SOUTHERN DISTRICT CIVIL PRACTICE ROUNDUP

Expert Analysis

'The New Black': Meditations on Metadata

The recent decision by Magistrate Judge Frank Maas of the U.S. District Court for the Southern District of New York in *Aguilar v. Immigration and Customs Enforcement Division*[1] marks a day of reckoning for those civil litigators who thought they could avoid learning about electronic discovery.

Although the eyes of many litigators glaze over when the topic turns to "load files" and "native format," as Judge Maas observed in *Aguilar*, "[m]etadata has become 'the new black.'" While hardly the sexiest aspect of 21st century litigation, time has run out for those who had hoped never to have to understand what "metadata" is, let alone to explain why they are entitled to obtain it or avoid its production.

Aguilar is a civil rights class action brought by Latino plaintiffs alleging that the Immigration and Customs Enforcement Division of the U.S. Department of Homeland Security (ICE) and individual ICE agents violated their constitutional rights by entering and searching plaintiffs' homes without warrants. Defendants conceded that they did not have probable cause for these searches, defending the lawsuit on the basis of their belief that plaintiffs had consented to the searches.

Although not discussed at the initial discovery conference mandated by Rule 26(f) of the Federal Rules of Civil Procedure, and not specified in plaintiffs' first request for production

---

By **Edward M. Spiro**

And **Judith L. Mogul**

EDWARD M. SPIRO *and* JUDITH L. MOGUL *are principals of Morvillo, Abramowitz, Grand, Iason, Anello & Bohrer, both concentrating in commercial litigation. Mr. Spiro is the coauthor of "Civil Practice in the Southern District of New York," 2d Ed. (Thomson West 2007.)*

of documents, plaintiffs subsequently informed defendants that they wanted various types of documents produced in particular formats that would provide "metadata"—data about the data—for the electronically stored information (ESI) subject to production.

By the time plaintiffs made that request, defendants had substantially completed collecting the documents for production. Magistrate Judge Maas held that having failed initially to request that documents be produced with metadata, plaintiffs had an "uphill battle" in seeking that information after almost all documents had been "harvested" without that data.

Metadata is secondary information that describes an electronic document's characteristics, origins, and usage, information that cannot be seen on the face of the document. Thus, when a hard copy of a document is produced, or documents are produced electronically but in what is essentially a snapshot form (such as a pdf or tagged image format (TIFF file)), the information provided is limited to the actual text or superficial content of the document. Metadata includes information such as the date on which a document was created, its editing history, blind copy recipients of an e-mail, or hidden codes or formulas used in creating the document.

As a prelude to his analysis of whether to order production of the specific information at issue in *Aguilar*, Magistrate Judge Maas provided a useful tutorial on the three basic categories of metadata. He explained that in general terms, metadata refers to data that describes the "history, tracking, or management of an electronic document," and is most usefully grouped into three primary types of information: substantive, system and embedded metadata. Relying on both the Sedona Principles which set out recommendations for best practices in discovery of electronic information,[2] and protocols for the discovery of ESI promulgated by the U.S. District Court for the District of Maryland,[3] Magistrate Judge Mass described the three types of metadata, and what their relevance or utility might be in discovery.

Substantive Metadata

Substantive metadata, or application metadata, is information created by the software used to create the document, reflecting editing changes or comments, and instructions concerning fonts and spacing. "Substantive metadata is embedded in the document it describes and remains with the document when it is moved or copied."[4] Such information is useful in showing the genesis of a particular document and the history of proposed revisions or changes. Because substantive metadata remains with the document when it is copied, but is not apparent on the face of the document, parties producing documents containing substantive metadata need to be alert to the hidden information they may be supplying to their adversaries.

System Metadata

System metadata reflects automatically generated information about the creation

or revision of a document, such as the document's author, or the date and time of its creation or modification. System metadata is not necessarily embedded in the document, but can be obtained from the operating system or information management system on which the document was created. Magistrate Judge Maas observed that system metadata is most relevant if a document's authenticity is at issue, or there are questions as to who received a document or when it was received.

Embedded Metadata

Embedded metadata is data that is inputted into a file by its creators or users, but that cannot be seen in the document's display. Common types of embedded metadata include the formulas used to create spreadsheets, hidden columns, references, fields, or internally or externally linked files. Embedded metadata is often critical to understanding complex spreadsheets which lack an explanation of the formulas underlying the output in each cell.

The two most common ways of producing metadata for ESI are to produce documents (i) in a TIFF or pdf format with an accompanying "load file" or (ii) in "native format." A load file contains coded data such as address fields, dates (on which the document was created, modified, sent or opened), text body and information about document length and attachments. Load files can be formatted for different software applications (such as Concordance or Ringtail) that permit the user to search for documents with specific properties or terms.

Creating a load file is time consuming and expensive. "Native format" is the format in which the document being produced was created and provides the recipient of the document all the information available to the original user. Documents can be produced more quickly and efficiently in native format, but cannot be bates numbered or redacted effectively, and may only be opened and used by the requesting party if it has the software on which the documents were created.

Operative Principles

Magistrate Judge Maas traced an evolution in the approach to evaluating requests for metadata, noting that the initial focus on the relevance of the information contained in the metadata had shifted and broadened over time to include consideration of the enhanced accessibility and functionality that metadata provides recipients.

