COMPUTER FORENSICS INC.™

04-CV-037
Request to Testify
1/12 San Francisco

December 7, 2004

Mr. Peter G. McCabe
Secretary of the Committee on Rules of Practice and Procedure
Administrative Office of the United States Courts
Washington, D.C. 20544

Re: Advisory Committee on Civil Rules Hearing
January 12, 2005

Dear Mr. McCabe:

I am writing to request an opportunity to testify before the Advisory Committee on Civil Rules at the January 12, 2005 Hearing in San Francisco. If you need additional information, please let me know at your earliest convenience.

Sincerely,

COMPUTER FORENSICS INC.™

Joan Feldman
President

1749 Dexter Avenue N.
Seattle, WA 98109
Phone: 206.324.6232
Fax: 206.322.7318
www.forensics.com

04-CV-037
Testimony

COMPUTER FORENSICS INC.™

December 22, 2004

Mr. Peter G. McCabe
Secretary of the Committee on Rules of Practice and Procedure
Administrative Office of the United States Courts
Washington, D.C. 20544

Re: Public Comment on Proposed Amendments to Federal Rules of Civil Procedure Relating to Electronic Discovery

Dear Mr. McCabe:

Please find enclosed my comments to the Committee regarding the proposed rule changes re electronic discovery.

I appreciate my inclusion in the San Francisco hearing, and look forward to participating on the 12th. In the meantime, please do not hesitate to write or call me if you have any questions.

Sincerely,

COMPUTER FORENSICS INC.™

Joan Feldman
President

Enclosure

04-CV-037
Testimony

MEMORANDUM

TO: COMMITTEE ON RULES OF PRACTICE AND PROCEDURE OF THE JUDICIAL CONFERENCE OF THE UNITED STATES

FROM: JOAN E. FELDMAN, PRESIDENT, COMPUTER FORENSICS INC.™

SUBJECT: COMMENTS ON PROPOSED AMENDMENTS TO THE ELECTRONIC DISCVOERY RULES

DATE: DECEMBER 22, 2004

The following comments are respectfully submitted in anticipation of testimony to be presented at the Civil Rules Committee hearing in San Francisco, California, on Wednesday, January 12, 2005.

1. Rule 26(b)(2) Discovery Scope and Limits - Limitations

Presumption against Discovery of Electronically Stored Information That Is Not "Reasonably Accessible"

Under the proposed amendments a party would not have to provide electronically stored information in response to a discovery request if it decides that the information is not "reasonably accessible."

It is important to note that a potential abuse of this rule may arise if a party "preserves" and/or deliberately changes originally active data to tape or another media deemed "not reasonably accessible". The party should not later be allowed to claim burden for reproducing or retrieving that data if it was originally available in active format at the time it was identified for discovery. This would include, among other things, data collected for other litigation matters.

Also, under the proposed rule, if the requesting party moves to compel the discovery of this information, the responding party would have to demonstrate that the information is not reasonably accessible. Reasonable accessibility often hinges upon the choice of system and operator(s) required to access the data.

I have seen a range of responses to this issue over the past decade. In many cases, producing parties solicit bids from internal and external service providers, and in many cases, they submit the highest dollar bids to the requesting party.

The requesting party should have some say in determining the method for data restoration – particularly for that of backup tapes and other media. Sampling techniques such as those proposed to test for relevance should also be applied to test the burdens and costs of production.

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

"Safe Harbor" from Sanctions

The proposed amendment would create a new subdivision (f) to protect a party from sanctions under the FRCP for failing to provide electronically stored information lost because of the "routine operation of the party's electronic information system." This "safe harbor" would not be available if the party violated a preservation order issued in the action, or if the party failed to take reasonable steps to preserve the information after it knew or should have known the information was discoverable in the action.

The danger inherent in the proposed language is that in many cases, what a party "should have known" about discoverable electronically stored information may be too loosely interpreted. For example, critical, responsive evidence stored in a database may be routinely purged by ongoing programmatic routines. Attorneys – many of whom are still struggling with conceptual issues of backup tapes and deleted files, often overlook this type of data and irreplaceable evidence is lost.

