RECEIVED
12/21/04

"Mike Heidler"
<michael.heidler@yale.edu>
12/20/2004 10:27 PM

To <peter_mccabe@ao.uscourts.gov>
cc
Subject Request to testify at February 11, 2005 Civil Rules Hearing on E-Discovery

04-CV-057
Request To Testify
2/11 DC

Dear Mr. McCabe,

I write to formally request the opportunity to testify at the public hearing on the proposed amendments to the Federal Rules of Civil Procedure scheduled for February 11, 2005, in Washington, D.C. I am one of four Yale Law School students who will also be submitting written comments on the proposed amendments.

I appreciate the opportunity to participate in this process.

Sincerely,

Mike Heidler
michael.heidler@yale.edu
111 Park Street #14-P
New Haven, CT 06511
(203) 668-5276

RECEIVED 2/3/05

04-CV-057

Testimony 2/11 DC

February 2, 2005

Michael Heidler
111 Park Street #14-P
New Haven, CT 06511
michael.heidler@yale.edu

Peter G. McCabe, Esq.
Secretary
Committee on Rules of Practice and Procedure
Judicial Conference of the United States
Thurgood Marshall Federal Judiciary Building
Washington, DC 20544

RE: Proposed Amendments to the Federal Rules of Civil Procedure for E-Discovery

Dear Mr. McCabe:

Thank you for the opportunity to comment on the proposed amendments to the Federal Rules of Civil Procedure for e-discovery. I can provide helpful suggestions that directly arise from my lifelong commitment to technology. I authored my first computer program at the age of four. Eighteen years later I completed college where I studied computer programming, database development and technology planning.[1] I spent the next four years managing and working on a range of technology projects for an information technology consulting firm.[2] My projects varied in size and complexity: at one point, I personally programmed a simple case management system for a non-profit legal services organization; I later managed a seven-person team that integrated four enterprise databases, interacted with two operating systems and constructed a networked enterprise software application for over 10,000 employees at a Fortune 500 pharmaceutical company. These hands-on encounters with technology, combined with my company's rigorous cost-estimating methodology, also enabled me to author dozens of proposals that estimated the costs required to complete technology endeavors.[3] These experiences inform my comments on e-discovery.

---

[1] University of Texas at Austin, Management Information Systems, 2000.
[2] Catapult Systems, Inc.
[3] The ability to accurately estimate technology costs was critical to my role. Most firms select technology vendors based on references and price. If a vendor underestimates the cost of a project, it will be unable obtain favorable references and will have trouble bidding on new proposals. If a vendor overestimates the cost of a project, then it will lose to vendor with a lower bid.

Rule 26 Should Explicitly Allow for Cost-Shifting and Should Include the "Reasonably Accessible" Standard.

Rule 26 should explicitly address and condone cost-shifting. The "reasonably accessible" standard should remain in the Rule, for it effectively balances the interests of requesting and responding parties.

The Proposed Amendment to 26(b)(2) Should Explicitly Allow for Cost-Shifting.

The proposed amendment to 26(b)(2) should allow discovery of data that is not reasonably accessible if the requesting party is willing to financially contribute to discovery. The Committee Note on the good-cause analysis states that "[t]he good-cause analysis would balance the requesting party's need for the information and the *burden on the responding party*."[4] (emphasis added) It is unclear whether this burden would account for cost-shifting; if a requesting party were willing to pay for discovery, the burden on the responding party would be negligible. It seems fair to allow the requesting party to discover data in this scenario. The proposed amendment would be stronger if it explicitly addressed and condoned this possibility.

The "Reasonably Accessible" Standard is Necessary for Electronic Discovery Because Unlike Paper Documents, Electronic Data Must be Restored, and the Technologies upon which it Depends Can Become Obsolete.

The "reasonably accessible" standard is necessary because accessing electronic information is inherently different from accessing paper documents. There are valid observations to be made about the enormous volume of data on sprawling corporate networks and the costs of halting or modifying the routine backup operations at big businesses. While these observations are important, I would like to consider the burden Rule 26(b)(2) could impose on small to medium-sized organizations if it lacked the "reasonably accessible" standard. As I explain below, my experience suggests that restoring data produced by an obsolete system could cost small to medium-sized organizations over $100,000 and thus make data not reasonably accessible to them.

