Joseph Masters
<joe@builderadius.com>
12/25/2004 09:59 PM

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

04-CV-063
Request To Testify
2/11 DC

Dear Mr. McCabe,

I am writing to 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 a group of Yale Law School students who will be submitting written comments on the rule amendments regarding electronic discovery. I wish to testify on the dangers of the proposed "reasonably accessible" standard and potential solutions.

I would appreciate the opportunity to be part of this process.

Sincerely,
Joseph Masters
joseph.masters@yale.edu
25 Lynwood Pl #25
New Haven, CT 06511
(617)331-1994

04-CV-063 January 20, 2005
Testimony 2/11 DC

Joseph T. Masters
25 Lynwood Pl. #3
New Haven, CT 06511
Joseph.Masters@Yale.edu

Peter G. McCabe, Esq.
Secretary
Committee on Rules of Practice and Procedure
Washington, DC 20544

Dear Mr. McCabe:

I appreciate being given the opportunity to comment on the proposed Amendments to the Federal Rules of Civil Procedure regarding e-discovery.

I think I can provide an important perspective on the Amendments, even though I am only a law student. Over the past ten years, I have received national press for several computer programs I have written. In 1999, I co-founded a software development and internet company and served as its Chief Technology Officer. I recently stepped down from that position to attend law school.

I have two concerns about the amendments as they are written: the creation of an unnecessary protection of discoverable information in Rule 26(b)(2), and the potential for electronic data obfuscation in Rule 34(b). In particular, I am concerned that technologically sophisticated parties will be able to use the amendments to the Rules to make the process of discovery significantly more difficult and expensive than it already is. In doing so, parties will require courts to be more deeply involved in complex technological disputes than they would have to be under the current Rules.

1. Unnecessary protection of discoverable information in Rule 26(b)(2)

    Under the current Federal Rules of Civil Procedure, discovery is limited by practical means. Rule 26(b)(2), as currently written, requires judges to weigh relevant factors to see if a particular piece of information requested should be discoverable. While motivated by the right goals, the proposed amendment to Rule 26(b)(2) would unintentionally allow a producing party to make discovery a much more costly process for a requesting party, and may even permit a producing party to hide information that would be discoverable under current rules.

    The process created by the proposed amendment would allow a producing party to declare certain information not "reasonably accessible." The requesting party would then have to file a motion to compel discovery, which the producing party could then successfully rebut by producing an expert to testify that the information is truly not reasonably accessible. It may be very difficult for a requesting party – which might not have access to the specific technologies that are used by the producing party and which necessarily knows the system less well than does the producing party – to show conclusively that the producing party's expert is wrong. This difficulty (and the expense) of the proposed process will likely be exacerbated by the need to communicate about complex technical issues with judges who, on the whole, are not experts in the relative substantive issues.

Thus, the cards are stacked heavily against the requesting party. At the very least, the party will incur significant costs. At the worst, if the producing party can find an expert witness who will testify that certain data is not reasonably accessible, then the information would only be discoverable if the requesting party can win a "good cause" hearing.

The concern that the "reasonably accessible" test is hoping to address is admirable. As the committee notes detail, some information may be 'legacy' data in obsolete systems, and is no longer used and may be costly or burdensome to restore and retrieve. Electronic data should be treated no differently than written information stored in an enormous warehouse, which may no longer be in use and which might be costly and burdensome to restore and retrieve. Even a situation where information may have been deleted in a way that makes it inaccessible without resort to expensive and uncertain forensic techniques is adequately covered by the existing language in Rule 26(b)(2)(iii), which states that where "the burden or expense of the proposed discovery outweighs its likely benefit," discovery will be limited.

It may be instructive to consider how cases cited by the committee notes on this section might have turned out differently had the "reasonably accessible" language been in the Rules at the time they were decided. In *Zubulake v. UBS Warburg*, it is likely that the emails on the backup tape would have been considered "not reasonably accessible." The "good cause" hearing described in the proposed amendment may have been based on the Judge Scheindlin's opinion, but the factors involved in a "good cause" hearing under the proposed amendment are not outlined in any way. It is, therefore, probable or even likely that *Zubulake v. UBS Warburg* would have been decided differently under the proposed amendments to Rule 26(b)(2). The same is true of the two other cases cited in the committee note to the amendment: *McPeek v. Ashcroft* and *Rowe Entertainment v. The Williams Morris Agency, Inc.*

The proposed amendment to Rule 26(b)(2) is a solution to a problem that does not exist. The proposed amendment will keep more information from being discovered than is true under the current rules. As it stands, the current Rule addresses the most significant and salient concerns raised by modern electronic discovery. The case law developing here and cited by the committee itself is proof. I therefore recommend that no changes be made to Rule 26(b)(2).

