View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Fordham University School of Law

Fordham Law Review

Volume 73 | Issue 1 Article 3

2004

Rules 33 and 34: Defining E-Documents and the Form of Production

David R. Buchanan

Adam I. Cohen

James C. Francis IV

Paul M. Robertson

Follow this and additional works at: https://ir.lawnet.fordham.edu/flr

Part of the Law Commons

Recommended Citation

David R. Buchanan, Adam I. Cohen, James C. Francis IV, and Paul M. Robertson, *Rules 33 and 34: Defining E-Documents and the Form of Production*, 73 Fordham L. Rev. 33 (2004).
Available at: https://ir.lawnet.fordham.edu/flr/vol73/iss1/3

This Article is brought to you for free and open access by FLASH: The Fordham Law Archive of Scholarship and History. It has been accepted for inclusion in Fordham Law Review by an authorized editor of FLASH: The Fordham Law Archive of Scholarship and History. For more information, please contact tmelnick@law.fordham.edu.

Rules 33 and 34: Defining E-Documents and the Form of Production

Cover Page Footnote

United States District Court Judge, Southern District of New York *Partner, Seeger Weiss LLP *Partner, Weil Gotshal & Manges LLP *United States Magistrate Judge, Southern District of New York *Partner, Bingham McCutchen LLP

This article is available in Fordham Law Review: https://ir.lawnet.fordham.edu/flr/vol73/iss1/3

CONFERENCE ON ELECTRONIC DISCOVERY

PANEL TWO: RULES 33 AND 34: DEFINING E-DOCUMENTS AND THE FORM OF PRODUCTION

MODERATOR

*Hon. Shira Ann Scheindlin**

PANELISTS

*David R. Buchanan***

*Adam I. Cohen****

Hon. James C. Francis IV†

Paul M. Robertson‡

JUDGE SCHEINDLIN: The topic to briefly discuss is the definition of e-data, and we are going to do that very briefly; that's a three-minute segment of this show.

We will then turn to the question of whether Rule 34 needs revision in order to refer to "data" or "information" rather than "documents,"[1] which as you just heard may be a passé concept in the 21st century. Listening to the last panel, I must say that relational databases and formulas for spreadsheets do not entirely sound like "documents."

Questions that we will cover in that segment will include such things as: In producing data stored on electronic media, should that production include all data stored or maintained as part of the electronic record? Just to whet your appetite.

Our next topic, our third of four, will be the form of production question. The question there is: Should Rule 34 require the requesting party to specify a particular form for producing the

---

* United States District Court Judge, Southern District of New York.
** Partner, Seeger Weiss LLP.
*** Partner, Weil, Gotshal & Manges LLP.
† United States Magistrate Judge, Southern District of New York.
‡ Partner, Bingham McCutchen LLP.

1. Fed. R. Civ. P. 34.

requested data; and should the Rule also talk about the grounds on which a producing party might object, such as inaccessibility?

Finally, we will briefly turn to whether Rule 33[2] needs to be amended to specify that interrogatories may be answered by the production of electronic data; and, if so, what responsibility might the producing party have to produce that data in some way that is actually usable?

Now, on each topic we have decided to go in this order: our author, Adam, will go first and try to give us the very briefest of backgrounds; David and Paul, who are set up a little bit to be sparring partners, a little bit of plaintiff/defense viewpoints, will then go next; and, as is always appropriate, the judge will get the last word on every topic, and of course that is Judge Francis, not your uncharacteristically quiet moderator.

MR. ROBERTSON: I guess a couple of thoughts just before I start this. The dichotomy that has been set up is between defendants' and plaintiffs' bar. As David and I talked in preparation, we found that on a lot of stuff there is some agreement here on the result that should be reached. We really wanted to make sure that we kept both questions in front of us at all times.

The first one was: Is there a problem; is there something that needs to be fixed? Only then did we get to the second one: Okay, if there is a problem, what is the proposed rule change? In all instances, even if I thought that there wasn't necessarily a problem, I thought it was important to at least propose a suggested fix, some suggested language. To the extent that a proposal was put forth, at least we had something to talk about.

In this instance, the first question, "Is there a problem?," is if we are going to put in some language in the rest of the Rules to talk about electronic discovery, do we need to define what the subject matter is at the starting gate?

If you take a look at some of the other states and federal district courts that have put in rules, none of them did so.[3] None of them defined electronic discovery. I think that looking through it, my thought after looking at what some of the other jurisdictions have done, and the general premise that definitions are not favored in the Federal Rules, I did not think that a definition was necessarily appropriate.

I think that if you talk to folks in the places where it has been put in place, when you talk about electronic discovery, most folks do not need to run to a dictionary to find out what it means.

I thought that to the extent, though, that we use a definition, I thought about the one that had been proposed, and I thought it was

2. Fed. R. Civ. P. 33.

3. *See, e.g.*, E.D. & W.D. Ark. Appx. R. 26.1.

an excellent start, and I molded mine working with that one. I had a couple of comments to it, though.

