CHAPTER ONE

Why Electronically Stored Information Is Different from Paper

The digital world differs materially from the paper world.

Everyone Is a File Keeper

In the paper world, documents typically are given to administrative staff for filing. In the digital world, every computer user who sends or receives e-mail, creates word-processing documents, prepares spreadsheets or information slides, or maintains databases decides whether to store files and has the ability to modify or delete a file. Even if the digital file keeper takes no action, it is likely that e-mail or other digital documents eventually will be moved to a backup tape that in many cases will be overwritten after some time period, causing the file to be lost forever.[1]

[1] An individual user can archive an e-mail in local storage media, and that may be the only place to find a document. *See* Hynix Semiconductor Inc., et al. v. Rambus Inc., 2006 U.S. Dist. LEXIS 30690, *27-28 (N.D. Calif. Jan. 5, 2006) (explaining that Rambus changed to a backup recycling schedule of three months and that employees should create their own

In the paper world, when an employee leaves a job, his documents, already archived, may remain in that state until a records-retention schedule calls for their destruction. In the digital world, when an employee leaves a job, her desktop or laptop hard drive (or both) may be reformatted, destroying all data on the drive(s) unless someone decides that there are litigation or business reasons to maintain the employee's digital status quo.[2]

In the paper world, when, say, a major construction project was completed, the paper associated with the project might be boxed and stored in a warehouse. In the digital world, the desktop and laptop computers used by everyone in the field may be moved to the next job. File management may be a function of project organization, or perhaps serendipity, depending on the individual file-keeping habits of each person on the job.

Metadata

A second key difference is the existence of metadata. The *Sedona Glossary*[3] defines metadata as "data typically stored electronically that describes characteristics of" electronically stored information (ESI) and explains that metadata "can describe how, when, and by whom ESI was collected, created, accessed, modified and how it is formatted." A pocket guide

archive copies of documents; for e-mail, that meant printing them or keeping them "on your hard drive").

[2] *See, e.g.*, Cache La Poudre Feeds, LLC v. Land O'Lakes, Inc., et al., 2007 U.S. Dist. LEXIS 15277 (D. Colo. Mar. 2, 2007) (wiping clean the computer hard drives of former employees, among other conduct, was sanctionable under the circumstances, but since the prejudice was not substantial, sanctions were limited to $5,000 and reimbursement of certain court-reporting costs).

[3] The Sedona Conference Glossary: E-Discovery & Digital Information Management 2nd Ed. (Dec. 2007), http://www.thesedonaconference.org/dlt.Form?did-TSCGlossary_12_07.pdf (*Sedona Glossary*). The *Sedona Glossary* continues by explaining that metadata can be "altered intentionally or inadvertently." It can be "extracted when native files are converted to image." Some metadata, such as file dates and sizes, "can easily be seen by users"; other metadata "can be hidden or embedded and unavailable to computer users who are not technically adept. Metadata is generally not reproduced in full form when a document is printed to paper or electronic image." The Sedona Conference working group series "is a series of think-tanks consisting of leading jurists, lawyers, experts and consultants brought together by a desire to address various 'tipping point' issues in each area under consideration." *See* http://www.thesedonaconference.org/. The Sedona Conference Working Group on Electronic Document Retention and Production has also published the second edition of Best Practices Recommendations & Principles for Addressing Electronic Document Production (June 2007). The document can be downloaded by going to the Sedona Conference website. *See* http://www.thesedonaconference.org/dltForm?did=TSC_PRINCP_2nd_ed_607.pdf.

provided to federal judges by the United States Judicial Conference gives this definition of metadata:

> Metadata, which most computer users never see, provide information about an electronic file, such as the date it was created, its author, when and by whom it was edited, what edits were made, and, in the case of e-mail, the history of its transmission.[4]

Yet another description appears in *Williams v. Sprint/United Mgmt. Co.*, 230 F.R.D. 640, 646 (D. Kan. 2005) (footnotes omitted):

