

COMMENTARY

Reprinted From E-Discovery: A Thomson West Report

Metadata and E-Discovery

By Ralph C. Losey, Esq.

What is metadata? Literally it means “data about data.” Many courts define the term by referring to the Sedona Conference’s glossary of commonly used terms for e-discovery and digital information management, which defines “metadata” as:

information about a particular data set or document which describes how, when and by whom it was collected, created, accessed, modified and how it is formatted. Can be altered intentionally or inadvertently. 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.

All computer files have metadata associated or within them that provide information about the files. For instance, e-mail software includes information in e-mail files about its author, creation date, attachments and identities of all recipients, including those who received a copy or blind copy. Metadata even tells you if an e-mail has been opened by a recipient.

The printout of an e-mail, which is essentially a TIFF version of the e-mail, may not show the blind copies and certainly will not tell you if it has been read or not. The metadata of an e-mail will also maintain the history of the e-mail and its conversation thread, such as who replied and who forwarded. Also, unless it is a Microsoft Outlook e-mail stripped out of its PST file into a MSG file, it will tell you in what folder the e-mail was filed by its custodian.

Some programs include information within the contents of files that is hidden until you instruct the software to reveal the information. This is called “embedded” information, and courts frequently refer to such information as metadata. Technically, it is not true “metadata” because it is not “data about data.” It is not information about the file itself.

Instead, it is information within a file, but hidden for some reason. It is information that the user has created, but is not visible without a command.

A good example of this is the “comments” feature in Microsoft Word. Comments can be inserted into a Word document that are not visible until you use the view command to show them. The comments are embedded into the file itself.

Another example is the formula that a user can place in an Excel or other spreadsheet to calculate values within a cell. The math used to calculate the value of a spreadsheet cell is embedded in the file. Although technically “embedded data” is not “metadata,” for purposes of legal analysis “embedded data” is treated as a form or type of “metadata” because most courts, and the legal profession at large, do not grasp the distinction.

When Should Metadata Be Produced?

The foremost metadata issue facing all parties and courts today is when metadata should be produced. The answer is found in a string of cases from 2005 to 2007.

Williams I

The key case on metadata is *Williams v. Sprint/United Management Co.*, 230 F.R.D. 640 (D. Kan. 2005). Here, terminated employees brought a class action and sought Excel spreadsheets with all metadata intact, including embedded formulas. The court held that under “emerging standards of electronic discovery,” metadata ordinarily visible to users of Excel spreadsheets “should presumptively be treated as part of the ‘document’ and should thus be discoverable.” *Id.* at 652.

The court reviewed case law, the pending new Federal Rules of Civil Procedure and commentary, and the Sedona Conference’s principles for electronic document production, especially Principle 12, which says 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.” The commentary to this principle at that time opined that “most of the metadata has no evidentiary value, and any time (and money) spent reviewing it is a waste of resources.” The commentary also set forth an important caveat or exception to principle 12: “Of course, if the producing party knows or should reasonably know that particular metadata is relevant to the dispute, it should be produced.”

The court accepted principle 12, with commentary, as an important part of the “emerging standard,” but rejected Sprint’s argument that this meant the Excel spreadsheets’ metadata should not be produced. Instead, the court found that the Excel metadata was material to the dispute, and Sprint should have known that and should have produced it. The actual holding then goes even further:

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.

The initial burden with regard to the disclosure of the metadata would therefore be placed on the party to whom the request or order to produce is directed. The burden to object to the disclosure of metadata is appropriately placed on the party ordered to produce its electronic documents as they are ordinarily maintained because that party already has access to the metadata and is in the best position to determine whether producing it is objectionable.

Placing the burden on the producing party is further supported by the fact that metadata is an inherent part of an electronic document, and its removal ordinarily requires an affirmative act by the producing party that alters the electronic document.

Until recently, it looked like, in spite of principle 12, metadata production was indeed to be the new standard, especially since new Rule 34(b)(2)(B) went into effect, which requires production of electronically stored information as is kept “in the usual course of business or in a form or forms that are reasonably usable.” The usual course of business is to keep files in their native format, because that is how they are used (*i.e.*, .doc files created in Word, .xls files created in Excel), and native files by definition include all metadata.