[4] Proposed Amendments to the Federal Rules of Civil Procedure, May 17, 2004, revised, August 3, 2004, p. 14.

Electronic Data Must be Restored.

A party would need to perform at least two steps to restore data archived on tape. The party would first need to restore the data from tape to disk, and *Zubulake v. UBS Warburg* warns that this process can require significant time and expense.[5] The party would then need to restore the disk data to a software application and hardware environment, and this process can be similarly burdensome.[6]

The "reasonably accessible" standard is difficult to evaluate because it is difficult to determine just how burdensome electronic discovery can be. The task is complex: the cost electronic discovery depends on hundreds of independent factors, any one of which could alter the overall cost by thousands of dollars. This is precisely the situation I faced when I professionally estimated the costs of technology projects. I found utility in the law of large numbers: any given factor could add or subtract from the overall cost, but when taken together, the sum of many small estimates proves fairly accurate. I will therefore identify the factors involved in the *second* restoration step for a medium-sized data restoration project and produce a range of possible costs. I will consider a scenario in which a party must restore data produced by a standard business system and backed up from a database management system.

A party in this situation would first need to restore its data from tape, and I do not consider that step here. After that, the party would need to load the data into an application environment, and I routinely performed this task as a technology consultant. To access the archived data, I would first acquire computer hardware to host the data.

---

[5] For example, see *Zubulake v. UBS Warburg, LLC*, 217 F.R.D 309, 314 (2003). There, UBS needed to restore 94 backup devices at a rate of five days per device. Alternatively, it could have hired technical experts "at greatly enhanced costs" to speed up the restoration. Moreover, this process was already streamlined because UBS used a modern archival technology with a special indexing feature. This feature enabled technical staff to determine, before any restoration took place, that the information they needed was on 94 specific backup devices. If a party's archival technology lacked this indexing feature, it could conceivably need to restore many more tapes just to determine which ones contain discoverable data.

[6] Users need software applications in order to view and search electronic data. Some data formats, such as ASCII text, are standardized and do not create an obstacle here. Data in other formats is relatively useless without specialized software applications, database management systems or both. For example, I have worked with complex client-server document management systems that store binary documents in large database systems. A data backup of such a system would contain only the data file created by the database management system. To interpret that data, a responding party would need to install the database software and restore the database that created data file. Because the internal document data is also binary, the party would need to install the client software designed to access the database, configure it to access the database, configure user accounts to access the desired documents, and transfer the documents to a client machine. If the documents were in a proprietary format (for example, Microsoft Word, WordPerfect or CAD), then the party would need to install and configure those software applications as well.

Often data is only meaningful in a specific database management system on a specific operating system, so I would then determine which foundational systems the data required; obtain software licenses for those systems; acquire the installation programs for the software; and install and configure the operating system and database management system.

Data can often be useless without the supporting application software.[7] The steps to install the application software would also vary from case to case. I would acquire, as needed, a second computer; a second operating system; a second operating system license; the application software; and the application software license. I would then install the application software and configure it to access the previously installed database system.

Business software is typically based on a user model that allows users to only view data that belongs to them. I would thus need either a list of user accounts or an administrative user account that could view all of the data. Either way, the necessary account passwords would be encrypted in the database, and I would need to find the encryption key and encryption algorithm to decrypt the password(s), login to the system and access the data.