> Some examples of metadata for electronic documents include: a file's name, a file's location (e.g., directory structure or pathname), file format or file type, file size, file dates (e.g., creation date, date of last data modification, date of last data access, and date of last metadata modification), and file permissions (e.g., who can read the data, who can write to it, who can run it). Some metadata, such as file dates and sizes, can easily be seen by users; other metadata can be hidden or embedded and unavailable to computer users who are not technically adept.[5]

Appendix I contains a good example of embarrassing metadata identified by a third party in a Microsoft Word document that the British government posted on June 30, 2003. The document was a dossier on Iraqi security and intelligence.[6]

Deleted Data That Do Not Die

A third key difference is that digital data can survive deletion, while discarded paper is not likely to be found again. The *Sedona Glossary* gives this definition of "deleted data":

> Deleted Data is data that existed on the computer as live data and which have been deleted by the computer system or end-user activity. Deleted data may remain on storage media in whole or in part

---

[4] Managing Discovery of Electronic Information: A Pocket Guide for Judges at 3 (2007), http://www.uscourts.gov/rules/eldscpkt.pdf.

[5] Examples of metadata that may be stored in Microsoft Word 2003 documents can be found at http://support.microsoft.com/kb/825576/ (e.g., your name, initials, company, or organization name; the name of your computer; the name of the network server or hard disk where you saved the document; other file properties and summary information; document revisions; document versions; template information; hidden text; and comments).

[6] Microsoft has since eliminated the software features that allowed the metadata in question to be discovered but the example serves as a good illustration of "metadata mining."

> until they are overwritten or "wiped." Even after the data itself has been wiped, directory entries, pointers or other information relating to the deleted data may remain on the computer. "Soft deletions" are data marked as deleted (and not generally available to the end-user after such marking), but not yet physically removed or overwritten. Soft-deleted data can be restored with complete integrity.

So, for example, a computer user moves data to "trash" or the "recycle bin." Until the trash or bin is emptied, the data remain fully restorable. Once the trash or bin is emptied, forensic experts may be able to reconstruct data fragments to recreate the deleted file, unless the storage media in question has been "wiped" (typically by software designed to achieve this aim).[7]

Multiple Sources of Data

A fourth key difference is the proliferation of data sources over paper. A "key player" in a particular dispute may have information stored in several places, and multiple copies of the same information may exist in multiple places. Consider these possibilities:

1. Office desktop storage media;
2. Office backup storage media;

---

[7] *See, e.g.*, Kucala Enterprises, Ltd. v. Auto Wax Co., Inc., 2003 U.S. Dist. LEXIS 8833 (N.D. Ill. 2003). In the course of this patent infringement case, Kucala installed and used software called "Evidence Eliminator" on a computer, just hours before it was to be examined by Auto Wax's computer specialist. The magistrate judge explained that "Evidence Eliminator" is a program designed to clean computer hard drives of data that may have been deleted by the user but still remain on the hard drive. Kucala also threw two other computers away during the litigation. He did so, he said, because they had crashed and were of no use to him. Kucala also admitted destroying documents, contrary to his attorney's advice, because he was afraid the defendant would not honor a protective order that was in place. Auto Wax's computer specialist inspected the computer on which Kucala had installed "Evidence Eliminator," and confirmed that the software had been used to delete and overwrite more than 14,000 files. Auto Wax filed a motion for sanctions alleging prejudice as a result of Kucala's destruction of one computer and deletion of relevant discovery from two others. Auto Wax sought a default judgment, attorneys' fees, expert fees, and costs. The magistrate judge found that Kucala had acted unreasonably, with gross negligence, and in flagrant disregard of the district court's order by deleting files just hours before Auto Wax's computer specialist was to inspect his computer. The magistrate judge recommended that the district court dismiss the action and require Kucala to pay the costs and attorneys' fees incurred by Auto Wax from the time Kucala deleted the files until the hearing.

3. Office laptop storage media;
4. Optical discs like CDs (compact disc) or DVDs (digital video disc or digital versatile disc);
5. Floppy disks;
6. Flash or "thumb" drives;
7. Home computer storage media, including external hard drives and portable drives;
8. Personal laptop storage media;
9. Office or home voice mail;
10. Cell phone voice mail;
11. Personal digital assistants (e.g., Blackberry);
12. Web-based storage;
13. Home or personal e-mail systems;
14. Devices that send or receive instant messages;
15. Printers, scanners, and copiers with computer memory;
16. Memory cards (e.g., from cameras); and
17. In appropriate cases, global positioning system devices.

