

Electronic Discovery Pragmatics Under the New Rules

Herbert L. Roitblat, Ph.D.

OrcaTec LLC

The Federal Rules of Civil procedure have been revised this year to better deal with the unique characteristics of electronically stored information in the discovery process. These rules officially go into effect on December 1, but they have already begun to impact practice.

Among the most significant changes are the requirements under Rule 26 that the parties meet within a few weeks after the commencement of a litigation, at least 3 weeks before the scheduling conference or issuance of a scheduling order, to discuss issues related to the discovery of electronically stored information. The goal of this meeting is to explicitly force attorneys to engage the topic of electronically stored information and its relevance to the issues of the case in advance of a scheduling conference. The rule raises the expectation that the parties will come to an agreement on discovery issues without court intervention.

The respective counsel are expected to explicitly describe any sources of electronically stored information that their client will use in support of its claims or defenses. They need to reveal where and how their clients have stored potentially responsive electronic information. The two sides need to reach preliminary agreements on the form in which the data will be produced. They must address the steps to be taken to preserve relevant information and agree on methods for asserting privilege and for dealing with the inadvertent disclosure of privileged information.

As the Advisory Committee Notes to the amendments state (Rule 26 (f)), “It may be important for the parties to discuss those systems [containing the electronically stored information], and accordingly important for counsel to become familiar with those systems before the conference.” The parties will have to know a great deal about their systems and strategies very early in the development of the case.

This paper is intended to provide some background to the kind of technical issues raised by the amendments. Dealing with these issues will require a collaboration among attorneys and IT people—a collaboration that will be needed very early on the development of a litigation.

The consequences of the new amendments are potentially enormous. The parties will be expected to commit to discovery plans and a schedule. It would be advantageous or even necessary, therefore, to have as complete a picture as possible of the client’s computer systems and their capabilities. How many discoverable repositories exist? What kinds of files are held in each repository? How accessible are they? What are the cost implications of the answers to these questions. Counsel will need a detailed map of their client’s entire information storage landscape.

On the one hand, the discovery planning conference requires counsel to gather a great deal of information in a very short time. On the other hand, it provides arguably the best opportunity available to control the direction, cost, and duration of discovery.

Among the key concerns are an identification of information sources and their accessibility, preservation issues, preservation of privilege, and production format. These amended rules place a new responsibility on counsel to be familiar at an early date with the information technology resources of their clients, with where electronically stored information may be kept, its accessibility, and other matters.

Thorough knowledge of these operations can also provide a strategic advantage. Under the new rules, there is less of an obligation to retrieve electronic information from inaccessible locations than, perhaps, had been true previously. The rules do require, however, an inventory of just what those inaccessible files are. It is your responsibility to identify sources that are not reasonably accessible. You may also be required to describe just how burdensome it would be to access these sources. Meeting these responsibilities may entail significant effort.

The terms negotiated at this discovery conference could have profound implications for the conduct and cost of the case. The attorneys will need to know to a reasonable depth the architecture of their client's electronically stored information and how accessible the information is. The attorneys will need to know the client's information retention policies as well as any other policies or practices that could affect how electronically stored information is organized, retained, and retrieved.

For example, an attorney may need to understand how a database works in general, and how his or her client's database works in particular, what its reporting capabilities are, how it gets updated and how often. This information could be critical for negotiating a reasonable schedule of production and range information to include in the production.

The same applies to understanding of the client's backup systems. What kind of backup schedule does the client employ (e.g., incremental daily backups with weekly full backups)? On what schedule are tapes recycled? What software is used to control the backup? How easy is it to retrieve just targeted files from the backup or does the whole tape have to be restored before data can be selected from it? Do indices of the backup media exist? How much control do they provide to limit the effort required to find responsive information in the backup volumes.

Software and storage media are constantly evolving. Older backup tapes may have been created using now obsolete software. Does the client have the software to retrieve information from these tapes? Similarly, the failure rate of backup tapes is often significant. What measures can be taken to deal with damaged backup tapes?