One, it talks about whether the information is "created, maintained, or stored in a certain capacity." I thought that it's okay to just simply say that it is stuff that is in a digital format.

I thought, too, the final part, the attempt to try to identify some of the sources from which this information could come, the definition was "computers, telephones, PDAs, media players, media viewers, etc." I thought maybe that might suffer from the fatality that Ken Withers had identified, that things move so quickly that if you talk about a PDA, in five years folks are not going to know what that is. You know, the techies tend to change these definitions before you have taken the equipment home.

So I tried with my definition, "electronic data is recorded information," and I thought it should say "recorded" because there is a danger I think that, although some of this stuff is becoming more abstract, that the abstraction shouldn't be removed from having it tangible. It is something that is kept somewhere, as compared to something that is an ethereal idea in a witness's head—"that is readable and available only through the use of electronic or other technological means." I put the "other technological means," and I thought that as we are moving along, maybe we do not want to limit it to electronic means, that for example biological and chemical data, although it sounds awfully farfetched today, I think some of the things that we talk about today sounded farfetched ten years ago.

So that was the proposal that I thought of.

JUDGE SCHEINDLIN: Since this really is our three- to five-minute segment, does anybody want to say anything more about that, or should we get right into Rule 34 and documents? Anybody want to comment on this one?

MR. COHEN: Just a couple of quick comments.

One, there is a problem with including documents that were created electronically as electronic information because that can be converted into paper and then it is not what we are thinking of as electronic information.

Also, I just want to point out there is a very interesting issue in terms of what is tangible when applied to data. Some of you may be familiar with all sorts of different cases, cases applying the "trespass to chattels" theory to documents, to electronic information; and cases dealing with whether insurance policies cover electronic information. So that is something we may all have different theories about in terms of use of the word "tangible" with respect to electronic information.

I think what is clear is that we are not talking about paper, we are not talking about oral testimony, and we are not talking about things like the cow in the "replevin for a cow" case that we all read on the first day of law school.

JUDGE SCHEINDLIN: Okay. I think we should probably turn to the big topic that we have for our panel, which is Rule 34.[4]

The question that we are really going to begin with, in the order that I mentioned earlier, is: Do we need to revise Rule 34 at all to define "data" or "information" and turn away from the concept of "document," which may be creating misunderstandings and causing problems? We are going to address that in the order we said. Adam, do you want to give us a start?

MR. COHEN: Okay. I am just going to try to set up some of the issues here and give a little bit of context.

The current Rule talks about "data compilations,"[5] which to us today probably sounds like a little bit of an odd formulation. It is not a phrase that we tend to use, although in 1970 it probably sounded almost like science fiction.

If you look at the notes[6] where that phrase was imported into the Rule, it is actually quite prescient, I think, in terms of recognizing changing technology, the requirement of using devices, which is similar to what we were just talking about in terms of electronic information, needing to use some kind of technology to look at it.

The last sentence is kind of funny in the conclusion there.[7] It's sort of what Ken was talking about, taking all the e-mails and printing them out. I think the way we look at this has changed.

---

4. Rule 34(a) reads:

> (a) Scope. Any party may serve on any other party a request (1) to produce and permit the party making the request, or someone acting on the requestor's behalf, to inspect and copy, any designated documents (including writings, drawings, graphs, charts, photographs, phonorecords, and other data compilations from which information can be obtained, translated, if necessary, by the respondent through detection devices into reasonably usable form), or to inspect and copy, test, or sample any tangible things which constitute or contain matters within the scope of Rule 26(b) and which are in the possession, custody or control of the party upon whom the request is served; or (2) to permit entry upon designated land or other property in the possession or control of the party upon whom the request is served for the purpose of inspection and measuring, surveying, photographing, testing, or sampling the property or any designated object or operation thereon, within the scope of Rule 26(b).

Fed. R. Civ. P. 34(a)

5. *Id.*

6. Fed. R. Civ. P. 34 advisory committee's note. The Advisory Committee's note to Rule 34 provides:

> The inclusive description of "documents" ... makes clear that Rule 34 applies to electronics data compilations from which information can be obtained only with the use of detection devices, and that when the data can as a practical matter be made usable by the discovering party only through respondent's devices, respondent may be required to use his devices to translate the data into usable form.

Id.

7. *Id.* ("Similarly, if the discovering party needs to check the electronic source itself, the court may protect respondent with respect to preservation of his records, confidentiality of nondiscoverable matters, and costs.").

There is also a recognition of the potential need to check the source itself, so even in 1970 recognizing that there may be information that you do not see when you print this stuff out.

I just want to point out that some local rules and state rules address whether electronic information is included within the scope of what is normally considered a "document" and whether it presumptively is or it is not. You have these rules in Texas[8] and Mississippi[9] where you have to specifically request electronic information and it will not be presumptively considered a document.

In Virginia, you have this rule dealing with subpoenas. It requires you to produce what they call a "tangible copy of [electronic] information."[10]

The central problem that I see, which was pointed out by the prior panel, is this issue of: Do we talk about "medium" or do we talk about "information" whatever the medium? There was a suggestion in the materials of a limited change, adding "data" or "data compilations in any media."

