

Getting to the Drive

Distinguishing between preservation and production can be a challenge.

By Craig Ball

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.

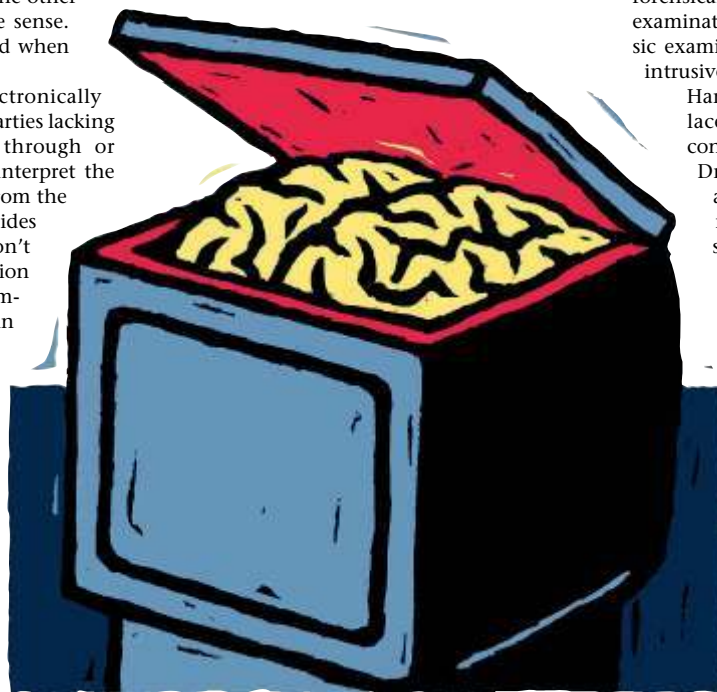
But sometimes, you've just got to "get to the drive." That means securing forensically qualified duplicates of relevant computer disk drives used by the other side, and having them examined by a qualified expert. Often lumped together, it's important to consider these tasks independently because each 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

See Ball Page 51

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. **LTN**

Craig Ball, a member of the editorial advisory boards of both LTN and Law.com Legal Technology, is a trial lawyer and computer forensics/EDD special master, based in Austin. E-mail: craig@ball.net.

McConnell from Page 45

trafficked areas of the firmwide intranet site — including each attorney's personalized home page, and every department page.

The entire project was developed within 10 months, (including approval time, testing, and graphics work), and went live in March, 2005. Measuring usage through hit counts has proved nearly impossible because of our (suc-

The site has played a strong role in the increase of hours.

cessful) efforts to inter-connect the site to other pages via open portals — particularly each attorney's intranet home page. Nonetheless, I received hundreds of "I'm interested" pings (and responded to at least that many informal inquiries generated by the site which are not easily tracked).

FACTORS

Ultimately, it is overly simplistic to say that any particular factor can be singled out as the cause for an increase or decrease in our pro bono productivity. Our commercial activity levels, compensation and pro bono policies, the quality of case referrals, and many other issues impact pro bono productivity.

Nonetheless, based on the activity levels on the sites, the firm is convinced that the site has played a strong role in the increase of our hours — from approximately 26,000 in 2002 to in excess of 41,000 in 2006.

And among the accolades we have received, we were delighted to be the recipient of the 2007 LTN Award for Most Innovative Use of Technology in a Pro Bono Project. **LTN**

Greg McConnell is the director of public interest law at Winston & Strawn, based in Chicago. E-mail: GMcConnell@winston.com.

IT director Chip Goodman, and senior programmer Cheryl Garrett contributed to this article.

Information: Reader Response card xx.

The Page Equivalency Myth

Describing a gigabyte is about as easy as counting angels dancing on pins.

By Craig Ball

When 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 .tif images and native formats — spreadsheets, word-processed documents, e-mail, compressed archives, maps, photos, engineering drawings — how sensible 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 gigabyte laptops. That means we will have to review 40 million pages per machine. Converting those pages to .tif images will cost Megacorp \$4 million 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 gigabyte 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



We talk about equivalency with all the credibility of an Elvis sighting.

Regarding Discovery of Electronically-Stored Information." Scholarly articles and reported decisions pass around the 500,000 pages per gigabyte value like a bad cold. Yet, 500,000 pages per gigabyte 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 about 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 life span 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 .tif images of 8½-inch x 11-inch documents scanned at 300 dots per inch 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 .tif 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 up to as many as 61 pages.

Billed by the GB, you'll pay almost five times more for the article as two .tif 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. **LTN**

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