02-ED-057

"Greg McCurdy (LCA)" <gmccurdy@microsoft.com>

03/08/2004 07:45 PM

To: <Lee_Rosenthal@txs.uscourts.gov>, <peter_mccabe@ao.uscourts.gov>
cc: "Tom Burt (LCA)" <tburt@microsoft.com>, "Rich Wallis (LCA)" <rwallis@microsoft.com>, "Gene Burrus (LCA)" <geneb@microsoft.com>, "Dawson, Martha (SEA)" <marthad@prestongates.com>
Subject: Microsoft Comments on E-Discovery & Civil Rules

Dear Judge Rosenthal and Mr. McCabe,

On behalf of Microsoft Corporation Gene Burrus and I would like to thank you again for the opportunity to participate in the Fordham Conference hosted by the Advisory Committee on Civil Rules. Attached are Microsoft's comments on some of the proposals before the Committee signed by Tom Burt, the Deputy General Counsel in charge of litigation. Microsoft offer these both as a litigant and as a technology company some of whose products are intricately involved in electronic discovery issues.

Please do not hesitate to call on us if can assist with any technical or legal issues. We take a great interest in the rules revision process and believe that this is a unique area where law and technology intersect. Technology has created great benefits and efficiencies along with some very significant difficulties and costs for lawyers and litigants. We believe that a sound analysis of technology is an important foundation for the development of the law in this area and any solutions for the challenges of the information age.

Best regards,

Greg McCurdy
Senior Attorney, Litigation
Microsoft Corporation
425-705-2724 (office)
206-355-4464 (cell)
425-936-7329 (fax)

Scan001.PDF

Microsoft Corporation
One Microsoft Way
Redmond, WA 98052-6399

Tel 425 882 8080
Fax 425 936 7329
http://www.microsoft.com/

Microsoft

The Honorable Lee Rosenthal
United States Courthouse
515 Rusk Street, Room 11535
Houston, TX 77002

RE: Advisory Committee on Civil Rules: E-Discovery Rule Discussion Proposals

Dear Judge Rosenthal:

Microsoft was very pleased to participate in the Fordham University conference on E-discovery. As both the creator of software that allows companies to work and communicate in the electronic world, and as a litigant in many major e-discovery cases, Microsoft has a unique perspective regarding the challenges and problems that companies face in the discovery of electronically created and stored information. Microsoft both applauds and supports the committee's on-going efforts to address the serious problems that have arisen and update the rules for the practical realities of the electronic age.

As discussed at the Fordham conference, the threshold question is whether changes to the Civil Rules are needed to address the problems surrounding e-discovery. Our view is that it is absolutely necessary in order to provide needed guidance to both litigants and the courts. Advances in computer software such as word processing and e-mail, as well as in the hardware needed for electronic storage of documents have changed the nature of discovery in a way that needs to be addressed in the rules of procedure, most of which were drafted to address the much simpler world of paper documents. Put simply, the products that Microsoft and its competitors produce have vastly increased the productivity of office workers. These new technologies allow each "knowledge worker" with a personal computer ("PC") to produce, replicate, disseminate and store vastly more documents and information, more quickly and more cheaply than ever before.

Each of the lawyers and judges involved in this process has witnessed this explosive development of technology and the consequent proliferation of potentially available documents and information. In general this has been a boon for productivity, but it also creates substantial litigation costs and inefficiencies that arise in searching, reviewing and producing vast numbers of electronic documents in discovery. It is the effects of the rising flood of e-documents that give rise to the challenges faced by all companies that are part of the "information economy."

It is not unusual for Microsoft to receive requests for "all documents" relating to a particular product. With over 55,000 employees worldwide, such requests could apply to all or a very large number of employees. As electronic storage has become virtually effortless for employees and substantially less expensive than it used to be, many major corporations are faced

with massive quantities of electronic documents—often retained by employees for no particular reason at all—that quickly dwarf the volume of documents that typically existed in the old, paper world. While an employee once was only able to store hard copy documents in one or two filing cabinets in their office[1], that same employee can now easily store the equivalent of millions of pages on their hard-drive or servers. In the paper world, documents were normally organized in subject matter files and required some physical effort to file and organize them. In the electronic world, retaining a document requires little or no effort by the employee and there is typically, in our experience, little or no file organization. Instead, all documents are in one or more gigantic folder[s] since it is so easy to "click and drag" to another folder, or allow email to pile up in an inbox and periodically dump it into a PC hard disk or on a network server. Unless managers of corporate IT systems take affirmative steps to prevent it, the indiscriminate storage of all email sent and received is often the default mode of operation for many employees. The vast majority of this material has little or only very transitory business value.