Where a client stores electronic information is another critical topic. Data are often stored on servers, on backup tapes, on desktop computers, and elsewhere. A list of common storage locations is presented later in this paper. The client's email policy may significantly affect the ease of conducting discovery, but in perhaps unexpected ways.

Many companies limit the volume of email and other files that a user can store on the central servers. This limitation has beneficial effects in that it reduces the expense of maintaining exorbitant amounts of storage and it limits the amount of information that

might have to be reviewed. On the other hand, the way that many users cope with this limitation is by offloading a substantial amount of information from the servers, where it is managed, to their own desktop computers, where it is not centrally managed. Rather than being able to rely on the central servers as authoritative sources for ediscovery, it may be necessary to visit each user and retrieve documents from his or own desktop computer—typically at substantial cost.

Asking the users to collect their own information is one way to try to cope with the problem of scattered storage, but it has its own problems. Expecting users to make relevance judgments about just what should be forwarded is problematic. Expecting them to be perfectly forthcoming and deliver potentially damaging information may be naïve. Finally, having them send (e.g., by email) or copy the stored files is likely to change the metadata of these files and may entail a risk of spoliation. A policy that is good for business—it limits the cost of operating the servers—may turn out to be a disaster for electronic discovery.

Another part of the discovery plan may be to identify the topics and time periods to be covered by the discovery. Consideration should then be given to how material appropriate to those topics will be identified. Will specific search terms or search methodologies be engaged? How will the data be selected or culled for detailed review? Judicious selection of topics and methods for identifying documents that are potentially relevant to those topics can substantially reduce the burden of electronic discovery.

Accessibility

The rules recognize a distinction between accessible and inaccessible data. Inaccessible data are subject to different discovery standard than are data that are more accessible. Ordinarily, “A party need not provide discovery of electronically stored information from sources that the party identifies as not reasonably accessible because of undue burden or cost” (Rule 26(b)(2)(B)).

Current email servers are usually considered accessible. Obsolete equipment and disaster recovery tapes may not be considered accessible. In between, whether something is considered accessible may depend on the particulars of the case, the amount in dispute, the effort required to retrieve the data, and the possibility of getting similar information from more accessible sources.

Encrypted data may be considered inaccessible, if the password is not available to decrypt it. Encryption scrambles a file’s content so that it cannot ordinarily be read without knowing the password that was used to encrypt it. If you don’t have the password, it may be difficult or impossible to view the file’s contents.

Sometimes, usually with older versions of programs, it is possible to bypass the need for the password and access the content without it. In other cases, the password has to be guessed, either from knowledge of the person who created them or by brute force methods (e.g., trying all possible letter combinations, trying all of the words in a dictionary of common passwords). Long passwords of more than about 7 characters may not be guessable by brute force and may be practically unbreakable.

The attorneys may also have to consider whether apparently inaccessible data is truly inaccessible. A client may choose to use backup tapes as a kind of archive, for example, and regularly access those tapes to retrieve “lost” files. It may be difficult to argue, then, that these are inaccessible resources, because they have obviously been accessed. Just how inaccessible a resource has to be before it is unreasonable to discover the information on it would depend on the circumstances of the case. But, in any case, it would seem to be essential to have some assessment of the situation in order to formulate appropriate arguments.

Preservation

The ability to comply with retention obligations also depends on the computer systems being used. The rules may recognize that some computer information may be lost in the ordinary course of business, but this may not be a blanket “get out of jail” card. The so-called safe harbor provisions (Rule 37) of the new rules apply only if data have been lost despite a good-faith effort to preserve them. If there are steps that could reasonably be taken to prevent the loss of critical information, and those steps are not taken, it may be difficult to convince the court that the data were lost despite a good faith effort to preserve them. What reasonable steps could the client take to preserve electronic information? Have these steps been taken?

According to the Advisory Committee Notes:

The good faith requirement of Rule 37(f) means that a party is not permitted to exploit the routine operation of an information system to thwart discovery obligations by allowing that operation to continue in order to destroy specific stored information that it is required to preserve. When a party is under a duty to preserve information because of pending or reasonably anticipated litigation, intervention in the routine operation of an information system is one aspect of what is often called a “litigation hold.” Among the factors that bear on a party’s good faith in the routine operation of an information system are the steps the party took to comply with a court order in the case or party agreement requiring preservation of specific electronically stored information.