Williams II

Several recent cases, along with a sequel to *Williams* itself, have, however, shown that the exact nature of the emerging standard is still in doubt. First, the sequel: *Williams v. Sprint/United Management Co.*, 2006 WL 3691604 (D. Kan. 2006). The spreadsheets in question in *Williams I* were produced in native format as the court ordered, but the plaintiffs wanted more. They returned to the court a year later to try to compel production in native format of all 11,000 e-mails produced that transmitted spreadsheets. The judge here used a golf analogy for “native” format: “Play it like it lies.” The plaintiffs argued that the original native files of the e-mails were needed in order for them to determine which e-mails transmitted which spreadsheets. The defendant had the burden to show why this native production should not be done, that it was permissible for it to have “improved their lie.”

The defendant met this burden, and the motion to compel was denied, primarily because the e-mails had already been produced on paper without objection, and a second reproduction at that point would have been very burdensome, especially since the e-mails contained many attorney-client-privileged materials. Actually, the plaintiffs had originally objected to paper production of the e-mails, but had withdrawn their objection during one of many discovery hearings based on the defendant’s assurances that it would provide spreadsheet reports that “would match up the transmittal e-mails with their respective attachments.”

The defendant argued it had done so as agreed, but the plaintiffs complained that the spreadsheet reports were deficient, and they were unable to match them up. The court disagreed that the spreadsheet reports were deficient, noted the apparent impossibility to redact privileged materials from native files, and held that since the plaintiffs had already received production in one format (paper), the new rules protected the defendant from having to produce them again in another format (native).

To continue the golf analogy, the court in effect applied new Rule 34(b)(iii) to prevent the plaintiffs from receiving a “mulligan,” a second request. The judge’s exact wording is instructive and, to a certain extent, explains and clarifies *Williams I*:

Federal Rule of Civil Procedure 34(b)(iii), as amended on Dec. 1, 2006, provides that “[u]nless the parties otherwise agree, or the court otherwise orders, ... a party need not produce the same electronically stored information in more than one form.” In this case, defendant has already produced the transmittal e-mails, as well as all the attachments to those e-mails.

Defendant has further created spreadsheet reports to correlate the transmittal e-mails to the attachments they transmitted. The court therefore finds that under Rule 34(b)(iii), defendant need not re-produce its ... transmittal e-mails together with their attachments in native format, as requested by plaintiffs.

Defendant raises legitimate concerns about producing the transmittal e-mails with their attachments in their native format, including whether production in native format would permit the redaction or removal of privileged information in the transmittal e-mail or the attachment.

Moreover, even assuming that defendant could produce the transmittal e-mails together with their attachments in native format with the privileged information redacted, plaintiffs have not sufficiently explained why they need the transmittal e-mails in their native format.

Previously, this court has ordered defendant to produce the ... spreadsheets in native format, but in that instance plaintiffs provided valid reasons for the spreadsheets to be produced in their native format. Namely, that the contents of the spreadsheet cells could not otherwise be viewed as the cells contained formulas.

Also, in many instances, the column width of the cells prevented viewing of the entire content of the cells. Here, other than arguing that ordering defendant to reproduce the transmittal e-mails together with their attachments in native format would be more helpful to plaintiffs in matching up the transmittal e-mails with their respective attachments, plaintiffs fail to provide any other reason why they need the transmittal e-mails produced in their native format. For these reasons, the court denies plaintiffs' request for defendant to produce all its ... transmittal e-mails in native format with all attachments in native format and attached to the transmittal e-mails.

The *Williams II* court did, however, say the plaintiffs could pose specific interrogatories to the defendant as necessary to decipher the spreadsheet reports and determine which e-mails matched with a particular spreadsheet.

Wyeth

Another court quoted *Williams I*, but reached an opposite result, disallowing production of native-format files in *Wyeth v. Impax Laboratories Inc.*, 2006 WL 3091331 (D. Del. 2006). The facts behind the decision are similar to those of *Williams II*. The court in *Wyeth* declined to require metadata production, relying primarily on the failure to request metadata before the production and a local rule making TIFF and JPEG the default formats of production.

