yourself into thinking that a weeklong boot camp will qualify you as a CF expert. In a battle between an experienced examiner and one with an advanced degree, juries may defer to the latter. Some jurisdictions require licensure to perform forensic investigations.

3. Experimentation...The ability to construct illuminating experiments and the patience to elicit data are hallmarks of a skilled examiner. If you need to know how metadata changes when a user touches a file, you'll be prepared to testify if you've proven your theory by competent experimentation. Experiment with systems, applications and operating systems to understand how they work.

4. Experience...There's no substitute for applying your skills and testifying in real cases. How can you get that experience? Apprentice to a veteran examiner or offer to perform a "shadow exam," to see if you find something he or she missed. Assist attorneys or local law enforcement at little or no cost.

5. Exchange...Every examiner benefits from the exchange of ideas with colleagues. Join industry associations, go to meetings, subscribe to online discussion groups and unselfishly share what you learn. Caveat: the CF community is very supportive, but other examiners may justifiably regard you as a competitor, so don't expect them to reveal all. Show respect by doing your homework. Be a learner, not a leech.

6. Equipment...Learn the tools and techniques suited to the task, and invest in them. Use quality hardware and properly license software. Keep applications up-to-date, test tools to insure they're reliable. Cross-validate results. Too many people confuse buying tools with acquiring skills. A well-trained examiner can do the job with a hex editor and a viewer. We use forensic suites, such as Guidance Software's EnCase or Access Data's FTK, to automate routine tasks, improve efficiency, and lower costs—but buying a program doesn't make you a ready expert.

7. Earning...The demand for examiners is growing, but it takes marketing skill and financial acumen to create a thriving business. You must attract and serve quality clients, and make ends meet, to transform opportunity into achievement. Consider a first job with established CF companies or law enforcement, not only for a steady income, but also for the training. Starting salaries average $50,000 to $75,000, but in the private sector, quickly rise to six figures as you gain experience and responsibility. (Examiners with J.D.s or network security skills command higher salaries.)

Many CF firms charge clients $250 to $600 per hour, so it's not unrealistic for entrepreneurial examiners to hang out their shingles after learning the ropes. Expect $25,000 in minimum startup costs for hardware, software and training. Overhead will vary on whether you operate from your home or offsite.

8. Essential Element — Character...The final "E" is the "essential element"—*character*. A successful examiner is at once, teacher and student, experimenter, skeptic, confidante, translator, analogist, and raconteur. He or she unearths the human drama hidden in the machine. So many qualities distinguish the best examiners—integrity, tenacity, technical skill, imagination, insatiable curiosity, patience, discretion, attention to detail and the ability to see both the forest and the trees. Ultimately, it's your character that will determine if you'll be a top computer forensics expert.

Vista Changes the View

by Craig Ball

[Originally published in Law Technology News, March 2007]

Vista, Microsoft Corp.'s long awaited re-invention of its Windows operating system, finally premiered with little fanfare. No Rolling Stones theme music this time, though users frustrated with ineradicable security holes in Windows XP could have made the case for "19th Nervous Breakdown" or "(I Can't Get No) Satisfaction."

While most businesses are taking a wait-and-see attitude about migrating, sooner or later, they'll make the move. Within two years, Vista will have made significant inroads against XP on business desktops and laptops, and in the home, Vista will dominate. Many Windows users will also upgrade to Office 2007, the latest release of Microsoft's four horsemen, Word, Outlook, Excel and PowerPoint. What does this inexorable Vista and Office creep mean for electronic data discovery and the nerdy little corner of EDD called computer forensics? Only time will tell, but dramatic changes are in store.

Versions

Remember that deluxe Crayola box you longed for as a kid—the one with the sharpener and colors like "flesh" and "periwinkle?" Well, Vista has nearly as many versions as that box had crayons. There's Vista Home Basic, Home Premium, Business, Enterprise, Ultimate and Vista with Retsyn (okay, I made that last one up). Though all affect EDD to some extent, as you move higher up the evolutionary ladder of Vista versions, you'll bump into features like BitLocker volume encryption that really complicate EDD and forensics.