Backup Tapes

Another key difference between the paper and electronic worlds is the existence of backup tapes,[8] which are typically used for disaster-recovery purposes. Backup tapes contain extraordinary amounts of information. To illustrate, consider this explanation from the United States Judicial Center's *Manual for Complex Litigation* of the volume of electronic information:

> The sheer volume of such data, when compared with conventional paper documentation, can be staggering. A floppy disk, with 1.44 megabytes, is the equivalent of 720 typewritten pages of plain text. A CD-ROM, with 650 megabytes, can hold up to 325,000 typewritten pages. One gigabyte is the equivalent of 500,000 typewritten pages. Large corporate computer networks create backup data measured

[8] The *Sedona Glossary* defines "backup tapes" as follows: "Magnetic tape used to store copies of ESI, for use when restoration or recovery of data is required. ESI on backup tape is generally recorded and stored sequentially, rather than randomly, meaning in order to locate and access a specific file or data set, all ESI on the tape preceding the target must first be read, a time-consuming and inefficient process. Backup tapes typically use data compression, which increases restoration time and expense, given the lack of uniform standards governing data compression."

> in terabytes, or 1,000,000 megabytes: each terabyte[9] represents the equivalent of 500 billion typewritten pages of plain text.

Manual for Complex Litigation (4th), § 11.446.[10]

One of the reasons backup tapes contain so much data is duplication. For example, after a thirty-day month, an entity that backs up daily, weekly, and monthly will have thirty daily tapes, four or five weekly tapes, and one monthly tape. The tapes will contain all the information stored at the time of the backup. Hence, a daily backup on a Tuesday will contain Monday and Tuesday's information. A weekly backup on Friday will contain whatever information that has been stored since the last weekly backup plus all the information contained on the prior weekly backup tape. The monthly tape will contain whatever has been stored since the prior month's backup tape and will duplicate much of the information on the daily and weekly backup tapes. Backup tapes are recycled after a period of time as well.

Typically, backup tapes are not reasonably accessible, as compared with "active data," which users can access easily.[11]

Perhaps as significant, backup tapes may be the only places certain documents reside. Unless they were printed, prior versions of a document may exist only on backup—they would be overwritten each time the user edits the file contained in active data storage. An individual that does not archive an e-mail on his or her individual hard drive will lose

---

[9] A "byte" is the basic measurement of "most computer data and consists of 8 bits." A "bit" is a "binary digit." A bit consists of either a 0 or 1. Generalizing, in computer code, for a word processing system, 0s and 1s (electronically switched off or on) are strung together to represent letters, numbers, and punctuation. There are 1,024 bytes in a "kilobyte," 1,048,576 bytes in a megabyte, 1,073,741,824 bytes in a gigabyte, and 1,099,511,627,776 bytes in a terabyte. *See* http://kb.iu.edu/data/ackw.html.

[10] Http://www.fjc.gov/public/home.nsf/autoframe?openform&url_l=/public/home.nsf/inavgeneral?openpage&url_r=/public/home.nsf/pages/470.

[11] One court has described the difference between data that are "accessible" and data that are "inaccessible." Data that are (1) "online" or archived on current computer systems (such as hard drives), (2) "near-line," such as that stored on optical discs or magnetic tape that is stored in a robotic storage library from which records can be retrieved in two minutes or less, or (3) "off-line" but in storage or archives, such as removable optical disc (e.g., CD-ROM or Digital Versatile Disc (DVD)) or magnetic tape media (e.g., Digital Linear Tape (DLT)), are readily accessible using standard search engines because the data are retained in machine readable format. Zubulake v. UBS Warburg LLC, 217 F.R.D. 309, 318-320 (S.D.N.Y. 2003). On the other hand, (1) routine disaster recovery backup tapes that save information in compressed, sequential, and nonindexed format, and (2) erased, fragmented, or damaged data, are generally inaccessible, because a time-consuming, expensive restoration process is required to obtain information. *Id.* at 319-320.