It is common for employees at Microsoft and similar companies with "knowledge workers" who spend much of their workday at their computers to send and receive hundreds of emails and other electronic documents each day. Many of these have little or no long-term business value since they are the functional equivalent and indeed the replacement for communications that were once casual phone calls and hallway conversation with co-workers. Nonetheless, many employees simply keep everything by default since it is so easy for them and it requires more effort to sort out the truly significant ones and delete the majority, which have no lasting business value.

It is this vast increase in the volume of material that must be collected and reviewed, coupled with lack of organization, and the challenges of handling documents in different formats and requiring different and ever-changing software, which make the scale and difficulty of the electronic world poles apart from the paper world. At Microsoft, our employees operate largely in a paperless world, making electronic discovery the central focus in our litigation.

Microsoft offers the following comments on the E-Discovery Rule Discussion Proposals. For ease of reference, we include the text of the discussion proposals HIGHLIGHTED below, followed by our comments. We follow the format of the materials provided for the February 20-21, 2004 Conference on Electronic Discovery.

In our view, there are four primary areas where meaningful changes to the rules are needed:

1. Scope: There should be a two tiered approach permitting discovery in the first instance of electronic information accessible in the ordinary course of business, and then requiring a showing of need to extend discovery to electronic information that is created or maintained for disaster recovery purposes, or to information that is not reasonably accessible in the ordinary course of business;

2. Cost: If a showing of need is made, the cost of accessing and producing that information should be allocated between the parties;

[1] Or approximately 30,000 to 60,000 pages.

3. Preservation: A "preservation safe-harbor" should be created that balances the operational requirements of complex information technology systems with the need to preserve discoverable materials;

4. Privilege: The inadvertent disclosure of privileged documents should be protected without creating new pressures for premature production.

A. Defining the Subject: A New Rule 26(h)

Rule 26. Duty to Disclose; General Provisions Governing Discovery

* * *

(h) Electronically-stored data.

(1) [Scope of] Electronically-Stored Data [defined].

Electronic data [Digital data?] [Computer-based data?] includes all information [created, maintained, or] stored in digital form, on magnetic, optical or other media, accessible by the use of electronic technology such as, but not limited to, computers, telephones, personal digital assistants, media players, and [or] media viewers.

Comment:

As noted by the Committee, definitions in the Civil Rules are not favored, and this is an area where the advances in technology raise the additional concern of whether a definition could be devised that would stand the test of time. Our view is that each of the alternative phrases (electronic data, digital data, computer based data) would work to generically describe the issue, but we suggest that "electronically stored information" is better suited, providing both the guidance to the court and the flexibility to deal with new technology that enters the market constantly. However, we do not believe that a new definition is necessary and suggest instead that this phase could be added to the existing Rule 34 in the appropriate place rather than creating a new paragraph in Rule 26.

B. Including Discussion of these Issues in Early Discovery Planning – Rules 26(f), 16(b) and Form 35

Rule 26

* * *

(f) Conference of the Parties; Planning

* * *

(3) ***Discovery Plan.*** A discovery plan must state the parties' views and proposals on:

* * *

(C) whether any party anticipates disclosure or discovery of electronically-stored data, and if so what arrangements should be made to facilitate [management of] such disclosure or discovery; and

(D) whether provision should be made to facilitate discovery by protecting the right to assert privilege after the [inadvertent] disclosure or production of a privileged document; and

* * *

[Form 35 and Rule 16 Text Omitted –Similar Amendments]

Comment:

There appeared to be broad support at the Fordham Conference for the addition of a broad statement that issues relating to electronic discovery should be addressed in the discovery plan. We would oppose a provision, adopted in some jurisdictions, that would require a company "to investigate and disclose" specific information regarding its computer systems for the simple fact that in most cases only a very small portion of those systems are even remotely relevant to the litigation. Also, as discussed below, we believe that this is the appropriate place for the parties to discuss and resolve issues relating to the "form of production," including requests for meta data or embedded data. Finally, to the extent revised rules contain a provision protecting against the inadvertent production of privileged documents, paragraph (D) may be redundant and unnecessary.

