



"Rudy Kleysteuber"  
<rudy.kleysteuber@yale.edu>

>  
12/14/2004 12:06 AM

To <peter\_mccabe@ao.uscourts.gov>

cc

Subject Request to testify at Washington, DC, rules hearing on 11  
Feb. 2005

RECEIVED  
12/14/04

04-CV-049  
Request to Testify  
2/11 DC

Dear Mr. McCabe,

I am writing to request formally an opportunity to testify at the public hearing on the proposed amendments to the Federal Rules of Civil Procedure regarding electronic discovery, scheduled for February 11, 2005, in Washington, D.C.

I also intend to submit written comments on the proposed rules. My personal testimony would concern the interaction between the Federal Rules and the costs of electronic discovery.

I appreciate the opportunity to be a part of this process. Please let me know if you need any more information.

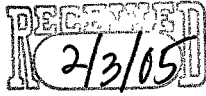
Thanks very much,

Rudy Kleysteuber

P.O. Box 200417  
New Haven, CT 06520-0417  
(203) 675-6828  
rudy.kleysteuber@yale.edu

Rudy Kleysteuber  
P.O. Box 200417  
New Haven, CT, 06510

04-CV-049  
Testimony 2/11 DC



2 February 2005

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

Dear Mr. McCabe and Members of the Committee:

I come to this process with a concern about how the perceived costs of electronic discovery have influenced and informed the drafting of the rule changes proposed in that area. My comments are inspired by my own experience with technology,<sup>1</sup> combined with Judge Rosenthal's appearance at Yale Law School in December. I am grateful to the Committee for taking my thoughts into consideration and hearing my testimony.

**THE PROPOSED REVISIONS TO RULE 26(B)(2) AND 37(F) ARE MOTIVATED BY COSTS THAT WILL CHANGE DRAMATICALLY AND UNPREDICTABLY IN THE FUTURE**

My basic point is that the costs of electronic discovery are neither (1) static enough nor (2) predictable enough to justify amending to the rules of civil procedure in a way that explicitly anticipates their influence. Although some of the proposed rule revisions regarding e-discovery are motivated by concerns of fairness and justice as well as costs, two revisions in particular — the “reasonably accessible” provision in Rule 26(b)(2) and the “safe-harbor” provision of Rule 37(f) — appear to be motivated primarily or even exclusively by apprehension regarding discovery costs.

**THE “REASONABLY ACCESSIBLE” CHANGE IN RULE 26(B)(2) IS UNDULY COST-FOCUSED**

The Report of the Civil Rules Advisory Committee<sup>2</sup> suggests to me that the motivation behind the change to Rule 26(b)(2) is primarily to avoid forcing the responding party to pay the *costs* of obtaining harder-to-get information. The Report speaks of information which is not “routinely” accessed or “easily” located — saying that disaster recovery data is “expensive to restore and

<sup>1</sup> I have been a “technophile” all my life. I experienced the technical side of coping with legacy computer systems when I worked as a programmer at the National Institutes of Health on a machine that was born before I was. (My task was to update an old program so that legacy data, stored on magnetic tapes, would still be useful on newer machines.) I have also worked in a computing security office for a major University, and in the licensing division of one of the world's leading software companies.

<sup>2</sup> COMM. ON RULES OF PRACTICE AND PROCEDURE, JUDICIAL CONFERENCE OF THE UNITED STATES, REPORT OF THE CIVIL RULES ADVISORY COMMITTEE (Aug. 3, 2004), *available at* <http://www.uscourts.gov/rules/comment2005/CVAug04.pdf>.

disorganized”; legacy data is “expensive and burdensome to restore”; and deleted data is inaccessible if it requires “significant cost, effort, and burden” to restore.<sup>3</sup>

The assertions that disaster-recovery data, deleted data, and legacy data are expensive to recover are *based on assumptions about cost and the current state of technology*. They also incorporate non-technological costs, such as the wages of IT professionals, which may similarly decline as technology evolves to require less human input.

Today’s technological capabilities and costs are probably equally bad predictors of our future ability to retrieve legacy data. New technologies are increasingly backwards-compatible, and at least some decision-makers elect to migrate old data to new storage formats as soon as they upgrade, to keep their data accessible. (That, in fact, was part of my task when working as a programmer at the National Institutes of Health.)

Although the goal of lowering the costs of discovery is admirable, the proposed Committee Note for Rule 26(b)(2) may work against that purpose by creating express categories — disaster-recovery, legacy, and deleted data — and permanently labeling those scenarios as “unreasonably accessible”. By making such a designation, this comment takes for granted that today’s balance of costs will persist in a few years’ time.