that e-mail to backup after a period of time. Hence, backup tapes may reveal whether an individual has deleted an e-mail.[12]

Retrieval of information from backup tapes can be costly. Costs can include separate charges for restoration, retrieval, and review. In *Zubulake v. UBS Warburg LLC*, 216 F.R.D. 280 (S.D.N.Y. 2003),[13] for example, there was a battle over the production of seventy-seven backup tapes. The district court ordered UBS Warburg to restore—at its expense—five tapes to give the district court an idea of the cost to restore and the relevance of the information contained on the backup tapes. The cost to restore five backup tapes was $19,003.43, and it resulted in the production of six hundred e-mails responsive to the plaintiff's request for production. UBS Warburg estimated that the cost to restore the remaining seventy-two tapes was $273,649.39, and that the cost to review the data before production would be $107,694.72.

One of the many reasons that review costs can be so high is that it is not easy to determine which backup data is protected by the attorney-client privilege. Privileged communications involving in-house counsel and outside-counsel's privileged communications with a client can be buried among millions of pages of documents on a backup tape. To identify privileged documents is both time-consuming and costly.

---

[12] Backup tapes will not capture an e-mail received by an individual and deleted the same day. To illustrate, in Zubulake v. UBS Warburg LLC, 217 F.R.D. 309 (S.D.N.Y. 2003), the district court explained UBS Warburg's electronic storage architecture: "UBS backed up its e-mails at three intervals: (1) daily, at the end of each day, (2) weekly, on Friday nights, and (3) monthly, on the last business day of the month. Nightly backup tapes were kept for twenty working days, weekly tapes for one year, and monthly tapes for three years. After the relevant time period elapsed, the tapes were recycled." *Id.* at 314. The district court explained why backup tapes might not contain certain e-mails: "Of course, periodic backups such as UBS's necessarily entails the loss of certain e-mails. Because backups were conducted only intermittently, some e-mails that were deleted from the server were never backed up. For example, if a user both received and deleted an e-mail on the same day, it would not reside on any backup tape. Similarly, an e-mail received and deleted within the span of one month would not exist on the monthly backup, although it might exist on a weekly or daily backup, if those tapes still exist." *Id.* at 314, n.25.

[13] Ms. Zubulake alleged she was a victim of gender discrimination and was eventually terminated. She then filed an additional claim that she was retaliated against for complaining about the employment practices of her supervisor. 216 F.R.D. at 281. The district court explained that under the Federal Rules of Civil Procedure, the presumption is that the producing party pays for production of accessible data. In addition, the district court held that the cost to review should always be borne by the producing party. With respect to the cost to retrieve, the district court evaluated each of seven factors identified by the district court as relevant to the determination of who should pay this cost, and decided to shift 25 percent of the cost to the requesting party, Ms. Zubulake. 216 F.R.D. at 283-90.

Key Players

In the paper world, there is not necessarily a premium placed on correctly identifying so-called key players—the people with knowledge or information about a claim—because many companies keep paper for a long time. In the electronic world, it is much more significant to identify the correct key players because any delay in doing so can result in the loss of relevant information.

For example, in *Consolidated Aluminum Co. v. Alcoa*, 2006 U.S. Dist. LEXIS 66642 (E.D. La. July 19, 2006), four key players initially were identified in November 2002 when Alcoa sent a demand letter to Conalco for costs associated with an environmental cleanup. Later, in 2003, Conalco decided to sue, seeking a declaration of nonliability. In 2005, Conalco issued a request for production that prompted Alcoa to identify eleven more key players. In the interim, however, the e-mails of these eleven individuals had been erased because of Alcoa's e-mail backup retention protocol.[14] Conalco moved for sanctions. The district court refused to grant punitive sanctions, but required Alcoa to pay the reasonable costs and fees Conalco incurred to bring the motion for sanctions, and also to pay the cost of redeposing up to thirteen people. The court also allowed Conalco to serve certain additional discovery requests.