The variable cost—the cost of the hourly labor required to restore the data—would be a product of the hourly rate and the number of hours required for restoration. I have worked for two technical consulting firms, and the hourly rates have varied from $85 to $350. Someone performing the steps mentioned above would need expertise in computer hardware, computer operating systems, database management systems, application software and encryption. A fair rate for someone with these skills is $135 per hour. The number of hours required to restore the data to an application environment would depend on the factors mentioned above, the quality of system documentation and the availability of the password encryption key and encryption algorithm. I typically would use this information to estimate the number of hours required to complete the assignment. I then would ask a colleague to do the same thing so I could compare our

[7] The databases accompanying commercial applications can be useless without the client software. One commercial system on which I worked relates data through numeric key values stored as comma-delimited strings in text fields; another system stores data in binary XML fragments in database fields; systems often store entire documents as binary data; application software often contains logic that performs necessary calculations on data, filters out old data or joins data in a meaningful way; and even professional, normalized data structures may not be descriptively named.

outcomes. In the proposed scenario, I would allow 85 to 125 hours to restore the data to its application environment. I asked a former colleague to review the hypothetical case; he independently estimated 60.5 to 150.2 hours, which closely approximates my estimate.[8] Therefore, using my figures and a rate of $135 per hour, the *variable* cost to restore data from medium-sized business system to a software application environment would be between $11,475 and $16,875.

The fixed costs—the costs of the computer hardware and software licenses—would vary depending on the size of the task. A small amount of data would require a small server, while a large amount of data could require a multiprocessor system with expensive operating system, database and application software licenses. The fixed costs could therefore range from nothing to over $50,000.[9] The overall cost of this step for a small to medium-sized organization would thus range from $11,475 to $66,875.

The full cost would also include the cost to restore data from tape. For example, if that cost were $30,000—which seems modest given the description in *Zubulake*[10]—then the total would range from $41,475 to $96,875.

The Technologies upon which Electronic Data Depends Can Become Obsolete.

The above steps assume that none of the technologies upon which the data relies are obsolete. This scenario is not representative of all cases, and the rules should account for other possibilities.

---

[8] Dan Holling and I have estimated and managed numerous technology projects together. He generated his estimate before he viewed mine. His primary concern is that one cannot produce a tight estimate with such little real information. For example, he correctly mentions that someone could spend between a week and a month installing and configuring an enterprise business system. He therefore also assumes that the consultant in the hypothetical scenario would install and configure a small to medium-sized business application that had professional system documentation.

[9] For examples of potential costs, see Oracle10g and SQL Server 2000 Price Comparison, http://www.microsoft.com/sql/evaluation/compare/pricecomparison.asp (last visited Jan. 26, 2005); Windows Server 2003 Pricing, http://www.microsoft.com/windowsserver2003/howtobuy/licensing/pricing.mspx, (last visited Jan. 26, 2005); Servers, http://h71016.www7.hp.com/dstore/SubFamMatrix.asp?oi=E9CED&BEID=19701&SBLID=&ProductLineId=431&FamilyId=1871&LowBaseId=11946&LowPrice=$499.00 (last visited Jan. 26, 2005); Servers, http://h71016.www7.hp.com/dstore/SubFamMatrix.asp?oi=E9CED&BEID=19701&SBLID=&ProductLineId=431&FamilyId=1256&LowBaseId=11684&LowPrice=$2,349.00 (last visited Jan. 26, 2005)

[10] In *Zubulake v. UBS Warburg, LLC*, 217 F.R.D 309, 314 (2003). the discovery request required restoring 94 backup devices at a rate of five days per device or hiring technical experts "at greatly enhanced costs" to more quickly restore the data.

First, data archival technologies rapidly evolved from the 1950's to the 1980's. The technologies changed in stages, and included simple reel-to-reel solutions; more sophisticated open-reel systems; helical scan devices; closed, single-reel cartridges; and other proprietary tape systems.[11] In *Zubulake v. UBS Warburg*, the discovery request covered data from 1999 to 2001, a period in which UBS used a single, modern archival technology.[12] The burden of responding to this request, while high compared to the burden of responding to requests for paper documents, was nevertheless low compared to the burden of responding to other possible requests for electronic data. Had the discovery request covered three years in the 1970's, 1980's or early 1990's, it would have implicated obsolete archival systems and required assistance from technical specialists with more narrow expertise.

Second, business systems have short lifespans.[13] When an organization replaces a business system, the data from the old system becomes legacy data. To access legacy data, organizations must acquire, implement and configure the correct versions of the old computer hardware, operating system, database management system and application software. A party could also be required to convert legacy data to a common format. The older the technologies, the more difficult it will be to restore the data.

