THE INCREASING IMPORTANCE OF METADATA IN ELECTRONIC DISCOVERY

By: W. Lawrence Wescott II*

Cite as: W. Lawrence Wescott II, *The Increasing Importance of Metadata in Electronic Discovery*, 14 RICH. J.L. & TECH. 10, http://law.richmond.edu/jolt/v14i3/article10.pdf.

[1] Metadata, by its nature, is a secondary class of data. Although commonly described as "data about data,"[1] a more formal definition has been given as "evidence, typically stored electronically, that describes the characteristics, origins, usage and validity of other electronic evidence."[2] The emphasis in the short history of electronic discovery has been on this "other electronic evidence," such that arguments were made, when drafting the electronic discovery amendments to the federal rules, that metadata should be excluded from discovery.[3] The January 2004 edition

* Senior Consultant, Kahn Consulting, Inc. M.B.A., Information Systems, University of Georgia, 1992; J.D., with honors, University of Maryland, 1985. B.A., *summa cum laude,* Washington & Lee University, 1982.

[1] *Lorraine v. Markel Am. Ins. Co.*, 241 F.R.D. 534, 547 (D. Md. 2007).

[2] Craig Ball, Beyond Data About Data: The Litigator's Guide to Metadata, at 2 (2005), *available at* http://www.craigball.com/metadata.pdf (last visited Feb. 23, 2008).

[3] Ken Withers, *Two Tiers and a Safe Harbor: Federal Rulemakers Grapple with E-Discovery*, 51 FED. LAW. 29, 33 (Sept. 2004).

> Several commentators and even members of the advisory committee made strong arguments that meta-data, system data, and other elements of electronic files that are not consciously generated by the user nor apparent to the reader in the ordinary course of business should be excluded from discovery under a restrictive Rule 34 definition.

Id. Mr. Withers also described efforts to limit electronic discovery to forms which could be rendered as closely as possible to paper documents, which by their very nature would virtually eliminate metadata discovery. Kenneth J. Withers, *Electronically Stored Information: The December 2006 Amendments to the Federal Rules of Civil Procedure*,

of *The Sedona Principles: Best Practices Recommendations and Principles for Addressing Electronic Document Production* took the position that "[u]nless it is material to resolving the dispute, there is no obligation to preserve and produce metadata absent agreement of the parties or order of the court."[4] However, metadata is playing an increasingly important role in electronic discovery. Far from excluding this often critical information, the practitioner is well-advised to preserve metadata as a regular practice, particularly in connection with complex litigation.

I. TYPES OF METADATA

[2] The "metadata universe" is actually much broader than its definition might indicate. The United States District Court for the District of Maryland, in its *Suggested Protocol for Discovery of Electronically Stored Information*, has identified three basic types of metadata:[5]

A. SYSTEM METADATA

[3] The *Suggested Protocol* defines "system metadata" as "data that is automatically generated by a computer system."[6] Examples of system metadata include "the author, date and time of creation, and the date a document was modified."[7] System metadata is what is most commonly meant when the term "metadata" is used. One commentator has noted that system metadata could have probative value because it is created

4 NW. J. TECH. & INTELL. PROP. 171, ¶ 75 (2006), *available at* http://www.law.northwestern.edu/journals/njtip/v4/n2/3.

[4] Sedona Conference Working Group, *The Sedona Principles: Best Practices Recommendations and Principles for Addressing Electronic Document Production,* Principle 12, at i (January 2004), *available at* http://www.thesedonaconference.org/content/miscFiles/SedonaPrinciples200401.pdf [hereinafter *Sedona Principles*] (last visited Feb. 23, 2008).

[5] United States District Court for the District of Maryland, *Suggested Protocol for Discovery of Electronically Stored Information,* at 25*,* http://www.mdd.uscourts.gov/news/news/ESIProtocol.pdf [hereinafter *Suggested Protocol*] (last visited Feb. 23, 2008).

[6] *Suggested Protocol, supra* note 5, at 25-26.

[7] *Id.* at 26.

automatically.[8] This type of metadata "may be more valuable in building or defending a case as it is often not consciously created by a user and is less vulnerable to manipulation after the fact."[9]

B. SUBSTANTIVE METADATA

[4] Substantive metadata, according to the *Suggested Protocol*, is "data that reflects the substantive changes made to the document by the user."[10] Substantive metadata "may include the text of actual changes to a document."[11] Substantive metadata has also been referred to as "application metadata," which moves with the file when it is copied (as opposed to the free-standing nature of system metadata).[12]

[5] Substantive metadata is the notorious version of metadata, which is responsible for some of the horror stories involving electronic documents. In one case, the Pentagon had posted a report online detailing an incident in which a U.S. soldier accidentally killed an Italian secret service agent in Iraq. Readers were able to access redacted, blacked-out information in the .PDF file by copying and pasting the confidential information into a Word document.[13] In other instances, "Google, Dell, Merck, the United Nations Secretary General, the Democratic National Committee, and others have recently made embarrassing and sometimes damaging revelations through inadvertent disclosures of metadata."[14] An adverse party was able to access a previous version of a document and learned that a suit by the SCO Group against DaimlerChrysler was originally intended for the Bank of America.[15]

---

[8] David K. Isom, *Electronic Discovery Primer for Judges*, 2005 FED. CTS. L. REV. 1, at 11, *available at* http:// www.fclr.org/2005fedctslrev1.htm.
[9] Isom, *supra* note 8, at 11(citing MICHAEL R. ARKFELD, ELECTRONIC DISCOVERY AND EVIDENCE 1-5 (2003)).
[10] *Suggested Protocol, supra* note 5, at 26.
[11] *Id.*
[12] Ball, *supra* note 2, at 3.
[13] Gene Koprowski, *NSA and the Dangers of Documents*, ECONTENTMAG.COM (April 2006), http://www.econtentmag.com/?ArticleID=15304 (last visited Feb. 23, 2008).
[14] Philip J. Favro, *A New Frontier in Electronic Discovery: Preserving and Obtaining Metadata*, 13 B.U. J. SCI. & TECH. L. 1, 4-5 (2007).
[15] J. Brian Beckham, *Production, Preservation and Disclosure of Metadata*, 7 COLUM. SCI. & TECH. L. REV. 1, 2 (2006).