A better approach would be to leave the balancing of potential costs and benefits of e-discovery up to the trial judge — an aim that is already accomplished by Rule 26(b)(2)(iii). If the Committee strongly believes that judges need more guidance than the current rule provides, I would urge two things. First, an amended rule should be as generally written as possible, avoiding mentioning particular kinds of records that may not always be expensive to recover. Second, in line with the comments of my colleague David Tannenbaum, the rule should not create a perverse incentive to keep data as inaccessible as possible. These objectives might be achieved by adding the following sentence immediately after part (iii) of Rule 26(b)(2):

*The court should pay special attention to the unique potential for technological barriers to increase the costs of discovery greatly and should seek specific information about those costs before deciding whether the burden or expense of the discovery outweighs its likely benefit.*

The committee could then attach the following Committee Note:

*In particular, when dealing with electronic records, judges should pay special attention to “the parties’ resources”, comparing the estimated costs of data retrieval to the amounts spent by the responding party on technology and electronic recordkeeping under normal circumstances.*

### **The “costs” of electronic discovery are not monolithic**

Different kinds of costs are considered by the set of proposed e-discovery amendments. Some of the amendments are clearly written with the costs of storage media in mind; other amendments anticipate search costs. To help illustrate the different kinds of costs that relate to each of the

---

<sup>3</sup> *Id.* at 11.

different amendments, I have broken down e-discovery costs into five categories: the price of storage media; the wages of IT experts who are paid to extract information; the time, hardware and software to find discoverable information or to search the records produced during discovery; the legal costs incurred when filtering out privileged information; and the private and social costs incurred by businesses that are forced to retain suboptimal amounts of data. As Table 1 demonstrates, not all of the proposed amendments address each of the possible “costs” of e-discovery.

**Table 1: Sources of increased discovery costs.**

Costs anticipated by amendment → Rule focus ↓	Cost of storage media	Costs of time spent by IT professionals	Time, software, hardware to search	Legal costs of filtering out privileged information	Costs of inefficient / suspended operations
<b>Form of production</b> (Rules 16(b), 26(f), and 34(b); Form 35)	✓	✓	✓		✓
<b>“Not reasonably accessible”</b> (Rule 26(b)(2))		✓	✓		✓
<b>Assertion of Privilege</b> (Rule 26(b)(5)(B))		✓ (if IT experts are reviewing)	✓	✓ (if lawyers are reviewing)	
<b>Limit on sanctions</b> for loss of discoverable information (Rule 37(f))	✓ (not buying new backup tapes)				✓

**Rules that seek to reduce costs should focus on the most significant costs**

In the comments submitted by the International Association of Defense Counsels, the author cites exorbitant costs as a motivation for rule change. “In my firm’s experience, our clients have incurred tremendous expense, which we do not believe they should have had to bear, due to the discovery requests of others that made my clients and I access data and documents that were not readily accessible. ... The cost of production has exceeded \$1,000,000 in one such case and we are just beginning our discovery efforts.”<sup>4</sup> A similarly bank-breaking scenario was presented in the conference on e-discovery at Fordham Law School, where one participant said she had been involved with a piece of complex litigation where one party was demanding 42,000 backup tapes and 20 hard drives — discovery that she proposed would cost “at a conservative estimate \$4 to \$5 million”.<sup>5</sup>

<sup>4</sup> Written submission of the International Association of Defense Counsel, J. Walter Sinclair at 2, *available online at* <http://www.uscourts.gov/rules/e-discovery/04-CV-004.pdf>.

<sup>5</sup> Joan E. Feldman et al., *Panel One: Technical Aspects of Document Production and E-Discovery*, 73 FORDHAM L. REV. 23, 27 (2004) (panel discussion). It is important to note that the absolute volume of the data was not specified in this conversation, and since tape prices vary according to their capacity, the figure of 42,000 tapes — without more specification about the size of the tapes — means very little about how much discovery we are actually talking about. If these were 10 GB tapes, then the total discovered data would be 420,000 GB, or 420 terabytes. For comparison, the World Wide Web in the summer of 2003 contained about 167 terabytes “on its surface,” (i.e. in