In effect, the court was relying on the "one format" production limitation of revised Rule 34(b), although the new rule had not yet gone into effect and was not mentioned in the opinion. The *Wyeth* court explained its ruling as follows:

Since the parties have never agreed that electronic documents would be produced in any particular format, *Wyeth* complied with its discovery obligation by producing image files. Further, neither party has argued that the need for accessing metadata was foreseeable or generally necessary. Finally, *Impax* has not demonstrated a particularized need for the metadata or database production it has requested. Therefore, this part of *Impax's* motion is denied.

Wyeth apparently tries to buttress the decision with a quote from *Williams I* that the "emerging standards of electronic discovery appear to articulate a general

presumption against the production of metadata.” The court correctly quotes *Williams I*, but does not point out that this was a summary of the defendant’s position, which *Williams I* rejected. (*Williams II* had not yet been decided.)

Kentucky Speedway

A better-reasoned case in a Kentucky district court explicitly rejects *Williams I*: *Kentucky Speedway v. NASCAR*, No. 05-00138 (E.D. Ky. 2006). This is an antitrust action against NASCAR in which the defendants had already spent \$3 million in five months responding to e-discovery requests. Then, with that background, the plaintiff for the first time asks for production of all metadata in documents already produced. (Again, note the similarities with *Williams II*, which had not yet been decided.) The plaintiff relied upon *Williams I* to try to justify this late request but failed to make “any showing of a particularized need for the metadata.”

The *Kentucky Speedway* judge rejected *Williams I* in this context and instead followed *Wyeth*, holding that:

In the rapidly evolving world of electronic discovery, the holding of the *Williams* case is not persuasive. Having the benefit of the newly amended rules, advisory notes and commentary of scholars, I respectfully disagree with its conclusion that a producing party “should produce the electronic documents with their metadata intact, unless that party timely objects..., the parties agree that the metadata should not be produced or the producing party requests a protective order.”

Here, the parties clearly had no agreement that the electronic files would be produced in any particular format. Plaintiff did not notify defendant ISC that it sought metadata until seven months after ISC had produced both hard copy and electronic copies of its documents.

To the extent that plaintiff seeks metadata for a specific document or documents where date and authorship information is unknown but relevant, plaintiff should identify that document or documents by Bates number or by other reasonably identifying features. Responding to a request for additional information concerning specific documents would be far less burdensome to defendant and far more likely to produce relevant information.

The opinion in *Kentucky Speedway* does not answer the question of whether the court would have reached the same result if the plaintiff had made the request for metadata from the beginning, and not waited until after the defendants had already spent millions of dollars to produce the same documents without metadata.

It is, however, clear from *Kentucky Speedway* that whenever a metadata production will create a substantial burden on the producing party (here, it would have cost NASCAR another \$500,000), the requesting party will have to provide good cause. The plaintiff’s reliance on *Williams I* to support the production of metadata in all circumstances, and without a good-cause showing, was misplaced and distorts the actual holding of *Williams I* (as shown for instance by *Williams II*).

Instead, both *Williams* rulings stand for the proposition that the producing party must object and show undue burden, and the burden then shifts to the requesting party to prove good cause. The argument on metadata production is essentially the same as the inaccessibility argument under Rule 26(b)(2)(B). If you can show a real need to see the metadata, as the plaintiffs did in *Williams I* (but not

Williams II), it may be possible to compel the production in spite of burden on the producing party. It will be a balancing test dependent upon the circumstances and following something like the seven factors recommended by the Rules Committee for 26(b)(2)(B) analysis,¹ and earlier in *Zubulake v. UBS Warburg LLC* for cost-shifting,² but with the added dimension of the debatable feasibility at this time of redacting privileged materials from native files.