[6] Another variant of substantive metadata has been identified by Judge Lee Rosenthal. According to Judge Rosenthal, "metadata is also increasingly recognized as including the software that assembles information from different databases and brings it together for the reader."[16] Within this context, obsolete legacy applications could be considered application metadata, as such programs would, in essence, need to be coupled with the legacy data file in order for the data file to be read.

C. EMBEDDED METADATA

[7] Embedded metadata is defined as "the text, numbers, content, data, or other information that is directly or indirectly inputted into a Native File[17] by a user and which is not typically visible to the user viewing the output display of the Native File on screen or as a print out."[18] Examples include:

> spreadsheet formulas (which display as the result of the formula operation), hidden columns, externally or internally linked files (*e.g.,* sound files in Powerpoint presentations), references to external files and content (*e.g.,* hyperlinks to HTML files or URLs), references and fields (*e.g.,* the field codes for an auto-numbered document), and certain database information if the data is part of a database (*e.g.,* a date field in a database will display as a formatted date, but its actual value is typically a long integer).[19]

[16] Lee H. Rosenthal, *A Few Thoughts on Electronic Discovery After December 1, 2006*, 116 YALE L.J. POCKET PART 167, 186 (2006).
[17] A "native file" is defined by the *Suggested Protocol* as "ESI [electronically stored information] in the electronic format of the application in which such ESI is normally created, viewed and/or modified. Native Files are a subset of ESI." *Suggested Protocol, supra* note 5, at 3.
[18] *Id.* at 27.
[19] *Id.*

[8] This variant of metadata has the potential to be more important than the primary data, as it may be able to explain the visible data. Failure to produce this data has exposed one litigant to the danger of sanctions.[20]

II. METADATA IN THE FEDERAL RULES

[9] Metadata is not explicitly addressed in the Federal Rules of Civil Procedure. The word "metadata" does not appear at all in the Rules, and appears only once in the Advisory Committee Comments to the Rules.[21] However, despite efforts to the contrary,[22] metadata is clearly included within the definition of "electronically stored information" contained in Rule 34.[23]

---

[20] In *Williams v. Sprint/United Mgmt. Co.,* 230 F.R.D. 640 (D. Kan. 2005), the data requestor sought sanctions against the producer for failure to provide spreadsheets with embedded metadata (spreadsheet formulas). The producer had provided spreadsheets with "locked" cells, so that the requestor could not view the formulas. The court declined to impose sanctions at that time, as "the Court recognizes that the production of metadata is a new and largely undeveloped area of the law. This lack of clear law on production of metadata, combined with the arguable ambiguity in the Court's prior rulings, compels the Court to conclude that sanctions are not appropriate here." *Id.* at 656.

[21] The Advisory Committee comments discuss metadata in the context of the Rule 26(f) conference. The Committee was concerned about possible waivers of privilege resulting from inadvertent production of electronically stored information. When reviewing documents for privilege review, attorneys could overlook electronic data not visible to the user. In that regard, the Committee stated:

> For example, production may be sought of information automatically included in electronic files but not apparent to the creator or to readers. Computer programs may retain draft language, editorial comments, and other deleted matter (sometimes referred to as "embedded data" or "embedded edits") in an electronic file but not make them apparent to the reader. Information describing the history, tracking, or management of an electronic file (sometimes called "metadata") is usually not apparent to the reader viewing a hard copy or a screen image. Whether this information should be produced may be among the topics discussed in the Rule 26(f) conference.

FED. R. CIV. P. 26(f), advisory comm.'s notes.

[22] *See generally* Ball, *supra* note 2.

[23] Rule 34 defines "electronically stored information" to include "writings, drawings, graphs, charts, photographs, sound recordings, images, and other data or data compilations--stored in any medium from which information can be obtained either directly or, if necessary, after translation by the responding party into a reasonably usable form." FED. R. CIV. P. 34(a)(1)(A).

[10] The December 2006 Amendments to the Federal Rules added new provisions relating to the production of electronically stored information. Rule 34(b) allows the requestor to "specify the form or forms in which electronically stored information is to be produced."[24] Therefore, if the information contains metadata, the requesting party can specify that metadata be produced along with the primary data. The producing party in its "response may state an objection to the requested form for producing electronically stored information."[25] If it objects to the requested form, the producing party must also "state the form or forms it intends to use."[26]

[11] If no form is specified by the requesting party, the responding "party must produce the information in a form or forms in which it is ordinarily maintained or in a reasonably usable form or forms."[27] This does not mean that electronic data must be produced in an electronic format. In *The Scotts Co. v. Liberty Mutual Ins. Co.*,[28] the court rejected the plaintiff's assertion that "as a matter of law, a party's discovery obligations are not satisfied by the production of computerized information in a hard copy format."[29] Note the disjunctive *or* in the rule. A party is not obligated to produce the information in the form "in which it is ordinarily maintained" if the data is "reasonably usable."[30] The producing party does not have total latitude in this area, however. The Advisory Committee Comments indicate that "[i]f the responding party ordinarily maintains the information it is producing in a way that makes it searchable by electronic means, the information should not be produced in a form that removes or significantly degrades this feature."[31]