C. Definition of "Document" in Rule 34

Rule 34. Producing Documents and Tangible Things, or Entering onto Land, for Inspection and Other Purposes

(3) Definition of document

(a)(1)(A) any designated documents – including writings, drawings, graphs, charts, photographs, sound recordings, and other data or data compilations in any [magnetic or other] media from which information can be obtained or, when necessary, be translated by the responding party into a reasonably usable form, [and including, for electronically-stored data, all data stored or maintained on that document {if the court so orders for good cause}.] or

Comment:

Any change to Rule 34(3)(a)(1)(A) that will both catch up with current technology and be forward looking in its ability to address whatever the future may bring must be generalized to address the concept of "information" in the definition of "document." Today, courts and parties struggle to shoehorn electronically stored information into the term "document" and this has led to great debate over what constitutes a document and things like "embedded data," "meta data" and "native formats." When stripped to what any electronic data actually is, "information," these debates resolve themselves. All text, digital images, meta data, and embedded data are just information, which, if relevant and non-privileged, may be subject to discovery. Reduced to that concept, courts and parties can do what they are long accustomed to doing, balancing burden and utility, relevance and need for production of certain types of information.

As a producer of software used by many people and businesses to create a great deal of electronic information, Microsoft may be uniquely positioned to comment on the nature of this information and help shed light on the limitations in the word "document." "Document" is an inherently arbitrary concept when it is applied to electronically stored information. Even things

that are commonly understood to be electronic documents are not really documents in the way that we would normally think of them. A letter written with Microsoft Word for example, is really just information that can be understood by Microsoft Word software that can then display some of the information on a screen or print it into what we commonly think of as a paper document. However, the computer file itself is simply information: information about the text of the letter, the font of the print, the margins of the page, and even "embedded" information about changes made to the letter, editors that worked on the letter, etc. It seems well settled that the computer file that contains that information, the ".doc" file, is a "document." However, the fact that all of this information is contained in a file ending in ".doc" is purely a design decision of Microsoft engineers. The information could just as easily be divided into multiple files that are simply associated with each other. For example, the text information could be in one ".doc" file, the formatting information in a ".fmt" file, and the editing information in a ".edt" file. Which file is the "document"? Suddenly, by virtue of a product design change at Microsoft or other software vendor, the commonly understood concept of electronic document may be affected and parties and courts must struggle anew at how to apply the old concept of "document" to the new world. This example also points out the futility of the proposed edit to include "all data stored or maintained on that document." Data is only stored "on" a document because of the way a software company has designed its product. It is all just information, regardless of where or how it is stored or maintained.

As the very name implies all ***soft***ware products are inherently flexible and easy to redesign to serve new purposes and be more effective. Just as car companies produce new models every year with updated styles, designs and technology software companies update and improve their products on a regular basis in order to meet consumer demand and keep up with competitors. This means that today, and increasingly in the future, electronic "documents" will simply be the arrangement and presentation of data from various sources, both inside and outside the "document" file itself. This is in essence what reports generated from relational databases[2] are today. While a printed report is certainly a document, it is impossible to think of the database itself as a "document." Until the data is arranged and presented, it is simply stored information, meaningless to almost anyone who attempted to view it in its raw form. Similarly, a letter created with Microsoft Word may display a date at the top, but the ".doc" file does not actually contain the text of the date at all. Instead, it contains an instruction to obtain the current date from elsewhere in the computer system and to display that current date at a particular place in the

[2]A database is simply a storage place for data. A database does not present information directly to a user; the user runs an application that accesses data from the database and presents it to the user in an understandable format. In a well-designed database, there are no duplicate pieces of data that the user or application must update at the same time. Related pieces of data are grouped together in a single structure or record, and relationships can be defined between these structures and records. When working with ordinary data files such as a Word file, an application such as Word must be designed to work with the specific structure of each data file. In contrast, a database contains a catalog that different applications can use to determine how data is organized. Generic database applications can use the catalog to present users with data from different databases dynamically, without being tied to a specific data format. A database typically has two main parts: first, the files holding the physical database and second, the database management system (DBMS) software that applications use to access data. The DBMS is responsible for enforcing the database structure, including:

- Maintaining relationships between data in the database.
- Ensuring that data is stored correctly, and that the rules defining data relationships are not violated.
- Recovering all data to a point of known consistency in case of system failures.

display or printout of the letter. In the future, it is easy to imagine a business presentation "document" being nothing more than a set of instructions on the presenter's computer: instructions to obtain formatting information from the company server, to get content from another server (perhaps even one controlled by a third party) and the second the on-screen presentation is concluded, all that remains on the computer is the stored instructions. Are those instructions included in the definition of a "document?" Certainly not in the way we think of a document today since no one ever viewed those instructions. More accurately, they are information from which one could begin to discover what was presented on the screen.