The big news in Vista is security, especially against prying eyes and careless keystrokes. Business, Enterprise and Ultimate editions include an automatic backup feature called Shadow Copy that invisibly saves your work to unused disk space to protect you from "Oh, No!" moments—like saving over an important file. Sounds great, except it salts away all prior versions, including those you *don't* intend to keep.

Unlike the hidden, fragmented forensic data the federal rules call examples of inaccessible ESI, the Vista shadow copy is a hardy survivor: complete, coherent and readily accessible. Vista grows the volume of discoverable ESI, perhaps significantly.

Little Brother

Not only does Vista do a better job hanging on to your work, it also keeps tabs on users as they work, through a feature called Transactional NTFS, or TxF. Bid goodbye to Last Access Times corrupted by peeking at the evidence or antivirus scans.

By default, Vista quits tracking access times as a file property. Instead, TxF logs file system activities, and a counterpart called TxR logs Registry activity. The bottom line is that a user's activity will be closely and constantly tracked, step-by-fateful-step. From the standpoint of investigating claims of evidence destruction, it's less a piecing together of fragments and more a "Let's look at that again in instant replay" situation.

This means a heck of a lot of new digital evidence out there to preserve and discover.

BitLocker
At the top of the Vista food chain, the Enterprise and Ultimate editions include a drive volume encryption feature called BitLocker—giving greater protection against data breaches from lost and stolen laptops.

BitLocker makes users access their data like $20 bills at the ATM. But to work, the protected machine must be equipped with a microchip called a Trusted Platform Module, or unlocked by a USB flash drive that serves as a key. Maybe Microsoft will hook up with the Stones again to market Jumpin' Jack Flash and Under My Thumb Drives.

The same encryption hardening a machine against identity theft and competitive intelligence can hopelessly frustrate forensic examination and emerging remote search and collection tools for EDD.

Absent a robust key escrow program, companies will have a harder time enforcing acceptable use policies or stealthily inspecting machines to identify candidates for litigation hold.

Rumors of a back door for law enforcement abound, but so far, Microsoft vehemently denies the existence of a way around BitLocker. Instead, the feature is reserved to only the most costly versions. Budget-minded criminals beware.

New Folders & Formats
Updates large and small will impact e-discovery. For example, Vista banishes spaces from standard folder names, so the Windows "My Documents" folder is now "Documents." Even so trivial a change can wreck havoc with automated or scripted collection protocols and trigger expensive do-overs if EDD systems and personnel aren't adaptable and vigilant.

There's some encouraging news, too. Vista and Office introduce fundamental changes to file formats that will, in time, dramatically impact EDD by lowering cost and complexity of both review and production. One of the persistent objections to native file production of ESI is the difficulty of redacting privileged content from native formats. Consequently, Microsoft's decision to store Office 2007 files in XML bodes a sea change for EDD, not only because it facilitates review but particularly because of the ease with which privileged content can be identified and redacted in XML. It opens the door to widespread use of native file formats in production and consigns the claim "native files can't be redacted" to "urban legend."

Search That Works?
A big EDD question mark hangs over Vista's enhanced search capabilities. Search in previous Windows versions was literally and figuratively a dog. But Vista's search tool rivals Google Desktop in speed, text search and metadata filtering.

It may yet prove an ally to litigation hold efforts if counsel circulates not only retention instructions describing targeted information, but also specific queries to be undertaken in Vista Search—a task potentially facilitated by Vista's ability to store programmatic searches in so-called "Search Folders."

The Big Picture

I've focused on a few of the many features of Windows Vista and Office 2007 likely to impact e-discovery, but these products are just the most visible components in a roll out of more than 30 Microsoft products that collectively promise to transform e-discovery and digital evidence in ways lawyers must anticipate and address. More wrenching changes will flow from Microsoft's embrace of collaboration and integrated messaging. The static, single-author printable document is evolving into something entirely new, organic and multiplayer. Collaborative construction means much more metadata playing a much more crucial role as the glue that holds "documents" together.

Integrated messaging shifts IM and voice to center stage and further undercuts paper and .tiff as viable forms of review and production. Both developments signify fresh challenges in every discovery phase.

For lawyers and litigants who feel like EDD has crept up and kicked them, you ain't seen nothing yet. Vista et al. paint a broad new horizon over rough seas.

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?