[24] *Id.* at 34(b)(1)(C).
[25] *Id.* at 34(b)(2)(D).
[26] *Id.*
[27] *Id.* at 34(b)(2)(E)(ii).
[28] No. 2:06-CV-899, 2007 U.S. Dist. LEXIS 43005 (S.D. Ohio June 12, 2007).
[29] *Id.* at *10-11,*13. Although the plaintiff argued that "some of the documents produced in hard copy form are not reasonably usable for the purpose for which they were requested since they cannot be searched for metadata," the court declined to rule on that argument because it was not clear that the parties had met and conferred on the issue. *Id.* at *14.
[30] FED. R. CIV. P. 34, advisory comm.'s notes.
[31] *Id.*

[12] The plaintiffs in *In re Payment Card Interchange Fee and Merchant Discount Antitrust Litigation*[32] took this precise approach in an attempt to avoid the production of electronic documents containing metadata.[33] They printed their electronic documents and scanned them into TIFF[34] files and then converted them into searchable files through the use of OCR (optical character recognition) software.[35] Therefore, the court found that "the Individual Plaintiffs have rather laboriously stripped their text-searchable electronic documents of metadata that would not appear in printed form, and then converted them back into text-searchable electronic documents without that subset of metadata."[36]

[13] The plaintiffs' gambit was ultimately not successful. Citing the Advisory Committee's comment that production of electronic data should not result in a degradation of the searchability of the data, the defendants objected that the ability to search the data electronically had, in fact, been degraded by the plaintiffs' conversion process.[37] However, because the plaintiffs had already produced a substantial amount of electronic documents using this method prior to the defendants' objection, the court concluded that it would be unduly burdensome to force the plaintiffs to produce the same material again in native format.[38] As the plaintiffs had

[32] No. MD 05-1720 (JG) (JO), 2007 U.S. Dist. LEXIS 2650 (E.D.N.Y. Jan. 12, 2007).
[33] *Id.* at *6-7.
[34] TIFF, or "tagged image file format" has been described as "a flexible and adaptable file format for storing images and documents used worldwide. TIFF files use LZW lossless compression without distorting or losing the quality due to the compression. In layman's terms, TIFF is very much like taking a mirror image of many documents in format that can be compressed for storage purposes." PSEG Power New York, Inc. v. Alberici Constructors, Inc., No. 1:05-CV-657 (DNH/RFT), 2007 U.S. Dist. LEXIS 66767, at *6 n.2 (N.D.N.Y. Sept. 7, 2007).
[35] The court defined OCR as "a computer software program that translates images of text into a format that can be searched or 'read' electronically." *In re Payment Card Interchange Fee and Merchant Disc. Antitrust Litig.*, 2007 U.S. Dist. LEXIS 2650, at *7 n.2.
[36] *Id.*
[37] *Id.* at *9.
[38] The court found that:

> [T]he Individual Plaintiffs provided a significant amount of discovery to the defendants, in several instalments [sic], in the form they prefer, and heard no objections for several months. While that history does not legally estop the defendants from insisting on a form of production more to their liking, it does suggest that as between the defendants and

conceded that the burdensomeness argument disappeared for all production thereafter, they would be required henceforth to produce future electronic documents in native format.[39]

III. METADATA IN THE HISTORICAL CONTEXT

[14] Perhaps the first case to appreciate the importance of metadata was *Armstrong v. Executive Office of the President*.[40] In *Armstrong*, the court decided that paper copies of electronic mail did not qualify as an "extra copy" for purposes of the Federal Records Act,[41] which would allow the originals to be destroyed, "because important information present in the e-mail system, such as who sent a document, who received it, and when that person received it, will not always appear on the computer screen and so will not be preserved on the paper print-out."[42] Although not explicitly referring to this type of information as "metadata," the *Armstrong* court clearly recognized that its value warranted preservation.

A. EVOLUTION OF THE OPPOSING TRENDS REGARDING THE IMPORTANCE OF METADATA

1. METADATA SHOULD NOT BE PRESERVED OR PRODUCED.

[15] As stated at the beginning of this article, the Sedona Conference originally took the position that "[u]nless it is material to resolving the dispute, there is no obligation to preserve and produce metadata absent agreement of the parties or order of the court."[43] In their commentary, the Conference authors recognized that metadata does have value in some

> the Individual Plaintiffs, it would be less fair to impose the costs of a second form of production on the latter.

Id. at *15-16.

[39] *Id.* at *16-17.

[40] 1 F.3d 1, 274 (D.C. App. 1993). *See* Favro, *supra* note 14, at 13 ("One of the earliest cases to recognize the significance of metadata in terms of document integrity is *Armstrong* . . . ."); *See also* Favro, *supra* note 14, at 5 n.23 (citing Momah v. Albert Einstein Med. Ctr., 164 F.R.D. 412, 418) (stating that the court "granted access to the computer list screen" so that the plaintiff could verify when certain documents were created).

[41] 44 U.S.C. § 3101 (2006).

[42] *Armstrong,* 1 F.3d at 1, 284.