The definition of "document" under Rule 34 has long lagged reality when it comes to electronically stored data. And change that is going to attempt to be forward looking will need to be more generalized to the concept of "information." We believe this is the appropriate location to add "electronically stored information" to the definition of a document in the rules. One possible approach would be a Rule 34 definition as follows:

> any designated documents or electronically-stored information -- including writings, drawings, graphs, charts, photographs, sound recordings, data bases, and other data compilations from which information can be obtained or, when necessary, be translated by the responding party into a reasonably usable form, or

In our view, the rule should not attempt to resolve issues relating to whether meta data or embedded data should also be produced. We believe those issues must be resolved on a case-by-case basis as they are just another piece of electronic information. We discuss this issue further in the context of the form of production.

D. The Form of Production – Rule 34(b)

Rule 34.
Form of Production

(b) Procedure.

(1) ***Form of the Request.*** The request must:

(A) describe with reasonable particularity each individual item or category, the items to be inspected; and

(B) specify a reasonable time, place, and manner for the inspection and for performing the related acts. The request may specify the form in which electronically-stored data are to be produced.

[Alternative]

(D) specify the form in which documents electronically-stored data are to be produced.

(2) ***Responses and Objections.***

* * *

(C) *Objections.* An objection to part of a request must specify the part and permit inspection and related activities with

respect to the remainder. A party may object to the requested form for producing electronically-stored data [and to production of electronically-stored data that are not {reasonably} accessible [without undue burden or expense] {reasonably available} in the usual course of the producing party's business {activities}].

(D) *Producing the documents.*

* * *

(ii) ***Electronically stored materials.*** A party producing electronically-stored data may produce them in the form in which they are ordinarily [created or] stored. Unless the court orders otherwise for good cause, a party producing electronically-stored data need only produce it in one form.

Comments:

While we agree that the requesting party should have the duty to advise whether electronically stored information should be produced in hard copy or electronically[3], Microsoft believes that the rules should not favor or specify any particular format of production and, in particular, that the rules should neither require nor favor the production of files in "native[4]" format, including section (D)(iii), which would give the producing party the right to produce in that format as a default. It is our experience that the format of production, particularly in complex cases, is almost always the subject of legitimate discussion because the parties have different IT systems and litigation support software and databases and different requirements for their systems. We suggest a rule that requires the requesting party to specify the form of production they prefer, and require the parties to meet and confer, and if they cannot resolve the issue, bring it to the attention of the court. Early discussion and resolution of these issues will save time and costs for the litigants.

The language of (D)(ii) provides that a producing party may always choose to produce documents in the form in which they were ordinarily created or stored. For electronically stored information, that is their "native" format—for example, documents created in Word are produced as Word files (document.doc). We are aware that some groups claim that production in native format is more desirable than a format that has more paper like characteristics, such as PDF and TIFF. However, this ignores the practical realities of depositions and trials and the deleterious

---

[3] We include as an electronic production TIFF and PDF files. TIFF is an acronym for Tag(ged) Image File Format. PDF is an acronym for Portable Document Format. TIFF and PDF are popular file formats that preserve the fonts, images, graphics, and layout of source documents, regardless of the application and platform used to create the source document.

[4] Native files are the original electronic format in which they were created. That means there are many different types of native formats which vary greatly in attributes that can affect their uses in litigation. Among Microsoft application software products, word processing documents are in Word format (and thus designated by the ".doc" extension on the "native" file name); presentations are in PowerPoint (.ppt); spreadsheets are in Excel (.xls) and email messages are in Outlook (.msg or .pst). Similarly, PDF documents created using Adobe's popular Acrobat software are designated with the ".pdf." file extensions.

effect unwieldy native file productions can have. It is difficult to fully understand all of the problems associated with native file productions. We identify some of them here.

First, we know of no way to place a unique document production number[5] on each "page" of many native file formats, which may become even more problematic as electronically stored information takes on characteristics that are less and less "document-like". "Pages" simply do not exist in many native file formats used for such common applications such as email, databases or spreadsheets. A "page" is an old concept from the paper world that has been loosely applied to electronic files that are "printed" on physical or virtual paper. Electronically stored information may be "printed" to paper, or converted (*i.e.*, printed) to electronic file formats such as TIFF or PDF files that share characteristics of paper documents such as page breaks and a strong resistance to alteration. Electronically stored information produced in native format are often easy to alter; in fact, formats such as Word or Excel are designed to be altered very easily.[6] Many native formats do not have page breaks; they are continuous files -- unless and until they are subdivided into pages for printing. Although it may in some cases be convenient and efficient to produce electronically stored information in their native format, courts and litigant's, particularly in large litigation, still use page numbers and the printed version of documents in depositions and at trial.