2. Potential for electronic data obfuscation in Rule 34(b)

The proposed amendments to Rule 34(b) include the following section:

> (ii) if a request for electronically stored information does not specify the form of production, a responding party must produce the information in a form in which it is ordinarily maintained, or an electronically searchable form. The party need only produce such information in one form.

The committee note on this section further explains that this change will allow parties to produce information in forms other than that in which they are maintained, as long as that form is electronically searchable. It also allows for the request of data in alternate forms if parties "cannot use" the information in the form in which it was produced. The note does not detail exactly what "electronically searchable" means. I read the text of the rule to mean that companies can produce, for example, ASCII text files instead of Microsoft Word files to avoid the production of metadata. Typically, a Microsoft Word file will contain metadata that includes formatting (e.g. fonts, sizes, spacing, margins), the name of the creator of the file, and also possibly text that was deleted from the document as it would be seen in Word. An ASCII text file contains just the actual text of the document. I believe the general idea here to be that producing parties can hand over requested documents without having to purge metadata that is not easily removed or seen from within Word.

I have two concerns with the amendment as written. First, it is not clear to me why the producing party is only required to produce data in a single format. This would encourage companies to keep data in obscure data formats that cannot be read by requesting parties, and then preclude relief from a requesting party because one form of the information was produced.

Second, there are ambiguities left by the current wording of the rule and committee notes. These affect both electronic data in the form in which it is ordinarily maintained, and potential alternate "electronically searchable" forms. Here are several potential scenarios that the proposed Rule permits:

A. The **expensive-to-read** scenario: The producing party delivers the information in a form that is electronically searchable for them, but would require great expense and/or expertise on behalf of the requesting party to even read it. For example, consider that the producing party runs Oracle as its database server and retains its data in a format. The requesting party would need to purchase Oracle in order to read the information. This is certainly not impossible, but it is expensive, because Oracle Database Enterprise Edition cost upwards of $40,000.00 (http://oraclestore.oracle.com/). This particular example may not occur frequently, but the more general problem is likely to be significant:

Different parties use different software programs, which, in turn, use different file formats in to store data. Some parties will use very expensive (and possibly custom) software programs that store data in proprietary formats. These formats are electronically readable by the receiving party only if it can read the file formats (which, as already noted, could be prohibitively expensive). The requesting party might find some relief in the "good cause" framework of Rule 34(b), but the committee note here suggests only that a requesting party may object if it "cannot use" the information provided.

B. The **sort-of-searchable** scenario: The current rule and notes do not define the word "searchable" to any specific degree. Technically, any file format is "searchable" using a hex editor, which is a program that examines a computer file bit-by-bit. By the comments, this part of the amendment appears to be intended to differentiate between an image of text and the text itself. An image of text is not easily searchable by a computer, but it could be if it were put through an Optical Character Recognition program. However, the wording as written is ambiguous and would not necessarily even cover the given example of image vs. text. Furthermore, the committee notes specify that if the producing party "ordinarily maintains the information in more than one form, it may select any such form." This might provide an incentive for a company to store all of its electronic data in both text and image format, and then provide only the image format in the course of discovery.

C. The **limiting-functionality-through-proprietary-program** scenario: Under the rule and notes as written, the producing party could supply the requesting party with the requested data rolled up into a computer program and encrypted. If this were to happen, the requesting party would have to use the supplied computer program to read the requested data. This would allow the producing party to put severe limits on how the requesting party could access the data – for example, by slowing the program down to only allow one page to be read per minute, or only allowing searching within a single document, not throughout all documents, so each document would have to be read separately, or not allowing the data to be printed. All of these limitations would technically be permitted by the Proposed Rule and notes. While the committee's vision of "electronically searchable" might be that of an ASCII text file or Microsoft Word file, nothing in the proposed Rule precludes the production of electronic data that makes electronic searches less efficient than skimming through paper documents.