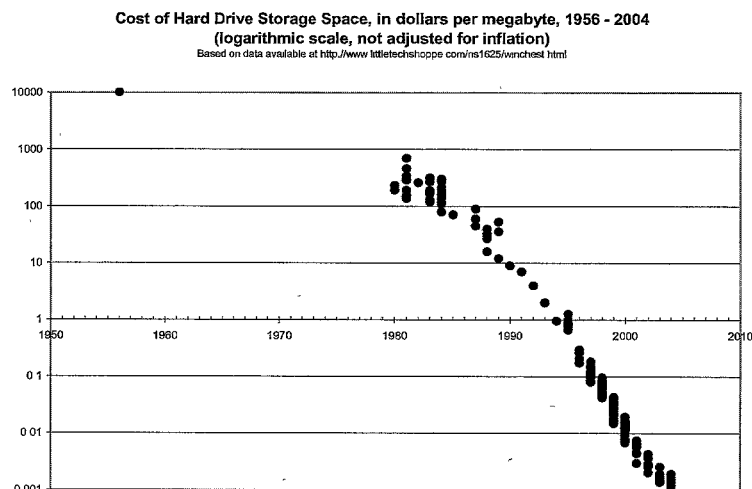
Both of these stories neglect to say how those costs are broken up. If hardware costs were the biggest concern for rulemakers, we could craft a rule forcing parties to make do with less equipment. If staff hours were the dominant cost factor, the rule should instead focus on encouraging as much automation as possible. Without better information about where the *actual costs come from* in e-discovery, any proposed rules will be under-informed.<sup>6</sup>

There is good reason to believe, for example, that the biggest cost involved in electronic discovery is the task of separating out privileged from non-privileged records.<sup>7</sup> If this is in fact the case, then the rules should focus on encouraging companies to better segregate their privileged records from non-privileged ones by providing some sort of reward to such companies during discovery.

If the predominant expense is the cost of storage media, then we may reasonably expect those costs to decline as technology improves, and if the driving factor in that cost is the time actually searching, then we may expect that cost to plummet as search software improves. However, if the predominant cost is the wages of IT professionals, those costs may be reasonably expected to increase with inflation, rather than decrease. I provide below two further illustrations of how the costs and capabilities of technology have changed and continue to change dramatically.

### **Magnetic storage is one example of a technology where costs have plummeted by many orders of magnitude**

Ten years ago my family paid \$1,000 for a hard drive that could store the then-unimaginable sum of *one gigabyte* of data. Back then, the going rate for hard-disk memory was about \$1 per megabyte. Today, the standard retail price is way less than \$1 per gigabyte (I found \$0.55 per gigabyte online today). Price has dropped by *several orders of magnitude*, without even accounting for inflation, in just a little more than 10 years. This anecdotal evidence can be quickly



publicly available content) according to a study called "How Much Information?" by Peter Lyman and Hal R. Varian, available at <http://www.sims.berkeley.edu/research/projects/how-much-info-2003/execsum.htm>.

<sup>6</sup> The comments by Microsoft, although helpful, suffer from the opposite problem: they discuss only the hardware cost resulting from not recycling backup tapes but do not mention the aggregate costs of e-discovery. Although Microsoft's costs seem very high — two weeks' worth of backup tapes cost about \$65,000, or about \$4,642 per day — those tapes hold "about 15 terabytes of data." Only a company with an extremely large IT budget could produce that volume of data, and it thus a cost of \$4,642 per day seems more reasonable in that context. This suggests the strength of the proposed Note language that I suggested above. See written submission of Greg McCurdy for Microsoft, at 2, available at <http://www.uscourts.gov/rules/e-discovery/04-CV-001.pdf>.

<sup>7</sup> E-mail from Nimrod Kozlovski, Yale Law School (Feb. 2, 2005) (on file with author).

verified on the Internet: a non-scientific collection of price data spanning several decades demonstrates the trend clearly.<sup>8</sup> The costs of other forms of storage — both optical, like CDs and DVDs, and magnetic, like tapes — have followed similar patterns.

**New technologies will continue to emerge that will change our conceptions of what is “reasonably accessible”**

A day is not far in the future when the words “backup tape” may sound the way “mimeograph” sounds today. Less than one month ago, InPhase Technologies unveiled its prototype of a holographic storage drive that reads disks holding 200 gigabytes. Later models in this family of drives will store up to 1.6 terabytes (about 667 of today’s DVD±Rs) on a single disk by 2010.<sup>9</sup>

In another few decades, we may see nanotechnology or biotechnology supplanting lasers and magnets as the way to store massive amounts of data, and it is pointless to guesstimate how much or how little such technology would cost. Basing our predictions on recent history, however, the only thing we can expect is that progress will continue to surprise just about everyone.

