

# InsideCounsel

A time to reap, a time to cull

By Clifford F. Shnier

June 12, 2008

Back in the “old days,” business records and communications were kept in the form of paper. Responding to a discovery request at the direction of in-house or outside counsel was straightforward. Paper files were stored in visibly labeled file cabinets and within those cabinets, in labeled folders. The folders that appeared relevant were piled into boxes and taken for photocopying.

Electronically Stored Information (ESI) is harder to find. It requires a team effort of inside counsel, outside counsel, people in the corporate enterprise with knowledge of the information infrastructure, and outside consultants who understand the infrastructure they’re being shown.

Once found, it must be collected properly. If collected incorrectly, ESI may be inadvertently changed, and subject to the objection of spoliation. Improper collection of electronic data is a common mistake made by companies and their counsel in discovery today.

The “gold-standard” of electronic data collection is forensic bit-stream imaging. This makes an exact “bit-by-bit” copy of the target hard drives of the servers of the individual desktops and laptops of the persons identified as the relevant custodians of electronic data. It captures everything on that hard drive—system files, deleted files not yet overwritten, program files, ordinary data files and even Internet browsing history. While not every case requires this form of forensic data collection, many do, and other cases require the advice of outside experts when the collection is performed.

Data collection without full forensic imaging is traditionally done by a technician going on-site and copying the identified target files from the identified machines, in a manner that is documented and defensible. While it is not as comprehensive as taking a complete bit-by-bit image copy, for cases where that level of authentication is not required, it is commonly done. But it still has to be done right. It also requires a lot of leg work to go from machine to machine to copy this data, almost as time-consuming as making a full bit-stream image.

Recently, the industry has seen the development of “appliances” (hardware with preloaded software) that “crawl” the corporate network, locating every server, PC or other device attached to the network and able to “see” what is on them. These devices can then collect the data from whatever custodians, machines, servers, folders, file shares, and file types that are designated, copying these to its own hard drive, which can range in size from 750 GB to 3 TB. These appliances are connected to the network “behind the firewall” within the corporate enterprise. They allow collection of data on a desktop in Des Moines by a technician working from Dallas. This is not forensic imaging of a hard drive, but it is an efficient way to copy active files. It is also an early opportunity to take a broad-brush reduction of the data, as you can set the copy parameters to file date, file type, etc. Some of these devices also permit keyword filtering at the time of data collection. A word of caution: the time of collection may be too early in the development of the case to have a good handle on what the keyword searches ought to be.

After the data is collected, the traditional next step in electronic discovery is called “processing.” Electronic data comes in many different formats and types: word documents, spreadsheets, e-mails, presentations, databases. If you tried to review electronic data using the native applications in which it was created you would have two problems. First, the mere act of opening up a file in its native application will change some of its

properties; second, it is unlikely you will have all the appropriate applications and versions on your system to permit this. Therefore you have to “process” that electronic data.

Processing electronic data involves turning each e-mail or document into a common format regardless of its original source application. That common format then permits the document to be loaded into the review software without causing any changes to the document itself so it retains its authenticity.

TIFF conversion was the standard for electronic discovery data processing for several years, and still is appropriate in many cases. However, native file review, without conversion to TIFF, is becoming more commonplace.

Processing native files for review involves taking a copy of each native file, as well as breaking apart each separate e-mail and any attachments—while maintaining the parent-attachment relationship—and then converting those native files and e-mails and attachments into a format where, while they remain “native,” viewing and notating is possible without changing the file itself.

The end product of the processing step, then, is

- A database record consisting of header information about that document: the date, from, to, and subject line of an e-mail, cross-references to its attachments, linked to
- The data itself in its native format, possibly rendered for viewing in HTML format;
- The full text of the document;
- And optionally with other renderings, accomplished by an actual conversion step, such as to TIFF or PDF.

The problem with processing everything that is collected (other than system and executable files that are easily removed and won’t process anyway) is that this can add up to a lot of money to process data that may largely be irrelevant. Some electronic discovery vendors perform keyword searching and culling on data after it has been processed in consultation with the client, so that some data volume reduction occurs prior to loading into the review software. This helps reduce review costs.

It is more economical if much of the irrelevant data is culled out prior to processing. Specialized software tools now permit this. Some of these reside within the new appliances referred to earlier. Others are programs that are used by consultants and e-discovery vendors, on the data prior to its being processed. One of them can “spear” its way into Outlook PST or OST files or Lotus NSF files and find only those e-mails or attachments that contain the search term. If you think of a PST file as a large ball of rubber bands, with each e-mail being one of the rubber bands, this tool can extract all of the green and blue rubber bands from within the ball, and leave the others. If desired, it can also deduplicate at that time. Another tool is able to index and search non e-mail files. These preprocessing technologies accomplish a reduction in the volume of data being loaded into the processing stage.

If you have an consultative e-discovery vendor who uses these preprocessing culling tools (and beware of those who don’t and tell you to process everything), you should look for staged pricing: one amount per gigabyte (or per hour) for performing the preprocessing culling on the incoming collected data, and then another amount per gigabyte for processing the smaller amount of data that remains after pre-processing culling.

The courts now accept and expect that search technologies will be used to reduce data to manageable proportions: *Zakre v. Norddeutsche Landesbank Girozentrale*, 2004 WL 764895 (S.D.N.Y. Apr. 9, 2004) adopted Sedona Principle 11: “A responding party may satisfy its good faith obligation to preserve and produce potentially responsive electronic data and documents by using electronic tools and processes, such as data sampling, searching, or the use of selection criteria, to identify data most likely to contain responsive

information.”

However, this doesn't mean that a quickly-brainstormed list is adequate. Proper search is a science and requires application of the scientific method. Courts have commented unfavorably on unilaterally-concocted and inadequate search strategies: *In re Seroquel Products Liability Litigation*, 244 F.R.D. 650, 661, 662 (D. Fla. 2007). In a multidistrict drug litigation, the plaintiffs sued defendants for claims arising from alleged injuries from ingesting a medication.

The Court extensively reviewed the defendant's discovery shortcomings and noted that “while key word searching is a recognized method to winnow relevant documents from large repositories, use of this technique must be a cooperative and informed process. Rather than working with Plaintiffs from the outset to reach agreement on appropriate and comprehensive search terms and methods, [defendants] . . . undertook the task in secret. Common sense dictates that sampling and other quality assurance techniques must be employed to meet requirements of completeness. If [defendants] . . . took such steps, it has not identified or validated them. . . . Examples include omitting Seroquel's generic name, acronyms for diabetes, hyperglycaemia spelled the British way, and endocrine. The search method apparently failed to include common misspellings or the singular forms of words and failed to make allowance for spaces or dashes.”

So, yes, the accepted strategies of the courts will be used to reduce data volumes, but those search strategies must be defensible and well-designed. If lawyers are going to rely on technology and search terms to help find the potentially relevant and privileged materials, they must understand how to develop a thorough and complete search protocol and they must be able to document it, in case the sufficiency of the production is questioned.

That's the subject of a future column.

*An integral actor in litigation support and e-discovery since the late 1980's, Clifford F. Shnier has been a consultant, owner of his own service bureau and a senior sales and management executive with major e-discovery and litigation support providers. Prior to that, Shnier practiced law in Toronto for eleven years, litigating complex commercial matters, torts and criminal cases, with extensive trial and appellate courtroom experience. Based in Scottsdale, Ariz., since 1994, he is active in the e-discovery arena on both sides of the U.S.-Canadian border.*

(c) 2008 *InsideCounsel*. A [Highline Media](#) publication. All rights reserved.