Do-It-Yourself Forensics

by Craig Ball

[Originally published in Law Technology News, June 2007]

All over America, vendors stand ready to solve the e-discovery problems of big, rich companies. But here's the rub: Most American businesses are small companies that use computers—and along with individual litigants, they're bound by the same preservation obligations as the Fortune 500, including occasionally needing to preserve forensically significant information on computer hard drives. But what if there's simply no money to hire an expert, or your client insists that its own IT people must do the job?

THE D-I-Y CHALLENGE

I challenged myself to come up with forensically sound imaging methods for conventional IDE and SATA hard drives—methods that would be inexpensive, use off-the-shelf and over-the-net tools, yet simple enough for nearly anyone who can safely open the case and remove the drive. In that vein, the safest way to forensically preserve evidence is to employ a qualified computer forensics expert to professionally "image" the drive and authenticate the duplicate. No one is better equipped to prevent problems or resolve them should they arise.

Further, when you open up a computer and start mucking about, plenty can go awry, so practice on a machine that isn't evidence until you feel comfortable with the process.

FORENSICALLY SOUND

When you empty deleted files from your computer's recycle bin, they aren't gone. The operating system simply ceases to track them, freeing the clusters the deleted data occupies for reallocation to new files. Eventually, these unallocated clusters may be reused and their contents overwritten, but until that happens, Microsoft Corp.'s Windows turns a blind eye to them and only recognizes active data. Because Windows only sees active data, it only copies active data. Forensically sound preservation safeguards the entire drive, including the unallocated clusters and the deleted data they hold.

Even lawyers steeped in electronic data discovery confuse active file imaging and forensically sound imaging. You shouldn't. If someone suggests an active data duplicate is forensically sound, set them straight and reserve "forensically sound" to describe only processes preserving all the information on the media.

PRIMUM NON NOCERE

Like medicine, forensic preservation is governed by the credo: "First, do no harm." Methods employed shouldn't alter the evidence by, e.g., changing the contents of files or metadata. But that's not always feasible, and the first method described departs from the forensic ideal.

METHOD 1: THE DRIVE SWAP COMPROMISE

Pulling the plug and locking a computer away is a forensically sound preservation method, but rarely practical. By the same token, imaging programs such as Symantec Corp.'s Ghost (www.ghost.com) or Acronis Inc.'s True Image (www.acronis.com) leave unallocated clusters behind and may alter the source. Our first do-it-yourself approach strikes a balance between practical and perfect by recognizing that users obliged to preserve the contents of unallocated clusters have no use for those contents. They use only active data. So, the first method

employs off-the-shelf cloning software to copy just active files from the original evidence drive to a duplicate of equal or greater capacity. The forensic twist is that you preserve the original drive and put the duplicate back into service.

Be sure that the drive you swap has the same size enclosure as the original (typically 2.5 inches for laptops and 3.5 inches for desktops) and that it connects to the computer in the same way, e.g., parallel ATA (a.k.a. "IDE") or Serial ATA. Pull the plug (for laptops, remove the battery too), then open the case to determine the type of drive interface before heading to the store. Buy the proper replacement internal drive in a gigabyte capacity at least as large as the original. Greater capacity is fine.

Accessing a laptop drive can be tricky, so check the manufacturer's website if you're uncertain how to remove and safely handle the drive. Another hurdle: laptops lack cabling to add a second internal drive, so you'll need an adapter to connect the target drive via USB port. A Vantec Thermal Technologies' (www.vantecusa.com) CB-ISATAU2 adapter cable runs about $25 at www.newegg.com, or find other adapters and suppliers by web searching "sata/ide usb adapter."

Follow the software's instructions, but never install the duplication software to the drive you're preserving because that overwrites unallocated clusters. Instead, run the application from a CD, floppy or thumb drive. It's critically important that you don't inadvertently copy the contents of the blank drive onto the original, so check settings, and then check them again before proceeding.

When the imaging completes, label the original drive with the date imaged, name of the user, machine make, model and serial number, and note any inaccuracy in the BIOS clock or calendar. Secure the original drive in an anti-static bag and install the duplicate drive in the machine. Confirm that it boots. The user should see no difference except that the drive offers more storage capacity.