Anyone actively involved in complex litigation will attest to the importance of the unique number on each page of the document. Page or "Bates" numbers help litigants track documents and ensure the integrity of their contents. Although it may be possible to place a number at the file level (*i.e.*, the meta data)[7] of the electronically stored information, it is often not displayed when printed, and for anything longer than one page, is not a unique number. Document numbers are particularly important to Microsoft and other litigants since they can also identify the specific employee from whose files the electronically stored information came.

Second, we know of no way to automatically place the protective order designation on each page of a native format document, and manually altering each native file to place a footer with this information is not practical. It is also standard practice in large cases for protective orders to have more than one tier of confidentiality of production, in some cases restricting certain categories of documents to review by outside counsel only. The inability to systematically place unique protective order designations on each page of the native file is a critical reason *not* to require or favor native productions. We therefore favor a rule that does not set a presumption for the format to either party, but rather requires the parties to identify and discuss the issues at an early stage in the litigation.

---

[5] At the Fordham Conference, it was stated that native files could be "wrapped" to deal with these concerns. On further inquiry, the "wrap" being referred to was PDF format with a tool that allowed users to switch between viewing the native document and the static document. We do not think that this solves the many issues associated with a native file production.

[6] While electronically stored information in many native file formats can be easily altered, both unintentionally and intentionally, printed versions of the same document, whether printed to paper, or electronic formats like PDF, or TIFF, are difficult to alter. As with paper documents it is difficult to say with certainty that it is impossible to alter or forge the content of an electronic document.

[7] Processes are available to rename the native file to include a unique number. For example, a Word document named "Letter to John.doc" could be renamed, at the file level, "Letter to John_1000005.doc".

Finally, there is no good way to use the average native file in depositions, in motion practice, or at trial. Litigants want to see and use paper, or the electronic equivalent of paper complete with pages, document page numbers and protective order designations.

That brings us to the question of meta data or embedded data which are data associated with electronically stored information that is not visible in most electronic formats and typically does not appear when printed. For example, a Word file may have meta data that specifies who created it, on what date, at what time and who edited it, and where it was saved. Certain meta data is useful because it reduces litigation costs by providing automated 'coding' for electronically stored information, for example, to, from, copyee, date, title, etc. Because of this, it can make sense for parties to agree to the exchange of TIFF or PDF files, supplemented by agreed meta data fields. With the exchanged meta data, text or OCR[8], TIFF and PDF files are electronically searchable, one of the main rationales for seeking native files in the first place.

Embedded data poses additional problems because the information is not readily viewable, and the added problem that it may contain privileged information in comments or edits in the document that must be separately reviewed for privilege. Reviewing embedded data for privilege is time-consuming and has all of the complications associated with native files—they are easily altered and the correct software must be used to open and view each document. We also believe there are very few cases where embedded data is really important to an issue in the case, and even then it is likely limited to a sub-group of electronically stored information actually produced.

Our view is that neither meta data nor embedded data should be separately included in the Rule 34 definition of "document" or automatically required for production in all matters. The addition of "electronically stored information" to the definition captures this information and the relevance of both meta data and embedded data should be discussed during the early discovery planning process where the issues can be addressed in the context of the particular litigation.

E. Interrogatories – Rule 33

Rule 33. Interrogatories to Parties

* * *

(c) Option to Produce Electronically Stored Information. If the answer to an interrogatory may be determined [by extracting, auditing, abstracting, or summarizing] [from] the responding party's electronically-stored data, and if the burden of determining the answer will be substantially the same for either party, the responding party may answer by:

(1) producing the electronically-stored data from which the answer may be determined; and

(2) giving the interrogating party sufficient information [and computer software] to enable it to derive or ascertain the desired information.

Comments:

[8] OCR means optical character recognition – the optically scanned text of the document.

Our view is that Rule 33(d) already adequately covers electronically stored information, and no addition or change is required. However, if an addition were made, it should not include the requirement that the answering party provide the same computer software that it uses to view the data, or that it be required to provide any other software reasonably available to the requesting party. There are major software licensing implications, and potentially substantial costs, associated with such a requirement. Licensees of commercial software typically do not have the right to transfer copies of the software without paying significant fees. The only requirement should be that the electronically stored information be provided in the format in which it is maintained in the ordinary course or in a format mutually agreed upon.

F. Addressing the Producing Party's Burden of Retrieving, Reviewing, and Producing "Inaccessible" Data – New Rule 26(h)(2)

Rule 26

(h) Electronically-stored data.