**THE SAFE-HARBOR PROVISION PROPOSED FOR RULE 37(F) IS MISLEADING AND FLAWED. AT THE VERY LEAST, BETTER LANGUAGE IS NEEDED TO DISTINGUISH BETWEEN ‘DELETION’ AND ‘ERASURE.’**

On its face, the proposed Rule 37(f) seems to address an important difference *in kind* with electronic discovery — “a distinctive feature of computer operations, the routine deletion of information that attends ordinary use.”<sup>10</sup>

But this entire proposed subdivision is founded on a premise that is at best misleading to judges and at worst just plain false. The fact is that computer systems *never* “routinely” delete data *without having been programmed to do so*.<sup>11</sup> The decision to delete files is a user-level decision. The decision to recycle backup tapes is an administrator-level decision. Some programs, it is true, create and then delete records “behind the scenes,” without the user ever even knowing that they were there, but even that involves a conscious decision on the part of the program designer to delete the information.<sup>12</sup>

---

<sup>8</sup> See <http://www.littletechshoppe.com/ns1625/winchest.html>. By using prices that are advertised in printed publications, the survey’s prices for that are somewhat higher than those found online, such as the price I quoted above of almost 50 cents per gigabyte today.

Of course, we have to deal with the problem that “if you build it, they will come.” As we become able to store more data, we find more data to store. But eventually, we may get to the point where we don’t really *need* to worry about *capacity* as a constraint on how much information we can store.

<sup>9</sup> See “Tech Weekend,” *The Dallas Morning News*, Jan. 22, 2005 at 9D, also available at LEXIS; see also <http://www.inphase-tech.com/news/firstholoproto.html>.

<sup>10</sup> Proposed Rules at 33.

<sup>11</sup> Anyone who has ever encountered a “Disk Full” error knows this to be the case.

<sup>12</sup> A classic example of this is your Web browser’s “cache” feature, which saves copies of your most recently visited Web sites so they will load faster, but deletes them after a certain size or time limit is reached.

This proposed subdivision would benefit from, at the least, a bright-line distinction between “deletion” and “erasure” — a distinction that is, in essence, a distinction based on intent.<sup>13</sup> The Note should define deletion as the *intentional* action of marking a file as no-longer-needed. Users of Microsoft Windows “delete” files when they place them in their Recycle Bin and then empty the bin.

As the Committee fully knows, however, when a file is “deleted,” it stays exactly where it is and is only marked invisible to the user and overwriteable to the computer. The computer then overwrites (“erases”) the file whenever it needs the physical space — recycling that space on the drive.

An exactly analogous thing happens with backup tapes. Administrators tell their backup systems how long they wish to keep a tape — and at what point it is no longer needed. They in effect decide that “this tape is ‘deleted’ after two weeks.” At this point, the data on the tape doesn’t just disappear; it stays there until it gets “erased” when it is overwritten in the next backup event.

Deletion is an intentional act by human agents, and it is *the decision to delete records* that the rules of discovery are concerned with; erasure is an everyday computer occurrence that is far less interesting. While it is true that some deleted-but-not-yet-erased data might be discoverable, the only way for any responding party to preserve this kind of data with confidence would be not to use their computers at all — to halt all operations completely. To prevent this absurd result, the wording of the rule could be changed to say the following:

*(f) ... a court may not impose sanctions under these rules on the party for failing to provide such information if:*

- (1) the party took reasonable steps to preserve the information after it knew or should have known the information was discoverable in the action; and*
- (2) the failure resulted from the unintended erasure of already-deleted data consistent with normal use of the electronic information system.*

The rule could then contain an explanatory note that described the difference between “deletion” and “erasure,” highlighted the importance of intent, or mentioned whatever other terms the Committee chose to communicate that important idea.

I prefer this language because by limiting the safe-harbor provision only to “already-deleted” data, it makes clear that users are expressly forbidden from making any kind of decision that would move data into the “deleted” category.

### **Recycling backup tapes is not a sufficient reason to craft the safe-harbor provision**

One of the reasons advanced in support of the safe-harbor provision by Microsoft and other commentators is that companies must be able to recycle their backup tapes to avoid paying potentially huge costs to preserve those tapes (see Table 1 and footnote 6).

---

<sup>13</sup> This is by no means a popular or universal distinction — I am proposing it as a new idea — but all that matters is that those who use it in this context understand the difference.

This is just another example of using a cost argument to justify significant rule change. Just because recyclable, expensive magnetic tapes are the current fashion in disaster recovery hardware does not mean that this will continue to be the case in the future. If future litigants have the option of saving many more backups without incurring significant cost, the cost-benefit balance indicated in Rule 26(b)(2)(iii) would be tipped toward compelling discovery, and we would not want a rule that gave companies an excuse to destroy data under the guise of the "routine operation" of computer systems.