This rule change would be optimally effective if tied to a mandatory 26(f) conference requiring full disclosure of systems, data stores, as well as stipulations regarding scope.

3. Rule 26(f) Conference of Parties; Planning for Discovery

Early Attention to Issues Relating to Electronic Discovery

Rule 26(f) would be amended to require parties to discuss during their meet-and-confer session: (i) any issues relating to disclosure or discovery of electronically stored information, including the form in which it should be produced; (ii) whether they can agree to production on terms that protect against privilege waiver; and (iii) any issues relating to the preservation of discoverable information (including electronically stored information).

I heartily endorse the adoption of this amendment, and offer the following additions:

(iv) any issues relating to the nature and volume of material to be produced including data sources, data types, date and time frames, and stipulations as to what constitutes duplicate or "near duplicate" data.

(v) use of a mutually agreed upon glossary of terms to be used throughout the discovery process.

I applaud the work of the committee to date, and look forward to participating in the continuing dialog regarding this important topic.

MEMORANDUM

TO: COMMITTEE ON RULES OF PRACTICE AND PROCEDURE OF THE JUDICIAL CONFERENCE OF THE UNITED STATES

FROM: JOAN E. FELDMAN, PRESIDENT, COMPUTER FORENSICS INC.™

SUBJECT: ADDENDUM TO COMMENTS ON PROPOSED AMENDMENTS TO THE ELECTRONIC DISCOVERY RULES

DATE: JANUARY 27, 2005

The following comments and suggestions are respectfully submitted to supplement testimony presented at the Civil Rules Committee hearing in San Francisco, California, on Wednesday, January 12, 2005.

Recommendations Regarding Rule 26(f) Meet & Confer Conference

The Committee recognizes that early discussion between the parties of issues surrounding the collection and production of electronically stored information will greatly facilitate the speed and efficiency of discovery. Rule 26(f) sets forth a recommended protocol for that dialogue, including an examination of issues relating to: (i) the disclosure or discovery of electronically stored information, including the form in which it should be produced; (ii) whether the parties can agree to production on terms that protect against privilege waiver; (iii) the preservation of discoverable information (including electronically stored information); ***(iv) the nature and volume of material to be produced, including data sources, data types, date and time frames, and stipulations as to what constitutes duplicate or "near duplicate" data[1]; (v) the drafting of a mutually agreed upon glossary of terms to be used throughout the discovery process.***

(i) Disclosure and Form of Production

It is important during the meet and confer process to have qualified technical personnel or experts not only advise, but also actively participate in the discussion. Even for those counsels who possess advanced technological skills, the utilization of in-house technical personnel, information technology specialists, and/or outside experts in the meet and confer process will ensure that the information exchange is accurate, usable, and complete. Because technology changes rapidly, and because parties may have computing issues unique to their environments, active participation by technology experts can

[1] Computer Forensics Inc. has added these two vital components to the recommended meet and confer topics.

facilitate development of an accurate inventory of potentially responsive data, as well as allow experts to address issues surrounding preservation, form of production, and accessibility.

Collection, review, and production of inventoried data are central elements of the meet and confer discussion. As noted in the Manual for Complex Litigation, (4th) § 11.423, the Court should encourage counsel to "... produce data in formats and on media that reduce the transport and conversion costs, maximize the ability of all parties to organize and analyze the data during pretrial preparation, and ensure usability at trial." Many factors bear on how to interpret this guidance, however, including how the data are maintained in the normal course of business, what is a practical and usable form of production, and each party's idea of how they can best organize and analyze the data. Ideally, parties will agree upon the format(s) of production appropriate to the case at hand, rather than attempt to apply a common standard to all data. For example, while conversion of electronic data to a common format (most typically TIFF or PDF images and text) may be desirable from a review perspective, it may also be cost-prohibitive because of volume, or impractical for certain types of data, such as databases. Additionally, production of the native (i.e., original) file is sometimes necessary to ensure complete and accurate production. This is especially true when dealing with certain types of files, including spreadsheets, complex/compound files (such as Lotus Notes and Microsoft Exchange), and any other files containing embedded data.[2]

(ii) Review and Privilege

Although the method of review has traditionally been the sole purview of the producing party, increasing efforts at cost sharing have opened the door for the requesting party to become actively involved in the methods, services, and tools used to review electronic data prior to production. Options such as native file review, keyword searching, concept-based review tools, and on-line repositories may be considered. Here again, technological advances will continue to shape the process, so parties have an opportunity to work together to minimize the cost and burden of review by discussing them early on. Consideration should also be given to mechanisms that will allow speedy review of electronic documents without creating an undue burden to review all metadata, including the option of a stipulation of non-waiver.