* * *

(2) Inaccessible electronically-stored data. In responding to discovery requests, a party need not include electronically-stored data [from systems] created only for disaster-recovery purposes, [providing that the party preserves a single day's full set of such backup data,] or electronically-stored data that are {not [reasonably] accessible without undue burden or expense} [accessible only if restored or migrated to accessible media and format] {not accessible [reasonably available] in the usual course of the responding party's {business} [activities]}. For good cause, the court may order a party to produce inaccessible electronically-stored data subject to the limitations of Rule 26(b)(2)(B), [and upon such conditions as the court deems just].

Comment:

We urge the committee to adopt a rule addressing two of the major problem areas associated with electronically stored information: (1) information which is created for disaster-recovery purposes (*i.e.*, backup tapes) and (2) information which is not reasonably accessible (*i.e.*, deleted files, backup tapes on outdated media, or legacy systems whose hardware or software requirements are no longer supported). We support the two tiered approach discussed at the Fordham Conference and believe that the following would adequately and appropriately address these areas:

(h) Electronically-Stored Information. In making required disclosures and in responding to discovery requests, a party need not include electronically-stored information created for disaster-recovery purposes, or electronically-stored information that are not reasonably available in the ordinary course of the responding party's business. For good cause, upon a showing of substantial need, the court may order discovery of such electronically-stored information [within the scope of Rule 26 (b) that is

otherwise producible] upon such conditions as the court deems just [in accordance with Rule 26 (b) (2)][9].

The phenomenon of backup tapes and deleted electronically stored information that *may* be retrievable only at great expense by forensic experts is unique to the electronic world—there is no equivalent in the paper world. We disagree with those who spoke at the conference who said that backup tapes are not really an issue because the technology is changing quickly and backup tapes will soon be outdated. Whether backup tapes or some other media is used, we believe that most businesses will continue to regularly duplicate data from servers onto mass storage tapes or other media for purposes of potential disaster recovery scenarios such as hard drive crashes, fires, floods or earthquakes. Because the purpose of a disaster recovery system is to recover data very soon after relatively rare disasters, backup systems do not constitute and were not designed to serve as archives from which particular documents or finite groups of documents may easily be retrieved, especially over long periods of time. Rather, the backups are by their very nature indiscriminate, capturing all information at a given moment from large numbers of servers that are backed up simultaneously using a number of tapes. As a result, the data on backup tapes is not organized for retrieval of individual documents or files, but rather for wholesale emergency uploading onto a computer system in order to wholesale "restore" the massive amount of data lost in the event of catastrophic failure that usually entails total memory loss or physical destruction of the server hardware. There is no legal requirement that a company or government agency have backup tapes per se, but most companies have backup policies because they are needed for business continuation in a disaster situation such as an earthquake, fire or terrorist attack.

In assessing the need for a rule regarding disaster recovery systems, we believe it is important to understand what data is involved. From a litigation perspective, backup tapes are very poor sources of electronically stored information not only because of the practical difficulties in retrieving specific documents, but also because as a storage medium they are both over-inclusive and under-inclusive. They are over-inclusive because most, if not all, of what is on a backup tape also exists on the server and/or PC that were used for its creation and storage. At the same time backup tapes are under-inclusive because when a full backup of a server is performed, only those documents that are on the server *at that very moment in time* can be backed up and not all of those are backed up. If a document is open and in use, it is not backed up. If a document was deleted from the server prior to the backup, it is not backed up. If a document was not yet created at the time of the backup, it is not backed up.

Backup tapes do not only duplicate documents that are in "live" storage or use on servers, but they also routinely overlap to a large extent with other backup tapes. It is fairly standard for companies to have backup policies that include creating daily, weekly or monthly copies of a complete backup of its primary servers (on a large number of tapes) and many of those are sent to offsite facilities for safekeeping and never touched again. These regular backup tape sessions largely overlap. The tapes from the first month contain those documents present at the moment of the backup; the tapes from month two contain all of the same documents that were on the first

[9] A comment to the rule could identify the various factors for the court to consider in cost-shifting. While Microsoft would support an amendment that required cost-shifting, we believe that this language, with comment, would give the courts the flexibility they need to address unique situations.

set of tapes that were not deleted in the interim, plus new documents. This repeats month after month for as long as the monthly backups are retained. Thus, there is massive duplication between the monthly backups. It is a common corporate practice to follow a backup schedule of daily incremental[10] backups that are recycled every seven days, once a week full backups that are recycled on a regular basis, often monthly. The retention periods vary by company, and can be anywhere from days to years, and may differ by server type. When, in the typical litigation, the events at issue occurred years before, the likelihood of finding important documents on recent backup tapes that are not already on the current server or the PCs of the individuals who created or received the document (and therefore easily retrievable) is extraordinarily rare.