Done right, this method hews close to a forensically sound image, the qualifier being that the cloning software and the operating system may make some (typically inconsequential) alterations to the source drive. The method combines the advantages of Ghosting (speed and ease-of-use) with the desirable end of preserving the original digital evidence with [most] metadata and unallocated clusters intact. Best of all, it employs tools and procedures likely to be familiar to the service techs at your local electronics superstore. Be sure they adhere to the cautions above.

Next month, I'll describe a do-it-yourself approach to *true* forensically sound imaging.

Do-It-Yourself Forensic Preservation (Part II)

by Craig Ball

[Originally published in Law Technology News, July 2007]

How does a non-expert make a forensically sound copy of a hard drive using inexpensive, readily available tools? That's the D-I-Y challenge. Last month, we discussed a nearly perfect way to forensically preserve hard drives that entails swapping the original drive for a Ghosted copy containing just active files.

But when it comes to crucial evidence, nearly perfect doesn't cut it. Last month's method made minor changes to the source evidence, didn't grab unallocated clusters (necessitating we sequester the original drive) and offered no means to validate the outcome.

Because a forensically sound preservation protects all data and metadata along with deleted information in unallocated clusters, think of the Three Commandments of forensically sound preservation as:

1. Don't alter the evidence;
2. Accurately and thoroughly replicate the contents; and
3. Prove the preceding objectives were met.

This month's method employs write blocking to intercept changes, software that preserves every byte and cryptographic hash authentication to validate accuracy.

Write Blocking

Computer forensics experts use devices called "write blockers" to thwart inadvertent alteration of digital evidence, but write blockers aren't sold in stores (only online) and cost from $150-$1,300. Hardware write blocking is best if timetable and budget allow. Manufacturers include Tableau, LLC (www.tableau.com), WiebeTech, LLC (www.wiebetech.com), Intelligent Computer Solutions, Inc. (www.ics-iq.com) and MyKey Technology, Inc. (www.mykeytech.com).

If you're running Windows XP or Vista, you may not need a device to write protect a drive. To hinder data theft, Windows XP Service Pack 2 added support for software write blocking of USB storage devices. A minor tweak to the system registry disables the computer's ability to write to certain devices via USB ports. To make (and reverse) the registry entry, you can download switch files and view instructions explaining how to manually edit the registry at http://www.lawtechnews.com/r5/showkiosk.asp?listing_id=1560974 (the contents of this web link follow on page 69).

You'll also need:

• **Imaging Machine**--a computer running Windows XP with Service Pack 2 and equipped with both USB 2.0 and IEEE 1394 (aka Firewire or i.Link) ports.

• **Forensic Imaging Application**--though forensic software companies charge a pretty penny for their analysis tools, several make full-featured imaging tools freely available. Two fine Windows-compatible tools are Technology Pathway's Pro-Discover Basic Edition (in the

Resource Center at http://www.techpathways.com) and AccessData's FTK Imager (http://www.accessdata.com/support/downloads/). I prefer FTK Imager for its simplicity and ability to create images in multiple formats, including the standard Encase E01 format.

• **Target Drive**--a new, shrink-wrapped external hard drive to hold the image. It should be larger in capacity than the drive being imaged and, if using software write blocking, choose a drive that connects by IEEE 1394 Firewire(as USB ports will be write blocked).

• [Software write blocking only] A **USB bridge adapter cable or external USB 2.0 drive enclosure** matching the evidence drive's interface (i.e., Serial ATA or Parallel ATA). Though you'll find drive enclosures at your local computer store, I favor cabling like the Vantec Thermal Technologies' (www.vantecusa.com) CB-ISATAU2 adapter cable because they connect to 2.5", 3.5" and 5.25" IDE and SATA drives and facilitate imaging without removing the drive.

Imaging the Drive

Here is a step-by-step guide:

1. It's important to carefully document the acquisition process. Inspect the evidence machine and note its location, user(s), condition, manufacturer, model and serial number or service tag. Photograph the chassis, ports and peripherals.

2. Disconnect all power to the evidence machine, open its case and locate the hard drive(s). If more than one drive is present, you'll need to image them all. Accessing a laptop drive can be tricky, so check the manufacturer's website if you're uncertain how to safely remove and handle the drive. Take a picture of the drive(s) and cabling. If you can't read the labeling on the face of the drive or comfortably access its cabling, uninstall the drive by disconnecting its data and power cables and removing mounting screws on both sides of the drive or (particularly in Dell machines) by depressing a lever to release the drive carriage.