Attorneys would seem to need to be familiar with the steps that were taken to comply with their retention obligations and be able to assess those steps in the context of the actions that could be taken. They may be called on to justify the reasonableness of their client’s actions.

Rules 34 and 45 allow the requesting party to specify the format in which the data should be produced. If the requesting party does not specify a format or if the producing party disagrees with the request, then the producing party must explicitly propose a production format. In the absence of an agreement, Rule 34 directs that information be provided in the form in which it is used in the ordinary course of business, in a reasonably usable form, or in a form that is mutually agreeable to the parties. Making appropriate choices about the form of production would seem to require knowledge of just how that information is stored in the ordinary course of business. Does it have to be translated into a reasonably usable form? If the receiving party argues for some particular form, it may be necessary to estimate the burdens or advantages that this alternative might present.

Rule 34(a) allows either party “to inspect, copy, test, or sample any designated documents or electronically stored information.” The committee notes make clear that the intention of this rule is not to provide unfettered access to the other party’s information systems and data, but they recognize that sometimes this may be appropriate.

Sampling may play a number of roles in discovery. It may test the effectiveness of the selection strategies used to identify documents to be reviewed. It may also short-circuit the use of broad requests, multiple rounds of discovery, or multiple depositions that would otherwise be needed to identify the truly responsive documents. Sampling may reduce the burden and expense of discovery and help to ensure that more complete responses are obtained to discovery requests.

Locating electronic documents

Electronic documents exist in a wide variety of forms and locations. They are not restricted just to a few file types on desktop or server computers. Many modern electronic gadgets can also hold discoverable electronic information and documents. Digital cameras and music players, for example, can be used as hard drives.

- Servers
- Backup tapes
- Desktops
- Laptops
- PDAs & Handhelds
- CDs and DVDs
- Jump and thumb drives
- iPods
- Home computers
- Digital cameras
- Cell phones
- Business application databases
- Building security systems
- Debit and credit card databases
- Voice mail systems
- Archives
- Closets
- Warehouses
- Obsolete computer equipment
- Instant message servers

Questions for discovery planning

The following questions are designed to elicit crucial information from the client to help insure that relevant files can be identified and analyzed. It may also be important to interview the custodians of these data, because people do not always follow the company’s policies faithfully. It is critical to identify what was actually done, rather than what should have been done. Data that were thought to have been destroyed, for example, may have been preserved by an employee who worried that management might “need it someday.” Talk to the “geeks” not just the “suits.”

These questions may also be of use to requesting parties. They may help to structure discovery requests as well as discovery plans. They may aid in interpreting what the other side is providing.

- Who has the knowledge, qualifications, and experience to support the discovery process?
- Have all appropriate steps been taken to preserve potentially responsive information?
- What records do you generate in the ordinary course of business?
- What records are you required to maintain in the ordinary course of business?
- Do you use a document management system?
- What records do you maintain in the ordinary course of business and for how long?
- What computer systems have you used during the relevant time period? Describe the brand and location of each computer, the size of its hard disk, the version of its operating system and any network software installed, and all installed software packages. Identify the custodian for each one.
- Are any other storage devices likely to contain relevant data?
- How accessible are the storage systems used to contain relevant data? Are any particularly difficult? What would be the level of effort needed to retrieve data from these difficult sources?
- How easy is it to retrieve just targeted files from the backup or does the whole tape have to be restored before data can be selected from it?
- Do indices of the backup media exist?
- Do you use any proprietary or atypical file formats in the ordinary course of business that could be responsive?
- What systems are in place to collect electronically stored information from desktop computers?
- What logs are maintained of network and server activity?
- What has been your network architecture during the relevant time period? Has it changed?
- What databases contain information that might be responsive? How can these data be accessed?
- List all current and former personnel who have had operational or maintenance responsibility for the computer systems. Don't just talk to supervisors. Employees sometimes follow their own rules rather than the official policy. Note when significant changes have been made to these systems.
- What has been done to determine whether other media are involved?
- What email systems (including versions) have been used in the relevant time frame? Are all versions available?