It is my experience from the software development world that all of these scenarios will happen if the proposed amendment is adopted. I have had to purchase a copy of Crystal Reports for over $500 to read an Oracle database in the course of my business (scenario A). I have written programs for my business that converted text to images so that users could not easily copy the text (scenario B). Finally, there are already highly restrictive protected ebook formats (e.g. eReader, Mobipocket, Microsoft's .lit format), and it would not require much work at all to only allow their use within a slowly-scrolling, barely-searchable "reader" program (scenario C).

These problems could be solved by changing the wording of Rule 34(b) and adding appropriate notes.

First, change the text of the rule to read:

> (ii) if a request for electronically stored information does not specify the form of production, a responding party must produce the information in a form in which it is ordinarily maintained, or in an electronically searchable common file format. The party need only produce such information in one format as long as that format is readable by the requesting party.

In order to address concerns that producing parties will originally store and then produce data in an **expensive-to-read** format, the only-one-such-format requirement should be relaxed to be satisfied only when the requesting party can read the format. By "readable", I mean that the requesting party must be able to interact with the file via computer in the same way that the producing party was capable of doing. This would mean that documents scanned in as images could be produced as images, but completely proprietary encrypted and/or compressed file formats would not be acceptable. The addition of the words "common" and "format" are designed to address the **expensive-to-read** scenario: by providing a format common to both parties, the electronic data will never be needlessly expensive to read. In the Oracle example above, the producing party would be required to produce the information in a plain text format. All text data that a company stores should be exportable to plain text – in my thirteen years of programming, I have yet to find a single application that did not allow exporting data to plain text. The expense of exporting is usually comparable to that of providing the data in original format. If it is not, then the balancing test provided in Rule 26(b)(2) will control. Furthermore, without providing incentives to companies to store data in obscure formats, they will naturally choose common file formats, like Microsoft Word. The addition of the word "file" will address the **limiting-functionality-through-proprietary-program** scenario: the producing party will be required to produce files, not proprietary programs with file data in them. This should be explicitly expressed in the committee notes.

Second, the committee notes should be amended to address the **sort-of-searchable** scenario. The sentence explaining that if a producing party "ordinarily maintains the information in more than one form, it may select any such form" should be changed. I propose: "if the producing party maintains information in more than one format, it must select a format that is electronically searchable if such a format exists." Thus, a party would only be allowed to provide an image of text if it did not also have a text version of the text data. It may be suggested that "most widely accessible format" is better language than "electronically searchable," but I believe that "most widely accessible" can be difficult to discern (e.g., is Microsoft Word or Adobe Acrobat more widely accessible?), whereas "electronically searchable" is clearer.

Third, the committee notes should contain further detail regarding the intended application of the amended Rule 34(b). It should note that the word "common" in the phrase "common file format" means common to the parties. Thus, if both parties own programs that can read the file format, the format is acceptable. If one party does not have access to a program that can read the file format and cannot obtain access to such a program at a reasonable cost, then the producing party must provide the data in a different, acceptable format.

Fourth, the committee notes should elucidate explanation of "readable" and "searchable." A program that can "read" a file format is a program that reproduces the requested data in a meaningful way, as opposed to a program that could technically "read" a file but not interpret the data in an appropriate way (i.e., a hex editor). A program that can "search" a file is a program that that can read the file (as explained in the previous sentence) and also find text strings within the text stored in the file.

Fifth, the committee notes should explain that requested data should be supplied in file format, not in a crippled proprietary format that limits the ability of the requesting party to work with the data.

The advent of electronic storage of information is not only a wonderful avenue to productivity – it has also ushered in many ways of hiding, encrypting, and obfuscating information. Many of these methods of hiding data have yet to be discovered, so it is extremely important to be clear at the outset that the Rules are not being amended to allow the hiding of information. The judges who read and interpret these rules *de novo* may not be the best equipped to know when parties are hiding information through the vagueness of the Rules. It is therefore critical that Rule 34(b) be clear and precisely crafted to avoid making discovery of electronic information more difficult and expensive than necessary.

I thank you once again for the opportunity to comment.

Respectfully,

Joseph T. Masters