

Zubulake – Toward a Path of Reasonableness in Computer Discovery

Deborah H. Juhnke, Vice President
Computer Forensics Inc.TM

The pursuit of digital data in civil litigation yields increasing volumes of information, and it is easy to become overwhelmed with both the cost and complexity of data review. Because the Federal Rules do not offer specific guidance for managing electronic discovery, we must rely on case law and common sense. In 2002, the decision in *Rowe Entertainment, Inc. v. William Morris Agency, Inc.*, 204 F.R.D. 421 (S.D.N.Y. 2002) paved the way for a multi-part test for cost allocation. The recent rulings¹ in *Zubulake v. UBS Warburg* bring a welcome refinement to the *Rowe* opinion and, most important, support the use of sampling in the discovery of what Judge Scheindlin terms “inaccessible data.”

The reasonableness of the sampling approach is another step in the right direction toward limiting the total amount of digital data reviewed in discovery. It is not uncommon to hear stories of hundreds of backup tapes being restored and reviewed, with the greatest benefit often going only to the vendor whose fees can approach the value of the litigation. By adjusting the multi-part test of *Rowe* and requiring that a sample of the backup tapes be reviewed first, the *Zubulake* court has helped put the brakes on indiscriminate data review.

Technical Perspective

The Court’s discussion of accessible versus inaccessible data formats and acknowledgement that large corporations are moving toward paperless environments helps support the idea that all data should not be treated equally when it comes to discovery. Although the facts of the case do not lead the Court to consider sources of email other than backup tapes and optical disks, the concept of selective review is appropriate to many other sources as well.

Digital discovery has become a battle of the sophisticates. A clear understanding of what data exists where and in what form must precede any judgment regarding its accessibility. Without the advice of a computer forensics expert or an attorney highly experienced in digital discovery, accessible sources of data may go untapped, while discovery arguments proceed over inaccessible sources.

The second test of the cost-shifting analysis then proves critical: Is the information available from other sources? Today, collaboration software, data warehouses, ISP-hosted email², and web-based content all present opportunities for both planned and indiscriminate archiving and dissemination of corporate information. In the context of

¹ May 13, 2003: Ordering a sampling of backup tapes, defining “inaccessible” format, and modifying the *Rowe* 8-point test to a 7-point test; July 30, 2003: Ordering that *Zubulake* share 25 percent of the costs of restoration, but that UBS pay for any costs incurred for review of data.

² For example, Yahoo mail, AOL mail, etc.

Zubulake, could these technologies also have presented opportunities for finding digital evidence from sources other than backup tapes? Perhaps not. The question, however, must be asked in any thoughtful discovery process. For example, Instant Messaging (IM) is becoming ubiquitous, is generally unmonitored, and offers a great way to circumvent email policies. Had IM been in use at UBS Warburg during the relevant time period, it could have been the most likely and easily retrieved source of candid intra-office discussions regarding Ms. Zubulake's tenure.

Presuming that the only available digital evidence is in an inaccessible format such as backup tapes, careful selection of the sample is paramount to providing an accurate representation of what might be contained in other tapes. This filtering may occur at multiple levels, depending on the entities involved and circumstances of the case. Name, date range, and backup protocol are three of the most common filtering mechanisms. When filtering at the file level (searching), it is also useful to think beyond the obvious (searching by name only) and consider that informal correspondence does not always identify the subject by name. "She" and "her" are just as descriptive when the subject is understood. Identifying and searching for other terms will often yield a more complete result.

The costs and time associated with computer-based discovery can be greatly minimized with a little prior planning. Careful selection of datasets, filtering, and sampling all offer ways to focus discovery efforts and limit unnecessary collection.

A development that may affect future application of "accessible vs. inaccessible" to data backups is that corporate America may be moving beyond backup tapes into the world of "data protection appliances." These are devices that can keep an initial copy of a protected drive and log changes at intervals as short as 30 seconds. Disk-based backups such as these may soon supplant backup tapes, with the singular goal of data recovery, not archiving. These backups will, by Judge Scheindlin's definition, be accessible yet fleeting, and preservation will become the greater issue. However, for now tapes continue to grow and by 2010 S-AIT tapes (super advanced intelligent tape) may hold as much as 4 terabytes³ per tape.

Can a sample alone be adequate to support the issues? Possibly yes. If, through sampling, both relevant and supporting evidence is found, then what should be the obligation to search further? To use a rough analogy—does shooting a dead body make it any more dead? In cases such as *Zubulake*, the sample will only serve to whet the appetite and point the way to more definitive evidence in the whole. In others, it may tell the story well enough that further restoration and review is unnecessary.

The third test—total cost of production compared to the amount in controversy—takes on a special meaning when coupled with sampling techniques and efficient digital discovery. What was once a cost-prohibitive process wielded as a weapon by defendants, can be made

³ A terabyte would yield approximately 50 billion printed pages.

more manageable. Although restoration and review of backup data can be costly, the option of sampling opens the digital discovery door for cases in which it would otherwise have been ignored. Cost-saving techniques such as early review of data in native format (E-triage) and creation of an e-discovery plan, and digital discovery becomes feasible for the small and mid-size case as well.