It is not unusual for a large company to end up producing thousands of backup tapes each month for its hundreds or thousands of servers. To reduce the costs of buying new tapes and storage most companies do not keep their daily or weekly backup tapes any longer than needed for their disaster recovery purpose. If companies were required to cease their recycling of backup tapes, this would increase the volume of tapes in storage immensely at great expense. Consistent with their purpose, most tapes are never even accessed. Rather, they are saved for the unlikely event of a disaster that happens infrequently.

From time to time technologically unsophisticated or malicious litigants make over-broad requests for "all" backup tapes for servers containing potentially relevant documents. This is an example of what the representative of the American Trial Lawyers Association called "weapons of mass discovery" at the Fordham conference. Due to the practical difficulty of identifying what files are on what tapes such requests quickly implicate thousands of backup tapes. When it can take up to 30 hours just to restore a single tape, this quickly turns into an immense burden. Most companies do not have the infrastructure to restore that volume of backup tapes, even assuming that they have the hardware and software needed to restore older tapes. But the cost of restoring the tapes, or creating a catalogue of the tape contents, pales in comparison to the cost of searching those documents (even using electronic search techniques) let alone the cost of reviewing the results for privilege or responsiveness. That burden is immense and, in our experience, little more than a fishing expedition that yields nothing of significance.

If Microsoft, or other large companies, were required to take a "snapshot" *i.e.*, preserve a "single day's full set of backup data" each time a lawsuit was filed, the result would be impractical and unworkable. The end result of such a requirement would likely require companies to essentially maintain a daily backup of every document in the company. It would require company-wide retention, even though the individuals involved in the issues of a case might be confined to one small location. It would require the retention of backup tapes of servers not even remotely connected to the issues in the case. For some companies, the cost for just the additional tapes would be hundreds of thousands to millions of dollars. All of this is unnecessary. A backup tape of a company's systems on the day a lawsuit is filed adds nothing of value where the company takes reasonable steps to place the appropriate personnel in the company on "litigation hold". Relevant documents are then maintained on the active server and are available for collection at the appropriate time. The company's normal disaster recovery policies protect those documents against a catastrophic failure of the server.

---

[10] Incremental backups capture only new material added since the last full backup.

Although the incremental evidence that could be gleaned from backup tapes is likely to be very small, the cost of restoring, searching and reviewing the data to the producing party is very large. This has a profound impact on the settlement value of a case. If a commercial dispute has a settlement value of "X", then a discovery request for backup tapes that will cost $1 million to restore, search, review and produce increases the settlement value to X plus $1 million. Even if the defendants' position is sure to win at trial, the threat of huge discovery costs enables plaintiffs to extort significant settlements. Unless this Committee acts, requests for data maintained in disaster recovery systems will continue to be a very effective way for plaintiffs to impose huge costs on defendants for the purpose of extorting settlements under the pretext of looking for evidence.

Beyond disaster recovery systems, the second issue raised in this proposal concerns inaccessible data, such as deleted files, older backup tapes where the hardware or software needed to restore them is no longer available, and legacy systems, such as email systems. It is simply the very rarest of cases where a forensic search for deleted files is necessary or justified. The need to use scientific methods to restore deleted materials is probably more relevant in criminal cases than in commercial litigation. The revised discussion proposal above recognizes that restoring deleted, *i.e.*, inaccessible data, is very difficult and expensive, yet provides litigants and the courts with the means of obtaining such data in appropriate and most likely rare circumstances.

G. Addressing Inadvertent Privilege Waivers

Rule 34. Producing Documents and Tangible Things, or Entering onto Land, for Inspection and Other Purposes

* * *

(b) Procedure.

* * *

(2) *Responses and Objections.*

* * *

(E) ***[Order Regarding] Privilege Waiver.*** [On stipulation, a court may order that] a party may respond to a request to produce documents by providing the documents for initial examination. Providing documents for initial examination does not waive any privilege or protection. The party requesting the documents may, after initial examination, designate the documents it wishes produced; this designation operates as the request under Rule 34(b)(1).

[ALTERNATIVE LANGUAGE]

(E) ***Inadvertent production of privileged material.*** When a party inadvertently produces documents that are privileged, that production does not waive any applicable privilege or protection if waiver would be unfair in light of

(i) the volume of documents called for by the request [given the time available for review of the materials produced]; and
(ii) the efforts the party made to avoid disclosure of the privileged materials; and
(iii) whether the party identified the privileged materials within a reasonable time after production and promptly sought return of the materials; and
(iv) the extent of the disclosure; and
(v) the prejudice to any party that would result from finding -- or failing to find -- a waiver; and
(vi) any other matter that bears on the fairness of waiver.]