[43] *Sedona Principles*, *supra* note 4, at i.

contexts: "[I]t is easy to conceive of situations where metadata is necessary to authenticate a document, or establish facts material to a dispute, such as when a file was accessed in a suit involving theft of trade secrets."[44] However, the authors continued that:

> In most cases, however, the metadata will have no material evidentiary value—it does not matter when a document was printed, or who typed the revisions, or what edits were made before the document was circulated. And there is also the real danger that information recorded by the computer may be inaccurate. For example, when a new employee uses a word processing program to create a memorandum by using a memorandum template created by a former employee, the metadata for the new memorandum may incorrectly identify the former employee as the author.[45]

[16] In the commentary to Principle 12, the Conference amplified its position. It acknowledged the benefits of metadata by stating:

> First, the preservation and production of metadata may provide better protection against inadvertent or deliberate modification of evidence by others. Second, preserving documents in their native electronic format usually preserves the associated metadata without incurring additional steps or costs. Third, the systematic removal or deletion of certain metadata may involve significant additional costs that are not justified by any tangible benefit. Fourth, the failure to preserve and produce metadata may deprive the producing party of the opportunity to later contest the authenticity of the document if the metadata would be material to that determination.[46]

On the other hand, the Conference noted that "[b]alanced against these factors is the reality that most of the metadata has no evidentiary value,

[44] *Id.* at 5.
[45] *Id.*
[46] *Id.* at 41.

and any time (and money) spent reviewing it is a waste of resources."[47] Viewing both sides, the Conference concluded that "[a]lthough there are exceptions to every rule, especially in an evolving area of the law, there should be a modest legal presumption in most cases that the producing party need not take special efforts to preserve or produce metadata."[48]

[17] In developing the electronic discovery amendments to the Federal Rules of Civil Procedure, the Civil Rules Advisory Committee of the Standing Committee of the Judicial Conference ("Advisory Committee") considered whether metadata should be produced under the alternative of producing data in the form in which it is originally maintained.[49] However, the Advisory Committee elected not to do so.[50] As a result, under either of the production alternatives set forth in Rule 34(b), "[n]either default form is intended to mandate production of metadata or embedded data."[51]

[18] Following this line of reasoning, the court in *Wyeth v. Impax Laboratories, Inc.*[52] denied the portion of the defendant's motion to compel production of documents requesting that electronic documents be produced in their native format, including metadata, rather than in the TIFF format in which they were produced.[53] The court followed the reasoning of the Sedona Conference comment regarding the lack of evidentiary value of most metadata, along with the comment regarding the emerging general presumption against the production of metadata.[54] In addition, "[t]he Default Standard for Discovery of Electronic Documents

[47] *Id.*
[48] *Id.*
[49] Civil Rules Advisory Comm: Minutes, 19 (Apr. 14-15, 2005), *available at* http://www.uscourts.gov/rules/Minutes/CRAC0405.pdf (last visited March 8, 2008).
[50] *Id. See also* Thomas Y. Allman, *The Impact of the Proposed Federal E-Discovery Rules*, 12 RICH. J.L. & TECH. 13, 15 (2006), *available at* http://law.richmond.edu/jolt/v12i4/article13.pdf (stating that "[t]he Advisory Committee discussed the competing concerns at some length but ultimately decided that the best course of action was to remain silent and leave the issue to individual case law development.").
[51] Allman, *supra* note 50, at 15.
[52] No. 06-222-JJF, 2006 U.S. Dist. LEXIS 79761 (D. Del. Oct. 26, 2006).
[53] *Id.* at *3.
[54] *Id.* at *4 (citing Williams v. Sprint/United Mgmt. Co., 230 F.R.D. 640, 651 (D. Kan. 2005)).

utilized in this District follows this general presumption. Paragraph 6 directs parties to produce electronic documents as image files (e.g. PDF or TIFF) if they cannot agree on a different format for production."[55] Nevertheless, "if the requesting party can demonstrate a particularized need for the native format of an electronic document, a court may order it produced."[56] The court observed that the parties had never agreed on a production format, nor did the defendant demonstrate a particularized need for the data.[57]

[19] The court in *Kentucky Speedway, L.L.C. v. National Association of Stock Car Racing, Inc*.[58] followed the reasoning of the *Wyeth* court in denying the plaintiff's request for production of metadata.[59] As in *Wyeth,* in *Kentucky Speedway* there was no agreement by the parties to produce metadata, and here, the request was made seven months after the defendant had produced data in both hard copy and electronic formats.[60] Similar to the decision in *Wyeth,* the court found that the plaintiff had not shown a particularized need for the metadata:

> Although plaintiff argues generally that it "needs document custodian information for the prosecution of its case" because "Kentucky Speedway has *no idea* of the origin of many of the documents" plaintiff does not identify any specific document or documents for which such information would be relevant and is not obtainable through other means.[61]

The court further observed that metadata may or may not provide the information the plaintiff requested: "Depending on the format, the

[55] *Wyeth,* 2006 U.S. Dist. LEXIS 79761, at *4-5.
[56] *Id.* at *5. The court, however, did acknowledge the obligation to preserve metadata in the event the requesting party could demonstrate a particularized need for it. *Id.*
[57] *Id.*
[58] No. 05-138-WOB, 2006 U.S. Dist. LEXIS 92028 (E.D. Ky. Dec. 18, 2006).
[59] *Id.* at *13.
[60] *Id.* at *23. This was the court's justification in *In re Payment Card Litigation* for denying defendants' request for metadata; that the request had been made after plaintiffs had already produced a significant number of documents. Going forward, however, plaintiffs were required to produce documents which included metadata. *See supra* note 38 and accompanying text.
[61] *Kentucky Speedway*, 2006 U.S. Dist. LEXIS 92028, at *23.