Handle the drive carefully. Don't squeeze or drop it, and avoid touching the circuit board or connector pins. If using a hardware write blocker, connect it to the evidence drive immediately and leave it in place until imaging is complete and authenticated.

3. Download and install FTK Imager on the imaging machine. If using software write blocking, initiate the registry tweak, reboot and, using a thumb drive or other USB storage device, test to be sure it's working properly.

4. Connect the evidence drive to the imaging machine through the hardware write block device or, if using software write protection, through either the USB drive enclosure or via bridge cable connected to a software write blocked USB port. ***Above all, be sure the evidence drive connects only through a write blocked device or port***.

5. If USB ports are software write blocked, connect the target drive via the IEEE 1394 port. Optionally, connect via USB port if using hardware write blocking.

6. Run FTK Imager, and in accordance with the instructions in the program's help file for creating forensic images, select the write protected evidence drive as the source physical drive, then specify the destination (target) drive, folder and filename for the image. I suggest incorporating the machine identifier or drive serial number in the filename, choosing "E01" as

the image type, accepting the default 650MB image fragment size and opting to compress the image and verify results.

Hash Authentication
Creating a forensically sound compressed image of a sizable hard drive can take hours. FTK Imager will display its progress and estimate time to completion. When complete, the program will display and store a report including two calculated "digital fingerprints" (called MD5 and SHA1 hash values) which uniquely identify the acquired data. These hash values enable you to prove that the evidence and duplicate data are identical. Hash values also establish whether the data was altered after acquisition.

7. When the imaging process is done, label the target drive with the date, the names of the system user(s) and machine identifier. Include the model and serial number of the imaged drive.

8. With the evidence drive disconnected, reconnect power to the evidence machine and boot into the machine's setup screen to note any discrepancy in the BIOS clock or calendar settings. Disconnect power again and re-install the evidence drive, being careful to properly reconnect the drive's power and data cables.

Whether you return the evidence machine to service or lock it up depends on the facts of the case and duties under the law. But once you've secured a forensically sound, authenticated image (along with your notes and photos), you've got a "perfect" duplicate of everything that existed on the machine at the time it was imaged and, going forward, the means to prove that the data preserved is complete and unaltered.

The safest way to forensically preserve digital evidence is to engage a qualified computer forensics expert because no one is better equipped to prevent problems or resolve them should they arise. But when there's no budget for an expert, there's still an affordable way to meet a duty to forensically preserve electronic evidence: ***do-it-yourself***.

Enabling and Disabling USB Write Protection in Microsoft Windows XP P2 and Vista

(This is the target page for the link in the preceding BIYC July 2007 column)

Windows XP machines updated with Service Pack 2 (SP2) acquired the option to enable write protection for removable storage devices connected to the machine via USB. You can still read from the devices, but you can't write to them. In my testing, it works as promised, preventing changes to the data and metadata of external USB hard drives and thumb drives. Though the Windows cache may make it seem that data has been written to the protected device, subsequent examination demonstrated that no changes were actually made. And you can't beat the price: it's free.

Still, software write protection has its ardent detractors (See, e.g., *The Fallacy of Software Write Protection in Computer Forensics*, Menz & Bress 2004), and because there's no outward manifestation that software write blocking is turned on and working, there's none of the reassurance derived from seeing a hardware write blocker play burly bodyguard to an evidence drive. Other downsides are that software write protection requires a geeky registry hack and lacks the selectivity of hardware write blocking. That is, when you implement software write blocking, it locks down all USB ports, including the one you'd hoped to use to connect an external USB hard target drive. Write blocked for one is write blocked for all.

Caveat: Software write protection of the USB ports only works in Windows XP with Service Pack 2 and Windows Vista. It can be implemented only by users with Administrator level privileges on the machine. Failing to disable write blocking may cause the loss of data you seek to store on external USB storage devices.

The Easy Way

To simplify software write protection, you can download a file from http://www.craigball.com/USB-WProtect.zip containing two .REG files that, when run (i.e., double clicked), serve as switches to enable and disable software write protection of the USB ports.