Then there is also a talking point to address the issue of metadata and embedded data, as to whether those are included in the definition of a "document."

JUDGE SCHEINDLIN: Okay. Dave?

MR. BUCHANAN: I guess when asked to consider the proposed amendment, the first thing that occurred to me is: What do I think we would all agree is information that should be disclosable in litigation? The last panel I think was pretty instructive in guiding us about the types of information that parties are wrestling with in terms of discovery disputes, and then, once we understand what we think should be disclosable in litigation, then make the definition fit the types of categories to make sure that we are at least broad enough.

The things that came out in the last panel were databases, relational databases, e-mail, spreadsheets, PowerPoints, embedded data, metadata, and backup tapes. These are all things that we are talking about as being sources of electronic data that may be disclosable.

Now, I am certainly not advocating a laundry list in a rule—I think that would be problematic—but the definition I think has to encompass those. The definition should not strike a balance between the relative burdens among the parties in terms of identifying or producing certain information. I think that is an important issue. That is an issue that needs to be addressed, though, elsewhere in the Rules, perhaps in Rule 26, or by the court in applying Rule 26.

---

8. Tex. R. Civ. P. 193.3, 196.4.

9. Miss. R. Civ. P. 26 (amended by Miss. Sup. Ct. Order 13, No. 89-R-99001-SCT, May 29, 2003).

10. Va. Sup. Ct. R. 3A:12(b).

The definition of "documents" has not caused problems for me in getting all the electronic data that I have needed. It has included relational databases, e-mails, metadata, and embedded data in very large litigation. So I think the Rule has been extended in such a way so that the definition encompasses those items.

That having been said, there are two items, embedded data and metadata, that present the thorniest issues under the current Rule. I would submit—and we'll talk about it in a little bit—that those should be items that are presumptively documents but perhaps not something that you get in every case.

But in thinking about what a "document" is, it certainly includes everything within the file. It includes the creation date, the edit dates, who did it, all that information that's all within the native file. It includes the embedded information within the file. I think it is the wrong place to strike the balance in Rule 34. If there are any issues of burden, that should be addressed elsewhere.

I could certainly address a proposal for the Rule, if you want to do that now.

JUDGE SCHEINDLIN: If you're staying in this part of it, sure.

MR. BUCHANAN: There has been a suggestion, and I think Adam highlighted it, that we should be talking about "information" or "data" that is "fixed in a medium." I think that eliminates the ethereal concept that we spoke about a moment ago, information that just crosses the wires, doesn't really register in any system, but yet it preserves the real object of a "document." There is something tangible, there is something physical. It is "information" or "data" that has been "fixed in a medium." Even if it changes over time, it has been fixed.

JUDGE SCHEINDLIN: All right. Paul?

MR. ROBERTSON: I guess I am in general agreement with David on this, that from the defense perspective we do not see that much of a struggle over whether a particular electronic piece of information is considered a "document." The struggle is always whether it is relevant to a particular case.

There are two issues that I think have been identified—and I look back. The first time I saw them identified is in the article that Judge Scheindlin and Jeff Rabkin did four years ago, which was extremely prescient in nailing some of these issues.[11]

One is, Is there a need to untie this to documents? A lot of this stuff doesn't really fit our old definition of "document." Things like cookies and other embedded information, does that really fit into the definition of "document"; shouldn't it be called "information"?

11. Honorable Shira A. Scheindlin & Jeffrey Rabkin, *Electronic Discovery in Federal Civil Litigation: Is Rule 34 Up to the Task?*, 41 B.C. L. Rev. 327 (2000).

The second excellent point was, look, these are data compilations. A compilation, if you look at the definition, is a heaping together, a collection of information from other places. Much of this data is not a collection or a heaping together; it is created in the first instance. I think of a cookie again as an example.

But I think that, again, although those are issues that have been identified, neither the practitioners nor the bench struggle with them. If you take a look at the *Anti-Monopoly, Inc. v. Hasbro, Inc.* case from several years ago, it really sets forth the law here, and I quote it: "[I]t is black letter law that computerized data is discoverable if relevant."[12] I think that has really become the issue.

So I do not see the need for a fix, even though there is a little bit of a discrepancy between what is being done in practice and what is actually written in the Rules. Given that everybody accepts that the definition described in the Rules today includes not only compilations of data but also data itself, there is not really a need for a fix.

To the extent of getting to the point if there were language to be included in the Rules, I think that adding the word "data" before the word "data compilations," so you simply say "data and data compilations," would serve that fix. I do not think that it would do any harm.

I do not think that you will find that it is a big-ticket item for either the defense bar or the plaintiffs' bar or the judiciary, but it would perhaps make the Rules consistent with what everybody's understanding is and it would clean up that confusion.

JUDGE SCHEINDLIN: All right. Judge?

JUDGE FRANCIS: I think as a judge one of my primary concerns is conflict avoidance. One way to avoid conflicts is to have clarity in the Rules, and particularly in the definitions.