metadata may identify the typist but not the document's author, or even just a specific computer from which the document originated or was generated."[62] However, if the plaintiff could identify specific documents for which identifying information was relevant, the court would be more receptive to the request.[63]

[20] A request to produce metadata was also denied in *Michigan First Credit Union v. Cumis Insurance Society, Inc.*[64] The court cited *Wyeth, Kentucky Speedway,* and the "emerging standards" statement of the Sedona Conference in its opinion.[65] The plaintiff had expressed concerns that metadata contained within the documents could reveal "who composed or received the message that might not appear in the PDF or hard copy."[66] The electronic mail PDFs produced by the defendant contained all of the metadata of the original electronic version with the exception of a character string which identified the message.[67] The Microsoft Office documents sought by the plaintiff were stored in paper format by the defendant in the ordinary course of business.[68] Based on these representations contained in defendants' affidavit, the court concluded that the metadata contained little value and would be unduly burdensome for defendant to produce.[69]

[21] The common thread running through these decisions is that in all of the cases, the requestors failed to demonstrate a need for the metadata. Had they been able to demonstrate that the metadata contained relevant information, the courts would have granted their requests. Furthermore, the requests came after the producers had created their data. Had the parties conferred in the production's early stages and the requestors had filed an immediate motion to compel prior to the data's production, the results could have been different.

---

[62] *Id.* at *24.
[63] *Id.* at *24-25.
[64] No. 05-74423, 2007 U.S. Dist. LEXIS 84842 (E.D. Mich. Nov. 16, 2007).
[65] *Id.* at *5-6.
[66] *Id.* at *6.
[67] *Id.* at *6-7.
[68] *Id.* at *7.
[69] *Id.* at *8.

2. METADATA SHOULD BE PRESERVED AND PRODUCED.

[22] Almost immediately after the publication of the Sedona Conference guidelines in January, 2004 (the source of the "emerging presumption" against the preservation and production of metadata), courts began to hand down opinions in contravention of that trend. Language in the previous Rule 34 requiring that data be produced as "kept in the usual course of business" was a key factor in the court's requirement that electronic mail be produced with metadata in *In re Verisign, Inc. Securities Litigation.*[70] Defendants had sought to produce the data in TIFF format, and argued that producing the data in the original .pst format, along with Bates numbers and redactions, would be unduly burdensome.[71] The magistrate had found that TIFF production was not sufficient, and that the production "must include metadata as well as be searchable."[72] In upholding the magistrate's ruling, the court stated that it "understands that it may be difficult for Defendants to incorporate their redactions and [B]ates numbers into the .pst format, but it is not convinced that the responsive documents are so replete with privilege redactions that such as task would transcend all reasonableness."[73]

[23] Metadata also played an important role in the magistrate's recommendation of a default judgment against the document producer PricewaterhouseCoopers for discovery violations in *In re Telxon Corp. Securities Litigation.*[74] The opinion is replete with examples of discrepancies between PwC's hard copy production and the contents of its electronic databases. One difference was that

> A hard copy of a document might give one person as the last individual to modify a document and the date of that modification while the metadata attached to the document might give an entirely different person and date for a later

[70] No. C 02-02270 JW, 2004 U.S. Dist. LEXIS 22467, at *8-9 (N.D. Cal. Mar. 10, 2004).
[71] *Id.* at *13.
[72] *Id.* at *7.
[73] *Id.* at *14.
[74] Nos. 5:98CV2876, 1:01CV1078**,** 2004 U.S. Dist. LEXIS 27296 (N.D. Ohio July 16, 2004).

> modification because the later modifier did not record the later modification on the document itself.[75]

Although PwC explained possible rationales for this difference,[76] the explanations "do not explain, however, why a modification not recorded on the document itself would have a date *later* than the last date of modification on the document."[77] The court also noted that electronic versions of the documents would contain links embedded as "popups."[78] These "popups" consisted of green boxes highlighting text in the documents.[79] When the boxes were clicked with a mouse, "a larger box of text appears to overlay the primary document and to provide information useful to the auditor."[80] This substantive metadata did not appear in hard copies of the documents.[81] In connection with other factors, the court concluded that "missing documents, missing attachments, missing metadata, and hard copies of documents in a version different from the versions on any of the electronic databases so far produced suggest that PwC may be withholding or has improperly destroyed discoverable information."[82]

[24] Metadata was an important factor in establishing a prima facie case of copyright infringement by downloaders of music from file-sharing servers in *Elektra Entertainment Group v. Does 1-9.*[83] Metadata of the

---

[75] *Id.* at *50-51.

[76] As PwC points out, however, the appearance of an individual's name in the Metadata as having modified a document may be misleading. In some cases, that individual may have prepared a document which served as a template for the document in question. In other cases, the appearance of an individual's name in the metadata as having 'modified' a document may indicate that the individual worked on the document in a previous year and the document was 'rolled forward' into the next audit year, carrying the individual's name in the metadata into the new audit. The fact remains that plaintiffs and Telxon cannot know why the name appears.

Id. at *57, n.18 (citations omitted).

[77] *Id.* at *51, n.15.

[78] *Id.* at *52.

[79] *Id.*

[80] *Id.*

[81] *Id.*

[82] *Id.* at *115.

