

Digital Discovery & e-Evidence

BEST PRACTICES & EVOLVING LAW

www.ddee.pf.com

Reprinted from Vol. 5, No. 5 | May 2005

Are Market Forces and Technological Advances Already Making the “Reasonably Accessible” Category Obsolete?

By Jake Frazier, Esq.

The proposed amendment to Federal Rule of Civil Procedure 26(b)(2) would establish a special category for electronic information that is deemed not to be “reasonably accessible.” Electronic information in this category would be somewhat protected from discovery requests as the requesting party would have a higher burden for establishing the value of this information to its case and the need for its production.

This proposed rule change is designed to restore balance to the electronic discovery landscape. In recent years, the costs associated with locating, restoring, reviewing and producing information from disaster recovery information systems have often proved prohibitive for the responding party. When faced with daunting discovery costs, the balance of power between plaintiff and defendant is often lost, and the litigation quickly devolves into a case about discovery and meta-discovery while the merits of the case take a back seat¹.

However, history has proven that a single e-mail is often the “smoking gun” or “white knight” in large civil litigation cases.² It is no surprise that e-mail stored in any format by a defendant is usually the most coveted target for plaintiff’s discovery requests. The public comments on the proposed rule changes are replete with excellent arguments both for and against the controversial establishment of a special category of information that is not “reasonably accessible.”

The hearings on the proposed changes primarily sought to clarify what criteria to use for categorizing information as “not reasonably accessible.” The advisory committee notes offer some guidance by referring to information that involves significant expense and effort to restore, often requiring the acquisition of special hardware and/or software.

The notes also explain that this information is most often one of three types: a) information that exists on backup media designed for disaster recovery, b) information that exists in outdated or legacy formats, and c) information that exists only in a deleted state. This article will focus on how emerging technology will affect the cost of recovering the first two types of information and the effect the proposed rules would have on this process.

Disaster Recovery Backup Systems

Backup tapes are probably the most common form of electronically discoverable information and are largely responsible for creating the uproar about the costs of electronic discovery. (See related story at p.10.) Disaster recovery backup systems were designed to create snapshots in time of what a computer system’s information store looked like. The vast majority of backup systems employ magnetic tape to store this information due to its relatively low cost. Other systems use DVD-ROM

or some other storage media, but the effect is largely the same.

Most backup systems were designed for disaster recovery and enjoyed the height of their popularity as the corporate world prepared for the anti-climactic Y2K “disaster” that wasn’t. Meant for restoring an entire network a day or two after a disaster, backup systems do not lend themselves to restoring information easily years after it was originally saved. They were definitely not designed for restoring subsets of e-mails, running keyword searches, or any of the tasks required in electronic discovery projects.

The real difficulty for those involved in electronic discovery stems from the specific manner in which backup systems encode information onto storage media. This coding can utilize one of two general procedures for backing up information: simple or complex. Simple coding essentially writes the files on the tapes sequentially: files A-D will be written as complete files and in their original order, A, B, C, D.

Restoration of files stored in sequential order from backup media is relatively simple, though many times a file being written to a backup tape carries a remainder when the tape is full. In that case, the remainder of the file is written to the next tape in a process called file “spanning.” Both tapes must be present to restore the original file in these circumstances.

Today, this simple or sequential coding is being replaced by a more efficient, but more complicated method. Newer backup technology uses complex or threaded storage.

Threading refers to the practice of placing small amounts of information on various tapes wherever there happens to be a good space for it. This is a much more efficient storage scheme and reduces the cost of backup media, which can be millions of dollars for larger corporations.

This complex encoding is not like bagging groceries, however. Whereas each grocery bag contains several intact items, threaded storage media act like a food processor, reducing all your food into the smallest possible size to save space and requiring you to reconstruct the individual items when you

get home. Perhaps it is not that difficult, but it is a system where Files A-D might be fragmented and written in this order: a1, d2, c1, a2, etc.

With complex backup systems like these, it is sometimes necessary to reconstruct the entire backup collection, just to restore one e-mail. This makes matters difficult for responding parties who have been ordered to produce information from backup systems that utilize “threading.”

“Reasonable Accessibility”

Non-native extraction is a relatively new technology that makes it possible to extract important data from various backup systems and legacy formats without the need to recreate the native environment. Non-native extraction identifies the digital code on the tapes and pieces together the file fragments. Because the process entails reading the code directly from the backup tapes, DVD, or other media, it is agnostic as to the software program originally used to author the data. Perhaps more importantly, it is also agnostic as to the operating system in which the data was originally created.

In addition, this technology is capable of extracting only non-duplicative information from the backup media, drastically reducing the storage costs associated with electronic discovery projects. Electronic discovery projects using non-native extraction can usually be completed in approximately 20 percent of the time required for native environment restoration.

This combination of savings from reduced software and hardware expenditures, reduced storage costs, and reduced personnel hours should be taken into account in determining how “reasonably accessible” data is.

This is not to suggest that non-native extraction is a panacea that will eliminate all difficulty in electronic discovery. Rather, it exemplifies how quickly technological enhancements can change the landscape of electronic discovery. As non-native extraction becomes more readily available, responding parties may be required merely to locate their backup tapes, extract the data into a database and run keyword searches to perform their initial cull. If the burden of recreating native environments is eliminated, then restoration of information from disaster

recovery systems becomes more “reasonably accessible” in the same way that determining the precedential value of a court decision became more “reasonably accessible” with the advent of online KeyCiting®³ and Shepardizing®⁴.

Market Forces

In addition to non-native extraction, two other technologies, “concept search” engines and “proactive discovery management” solutions, are easing the burden and reducing the costs of electronic discovery production.