Comment:

The first alternative is often referred to as the "Quick Peek" or "Grab Back" approach. A rule that provides for involuntary court ordered production of documents that have not been reviewed is an unacceptable invasion of the parties' legitimate interests and rights in protecting its privileged and work product documents.

Microsoft supports the alternative language addressing the "Inadvertent Production of Privileged Material." We believe this is a permissible procedural rule change providing guidance to the courts and litigants on the factors to be considered in determining whether a waiver has occurred. The criteria set forth in this alternative reflect the current trend in the majority of federal circuits, as well as the state courts. Application of these factors in evaluating waiver[11] appropriately recognizes the volume and complexity of document requests in modern large-scale litigation, as well as the time pressures faced by litigants in responding to these requests, both factors which inevitably result in the occasional unintentional production of a privileged document.

[11] Frequently, other versions of that document, from the files of different custodians, are properly withheld and appear on the privilege log provided to the opposing party. Microsoft would also support a comment to this amended rule that recognizes that treatment of similar privileged materials would be considered as a mitigating factor against waiver under section E (ii) of this rule.

H. Preservation, "Safe Harbor," and Sanctions

Rule 34.1. Duty to Preserve

Upon [notice of] commencement of an action, all parties must preserve documents and tangible things that may be required to be produced pursuant to Rule [26(a)(1) and] (b)(1), except that materials described by Rule 26(b)(2) need not be preserved unless so ordered by the court for good cause]. Nothing in these rules requires a party to suspend or alter the operation in good faith of disaster recovery or other [computer] systems {for electronically-stored data} unless the court so orders for good cause, [providing that the party preserves a single day's full set of such backup data].

Comment:

We strongly support the presumptive exclusion of disaster recovery systems and inaccessible electronically stored information from discovery. We also favor a clear statement that a party is not required to suspend or alter the good faith operation of their disaster recovery systems which includes the regular recycling of tapes in the daily, weekly or monthly rotations that are commonly used to back up servers in corporate networks. For the practical and technical reasons discussed above, the requirement that a party preserve a single day's full set of backup data is unworkable and would impose a significant additional burden on companies already overwhelmed by discovery in the electronic world. It is also unnecessary since backup tapes by definition contain redundant copies of electronically stored information that are far more easily retrieved from live servers and PCs.

Rule 37.
Failure to Make Disclosure or Cooperate in Discovery; Sanctions.

(f) Failure to Produce Electronically-stored Data. A court may not impose sanctions on a party [under Rule 37(b)] for failure to produce electronic documents unless [the court finds that]

(1) the party deleted, destroyed, or otherwise made unavailable electronically-stored data after service of discovery request that were described with reasonable particularity in a discovery request, or [and]

(2) the party willfully or recklessly deleted, destroyed, or otherwise made unavailable electronically-stored data in violation of [Rule 34.1] {Rule 26(b)(3)}.

Comment:

An amendment protecting litigants from the threat of sanctions for spoliation resulting from the normal operation of backup systems would be most welcome. There have been examples of plaintiffs accusing defendants of "intentionally deleting" email because it recycles its email server backup tapes on a regular basis. That argument, of course, totally ignores the limited disaster recovery purpose of backup tapes, but nonetheless places companies in the untenable position of later facing serious sanctions for merely continuing its regular disaster recovery processes. The Committee should clarify that proposed Rule 37(f)(1) does not implicate the routine recycling or disposal of backup tapes in the ordinary course of business. That could

be achieved by enacting proposed subparagraph (2) rather than (1). Obviously, the willful or reckless destruction of data should be sanctionable and a comment should point out that the routine recycling of backup tapes would not per se be willful or reckless.

We do not believe that the safe harbor provision should be limited to Rule 37(b) sanctions – the rule should apply to all forms of sanctions. One concern with sub-paragraph (1) is that it would appear to permit a litigant to describe with "reasonable particularity" backup tapes or other inaccessible data, thereby circumventing the purpose of the draft proposals. At a minimum, the rules should make it clear that the finding must be unrelated to Rule 34.1 / 26(h)(3) issues.

* * *

We thank you for the opportunity to provide comments on the draft discussion proposals. Please feel free to contact me if we can be of any assistance to the committee.

Very truly yours,
Microsoft Corporation

By Thomas W. Burt
Thomas W. Burt
Vice President & Deputy General Counsel