Similarly, in *E*Trade Securities LLC v. Deutsche Bank AG*, 2005 U.S. Dist. LEXIS 3021 (D. Minn. Feb. 17, 2005), in 2001, defendant NSI put a litigation hold on backup tapes, but not on e-mail messages. The reason was that NSI had a practice of backing up all e-mail messages. However, NSI also recycled its backup tapes after three years, a policy that the company did not change even though the litigation continued more than three years beyond the date of the litigation hold. When it became clear in 2004 that additional e-mail boxes needed to be searched beyond those of the initial key players, the e-mails were no longer available. As a sanction, the court approved an adverse inference instruction.

---

[14] Alcoa submitted an affidavit describing the protocol: "Once every week, all messages older than thirty (30) days in a user's Exchange mailbox are moved to a 'System Cleanup' folder. At the same time, all messages older than fifteen (15) days (forty-five (45) days total) in a user's System Cleanup folder are deleted and are no longer directly recoverable by the user . . . . In addition, Alcoa's disaster recovery system retains email for the trailing six (6) months." 2006 U.S. Dist. LEXIS at *19, n.12. That prompted the magistrate judge to say: "Thus, it is possible that relevant emails for the six (6) months prior to November 2002 could have been retrieved, had Alcoa properly suspended its routine document destruction policy when it became aware of potential litigation with Consolidated in November 2002." *Id.* at *19.

Forms of Production

In the world of electronically stored information, parties have choices about the form of production. A requesting party may seek production in "native" format, which refers to the file as it exists on the other party's storage media, together with associated metadata. A producing party may prefer to produce documents in "tagged image file format" (TIFF)[15] or "portable document format" (PDF)[16] in order to, among other things, bates-label the documents. Vendors should be able to link meaningful metadata to an associated TIFF or PDF image depending on the parties' agreement or the scope of a court's order on production of electronically stored information.[17] Of course, there is always paper as well. Electronic files might be printed for production in paper format.

---

[15] The *Sedona Glossary* defines TIFF as: "A widely used and supported graphic file formats [sic] for storing bit-mapped images, with many different compression formats and resolutions. File name has .TIF extension. Can be black and white, gray-scaled, or color. Images are stored in tagged fields, and programs use the tags to accept or ignore fields, depending on the application."

[16] The *Sedona Glossary* defines PDF as: "An imaging file format technology developed by Adobe Systems. PDF captures formatting information from a variety of applications in such a way that they can be viewed and printed as they were intended in their original application by practically any computer, on multiple platforms, regardless of the specific application in which the original was created. PDF files may be text-searchable or image-only. Adobe® Reader, a free application distributed by Adobe Systems, is required to view a file in PDF format. Adobe® Acrobat, an application marketed by Adobe Systems, is required to edit, capture text, or otherwise manipulate a file in PDF format."

[17] According to the *Sedona Glossary* definition of "native format," "static" formats such as TIFF or PDF, "are designed to retain an image of the document as it would look viewed in the original creating application but do not allow metadata to be viewed or the document information to be manipulated." *Cf.* White v. The Graceland College Center et al., 2008 U.S. Dist. LEXIS 63088 (D. Kan. Aug. 7, 2008). In this matter, the magistrate ordered the producing party to produce emails and their attachments in native format after they had been produced in paper form. The requesting party had not specified a form of production thereby triggering Rule 34(b)(2)(E)(ii) which provides that, in such a case, the producing party has the option to produce in a form or forms in which the information is "ordinarily maintained or in a reasonably usable form or forms." Reduced to its essence, however, the magistrate's ruling was premised on the importance that metadata played in the requesting party's case. To get to this result, the magistrate said that the producing party should have produced the electronic information in native format since that was "a reasonably usable form" and—because of the need for the metadata—paper was not such a form on the facts of this case. *Cf.* Williams v. Sprint/United Management Co., 2006 WL 3691604, *14 (D. Kan. Dec. 12, 2006), where the same magistrate judge refused to require production in native format because no need for metadata was shown.

Parties wishing to perform electronic searches will want data in a searchable format. In addition, there may be reasons to obtain electronically stored information in different formats. For example, word-processing documents could be imaged with word-search capabilities while spreadsheets may be produced in native format to view the formulae used in the spreadsheet.