[83] No. 04 Civ. 2289 (RWS), 2004 U.S. Dist. LEXIS 23560 (S.D.N.Y. Sept. 7, 2004).

music files at issue "often reveal who originally copied a particular sound recording from a CD to a computer disk (a process called 'ripping') and provide a type of digital fingerprint, called a 'hash,'[84] that can show whether two users obtained a file from the same source."[85] Thus,

> Using the metadata associated with the music file that Doe No. 7 was offering for distribution on Kazaa, plaintiffs have determined that many sound recordings were ripped by different people using different brands of ripping software. Such information creates a strong inference that Doe No. 7 was not simply copying his or her own lawfully purchased CDs onto a computer, but had downloaded those files from other P2P[86] users. Because "the use of P2P systems to download and distribute copyrighted music has been held to constitute copyright infringement," plaintiffs have adequately pled copyright infringement to establish a *prima facie* claim.[87]

[25] Metadata in *Elektra Entertainment Group* made the difference between continuation of the case and dismissal. Metadata was also critical in establishing the plaintiffs' case in *Experian Information Solutions, Inc. v. I-Centrix, L.L.C.*[88] In *Experian*, plaintiffs sought to make an imaged copy of the hard drive of one of the defendants.[89] The case involved misappropriation of trade secrets, and plaintiffs' theory was that the defendant's computer had been used to write infringing computer code.[90] Plaintiffs "wish to discover information about the use of Fortran files on [defendant's] personal computer, including the number of Fortran files that exist, or once existed, on [defendant's] computer and the frequency

[84] *See generally* Ralph C. Losey, *Hash: The New Bates Stamp*, 12 J. TECH. L. & POL'Y 1 (June 2007) (discussing the "hash" concept).
[85] *Elektra Entm't Group,,* 2004 U.S. Dist. LEXIS 23560, at *11.
[86] *Id.* at *2 ("A P2P network is an online media distribution system that allows users to have their computers function as an interactive Internet site, disseminating files for other users to copy.").
[87] *Id.* at *11-12 (citations omitted).
[88] No. O4 C 4437, 2005 U.S. Dist. LEXIS 42868 (N.D. Ill. July 21, 2005).
[89] *Id.* at *1-2.
[90] *Id.*

with which those files were edited, printed and copied."[91] Metadata could establish the extent to which infringing activity took place.

[26] One of the most significant cases involving the production of metadata is *Williams v. Sprint/United Management Co.*[92] Favro cites *Williams* as a "watershed" case "because it represents the first instance in a published case where a party was specifically compelled to produce metadata."[93] At issue in *Williams* was a series of Excel spreadsheets allegedly used by the defendant to determine which of its employees would be laid off during a reduction-in-force.[94] Plaintiffs alleged that age was invalidly used as one of the criteria.[95] Originally, the parties had agreed that the spreadsheets would be produced in TIFF format.[96] Subsequently, plaintiffs requested the actual spreadsheets so they could perform statistical analysis without being required to re-key all of the spreadsheet data.[97] The court asked why the spreadsheets could not be produced in their original form, and the defendant replied that at that point, it was still reviewing for privilege.[98] The court then took the position that the only issue affecting production was privilege:

> What I'm talking about is if you're talking about documents maintained on Excel, you've got that in some form, whether it's on disk or paper, whatever it's on. It's an electronic form of Excel containing the data. The only thing you would have to do is review it for privilege and then give it to them.[99]

At this point, the court implicitly assumed that the entire file would be produced or it would be withheld from production as privileged. In a subsequent discovery conference, the court made its position clearer:

---

[91] *Id.* at *4 n.2.
[92] 230 F.R.D. 640 (D. Kan. 2005).
[93] Favro, *supra* note 14, at 15.
[94] *Williams*, 230 F.R.D. 640, at 641-42.
[95] *Id.* at 641.
[96] *Id.* at 643.
[97] *Williams,* 230 F.R.D. at 642-43.
[98] *Id.* at 643.
[99] *Id.*

> THE COURT: Okay. Before we get much further here, I thought it was clear from the last time we discussed this electronic issue, that you [Defendant] were looking for them and you were going to produce them. It's not an issue that you're not going to do it. It's a question of when.[100]

This frame of reference is important in understanding the court's reaction when the defendant actually produced the spreadsheets. The plaintiffs discovered that the metadata had been removed from the spreadsheets, and certain cells had been locked so that the plaintiffs could not access them.[101] Although defendants argued that the scrubbed metadata "is irrelevant and contains privileged information,"[102] the court ordered the

> Defendant to show cause why it should not be sanctioned for not complying with "what at least I understood my Order to be, which was that electronic data be produced in the manner in which it was maintained, and to me that did not allow for the scrubbing of metadata because when I talk about electronic data, that includes the metadata." The Court then gave Defendant seven days to show cause why it had scrubbed metadata and locked data, "because my intent from the two previous Orders was to do as I said, produce it in the format it's maintained, not modify it and produce it." The Court advised Defendant that if it could show justification for scrubbing the metadata and locking the cells, the Court would certainly consider it, but cautioned that "it's going to take some clear showing or otherwise there are going to be appropriate sanctions, which at least will be the production of the information in the format it was maintained."[103]

[27] In its response to the court's show cause order, the defendant explained that the metadata had been deleted "to preclude the possibility that Plaintiffs could 'undelete' or recover privileged and protected

[100] *Id.*
[101] *Id.* at 644.
[102] *Id.*
[103] *Id.*

information properly deleted from the spreadsheets and to limit the information in the spreadsheets to those pools from which it made the RIF decisions currently being litigated."[104]

[28] The *Williams* opinion is notable for its thorough discussion of metadata. It cited extensively from the Sedona Conference's writings on the topic, and explicitly considered whether the defendant's contention "that emerging standards of electronic discovery articulate a presumption against the production of metadata" was accurate.[105] Looking at the language of then current Rule 34 as well as the then proposed 2006 amendment and the advisory committee language, the court concluded that the language of the amended rule provided "no further guidance as to whether a party's production of electronically stored information 'in the form or forms in which it is ordinarily maintained' would encompass the electronic document's metadata."[106] Although noting the orders in *Verisign* and *Telxon*, in which metadata had been ordered produced, the court found that they did not address the question of "whether metadata should ordinarily be produced as a matter of course in an electronic document production."[107]