Concept searching is a relatively new method of conducting a review within a responsive information set. After the responsive set is determined by using keywords and Boolean logic to perform the initial cull of the overall information store, concept searching can drastically reduce the amount of time needed to conduct a privilege review. Several firms offer proprietary technology that allows lawyers to identify those documents that might be privileged and need further review.

For example, assume a defendant corporation has privileged communications being exchanged between counsel and various members of the party regarding the strategy of using the Diamond case as a precedent for a defense, or the analysis of potential testimony from a Mr. Diamond. However, what if that party is also a jewelry distribution company and has millions of e-mails about diamonds in its information store. If the defendant corporation were ordered to review and produce from its active data, that alone would make the process very labor intensive. If, however, it were ordered to review and produce from its backup tapes as well, the workload would explode exponentially.

Fortunately, with some of today’s more ambitious search engines employing capabilities and technologies such as “artificial intelligence,” “concept searching,” “linguistic pattern matching,” and “neural networks,” it may be possible for attorneys to construct a search that will “know” to exclude those e-mails that are about diamonds, the gemstones, leaving only a fraction of the initial responsive set for review. It may be too early in the lifespan of this technology to make any bold predictions, but it could very well ease the burden of production in the future.

Another promising approach, the proactive creation of discovery retention systems, is designed to create information stores completely separate from backup systems. This proactive tactic is emerging in the marketplace as an attractive option to the crisis-response paradigms now being employed by corporations. With the largest corporations reporting hundreds of lawsuits being litigated at any given moment it is entirely possible that the preservation holds affecting the company will overlap so much as to almost dictate the inadvertent creation of a *de facto* litigation repository.

Some forward-looking companies have analyzed this situation and found it to be less expensive to proactively maintain a litigation database of electronic discovery that is ready to be searched and culled at any time. It only takes a few productions from this repository to realize a full return on investment, and it is an extremely effective insurance policy against sanctions and spoliation claims.

Non-native extraction, concept searching, and proactive preservation archiving are all technologies that have only risen to prominence because of increased demand. Corporations require tools for responding to discovery requests and the market is providing them. Because market forces can, and historically do, play such a large part in restoring balance to inequities such as those created by leveraged electronic discovery requests that are merely “fishing expeditions,” then rule amendments need to be written with care not to impede these market forces’ progress. If the rules accord information on “backup tapes” an extra layer of protection from discovery, for example, then perhaps corporations might be encouraged to use antiquated techniques for a strategic advantage. Corporations would have an incentive for keeping the backup systems that are the most difficult to access, which would impede the market’s progress in some of the most innovative and promising technologies.

Allowing the market and the test currently outlined in FRCP 26(b)(2)(iii) to correct the imbalance will ultimately improve the quality of productions. If the market is allowed to evolve, the enhanced technology will reduce the cost of electronic discovery while yielding much more responsive information, more thorough exculpatory and incrimi-

nating evidence, a reduced amount of privileged information accidentally divulged, and a reduced amount of information being destroyed as a “routine business practice.”

Future Outlook

If the proposed changes to the FRCP are eventually enacted, it is vitally important that they be designed to withstand the test of time. To do this, they will need to be flexible enough to be applied in environments and situations with variables that are not yet contemplated. However, much of the commentary about proposed rule 26(b)(2) calls for *more* specificity. If the rules are to move gracefully through a constantly changing technological environment, over-specificity is a potential snag that could derail them and make another round of amendments necessary.

A relatively straightforward pitfall would be automatically considering “disaster recovery” or “backup tapes” as classes of information that presents an extra burden for the requesting party. This concept could become outdated in just a few years, a few months, or even sooner.

Non-native extraction aside, what if Storage Area Networks (SANs) experience an exponential growth in capacity due to technological breakthroughs; would we need to draft new rules again, because virtually all information would then be “reasonably accessible”? Rule 26(b)(2) as currently written could withstand that change (the second tier would just become moot). However, the commentary and the dicta in the case law would be too rigid and would

be inapplicable absent some alteration. It would be as if the rules in 1993 absolved corporations with PCs containing 386 speed CPUs from ever producing graphics (because their machines couldn’t accommodate the size of graphics files) but required corporations with PCs containing 486 CPU’s to produce them. If the rules are dependent on how today’s technology defines what is not easily accessible, will future courts and litigators be forced to recreate the “native legal environment” of 2005 in order to interpret the rules in light of future technological breakthroughs?

Judges are the best tool the legal system has for correcting any imbalances currently presented by electronic discovery. Just as *Daubert*⁵ required the bench to become mini-experts in scientific evidence, electronic discovery is doing the same with regard to information technology. It would seem that the tools needed to ensure a balanced yet flexible body of case law exist already in FRCP 26(b)(2)(iii).

Endnotes

¹ See Generally *Coleman (Parent) Holdings inc. v. Morgan Stanley & Co. Inc.*, 2005 WL 674885 (Fla.Cir.Ct.).

² See Generally *U.S. v. Arthur Andersen, LLP*, 374 F.3d 281, C.A.5 (Tex.),2004.

³ Keycite® is a registered trademark of The Thomson Corporation. For more information, visit www.keycite.com.

⁴ Shepardize® is a registered trademark of Reed-Elsevier, PLC. For more information, visit www.lexisnexis.com/shepards.

⁵ See *Generally Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 43 F.3d 1311, C.A.9, 1995 (requiring judge to use some objective scientific analysis).

Jake Frazier, Esq. is a licensed attorney and the senior manager of Legal Research and Education at RenewData in Austin, Texas. Mr. Frazier has extensive experience in the legal research and education industry and also received the International Academy of Trial Lawyers Award for Advocacy and the Willard H. Pedrick Scholar Award while earning his Juris Doctorate from Arizona State University. Mr. Frazier can be reached at jfrazier@renewdata.com.