I think that while there has not been a massive problem with the definition of "documents," for the reasons that my colleagues have described, I think it may well be advisable to bring the definition into conformity with actual practice, particularly because the definition of "document" basically creates a default position. In the absence of judicial gloss on this, people look to the Rules. "Document" I think suggests paper, and I think it may be helpful to expand that.

I think it has implications for other parts of the Rules. For example, when a party is going to respond to a document request, are they going to search for everything but then respond in paper because the Rule currently talks about "documents"?[13] So I think [an expansion of the definition of "documents" in the Rules is needed] in order to provide some clarity and to bring things in line with real practices.

12. Anti-Monopoly, Inc. v. Hasbro, Inc., No. CIV.A.94-2120, 1995 WL 649934, at *2 (S.D.N.Y. Nov. 3, 1995).

13. *See* Fed. R. Civ. P. 34(a).

And also I think to anticipate the future. We may agree that everybody understands now that computerized information is a "document," but when we go on beyond computers and we talk about biological information and so forth, is that going to be encompassed within the information that would be discoverable under Rule 34? I think we should adapt to that as well.

JUDGE SCHEINDLIN: Before we turn to our next topic, which is metadata, let me just ask you all one question. There is information or data that is stored and never reduced to a document, such as transient information, like spreadsheets, and they change every time the parameters are changed, or a daily example might be an e-ticket that is never a document unless it becomes printed. So there is information in data that is simply stored on a medium but is not yet a document. Does that question make you think that that needs to be addressed in this definitional Rule 34?

MR. BUCHANAN: The important point I think with an e-ticket, for example, is there is a database behind that e-ticket that contains all the parameters. There is something electronic in nature that has been fixed in a form that contains all the parameters of that e-ticket.

The same with the spreadsheet that you highlighted. While it may change day-to-day and you have multiple versions of the document, the formula, for example, within the document is the same perhaps, or maybe that changes over time, too, the resulting numbers.

JUDGE SCHEINDLIN: But as the last panel said, if you printed it out, you would never see that formula. So the question is, Can you obtain that data when you think of the term "four-cornered document?" That is the question I am asking.

MR. BUCHANAN: I agree. I think that is more of a production issue in my mind, the format in which it is delivered to the other side.

JUDGE SCHEINDLIN: Okay. Anybody else want to address that?

MR. COHEN: I just want to say that it seems that with the types of electronic information that we have these days and that are becoming more and more prevalent, such as transient data, instant messaging, and digitized voicemail, we are moving closer and closer to what is more like oral communication in how evanescent it is.

We might ask ourselves: If we are going to require data like this to be captured and produced, does this mean now that when we have oral conversations about a case when we are under a duty to preserve we should be recording it all?

JUDGE SCHEINDLIN: Let's turn to an issue that at least in the Advisory Committee we spent a lot of time thinking about, and that is the question very specifically now of the production of metadata and embedded data. I shouldn't have thrown in the word "production." Put that aside for a minute. Just whether Rule 34 conceptually would call for the production of metadata and then later embedded data. I

would like to take those separately because they are different concepts.

Let's talk for a minute about metadata, starting with Adam.

MR. COHEN: Okay. Just to set up the issue, Do you make this a routine requirement of production; do you make it a permissive requirement?

What are the positions on either side? You know, on the one hand, opposition to routine requirement would be based on the notion that there is not really a likelihood that it is going to be terribly material. It is going to add costs. On the other hand, there are situations where you are going to be adding more costs by stripping that data out—and believe me that happens a lot in real life, oddly enough.

On the other hand, you might need the metadata to facilitate the searching, the manipulation, the kinds of litigation databases that people use right now to handle large amounts of documents. Some of the formats that people produce their documents in, these image formats without the metadata, require a lot of work before they are actually usable in one of these databases.

So the question becomes, Should this be presumptively something that gets produced or is it only available by special permission? We have some positions on that that have been taken by members of the bar and the judiciary.

The Sedona Conference[14] document shows a position where this type of information is presumptively not something that is included in a production unless there is separate analysis on a case-by-case basis.

The ABA talks about "duty to preserve" in a very broad way, specifying it as "media" rather than the type of information.[15]

That brings us to form of production, so why don't I let the panelists talk about metadata?

JUDGE SCHEINDLIN: We are going to hold off on form for a little while. Let's just talk about the concept of metadata as something that ought to be produced with the information, or not. Dave?

MR. BUCHANAN: Again, I think, focusing within Rule 34, the

14. The Sedona Conference is accurately described as "a research and educational institute dedicated to the study of law and policy . . . [which] . . . has developed best practices recommendations . . . for addressing electronic document production." Albert Barsocchini, *EDD Services' Growth Rate is Staggering*, The Legal Intelligencer, Oct. 1, 2003, at 5, 5. The Sedona Principles, the prior working draft, and related documents can all be accessed online at http://www.thesedonaconference.org/publications_html (last visited July 19, 2004).