The Geeky Way

If you'd rather make the registry changes manually, here's how:

Caveat: It's prudent to create a system restore point before editing the registry. To do so, click Start > All Programs > Accessories > System Tools > System Restore. Select "Create a restore point," then click "Next." Type a brief description for your restore point (e.g., "Before adding write protection"), then click "Create."

Enabling Write Protection

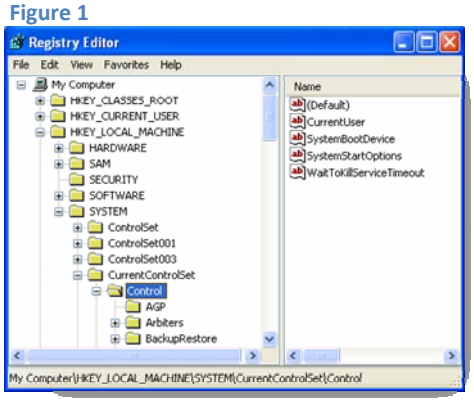
To block the computer's ability to write to a removable storage device connected to a USB port, begin by calling up a Windows command dialogue box:

Press the Windows key + R to bring up the Run dialogue box (or click Start > Run).

Type regedit and click “OK” to activate the Windows Registry Editor.

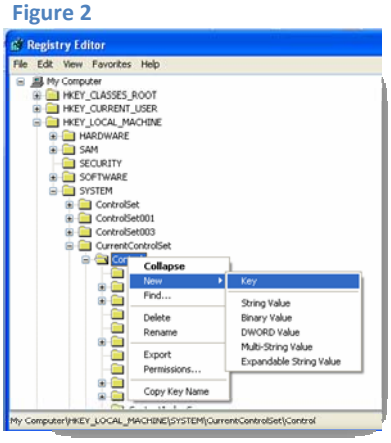
Click the plus sign alongside HKEY_LOCAL_MACHINE, then drill down to SYSTEM\CurrentControlSet\Control. [Fig 1.]

Figure 1



Examine the tree under Control to determine if there is a folder called “StorageDevicePolicies.” If not, you need to create it by right clicking on Control and selecting New > Key. [Fig. 2]

Figure 2

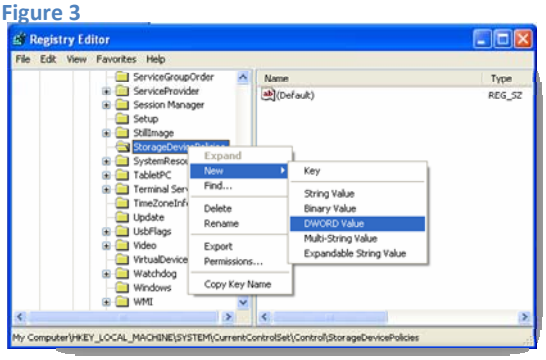


Name the key “StorageDevicePolicies,” (All one word. Match capitalization. Omit quotation marks) then right click on the key you’ve just created and select New > DWORD value [Fig. 3]

Name the new DWORD “WriteProtect” and hit Enter.

Right click on the new DWORD value and select “Modify.” Set the WriteProtect DWORD value to 1. [Fig. 4]

Figure 3



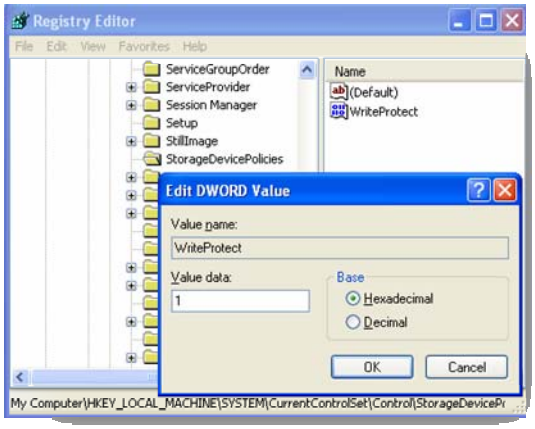
Exit the Registry Editor and reboot the machine. The USB ports should now be write protected.

Disabling Write Protection

To restore the system’s ability to write to USB media, navigate to the WriteProtect key as above and either delete it or change its value to 0.