Vendor Contracts

Because of the different forms of production for different forms of data, one can readily see that in major productions of electronically stored information, litigation counsel will need help.[18] There is no shortage of firms willing to provide e-discovery assistance. Contracts with these firms must take into account a number of issues, including the following:

- What data will be collected from what locations or types of storage media?
- How will the collected data be transferred to the vendor and will it need to be encrypted or sent via other secured method?
- How will the data be extracted, and what steps will be taken to address—for example—file integrity, de-duplication of data, maintenance of attachments with e-mails, proper separation of documents, preservation of metadata, extraction of embedded information, processing of foreign language documents, identification or labeling of documents, and chain of custody?
- Are there any issues with collecting, transferring, or processing documents from a country subject to data protection, privacy, or blocking laws?
- Where and how will the data be stored and, if hosted by the vendor, at what price?
- What encryption, virus protection, or other security measures will be taken to protect the stored data?[19]
- How will the data be accessed by the client?
- Will any storage media have to be imaged?

[18] Corporate lawyers setting up large data banks for a deal may also need help.

[19] *See* New Jersey Ethics Opinion 701 (2006) (lawyer must take reasonable care to protect the confidentiality of electronic files and where the lawyer has entrusted the files to an outside provider, there must be "an enforceable obligation to preserve confidentiality and security," and "available technology to guard against reasonably foreseeable attempts to infiltrate the data" must be used.) http://www.judiciary.state.nj.us/notices/ethics/ACPE_Opinion701_ElectronicStorage_12022005.pdf.

- Will forensic examination of any storage media have to be undertaken?
- What search and filter tools will be utilized, are there any additional charges for their use, and who will make this decision?
- How and at what cost will information be reviewed and processed if the vendor is unable to do so using its normal processes?[20]
- In what form will the data be produced to a requesting party and who will do the quality control check on the production?
- How will production be tracked and will the tracking account for special productions or production in various formats?
- If the form of production changes, will contract costs be affected? If so, under what circumstances?
- What is the pricing structure? Fixed fee? Price per gigabyte? Variable rates for various tasks?
- Are there any special costs for processing computer images or backup tapes, or for any other activities?
- Under what circumstances will change orders on price be permissible, if at all?
- If there are problems with data extraction or data production, who bears responsibility for any additional costs associated with the resolution of those problems?
- Who bears the risk of software failure, file corruption, storage media deterioration, and hardware failure or replacement?
- If sanctions are awarded because of the failure of a vendor to exercise the appropriate degree of care, who pays the sanctions?
- Will you include a confidentiality term? Assuming so, what should it say? How will confidential data be treated in the event of litigation with the vendor?
- Will a limitation-of-liability clause be permitted? If so, at what limit?
- Should there be an indemnity provision and who should it favor?

[20] My colleague and friend, Arlen Tanner, Esq., is an extraordinarily talented member of Shook, Hardy & Bacon's E-Discovery Working Group with rich, technical experience. He reports that there may be files that a vendor cannot open because of the vendor's software. When there are resistant-to-processing documents like this, they may be viewable using alternative software and can then be imaged, reviewed, and included in the production if responsive. Vendors may not be motivated to process these files because they require individual attention and the identification of viewing software. Personal Communication with Arlen Tanner, Esq. (April 21, 2008).

- Will insurance play a role in allocating risk under the contract?
- What will the contract say about termination?
- Is the client in a position to receive the data immediately in the event the contract is terminated?
- Are there "data-return" charges in the event of termination? What data will be returned, at what cost, if any, and in what format?[21]
- Should there be dispute resolution mechanisms built into the contract? If so, which ones?
- What law should govern?
- Should there be a venue clause in the event of judicial dispute resolution?
- Should there be a waiver of jury trial?
- Should attorneys' fees be recoverable by a prevailing party if a dispute requires judicial or arbitral resolution?

Not every e-discovery obligation will require a vendor contract, but many e-discovery obligations will require a vendor. As a result, one or more of these questions may still apply.

[21] The data might be returned in a manner that requires a new vendor to reprocess it. Sometimes review, redaction, de-duplication, and production data are lost. Or information is stored by the vendor in a proprietary format or costs a significant amount to recover or is not defined as the client's "data" for data-return purposes. Personal Communication with Arlen Tanner, Esq. (April 21, 2008).