[29] The court then turned to the *Sedona Principles for Electronic Document Production*. It found two principles to be "particularly helpful in determining whether Defendant was justified in scrubbing the metadata from the electronic spreadsheets. Principle 9 states that 'absent a showing of special need and relevance a responding party should not be required to preserve, review, or produce deleted, shadowed, fragmented, or residual data or documents.'"[108] The comment to Principle 9 suggested that a Rule 34 "document" should be defined in terms of what is visible to the user when the document is viewed; thus, data not visible, such as metadata, should not presumptively be considered part of the document.[109] On the other hand, there could be circumstances where metadata should be preserved and produced.[110] The other helpful principle was Principle

[104] *Id.* at 645.
[105] *Id.* at 648.
[106] *Id.* at 649.
[107] *Id.*
[108] *Id.* at 650.
[109] *Id.* at 650-51.
[110] *Id.*

12.[111] Applying the Sedona Principles to the instant case, the court concluded that:

> [E]merging standards of electronic discovery appear to articulate a general presumption against the production of metadata**,** but provide a clear caveat when the producing party is aware or should be reasonably aware that particular metadata is relevant to the dispute. Based on these emerging standards, the Court holds that when a party is ordered to produce electronic documents as they are maintained in the ordinary course of business,…the producing party should produce the electronic documents with their metadata intact, unless that party timely objects to production of metadata**,** the parties agree that the metadata should not be produced, or the producing party requests a protective order.[112]

[30] The defendant argued that metadata was not requested by the plaintiffs and was never mentioned during discovery conferences.[113] The court responded that "Defendant should reasonably have been aware that the spreadsheets' metadata was encompassed within the Court's directive that it produces the electronic Excel spreadsheets as they are maintained in the regular course of business."[114] Furthermore, the court noted that,

> [T]aken in the context of Plaintiffs' stated reasons for requesting the Excel spreadsheets in their native electronic format and the Court's repeated statements that the spreadsheets should be produced in the electronic form in which they are maintained, the Court finds that Defendant should have reasonably understood that the Court expected and intended for Defendant to produce the spreadsheets' metadata along with the Excel spreadsheets.[115]

---

[111] *See supra* note 43 and accompanying text.
[112] *Williams*, 230 F.R.D. at 652 (footnote omitted).
[113] *Id.* at 654.
[114] *Id.*
[115] *Id.*

[31] The *Williams* court, considering the same Sedona Conference Principles as the *Wyeth* court[116] (and its progeny, *Kentucky Speedway* and *Michigan First Credit Union*), came to the exactly opposite conclusion. The requestors' failure to ask for metadata, an important factor in the latter courts' decisions, had no impact on the court's decision in *Williams*, which found instead that the defendant producer should have known that metadata was included within its directives.

[32] Despite the Sedona Principles' "emerging standard," metadata would continue to play an important role in electronic discovery issues. The format of electronic document production was at issue in *Hagenbuch v. 3B6 Sistemi Elettronici Industrali S.R.L.*,[117] a patent infringement case. Among the plaintiff's objection to the defendant's production of TIFF documents was that the documents "lack metadata that track when a document was created or modified and whether e-mails contained attachments and to whom they were sent."[118] Responding to the defendant's arguments that although the documents were not identical to the originals, they were reasonably usable, the court observed that "unlike the original electronic media, the TIFF documents do not contain information such as the creation and modification dates of a document, e-mail attachments and recipients, and metadata."[119] Furthermore, the plaintiff claimed that:

> [T]he information contained in the designated electronic media is relevant to his infringement claims and will allow him to piece together the chronology of events and figure out, among other things, who received what information and when. Because the information sought by Plaintiff may be relevant at the discovery stage, and because 3B6 USA does not suggest that the electronic media contain privileged or classified information, Plaintiff is entitled to that information.[120]

[116] Ironically, *Wyeth* cited *Williams* as its source for the Sedona Conference statement regarding the "emerging general presumption" against the production of metadata. *See supra* notes 54-55.
[117] No. 04 C 3 109, 2006 U.S. Dist. LEXIS 10838 (N.D. Ill. Mar. 8, 2006).
[118] *Id.* at *4.
[119] *Id.* at *8-9.
[120] *Id.* at *9.

The metadata would allow the plaintiff to establish critical timelines by demonstrating when the defendant possessed documents containing information relevant to the plaintiff's infringement claims.

[33] Metadata evidence contributed to the award of a default judgment in favor of a counter-defendant in *Krumwiede v. Brighton Associates, L.L.C.*[121] On the date the plaintiff was directed to return his laptop to the defendant or shortly thereafter, subsequent metadata analysis demonstrated that thousands of files were accessed, moved, or deleted.[122] In addition, metadata indicated that files had been transferred from the laptop to another destination.[123] The court also observed that changes made to the file metadata prejudiced the counter-defendant's ability to prove its case, since those changes made the authenticity of the underlying files suspect.[124] Similarly, in *Plasse v. Tyco Electronics Corp.*,[125] metadata demonstrated that résumé files material to the litigation had been modified after the defendant had filed a motion to compel production of the plaintiff's computer.[126] The plaintiff also changed the system date and opened files after he had done so, two days before he was to turn the computer over.[127] The court concluded that the plaintiff had,

> [D]irectly flouted this court's authority by destroying or modifying documents after the court specifically invited Defendant to obtain an inspection of Plaintiff's computer and disks. Plaintiff not only concedes that he "may have" deleted one such document, but appears to believe that his actions were insignificant. Under these circumstances, dismissal is the appropriate sanction.[128]

[121] No. 05 C 3003, 2006 U.S. Dist. LEXIS 31669 (N.D. Ill. May 8, 2006).
[122] *Id.* at *11-12.
[123] *Id.* at *13.
[124] *Id.* at *30.
[125] 448 F. Supp. 2d 302 (D. Mass. 2006).
[126] *Id.* at 306.
[127] *Id.* at 309.
[128] *Id.* at 311.