Reminder: WriteProtect = 1 [ON]
WriteProtect = 0 [OFF]

Figure 4



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.

Re-Burn of the Native

by Craig Ball

[Originally published in Law Technology News, September 2007]

I could hear the frustration in her voice. "We keep going back and forth with the plaintiff's lawyer. I don't understand what he wants. Can you help us?"

Defense counsel was trying to satisfy an opponent bent on getting e-mail in "native file format." With each disk produced, the plaintiff's lawyer demanded, "Where's the e-mail?" Now he was rattling the sanctions saber. Poring over copies of what she'd produced, defense counsel saw the e-mail. "Why can't he see it?"

Reviewing the correspondence between counsel, I spotted the problem. The e-mail was there, but in Rich Text Format. Like many lawyers new to e-discovery, defense counsel regarded electronically stored information and native data as one-and-the-same. They're not.

The IT department had dutifully located responsive e-mail on the mail server and furnished the messages in a generic format called Rich Text Format or "RTF." It's a format offering full access to the contents of the messages, and it's electronically searchable. Any computer can read RTF files. So, it's a pretty good production format.

But, it's not the native format.

Container Files

The native format for virtually all enterprise e-mail is a *container file* lumping together relevant, irrelevant, personal and privileged communications, along with calendar data, to-do lists, contact information and more.

The precise native format depends upon the e-mail client and server. The prevailing enterprise e-mail application, Microsoft's Exchange Server, uses a container file with the file extension .EDB. Lotus Notes stores its e-mail on a Lotus Domino server in a container file with the extension .NFS. These containers are the "native file format" for server-stored e-mail, but they hold not only all then-existing e-mail for a specific user, but also the e-mail and other data for ALL users. Furnishing these files is tantamount to letting the opposition rifle every employee's desk.

When enterprise e-mail is stored locally on a desktop or laptop system, it's almost always in a container file, sometimes called a *compound file*. For users of Microsoft's Outlook e-mail program (a "client application" in geek speak), the local container file is typically called "Outlook.PST" or "Outlook.OST." There may also be a file holding older e-mail called "Archive.PST." Collectively, these data are commonly referred to as a user's "local PST."

Like their counterparts on e-mail servers, local container files weave together the user's responsive and non-responsive items with privileged and personal messages; consequently, they're more like self-contained communications databases than paper correspondence folders.

Conundrum

Because the native file format for enterprise e-mail is bound up with information beyond the scope of discovery, it's the rare case where e-mail should be produced in its native format. Litigants must also be wary of producing native e-mail container formats because, until those containers are compacted by the client application, they hold information (like double deleted files) invisible to users but potentially containing privileged and confidential material. It's possible to "mine" local PSTs for hidden data, and metadata scrubber tools offer no protection.

How, then, do we realize the considerable benefits of native production for e-mail? The answer lies in distinguishing between production of the native container file and production of responsive, non-privileged e-mail in electronically searchable formats that *preserve the essential function of the native source*, sometimes called *quasi-native* formats.

Quasi-Native Production

Chockablock as it is with non-responsive material, there are compelling reasons not to produce "the" source PST. But there's no reason to refuse to produce responsive e-mails and attachments *in the form* of a PST file, so long as it's clearly identified as a reconstituted file containing selected messages and the contents fairly reflect the responsive content and relevant metadata of the original. Absent a need for computer forensic analysis or exceptional circumstances, a properly constructed quasi-native production of e-mail is an entirely sufficient substitute for the native container file.

It doesn't have to be in PST format. There are several generic e-mail formats well suited to quasi-native production (e.g., .MSG and .EML formats). Even RTF-formatted production may suffice when paired with attachments, if the parties don't need to search by discrete header fields (i.e., to sort by To, From, Subject, Date, etc.).

Talk to Me

In the case at hand, the problem isn't one of intent or execution. It's miscommunication and misunderstanding. Plaintiff counsel saw only that he hadn't gotten the format he wanted. Defense counsel saw e-mail in an electronic format and assumed that it must be the right stuff. One fixed on form and the other on content. In e-discovery, both matter.

Accordingly, defense counsel will burn new disks containing the responsive e-mail in PST format.