For example, I once worked on a project that required a colleague to convert legacy data from an obsolete database management system. She could only find one software product (known as a "database driver") that could access the obsolete system. The company that produced the software was no longer in business, and the software unexpectedly crashed every time my colleague attempted to transfer the data. She thus had to transfer the data in small pieces, closely monitor each transfer and reassemble the data on the other side. A task that would normally take one or two days took over three weeks due to software obsolescence.

---

[11] See A brief History of Tape, http://www.exabyte.com/support/online/documentation/whitepapers/history.pdf (last visited Jan. 21, 2005)

[12] *Zubulake v. UBS Warburg, LLC*, 217 F.R.D 309, 314 (2003)

[13] Business systems implicate computer hardware, operating systems, network protocols, application software and database management systems. These variables are constantly changing. Computer hardware manufacturers create newer and faster chips, making the old ones obsolete and unused. Database vendors create new versions of their products with faster retrieval rates, stronger security and more advanced configuration options. Vendors usually stop supporting software when it is superseded by newer versions. Computer operating systems change in the same way. As these components age, the business systems that rely on them become difficult to maintain, and companies are forced to replace them with new, maintainable systems.

A three-week delay would add $20,250 to the cost of the project described above, which would bring the total for the second step to between $31,725 and $87,125. If the first step, which could require restoring data from an obsolete archival system, cost $50,000, then the total would range from $81,725 to $137,125 for a small to medium-sized organization to restore data into a business system.

Unlike paper documents, electronic data must be restored, and the technology upon which it relies can become obsolete. These factors are prevalent in the electronic world and can make data not reasonably accessible to small and medium-sized organizations.

Any Unwanted Incentives Created by the "Reasonably Accessible" Standard Would Not be Actionable.

The "reasonably accessible" standard has been challenged for creating incentives to implement inadequate archival systems and create legacy data by replacing business systems more often than necessary. Both of these concerns are unfounded.

Businesses Would Not Intentionally Implement Inadequate Archival Systems.

Critics have argued that the "reasonably accessible" standard creates incentives for potential responding parties to implement inadequate archival systems in order to thwart discovery requests. This concern is overstated. Businesses archive data to satisfy business needs—such as the need to recover from disasters—and adhere to legal requirements.[14] Therefore, if a business lacked these requirements, it would not implement *any* archival system, so the concern over inadequate systems is misplaced; alternatively, if a business were bound by these requirements, the requirements would naturally check the desire to make data inaccessible for discovery.

Businesses Would Not Replace Business Systems More Often than Necessary.

The "reasonably accessible" standard has been criticized for encouraging potential defendants to frequently replace their information systems and thereby turn otherwise discoverable data into "legacy data." This scenario is implausible. Business

[14] See The Basic Backup Guide, http://www.exabyte.com/support/online/documentation/whitepapers/basicbackup.pdf (last visited Jan. 21, 2005)

system migration projects are costly and comprehensive.[15] Business systems are usually implemented in the form of interlocking *modules*, with each module representing a discrete business function.[16] A representative business systems vendor currently offers software modules for sales; customer service; orders; marketing; and human resources. Each module often depends on other modules.

For example, when a customer orders a product on-line, she places her order through a single web page, but the retailer activates several software modules behind the scenes. The retailer's *customer service module* saves her profile information (e.g., name, address and phone number), while the *order module* saves her order information and begins processing her order. When the customer then calls to request the status of her order, the operator opens a window to the customer service module. The customer service module communicates with the orders module, allows the operator to simultaneously view the customer's name and order status, and enables the operator to respond, "Ms. Doe, your product should arrive in three days."

If the retailer in this scenario decided to replace a single module, the interaction among all of modules would require the company to replace or modify several of them at once. Moreover, replacing these modules would require a significant commitment from the managers and employees in the corresponding business units (e.g., customer service and order processing). Employees in these situations commonly set aside many day-to-day activities to participate in meetings and systems tests with technical personnel.[17]

While a company might prefer to hide its data from future litigants, it must consider numerous factors in system implementation decisions, including cost, schedule, business disruption, company morale, data loss and system performance.[18] A desire to harm potential litigants would be insufficient to justify the enormous resource commitment required to replace a business system.