B. RECENT DEVELOPMENTS TOWARDS RECOGNITION OF THE IMPORTANCE OF METADATA

[34] While the impact of metadata is not necessarily as significant as in *Krumwiede* or *Plasse*, many courts consider it useful.[129] Significantly, the Sedona Conference, in the second edition of its *Best Practices Recommendations & Principles for Addressing Electronic Document Production*[130] has revised its guidelines regarding metadata. Principle 12 reads:

> Absent party agreement or court order specifying the form or forms of production, production should be made in the form or forms in which the information is ordinarily maintained or in a reasonably usable form, taking into account the need to produce reasonably accessible metadata that will enable the receiving party to have the same ability to access, search, and display the information as the producing party where appropriate or necessary in light of the nature of the information and the needs of the case.[131]

The Conference's position reflects an enhanced understanding of the potential value of metadata. For example, in its commentary on Principle 12, it stated that,

---

[129] *See, e.g.*, Vennett v. Am. Intercontinental Univ. Online, No. 05 C 4889, 2007 U.S. Dist. LEXIS 92891, at *7 (N.D. Ill. Dec. 13, 2007) (indicating that metadata established existence of prior versions of memo); ACMG of Louisiana, Inc. v. Towers Perrin, Inc., No. 1:04-CV-1338-RWS, 2007 U.S. Dist. LEXIS 91291, at *4-5 (N.D. Ga. Dec. 11, 2007) (stating that metadata showing dates of file transfer, deletion or modification are relevant to litigation); Klein-Becker usa L.L.C. v. Englert, No. 2:06CV00378 TS, 2007 U.S. Dist. LEXIS 45197 at *10 (D. Utah June 30, 2007) (stating that the plaintiffs' case would have been facilitated had defendant provided discoverable data in electronic format, along with metadata which would have assisted its searchability); PML N. Am. L.L.C. v. Hartford Underwriters Ins. Co., No. 05-CV-70404-DT, 2006 U.S. Dist. LEXIS 94456, at *13 (E.D. Mich. Dec. 20, 2006) (stating that metadata indicated that folder was accessed after complaint was filed, in contravention of CEO's denial of file's existence).

[130] Sedona Conference Working Group, *The Sedona Principles: Best Practices Recommendations and Principles for Addressing Electronic Document Production*, Principle 12, at ii (June 2007).

[131] *Id.* at ii.

> [I]t should be noted that the failure to preserve and produce metadata may deprive the producing party of the opportunity later to contest the authenticity of the document if the metadata is material to that determination. Organizations should evaluate the potential benefits of retaining native files and metadata (whether or not it is produced) to ensure that documents are authentic and to preclude the fraudulent creation of evidence.[132]

[35] In a forum held at Fordham Law School, Judge James C. Francis IV summarized some of the benefits of metadata which have emerged in electronic discovery:

> There are also less obvious ways that metadata may be both relevant and discoverable. What about the authenticity of documents? How do you demonstrate that an e-mail that you have now printed out is authentic? You may need to get the metadata to demonstrate where it came from, what its genesis was, and what its path was throughout a particular organization, in order to make your admissibility argument at trial. So there is an argument to be made that all of that metadata is critical to the authenticity issue.
>
> The metadata may be critical to either supporting or challenging a claim of privilege. For example, in order to determine whether any claim of privilege may have been waived, it is important to know to whom the document was distributed, even if it does not appear on the face of the document. Was it distributed to somebody's nanny for some reason, or to somebody outside any reasonable view of the attorney-client privilege?
>
> Finally, there is a question of whether metadata may be important for searchability purposes. A normal word search may or may not need metadata to provide additional words that can link you to the document. However, now there are conceptual search regimens which make use of

[132] *Id.* at 61.

> the metadata in order to determine how different documents may be linked, and therefore whether they may be conceptually related to a particular discovery inquiry. So if a party strips off the metadata, there may not be a direct relevance issue, but that may make it harder for the requesting party to search the information.[133]

IV. CONCLUSION

[36] The potential value of metadata has always been recognized by the commentators and the courts. Initially, due to preliminary concerns regarding data volumes in electronic discovery and the inherent secondary nature of metadata, the presumption against preservation and production was established. Ironically, however, the issue of data volumes actually militated in favor of access to metadata, as metadata enhanced the searchability of large amounts of data. Use of metadata in authenticating electronic documents and establishing privilege claims came to be recognized. A combination of all of these factors undoubtedly influenced the Sedona Conference in eliminating any reference to a presumption against the preservation and production of metadata. As the amount of electronic documents continues to increase overall, counsel will require as many tools as possible to help them distill and validate operative facts from the mass of data. Metadata has been, and will continue to be, an important device to aid in this effort.

[133] *Managing Electronic Discovery: Views from the Judges*, 76 FORDHAM L. REV. 1, 22 (2007).