So, talk to each other, and don't rely on buzzwords like "native file format" unless your meaning is clear. You'll be amazed how often the question, "What do you mean by native file format?" will be answered, "I have no idea. I just heard it was something I should ask for."

CRAIG BALL

Trial Lawyer & Technologist
Computer Forensic Examiner

1101 Ridgecrest Drive
Austin, Texas 78746
E-mail: craig@ball.net
Web: craigball.com
Office: 512-514-0182
Fax: 512-532-6511

Craig Ball is a Board Certified trial lawyer and computer expert He has dedicated his career to teaching the bench and bar about forensic technology and trial tactics. After decades trying lawsuits, Craig now limits his practice to serving as a court-appointed special master and consultant in computer forensics and electronic discovery, and to publishing and lecturing on computer forensics, emerging technologies, digital persuasion and electronic discovery. Craig's award-winning e-discovery column, "Ball in Your Court," appears in Law Technology News. Named as one of the Best Lawyers in America and a Texas Superlawyer, Craig is a recipient of the Presidents' Award, the State Bar of Texas' most esteemed recognition of service to the profession and of the Lifetime Achievement Award in Law and Technology.

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 Board, Law Technology News (American Lawyer Media)
Board Member, Georgetown University Law School Advanced E-Discovery Institute
Special Master, Electronic Discovery, Federal and Harris County (Texas) District Courts
Member, Sedona Conference Working Group 1 on Electronic Document Retention and Production
Instructor in Computer Forensics, United States Department of Justice
Special Prosecutor, Texas Commission for Lawyer Discipline, 1995-96
Council Member, Computer and Technology Section of the State Bar of Texas, 2003-
Chairman: Technology Advisory Committee, State Bar of Texas, 2000-02
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.
Member, Texas State Bar College
Member, Continuing Legal Education Comm., 2000-04, Civil Pattern Jury Charge Comm., 1983-94,State Bar of Texas
Life Fellow, Texas and Houston Bar Foundations
CLE Course Director: E-Discovery A-to-Z (NY, Chicago, SF, Boston, Washington, D.C., Minneapolis, Miami, Houston, Seattle) 2004-6; Electronic Evidence and Digital Discovery Institute 2004-6; Advanced Evidence and Discovery Course 2003; 2002; Enron—The Legal Issues, 2002; Internet and Computers for Lawyers, 2001-02; Advanced Personal Injury Law Course, 1999, 2000; Preparing, Trying and Settling Auto Collision Cases, 1998.
Member, SBOT President's "Vision Council" on Technology, 1999-2000; Strategic Planning Committee Liaison, 2001-02; Corporate Counsel Task Force 2001-02

ACADEMIC APPOINTMENTS AND HONORS

2006 Recipient of the State Bar of Texas CTS Lifetime Achievement Award for Law and Technology
The March 2002 CLE program planned by Mr. Ball and Richard Orsinger entitled, "Enron—The Legal Issues" received the Best CLE of 2002 award from the Association for Legal Education
National Planning Committee, Legal Works 2004 (San Francisco)
Recipient, State Bar of Texas Presidents' Award (bar's highest honor), 2001
Faculty, Texas College of Trial Advocacy, 1992 and 1993
Adjunct Professor, South Texas College of Law, 1983-88

Listed in “Best Lawyers in America” and Selected as a “Texas Super Lawyer,” 2003-2006
Rated AV by Martindale-Hubbell

LAW RELATED PUBLICATIONS AND PRESENTATIONS

Craig Ball is a prolific contributor to continuing legal and professional education programs throughout the United States, having delivered over 450 presentations and papers. Craig’s articles on forensic technology and electronic discovery frequently appear in the national media, including in American Bar Association, ATLA and American Lawyer Media print and online publications. He also writes a monthly column on computer forensics and e-discovery for Law Technology News called "Ball in your Court," which received which is the 2007 Gold Medal honoree as “Best Regular Column” as awarded by Trade Association Business Publications International. It’s also the 2007 Silver Medalist honoree of the American Society of Business Publication Editors as “Best Contributed Column” and their 2006 Silver Medalist honoree as “Best Feature Series” and “Best Contributed Column.” The presentation, "PowerPersuasion: Craig Ball on PowerPoint," is consistently the top rated educational program at the ABA TechShow.