- Do you employ a hierarchical storage system? If you do, what policies govern retention in each storage medium?
- Describe the policies and procedures that governed data backups. What schedule and method was used (e.g., incremental nightly backups, weekly complete backups)?
- How long were tapes kept?
- Do you have an estimate of how often backup tapes turn out to be damaged?
- What software was used to make the backup tapes during the relevant time period? Are all versions of this software available?
- Are any of the servers or backup systems shared with other business units?
- Were desktop files backed up on a regular basis?
- Where are backups physically located?
- Describe policies and procedures for email retention and deletion.
- Were employees allowed or encouraged to maintain their own PSTs or similar archive files?
- Have retention policies been consistently enforced? How do you know that employees have complied?
- What outside electronic services have been used for email access (e.g., gmail, Yahoo mail, AOL, ISPs)?
- Are employees allowed to access instant messaging or web-based email accounts? What steps have been taken to preserve relevant information from these accounts?
- Do employees use personal digital assistants (PDAs, such as Blackberry, Palm) for business purposes?
- Are employees allowed to work on company documents at home? On their home computers?
- What relevant phone or voice mail records exist?
- Are files password protected? Are passwords available?
- How will you ensure the chain of custody for these documents?
- Are you prepared to provide someone who can testify to the adequacy of the procedures and the authenticity of the data?
- How would you prefer to produce these data to the other side?
- Will you be engaging a vendor to collect or process these data?

Common document and archive file types

The following list shows common file types that often contain useable text. Image files, such as PDF may contain pictures of text, but these are generally not indexable without

converting the pictures into text using OCR (optical character recognition). The first column shows the file name (e.g. mypaper.DOC, where the part after the final dot tells Windows computers what kind of file it is). This list may help you to identify the kinds of files that need to be processed. Bear in mind, though, that new file types are constantly emerging. For a more complete listing of file types and their corresponding extensions, see <http://filext.com/>.

Extension	Description
ASP	Microsoft Active Server Page
BIN	Macintosh Binary archive
CSS	Cascading style sheet for web pages
CSV	Comma Separated Values, sometimes used as a spreadsheet format or for import/export of data
DBF	dBASE
DOC	MS Word document format
DOT	MS Word template format
DWG, DXF	Autocad drawing
EML, PST, OST, MSG	Email messages
HQX	Macintosh BinHex archive
HTM, HTML, SHTML	Hypertext Markup Language, World Wide Web document format
LHA, LHZ	A compressed file archive
LWP	Lotus WordPro
MAQ, MAR	Microsoft Access queries and reports
MDB	MS Access database
MW	MacWrite
NSF	Lotus Notes
ODS, SXC	Open Office Spreadsheet
ODT, OTT, SXW	Open Office Documents
PDB	Aportis (Palm) document
PDF	Adobe Portable Document Format (e.g., Acrobat)
PPS, PPT	MS PowerPoint presentation

Extension	Description
PS, EPS	Postscript files
PUB	Microsoft Publisher
PXL	Pocket Excel
RTF	Rich text file (readable by MS Word and others)
SDW	StarWriter
SGML	Standard Generalized Markup Language
SHW	Corel Presentations presentation format
SIT, SEA	MacIntosh StuffIt archive
SNM	Netscape mail
SWF	Shockwave Flash
TAR, TAR.GZ, TAR.Z, Z	Unix archive
TXT, ASC, PRN	Text File format
WB3	Corel Quattro Pro spreadsheet format
WK1, WK2, WK3, WK4, WK5, WK1, WKS, WKU	Lotus 1-2-3
WKS	MS Works spreadsheet
WPD	Corel WordPerfect document format
WPS	MS Works document
WRI	MS Write
XLS	MS Excel spreadsheet format
XLT	MS Excel spreadsheet template
XML	Extensible Markup Language (many programs)
YMG	Yahoo Messenger
ZIP	PC Archive format