15. Litig. Section, Am. Bar Ass'n, Civil Discovery Standards § VIII (a), at 49 (Aug. 1999), *available at* http://www.abanet.org/litigation/civiltrialstandards/home.html ("A party's duty to take reasonable steps to preserve potentially relevant documents . . . also applies to . . . electronic medium or format . . . word-processing document[s] [and] . . . electronic mail.").

conclusion I reach is that this is supposed to talk about the types of forms of documents you can request or the types of forms of information that can be requested in litigation.

I think metadata unquestionably can be relevant to a claim. We have seen—well, how about in paper productions of years past a file routing slip on the top of a document that showed when a document went to somebody, when it moved to the next person, or a revision history that tracked changes to a contract over time? These types of things were discoverable. They were affixed to a document or to a file.

Now we have electronic documents that have different flavors of similar concepts. Rule 34 needs to contemplate that those types of documents are documents or information discoverable in litigation.

If there needs to be a balance struck, it should be struck elsewhere.

One thing that I think is important to note is there are other proposed changes in Rules 26 and 16 that require the parties to talk.[16] What I heard from the last panel, and I think it is an important issue, is that the parties need to talk. I would expect that metadata and embedded data would be something that would be discussed during those early planning conferences both privately and with the court.

So I think Rule 34 is not the place to limit this. Rule 34 should be encompassing, though, of metadata and embedded data. The question is: Is rule reform necessary to accomplish that?

This is the only area in my practice I think where there is any debate with defendants about whether metadata or embedded data is a "document." So I do believe that clarification would be helpful in that regard, but it should not be on a showing of good cause within Rule 34.

JUDGE SCHEINDLIN: So in other words, you don't think it's second tier; you think it is presumptively part of the data?

MR. BUCHANAN: It is. It's within the file wrapper.

JUDGE SCHEINDLIN: Okay, it's within the file wrapper. A little patent law. Okay. Paul?

MR. ROBERTSON: I think that the first thing to do—you know, the issue of whether there is a problem, I think there is a need to unpack embedded data and metadata for a second because they really are different things.

Folks talk about metadata and they quickly say it's information that is embedded in the document. If you look at some of the articles and some of the writings on this, the excellent article by Judge Scheindlin and Jeff Rabkin talked about embedded data as being metadata;[17]

16. At its April 15-16, 2004 meeting the Advisory Committee on Civil Rules approved proposed amendments to Civil Rules 16 and 26, among others. *See* http://www.uscourts.gov/rules/ (last visited Jul. 19, 2004).

17. *See* Scheindlin & Rabkin, *supra* note 11, at 338.

they used the terms interchangeably, as a lot of folks do. Sedona talks about metadata being embedded data.[18]

But they really are different things. I think that metadata is the information about the information in the document—things like, in an e-mail, the code that tells how the e-mail is to be delivered, how it is to be routed; the information in a Word document, paragraph shifts; and information in a spreadsheet about how calculations are to be made.

I think, on the other hand, when we talk about embedded data, it is a very different animal. It is typically user-created edits or information that is put into the document purposefully—things like track changes; things like a sticky note that you put underneath; things like other versions of the document that are hidden beneath it. I think that those are very different things.

I think that when you are talking about embedded data, the way that I understand it more easily is to think about embedded edits. I think that edits to a document certainly are in certain circumstances presumptively discoverable as a type of draft of the document.

I think, on the other hand, in 99.9% of the cases metadata is irrelevant because it is not even the envelope that you are sending the e-mail in—and most of this stuff, by the way, is about e-mail—it is not the envelope information, who the e-mail is from, to whom it is being delivered, but instead, it is instructions that you have given to the mailman about how to take it, how it is to be routed, and then information about how that e-mail was actually delivered.

In most instances you don't keep the FedEx package, you don't keep the instructions telling the FedEx man or woman to go to this certain place. It is not typically relevant. It is the equivalent of having to go back and say, after doing a document production, "I want to fingerprint your data room to find out who was in there and who was not."

That said, I would certainly agree that in some cases it is very relevant. Martha Stewart is an example of a case where you wanted to find out about who edited this document and when.

But the question I have next is: Is there a dispute about whether that is considered a "document"? I think that again the Rules do a very handy job of this. I don't see any cases out there where a court has said, "You can't have it because it's not a document." The issue becomes, "You can't have it because it is not relevant."

Even in those cases where you do need metadata—information about information—it is usually targeted to very few spreadsheets, a few e-mails, and in most cases a requesting party does not want to get with each document sometimes 800 pieces of information about that e-mail that neither do they need nor they understand.

---

18. *See, e.g.*, Abigail E. Crouse & Stephen J. Snyder, *Applying Rule 1 in the Information Age*, 4 Sedona Conf. J. 165 (2003).

So I don't think that there is a problem with respect to metadata. I think that most folks understand that it is a "document," but the question is whether it is relevant.

JUDGE SCHEINDLIN: And the judge?

JUDGE FRANCIS: I love seeing so much agreement between plaintiffs and defendants.