(iii) Preservation

A comprehensive inventory is the base component for discussion of preservation issues, and is also a key component in full disclosure by the responding party. Failures to agree on a preservation protocol, or claims of spoliation, have in many instances been caused simply by an inability of one party to understand the scope and nature of the data collection. The inventory should contain enough detail regarding retention policies and procedures to enable the parties to reach common ground on preserving what is necessary

[2] Embedded data is information that is generally hidden but is an integral part of a computer file, such as "track changes" or "comments." Metadata, on the other hand, is information *about* a file, including its creation date, size, etc. While some metadata is routinely extracted during a conversion process to TIFF and text, embedded data is not. Therefore, it is only available in the original, native file.

while not unduly disrupting the responding party's ability to do business. This is particularly true when the responding party is a larger business entity. In this case, knowledge of enterprise computing systems[3] is also a pre-requisite to any discussion of burden.

(iv) Nature and Volume of Data

A comprehensive inventory of potentially responsive data lays a critical foundation for subsequent dialogue regarding timing of discovery, collection methods, and costs. The inventory should be focused on the issues, parties, and time frames pertinent to the litigation, and designed to capture information regarding all likely sources and types of relevant data subject to discovery. Time is of the essence in compiling this inventory, as computer data is volatile and can be transitory. Components of an electronic discovery inventory might include identification of both custodial and non-custodial data[4], email, shared servers, web servers, desktop and laptop computers, databases, voice messaging, instant messaging, backup tapes, and other media, software, and data sources upon which the responding party relies in its daily business, and as it relates to the litigation.

Identification of the nature and volume of data is critical in the formulation of a realistic discovery plan. For example, an early inventory should form the basis for cost and timing estimates for collection and review of the data. A vendor estimate that is based upon verifiable assumptions will be more reliable in both cost and timing than one based on speculation or ill-informed assumptions. The parties should consider, however, whether the vendor estimate—or, preferably, estimates—should be based on the universe of potentially responsive data or upon a subset of that data. As we have seen in recent cases, there may be occasions when a sampling of data is more desirable and appropriate. *See, e.g., Zubulake v. UBS Warburg LLC*, 217 F.R.D. 309 (S.D.N.Y. 2003). Though cost allocation issues have historically arisen post-production, the increasing volumes and complexity of electronic discovery demand earlier attention.

Duplicate Documents

The need to deal with duplicate documents in the collection and review of electronic data is far greater than it was when discovery was primarily paper-based. The volume of electronic documents presented for review can easily increase as much as ten- or twenty-fold if duplicates and near-duplicates are not eliminated early. Duplicate documents are typically defined as exact copies. The source of the document may or may not be taken into account. Studies have shown that roughly 70 percent of documents in a corporate environment are duplicate or near-duplicates. Technologies exist to eliminate both exact and near-duplicates, such as those that result from the common practice of adding a superfluous 'thank you' comment to an original email and passing it along. Parties are,

[3] An enterprise computing system is most often a mainframe- or mini-computer-based system designed to manage major business processes. For example: transaction processing for a national retailer, personnel systems for multi-office corporations, or banking systems.

[4] A custodian is typically a person. Non-custodial data is that data which is either shared by a number of persons or is "owned" or managed by the corporate entity. Examples of non-custodial data include enterprise databases, management software, and workgroup libraries.

therefore, encouraged to examine ways to cull out these additional documents well in advance of further processing and review.

Accessible vs. Inaccessible Argument

The parties should attempt to agree on what constitutes accessible data. Whether data are reasonably considered inaccessible depends in part upon the timing and reason for a transfer to backup media, as well as the intended use and/or frequency with which it is retrieved in the normal course of business. As data storage technologies evolve, media that might have been considered inaccessible according to the *Zubulake* standard may become accessible, so care should be taken to define accessibility according to use and purpose, rather than by type of media. Though there are certainly instances where discovery of the inaccessible, unallocated space on a computer hard drive is not appropriate, there are others where these data are clearly relevant.