---

[15] See Migrating Unix ERP Installations to a Windows Server Environment: A Qualitative Assessment of Business Impact, http://www.metagroup.com/webhost/ONLINE/777164/wp2093.pdf (last visited Jan. 21, 2005). In a survey of 24 companies that were replacing their business systems, respondents reported that the migration required seven months and affected at least 50% of organizational employees.

[16] *Ibid.*

[17] *Ibid.*

[18] *Ibid.*

Rule 34 Should Require Requests for Electronic Data to Specify at Least One Data Format.

The proposed amendment to Rule 34 potentially allows the responding party to unreasonably burden the requesting party by supplying electronic data and information in a cumbersome format. The proposed amendment to 34(b) does not require the requesting party to specify a data format. If the requesting party does not specify a format, 34(b)(ii) allows the responding party to produce the information "in a form in which it is ordinarily maintained."[19] This could create an incentive for parties to store information multiple formats, some of which would be extremely cumbersome to use, so they could produce data and information that would be of limited value.

The Committee could remove this possibility with a small change to lines 23 and 24 in 34(b). The relevant portion now reads, "[t]he request may specify the form in which electronically stored information is to be produced."[20] If this sentence read, "[t]he request *shall* specify the form *or forms* in which electronically stored information is to be produced," the sentence would then require the requesting party to specify at least one form for the information and would likely spur beneficial discussions between the requesting and responding parties. (emphases added)

Rule 37 Should Not Include a Safe Harbor.

The proposed amendment in 37(f) apparently represents a concern that computer systems will automatically delete discoverable data. I have worked with a variety of computer systems that serve a multitude of purposes, and my experience with these systems suggests the opposite. Most computer systems do not delete data, but archive, backup and maintain it. In addition, most systems that appear to delete data actually do not; they merely hide it from users without removing it from the underlying system.[21]

---

[19] Proposed Amendments to the Federal Rules of Civil Procedure, May 17, 2004, revised, August 3, 2004, p. 27.

[20] Proposed Amendments to the Federal Rules of Civil Procedure, May 17, 2004, revised, August 3, 2004, p. 26.

[21] In my experience, most companies use archival systems, so even deleted data would not usually be inaccessible for the reason mentioned in the proposed amendment. This was the case in *Zubulake v. UBS Warburg LLC*, where deleted e-mail messages were still available on backup tape. Moreover, every business system I have every used, developed or maintained has not actually *deleted* data. These systems rather *mark* data as hidden and thereby make it invisible to business users. A system administrator could always *unmark* the data to make it appear again.

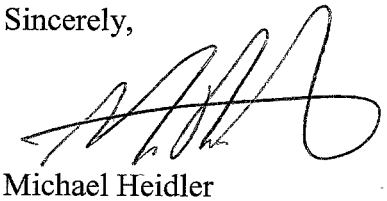
Adding the safe harbor provision would protect the occasional owner of a poorly-designed system that deletes useful data. It would also create incentives for parties to intentionally delete useful data under the guise of computer malfunction. Removing the safe harbor provision would hold parties accountable for deleting useful data, whether through accident or design, and I find the resulting penalties and incentives fair and desirable. Alternatively, if the Rule is based on known instances in which correctly functioning computer systems routinely delete data, the Committee could improve the Rule's text by narrowly tailoring it to the class of instances to which it is intended to apply.

Rule 45 Should Require Requests to Include at Least One Data Format.

For the reasons discussed in my comment on Rule 34, I suggest modifying lines 20-21 in 45(a). The relevant sentence currently reads, "[a] subpoena may specify the form in which electronically stored information is to be produced."[22] I suggest that the line read, "[a] subpoena *shall* specify the form *or forms* in which electronically stored information is to be produced." (emphases added)

I again thank you for the opportunity to comment on the proposed amendments.

Sincerely,



Michael Heidler

[22] Proposed Amendments to the Federal Rules of Civil Procedure, May 17, 2004, revised, August 3, 2004, p. 38