I think there is agreement that metadata and embedded data are information and that they are at least potentially relevant and therefore come within, or should come within, the definition of "document"[19] or "discoverable information,"[20] however we characterize it.

I think the tougher issue is whether there should be some good cause requirement imposed before a requesting party has access to that information. There I would point out that as a judge one of the values that I try to embody is doing justice, and that means being able to adapt the law to the facts in a particular case. The more constraining the Rules are, the more difficult it is for me to do that adaptation.

If there is a good cause requirement, it is a thumb on the scales of justice, and somebody is going to have to overcome that presumption in order to get what may ultimately be relevant discoverable information.

I think it is preferable to leave that to be determined on a case-by-case basis. I think the ABA's approach to putting the burden on the requesting party to ask for that kind of information[21] is fine, but to place a burden of persuasion on that party I think would probably be a mistake.

JUDGE SCHEINDLIN: I think we are going to talk more about metadata and embedded data when we move to form of production, so don't worry that we have left it behind. We are going to get a second round of hearing about it.

We are now going to turn to the form of production question. The question there is whether Rule 34 should be amended to require, either permissively or mandatorily, that the requesting party state the form in which the e-data is to be produced. If so, should that request be as simple as "I want paper" or "I want an electronic mode of production"? Or should it be more complicated, such as, "I will be satisfied with a mere .tiff image," or "I want a .pdf-searchable," or "I want native digital information produced in a specific format, like a

19. *See* Fed. R. Civ. P. 34(a).

20. *See* Fed. R. Civ. P. 26(a)(1)(A).

21. *See* Electronic Discovery Task Force, Litig. Section, Am. Bar Ass'n, Memorandum to Members of the Bench, Bar and Academia (Nov. 17, 2003), *at* http://www.abanet.org/litigation/taskforces/electronic/document.pdf.

DVD, and it has to be compatible with my Windows operating system"?

So the question is what level of specificity should the requesting party have to express if they should have to make a choice at all?

Then, of course, the flip side of that question is if they don't specify, is there a default mode of production?

The third question, I suppose, is what is the producing party's ground of objection there? Can the producing party say, "I shouldn't have to produce it all because it is inaccessible"?

So it is that series of questions that we are about to address. And I think, inevitably, in addressing those we are going to get back to the metadata and embedded data because how you produce it may mean whether or not you include those types of information.

So with that quick background—maybe I did too much—Adam?

MR. COHEN: Okay. I think that is right. There is the segue right there—and we should probably talk to some of the technology people about this—but the parties' determination of whether they are going to resist production of metadata or embedded data may depend largely on what format they have their documents in and they are ready to produce them in. If they have gone and printed out all the e-mails and scanned them in and created electronic images that are stripped from the metadata, then they are not going to want to produce the metadata.

Why require or permit a specification of the form? Well, if you ask for documents in a certain form, this should preclude you later on from coming in and saying, "No, no, I want something different." On the other hand, making it optional may make sense because at an early stage in the case when you are formulating your request you may not know what format you need or what the other side uses or what is going to make the most sense in general. In any case, there is always going to be a need to balance the burden of producing in a certain form against the utility to the other party.

Another issue that has come up—I don't know how common this issue is—is that there are certain proprietary aspects to certain formats. For example, parties have data that can only be viewed with certain proprietary software and generally will resist producing that type of software. At any rate, it seems like it would make sense to discuss it in the initial conference.

If you look at the Sedona Conference, the position that they represent, they talk about the importance being the substantive information content, that you should not have to produce documents in more than one format. They suggest that "production of electronic data in a commonly accepted image format should be sufficient."[22]

---

22. Jonathan M. Redgrave, *The Sedona Principles: Best Practices Recommendations & Principles for Addressing Electronic Document Production*, 4

Now, that has implications of course for metadata and embedded data. "Data that is not ordinarily viewable or printed when performing a normal print command need not be produced."[23]

At the same time, there is a recognition that electronic formats may be preferable in many cases; whatever format is chosen should deal with the genuineness/authenticity issues; and that there should never be a requirement to produce in both hard copy and electronic form. I know this is something that is often the subject of debate based on the case law that is out there already.

The ABA has said that you should consider asking for production in electronic form—you should consider asking for production in a form that gives you the ancillary information.[24]

And then there are some of the cases that were talked about that deal with these issues in different contexts.

Here the *Bristol-Myers* case[25] shows what happens when a party goes ahead and scans all these paper documents into images and then they want to produce them back in paper as per the ancient past. This was a case where I guess no one had said anything about the fact that these documents were available electronically, and were trying to get somebody to pay the cost of a normal paper production when that wasn't really necessary. In that case, they were required to produce in electronic format.[26]

Interestingly, and probably most controversially, there was no requirement that the other side, the requesting party, pay for any of those costs that were involved in creating the electronic format.

The issue of the proprietary format came up in the *Honeywell* case,[27] where PricewaterhouseCoopers stated that production of these documents in a usable form would require the use of proprietary software or large cost. The court basically gave them the option of either producing the proprietary software, the proprietary format, using the protective order, or paying for it themselves.[28]

And then finally—and this shows another aspect of this issue—the *McNally* case,[29] which shows no presumption that you get the computer files when you've got the paper production because you need to show some sort of special basis for it.