(v) Glossary of Terms

The benefit of a common glossary of terms is self-evident. Parties may find themselves unable to reach agreement on discovery issues merely because they do not fully understand technical terminology. By first ensuring this common understanding, parties may move more quickly through the remaining elements of the meet and confer protocol. Once a basic vocabulary has been established, new terms may be added during the course of discovery as the facts and nature of the case warrant. ***A sample glossary is attached.***

Computer Forensics Inc.™

GLOSSARY OF TERMS

Active Data	User or system data that can readily be seen or accessed. Typical examples include word-processing, spreadsheet, PDF, and database files.
Backup Tape	Used to archive user and email data, most often from servers. Data must be restored from tape before it can be searched.
Blade Server	A thin circuit board, containing one or more processors and memory, which can be easily inserted into a space-saving rack. Representative of newer server architecture, which can accommodate vastly larger amounts of data.
Cache	Computer memory that is used to store data as it is transmitted from a web site, or as data is used by software programs. Web site contents may reside in cached storage on a hard drive.
Cookie	Computer code that is downloaded to a user computer by a web site to track usage and personal information. Can be an indication of user activity.
CRM	Customer Relationship Management: software programs that help manage clients and contacts. Used in larger companies and a significant repository of sales, customer, and sometimes marketing data.
Database	Collection of data organized for easy computer retrieval. Databases are comprised of tables (files), records (group of fields containing related information), fields (single types of data, such as author, date), and values (the data itself, such as John Smith, 11/12/04). Databases may be small enough to fit on a CD or so large they must be stored on a mainframe computer. Output from a relational database (one comprised of multiple, related tables) is typically in the form of a report or an "export."
Data Mining	Finding hidden information or relationships between data sets, most often in larger companies using specialized software. Also known as "executive dashboard," giving a snapshot of business activity and forecasts. Sometimes erroneously referred to as locating relevant data.
De-Duplication	Removing duplicate records. Definition of duplicate must be agreed-upon, i.e., whether an exact copy from a different location (such as a different mailbox, server, tape) is considered to be a duplicate.
De-Frag	A computer utility that reorganizes files on a hard disk. De-frag is used to optimize the operation of the computer and will overwrite information in unallocated space.
Diskwipe	A computer utility that overwrites existing data. There are various wiping programs with varying efficiencies. Though all will wipe named files, only some will wipe unallocated space of residual data. Unsophisticated users who try to wipe evidence may leave behind files of which they are unaware.
DAT	Digital Audio Tape: type of backup tape. Holds up to 40 GB, or approximately 60 CDs if all user data.
DLT	Digital Linear Tape: type of backup tape. Holds up to 80 GB or 120 CDs if all user data.
Domain	A sub-network of servers and computers within a LAN. Domain information is used when restoring backup tapes, particularly of email.
Encryption	Conversion of plain text into unreadable code. Used to secure computer files and documents.
ERP	Enterprise Resource Planning: software designed to integrate a variety of company functions, including human resources, inventories, and financials while simultaneously linking the company to customers and vendors. Typically found in larger companies, and often a rich source of data.
Evidentiary Image	A compressed bit-by-bit copy of a drive. Captures active and residual data by copying the entire surface of the drive. Compresses and encrypts copy to ensure authentication and protect chain-of-custody.
Extranet	An Intranet that allows restricted access by outside users.
FAT	File Allocation Table: index to files on a computer hard drive. If the FAT is