Implications for E-Risk Management

One of the greatest costs to corporations in digital data discovery is first *finding* the sources of potentially relevant data, then gathering and reviewing that data. Judge Scheindlin's July 30, 2003 ruling makes it clear that cost-shifting is not to be assumed and that corporations can be compelled to produce what they keep. Well-defined and enforced data retention and categorization policies then become even more important as ways to minimize the amount of superfluous data available for review and production. It is the responsibility of every corporation both to protect its data and to minimize the risks and costs associated with keeping it. Yet the tacit policy of most corporations seems to allow retention of far too much data.

Why keep what can't be accessed? Maintaining corporate data on media or in formats that can't easily be read is akin to maintaining a library of 78-rpm records. Some companies have run into discovery troubles when they were compelled to produce data that was not easily retrieved. The likelihood that data must be produced if maintained may help create incentives to develop better systems, and to update document retention policies and backup systems. This is particularly true for unregulated industries, whose recordkeeping is not mandated by statute.

Segregation and categorization of backup data is a good place to start. Because of the special rules governing securities trading, UBS Warburg chose to archive certain email on optical disks. Although this procedure captured necessary correspondence, it also captured "all emails sent to or received from outside sources." Arguably, not all email is business-related, thus this type of system will end up capturing much more than necessary. Because optical disks can't be erased, care should be taken to archive only business records on optical media and to segregate them according to retention requirements.

One may argue that there is little benefit in making data more accessible to the opposition. In fact, one can draw a parallel between our haphazard digital record keeping and the warehouse and boxcar discovery practices of the late 70's. Although there may be little apparent incentive to keep data in a tidier, more searchable form for potential litigation, doing so can offer enormous benefits for corporate management. These include reduced storage costs and increased efficiencies when seeking data for day-to-day business purposes.

Implications for Discovery Practice and Litigation Support

It is difficult to argue tactics in a vacuum. Digital discovery will be driven by a number of factors, including the sophistication of the parties, the perceived availability of digital evidence, and the overall tenor of the litigation. That said, taking a “bull in the china shop” approach using unedited boilerplate discovery requests is a certain path to expensive discovery motion practice and potentially punitive rulings. *Zubulake’s* first test—specificity of the request—should provide incentives to take care when drafting discovery pleadings. Instead, we have continued in this digital age to follow the same practices used for decades in paper-based discovery. We request “all” and “any,” without regard to the possibility that our requests may now be too broad. A better approach toward self-policing our discovery would be to focus on finding all the *relevant* data, not just all data.

Fun with Numbers

Test number six—the relative ability of each party to control its costs and its incentive to do so—goes to the heart of historical practices that lead us to collect *and manipulate* too much data. In many ways, digital data is cheaper and easier to produce than paper. Yet many insist on immediately converting that nice, searchable digital format into its paper equivalent—TIFF or PDF images—at great expense. There is a time when such conversion is appropriate if not necessary, but it is typically long after significant filtering has been done.

The threat of high production costs, sometimes inflated, has been used for years as leverage in discovery. Unfortunately, there is little incentive on the part of litigation support providers and their clients to control cost estimates, which are often used to gain early compliance or concessions, before any processing is begun. It is easy to produce an estimate that says a tape restoration project will cost millions of dollars, when in fact a more careful review of the project would dramatically limit the costs. A tactical approach? Yes, but one not necessarily grounded in forward-thinking and responsible computer discovery practice, and one that can come back and bite you in the end.

Care should be taken, therefore, when analyzing or extrapolating costs. Restoration and review of the first sample of backup tapes will generally include costs not associated with review of subsequent tapes. For example, in her opinion and order dated July 30, 2003, Judge Scheindlin calculated that restoration of the remainder of the tape collection would cost roughly \$3,800 per tape—a number derived by dividing the cost of the sample (\$19,003) by 5. In practice, however, there are at least three areas where costs could actually be lower.

First, the initial sample included development time for the search script. Most of this time would not be required for processing subsequent backup tapes.

Second, presuming the initial restoration costs included some set-up and configuration time for the computing environment (servers, software, etc.), this component would also

not be necessary for subsequent backup tapes, since the computers would already have been configured.

Third, there are economies of scale in mass processing of tapes, thus the total number of hours required of both the computers and personnel might be reduced by up to 20 percent.

Judge Scheindlin recognizes that an outside vendor will perform this restoration work, and that “once a vendor is chosen the costs are not within the control of either party.”

Vendors will, however, have differing incentives, pricing models, and sophistication, thus the initial selection of a vendor becomes even more critical. Take, for example, a situation where a vendor is approached with a restoration project for 100 tapes. If they are commodity vendors, their incentive will be to process all tapes equally and they may not be able or willing to recommend lower-cost alternatives.

We must develop an understanding of hardware and software beyond our personal experience to adequately pursue or defend computer discovery in litigation. It is likewise easy to take a “been there, done that” attitude computer discovery, but the times are quickly changing. Case law will always reflect the facts, and until court rules are modified to address advances in technology as they relate to discovery, we will be continually challenged to plough new ground. Judge Scheindlin’s insights into the world of computer discovery are a welcome step in the right direction.