Sedona Conf. J. 197, 229 (2003).

23. *Id.*

24. *See* Electronic Discovery Task Force, *supra* note 21.

25. *In re* Bristol-Myers Squibb Sec. Litig., 205 F.R.D. 437 (D.N.J. 2002).

26. *Id.* at 442-43.

27. *In re* Honeywell Int'l, Inc. Sec. Litig., No. M8-85, 2003 WL 22722961, at *1 (S.D.N.Y. Nov. 18, 2003).

28. *Id.* at *2.

29. McNally Tunneling Corp. v. City of Evanston, No. 00-C-6979, 2001 WL 1568879, at *1 (N.D. Ill. Dec. 10, 2001).

MR. BUCHANAN: I suspect this will be more of a point of departure between the plaintiffs and the defendants, and that is the form of production.

There is no question that plaintiffs prefer as a general matter native production of electronic files. That provides all the embedded data, the metadata to the extent it has been appropriately preserved. It gives you the opportunity to quickly search for terms. In short, it puts you on the same playing field as the defendants, or the company at least, in accessing their own data. Those are the arguments plaintiffs use to get native production.

But you may not want a native production in all cases, and that is why it is important I think for there not to be a presumptive production format of native, because we talk about proprietary formats. Or even if we're talking about relational databases, if I have to receive all of your databases in a native format, I may not have the capability of rebuilding that, as opposed to me meeting with you and discussing the appropriate searches to run on the data, extracting the data, running it in reports, and producing the electronic versions of the reports that I can then load into my database.

So I think again this is something that is in Rule 34, but I think it is something that will be addressed quite specifically by the parties at their 26(f) conference and at the Rule 16 status conference as to how to treat non-paper documents: How are we going to treat electronic data? How are we going to produce it? How are we going to preserve it? What are we going to do with the embedded data? And what are we going to do with the metadata?

The Rule needs to contemplate the production of native data. That is the most easily usable form for litigants as a general matter. That statement can be thrown completely out the window, though, when it comes to large proprietary systems where a smaller plaintiff, or even a large plaintiff, doesn't have access to the software to view it.

JUDGE SCHEINDLIN: Let me just ask you a few quick questions. So do you favor a rule that mandates the requesting party should select the form of production it wants? You are usually a requesting party. Should you have to state what you want?

MR. BUCHANAN: I will, and I do, and I will do that in the Rule 26 conference and I will do it in the Rule 16 conference. I think making it permissive to do so and making it permissive for the other side to object to the form requested is fine. But I think having a presumptive form of production would tilt the scales in favor of something that may not work across a large-scale litigation.

JUDGE SCHEINDLIN: That was only my first question. Should you have to specify to avoid confusion?

MR. BUCHANAN: I think it should be permissive.

JUDGE SCHEINDLIN: Permissive, okay.

The other question I have for you is should the Rule talk about "the

data should be produced in the form in which it is created, in which it is ordinarily created"? Should that be the fallback, presumptive form?

MR. BUCHANAN: Here's what I want. I will let people who are good with language and the Drafting Committee tell me the best way to implement it. What I want is information that is as accessible or as usable as on the defendant's system. I mean that is what I want. In many cases that is native files. In other cases with complex databases, it may be an extract of the data from the databases.

JUDGE SCHEINDLIN: Okay, Paul?

MR. ROBERTSON: I hate to disappoint once again, and I think it is a function of how reasonably David approaches most of these issues, but I don't substantially depart from what he is saying.

I think that, again, the first question is, "Is there a problem?" As he said, "Look, in some cases I want the data in its native format," and I think that is absolutely right. In some cases, there are issues where the data in its native format is relevant. I think that in other cases you don't want that.

The Sedona Principles took the approach that, in most cases, production in paper or .tiff images is acceptable.[30] That draws gasps from a lot of plaintiffs, and rightfully so in these mega-document cases. But I think that sometimes we forget, with all of these numbers of terabytes and petabytes, that in most cases the typical sides are not looking at that kind of volume of documents, they're looking at a smaller volume of documents.

So when you create a default position that says things like, "you have to express how you want electronic documents to be delivered to you," often it is only 1000 pages or 2000 pages, and so getting electronic documents isn't necessarily useful.

Is there a presumption that a party should have to produce things in its native format? I think that the answer is there should not be such a presumption because it is a very fact-driven issue.

JUDGE SCHEINDLIN: Should the requester have to ask up-front and specify [the format]?

MR. ROBERTSON: I think that the answer is this. A caveat, though, is of course if you specify electronic documents—that doesn't get you to where you want to go, by the way, because you will sometimes get a .tiff image, and that still is the equivalent of getting a hard-copy document. So I think that, as Dave was saying, you want something that is both electronic but then searchable in the same way that the defendant had it.