GLOSSARY OF TERMS

	corrupt, a drive may be unusable, yet the data may be retrieved using forensic methods.
Field	A single unit of data in a database, such as date or part number.
File Fragment	Portions of deleted data that remain in unallocated space on a hard drive. A fragment may be an entire page of text.
Format	Makes a drive ready for first use. Erroneously thought to "wipe" drive. Typically, only overwrites FAT, but not files on the drive.
Ghost	A commercial software product used to create a copy of a hard drive. Not forensically-sound, and may or may not capture unallocated drive space, depending on settings.
Hash	A mathematical algorithm that represents a unique value for a given set of data, similar to a digital fingerprint. Common hash algorithms include MD5 and SHA.
Hidden Data	Information embedded within a file, such as comments in word-processing file, or formulae in a spreadsheet. Is not extracted when native files are converted to TIFF or PDF.
HTML	HyperText Markup Language. The language used to author most web pages. HTML files will reside on a computer hard drive after Internet surfing and can contain both web site data and ISP mail.
IM	Instant Messaging: available to consumers through AOL, Yahoo, and Microsoft (among others). Used in corporations unofficially (through employees' private accounts), and officially (through corporate-sanctioned tools). Depending on the system, there may or may not be a record of the message.
Intranet	Private network that uses web browsers to index and access information accessible only to the users of the network. Data is maintained on servers.
IP Address	Internet Protocol Address: a unique identifier for the physical location of the server containing data. Example: 206.11.34.123 may identify an internal address or an external (Internet) address.
ISP	Internet Service Provider: examples are Yahoo, Earthlink, AOL. ISP mail may be a source of computer evidence either through files stored on local PCs or through files stored on ISP servers.
LAN	Local Area Network: comprised of servers and client computers.
Legacy Data	Data stored on older computer systems or in older file or database formats. Often remains behind as the legacy of outdated technologies, such as older mainframe systems that have been supplanted by PC technologies. In discovery, it can present a formidable challenge because it may be useful but difficult to extract.
LTO	Linear Tape-Open: newest tape format. Can hold as much as 400 GB of data, or 600 CDs if all user data.
Metadata	Information about a file, such as author, date created, size. May provide information regarding the source and history of a file, but may be changed and is not always conclusive. May be extracted when native files are converted to TIFF or PDF.
Migrated Data	Data that has been moved from one format to another, e.g., from a legacy system to a newer system.
NAS	Network Attached Storage. Servers that store data and are accessible through an IP address. Used to optimize systems by separating data storage from application (software) storage.
Native Format	The original form of a file, e.g., MS-Word, Lotus Notes, NSF file, Outlook .PST file, etc. How computer data is stored in "normal course of business."
.NSF	A Lotus Notes email store. Multiple .nsf files may exist and contain archived mail.
OCR	Optical character recognition. Used to read a bitmapped image (such as TIFF) into a text format that may be searched. Sometimes used unnecessarily when

GLOSSARY OF TERMS

	original (native) text already exists.
Operating System	Master program that runs the computer. Common O/S in use are Microsoft Windows and Unix.
PDA	Personal Digital Assistant: multifunction, handheld computer device that typically supports calendaring, contact management, and email (e.g., PalmPilot and Blackberry).
Post Office	Central repository of email user accounts/mailboxes. Large companies may have multiple post offices within their email system.
.PST	A Microsoft Exchange email store. Multiple .pst files may exist and contain archived mail.
RAID	Redundant Array of Independent Disks (or inexpensive drives): a method of storing data on servers that builds in redundancy and backup capability. Provides protection from drive failure, but more complicated to copy and restore.
Record	A group of fields that stores related information.
Residual Data	Hard drive data accessible only using special tools. Found in unallocated space on computer hard drives. May contain copies of previously-deleted files, Internet files, Internet email, and file fragments.
SAN	Storage Area Network: multiple servers connected to central pool of disk storage. Used to store large collections of data.
SDLT	Super DLT: type of backup tape. Holds up to 220 GB or 330 CDs if all user data.
Server	Central computer that is accessed by many users. Servers may be used to store email, shared files, application files (software), and web files.
Thin Client	User computer that acts as a terminal only and stores no applications or user files. May have little or no hard drive space.
Thumb Drive	Also known as jump drive. Portable, USB storage device that can hold up to 1 GB of data, thumb-sized. Can be imaged and may contain residual data.
VoIP	Voice over Internet Protocol. Provides telephonic capability across an IP connection. Increasingly used in place of standard phone systems.
VPN	Virtual private network. A secure network that uses public connections (typically the Internet) to connect remote computers. Often used to connect business offices.
WAN	Wide Area Network: comprised of 2 or more LANs. Found in larger companies.