In re Payment

The latest word in this controversy of “emerging standards” comes out of a consolidated group of class actions styled *In re Payment Card Interchange Fee and Merchant Discount Antitrust Litigation*, No. MD 05-1720, 2007 WL 121426 (E.D.N.Y. Jan. 12, 2007). As in *Kentucky Speedway*, *Wyeth* and *Williams II*, the defendants here sought the production of the metadata for documents already produced without metadata. (Actually, the defendants never filed a motion to compel; they just raised the issue at a conference, and that was part of the problem.) In addition, the defendants wanted the plaintiffs to produce all metadata on documents they had not yet produced.

As to the previously produced documents, the holding here follows *Kentucky Speedway*, *Wyeth* and *Williams II*. *In re Payment* holds that the defendants waited too long to complain of the metadata-stripped production and implied that there had been a waiver. The lesson here is that in order to obtain metadata you may need, you should specifically ask for that metadata to begin with, and if the production is later stripped of the metadata requested, you should immediately and vigorously object.

In re Payment also followed *Williams I* to a certain extent, in that it ordered all future productions by the plaintiffs to include metadata. The court explained its reasons for requiring full metadata production:

The defendants object to individual plaintiffs’ production protocol on the grounds that, by failing to supply metadata, it does not comply with amended Rule 34. The Advisory Committee on Civil Rules, in its notes to the 2006 amendment to Rule 34, wrote that a party responding to a discovery request may elect to produce a “reasonably usable” form of electronic data rather than produce the information as kept in the ordinary course of business. FED. R. CIV. P. 34(b), 2006 Amendment, Advisory Committee’s Note. That is precisely what the individual plaintiffs have done. By making that choice, however, they have run afoul of the Advisory Committee’s proviso that data ordinarily kept in electronically searchable form “should not be produced in a form that removes or significantly degrades this feature.” *Id.*

However, the same cannot be said of prospective discovery, meaning the electronic documents that the individual plaintiffs have not yet substantially prepared for production as searchable TIFF images. Now that the individual plaintiffs are aware of the defendants’ objections, their argument of undue burden is weaker; indeed, they have conceded that their concerns about the burdens of producing electronic documents in native format largely disappear with respect to the documents they have not yet processed for production.

Thus this new metadata case, like *Williams I* and *II*, supports the proposition that the emerging standard requires metadata production when requested and not objected to, but at the same time emphasizes the need for early, clear requests and prompt objections if the metadata is not provided.

Ralph C. Losey is a 1979 *cum laude* graduate of the University of Florida College of Law, with a B.A. from Vanderbilt University. He is a practicing attorney and co-chair of Akerman Senterfitt's electronic discovery practice group. Mr. Losey was the first attorney in Florida with a Web site and has been involved with computers and technology since 1978. He currently maintains two law-related Web sites and a blog on e-discovery: Ralph Losey's Law Web, www.floridalawfirm.com; the e-DiscoveryTeam site, www.e-discoveryteam.com; and the e-DiscoveryTeam blog, <http://ralphlosey.wordpress.com>.

■ Notes

- ¹ The committee listed seven factors similar to those delineated in *Zubulake I*, 217 F.R.D. 309 (S.D.N.Y. May 13, 2003). The committee said, "Appropriate considerations may include: (1) the specificity of the discovery request; (2) the quantity of information available from other and more easily accessed sources; (3) the failure to produce relevant information that seems likely to have existed but is no longer available on more easily accessed sources; (4) the likelihood of finding relevant, responsive information that cannot be obtained from other, more easily accessed sources; (5) predictions as to the importance and usefulness of the further information; (6) the importance of the issues at stake in the litigation; and (7) the parties' resources." Fed. R. Civ. P. 26, Advisory Committee Notes 2006 Amendments Subdivision (b).
- ² "In conducting the cost-shifting analysis, the following factors should be considered, weighted more-or-less in the following order: (1) the extent to which the request is specifically tailored to discover relevant information; (2) the availability of such information from other sources; (3) the total cost of production, compared to the amount in controversy; (4) the total cost of production, compared to the resources available to each party; (5) the relative ability of each party to control costs and its incentive to do so; (6) the importance of the issues at stake in the litigation; and (7) the relative benefits to the parties of obtaining the information." *Zubulake*, 217 F.R.D. 309.