I think that there are three ways to go about this: education is one; two is putting something in the Rule 26(f) conference; and three is

30. *See* Redgrave, *supra* note 22, at 223.

putting it in Rule 34. I think that the first two answers are the way to go. I think that this is something where education is needed, where places like the *Manual for Complex Litigation, Fourth*[31] has some language to encourage parties to talk about this. I think it is important to put it in the Rule 26 checklist,[32] to make both parties talk about these kind of things, or suggest that they do, so that they can avoid these situations in cases where they are relevant.

I don't think it is appropriate to put it in Rule 34 for a couple of reasons. One is that there are good reasons that defendants do not produce things in their native format, and it is not simply to hide things. It is because, for example, you can't Bates stamp things, they are manipulatable by the discovering party, and they can be changed so you show up in court and something that you produced in one format looks much different than it otherwise did.

I think the other problem is that if you set up in Rule 34 the suggestion that one party "must" or "may" specify and the other party has the right to object, you create a sort of presumption that there is this right to get things in a native format. And I don't think that anybody is going there. I think people are saying it is a communication problem, which I think is best handled with 26(f).

JUDGE SCHEINDLIN: How does that play into our discussion of metadata and embedded data, though? I mean, if you are doing a .tiff image, you are presumptively not getting it. If you are doing paper, you are presumptively not getting it. I think Dave said he thinks he presumptively should get it.

MR. ROBERTSON: I think that—and I'd be willing and eager to hear David's comments on this, of course—in many cases the metadata is not relevant. When you say metadata, I think that you want to have it searchable.

You want to get an e-mail that even if it is produced in .tiff, you have a concomitant list of searchable data that allows you to organize it by sender, by recipient, by date. That is important. But when I think of metadata, I think about pages and pages of code about how the e-mail got from Tallahassee to Gainesville via some server out in the western part of the country. I don't think anybody wants that and it's very rare that it is needed.

JUDGE SCHEINDLIN: Given our time constraints, we are going to the judge.

JUDGE FRANCIS: This is a series of issues where I think I am firmly ambivalent. I think that I disagree with Dave and believe that it probably would be helpful in avoiding conflict to require the requesting party to identify the form of production. Now, I do not think that that needs to be done in the Rule very specifically, but the

31. Manual for Complex Litigation (Fourth) (2004).
32. *See* Fed. R. Civ. P. 26(f).

Advisory Committee notes might point out that the greater the specificity, the more likely we will be to avoid future problems.

If there is such a requirement, should there be a default mode identified? I think that is important. If there is no default mode, then judges are left with the question: "Well, he didn't identify the form of production; that means I should impose a default mode, or it means the request should be stricken?" I don't think that provides enough guidance, so I think there needs to be a default mode.

Which brings us to the $64,000 question, which is what is that default mode? There I am truly at sea. If I were to write a rule for today, I think I would say paper production, because that is what everybody is capable of doing, everybody who receives it is capable of analyzing it. It is cumbersome, it is burdensome, but everybody can deal with it.

But as I hear our technological people tell us that paper is going to disappear, I think such a rule would be quickly archaic. So I am looking for something that would be a reasonable default position. But I think that there needs to be a default position.

And finally, in terms of whether there should be an identification of the responding party's right to object because it is inaccessible or hard to produce, I think that is in the Rules. I don't think there is a necessity for electronic or other kinds of information to specify the opportunity to object.

Rule 26(b)(2)[33] sets out terrific guidelines for weighing factors to determine whether a document production is too burdensome, too costly, and so forth, and those Rules encompass the question of accessibility. I don't think there is anything necessary to be done there.

JUDGE SCHEINDLIN: Given our time constraints, our last topic, responding to interrogatories, will receive the shortest treatment.

MR. COHEN: And we have some suggestion as to what we could do to sort of tweak the current Rule to deal with electronic information. That would include producing the electronic information and, I suppose, identifying it as well, and possibly giving computer software so that you could derive the answer to your interrogatory from that electronic information.

Questions that have been raised are:

- Whether we need to include this option of giving the "computer software," or whether we stay with, I guess, the more general solution of giving sufficient information to find what you are looking for.

- And a question as to whether parties are employing Rule 33(d) with regard to hard-copy and computerized files. In my experience,

33. Fed. R. Civ. P. 26(b)(2).

they are. This is one of those situations where you might invite somebody over to come run queries on your database.

- And how does the fact that in many cases data produced is prepared for the purpose of responding to an interrogatory—how does that mesh with the obligation imposed under Rule 34?

JUDGE SCHEINDLIN: I think the real question here is: if a producing party takes the option of producing in this way, you would think the requesting party wants to be able to use it; so if you are going to produce it, do you have to produce enough to make it usable, which may mean the software or other material that goes with it?

Do you want to say just one [final] thing?

MR. BUCHANAN: The premise of this provision is that it is as easy for the receiving party to access the data as it is for the defendant or for the producing party. If you don't have the software tools to access the data, you don't have the same ease to access the data that the defendant does. So I think any production of electronic data pursuant to an interrogatory request has to be accompanied by the tools to access the data.

Notes & Observations