He cited the commentary to the influential Sedona Principles which advises parties to consider "(i) what metadata is ordinarily maintained; (ii) the relevance of the metadata; and (iii) the importance of reasonably accessible metadata to facilitating the parties' review, production, and use of the information."[5]

The judge explained that in general terms, metadata refers to data that describes the "history, tracking, or management of an electronic document," and can be grouped into three primary types of information: substantive, system and embedded metadata.

As Magistrate Judge Maas observed, the Sedona Principles, the Federal Rules, and applicable case law all contemplate that electronic discovery will be a "party-driven process."[6] Rule 26(f) which requires the parties to meet and confer in order to develop a discovery plan, specifically directs that the plan address issues about disclosure of ESI, including the form or forms in which ESI is to be produced.[7]

Rule 34 sets out a detailed procedure for resolving how ESI is to be produced, putting the onus on the parties to address and narrow issues relating to electronic discovery before its production, and before involving the court in any related dispute.

The rule provides that the party seeking discovery specify the form or forms in which the information should be produced, after which the responding party must indicate the extent to which it will comply with that request and those aspects, if any, to which it objects.

If the responding party objects to producing any ESI in the form requested, it must indicate the form in which the requested information will be produced. The burden is then on the requesting party to bring a motion to compel production of the ESI in the requested form.

In resolving a motion to compel, the court balances the probative value of providing the information in the requested form against the burden and expense of its production.[8]

The Advisory Committee comments to the amendments to Rule 34 dealing with discovery of ESI caution that a responding party who simply produces ESI in a form other than that requested, without first resolving disputes as to how the information should be produced, runs the risk of having to produce that information again following litigation over the form of production. By the same token, as *Aguilar* makes clear, a party who does not specify the form in which requested information should be produced, may well have to make do with the format chosen by its adversary.

'Aguilar' Analysis

Magistrate Judge Maas noted that the case law establishes a clear pattern correlating the outcome of a request for production of metadata with the timing of the request for that information. Where the request for metadata is sought in the initial production, before documents have been produced in any form, such requests are generally granted.[9] Similarly, when the request for metadata is delayed, particularly until after documents have been produced in some other form, courts generally refuse to compel the production of metadata.[10]

In *Aguilar*, plaintiffs informally and "in passing" first mentioned that they wanted metadata two months after the initial rule 26(f) conference took place and a month after they served their written discovery requests. This informal request was not formalized for another two months, by which time defendants had largely completed their collection efforts and had already produced many of the requested documents in pdf format, without accompanying metadata.

Magistrate Judge Maas observed that this timing presented plaintiffs with an "uphill battle" in seeking to compel production of metadata. But the delayed request for metadata was not dispositive of plaintiffs' motion. Rather, he measured the importance of the metadata to plaintiffs against the burden its belated retrieval would impose on defendants, taking into consideration the fact that those burdens would have been lessened, and more appropriately borne by defendants, if the request for metadata had been timely made.

Specifically, plaintiffs sought load files for a universe of approximately 500 e-mails that would disclose, among other information, recipients of blind copies of the e-mails, and would permit plaintiffs to search the e-mails more efficiently.

Defendants' counsel had collected some of those e-mails in a manner which preserved the underlying metadata, while other e-mails were forwarded to the collecting attorney in a way that altered the original metadata.

Defendants offered to re-produce a portion of the e-mails for which the metadata had been preserved, but declined to do so for more than half the documents they had produced earlier, citing the burden of that review.

Magistrate Judge Maas declined to have defense counsel retrace their steps, citing plaintiffs' delay, the fact that it was far from clear that the requested metadata would reveal much useful evidence, and that the universe of documents (500) was small enough that it could be effectively searched without metadata. He also denied plaintiffs' request that defendants search their back-up tapes to retrieve e-mails for which the metadata had been altered, finding that the high cost of that endeavor outweighed its likely low benefit.

Magistrate Judge Maas also took a dim view of plaintiffs' request for system metadata for various Word, PowerPoint and Excel documents—information they claimed would facilitate their searches of the documents and reveal "who knew what when." He noted that although produced without metadata, the documents were produced in text-searchable form, and comprised just over one box, observing that plaintiffs' concerns about search efficiency would carry more force if hundreds of boxes of documents were involved.

The judge also found that plaintiffs' relevance claim was diminished by defendants' concession that they lacked probable cause for the searches at issue. He observed that "[i]f establishing probable cause…were important, there might be a need to know what information each officer had learned by the time of a search," but that "[w]ith probable cause generally out of the picture, 'who knew what when' is, at best, only marginally relevant."[11]

Rule 34 provides that the party seeking discovery specify the form or forms in which the information should be produced, after which the responding party must indicate the extent to which it will comply with that request and those aspects, if any, to which it objects.

He predicted that the probative value of the metadata information was likely outweighed by the burden of requiring defendants to make a second production in response to plaintiffs' delayed request. Magistrate Judge Maas concluded, however, that since there was some potential relevance to the metadata, and some increased utility in using it to search the documents, he would grant plaintiffs' motion to compel this metadata on the condition that plaintiffs bear the cost for the second production.

Magistrate Judge Maas was more favorably inclined toward plaintiffs' request for the metadata associated with various spreadsheets reflecting the dates that various field operations were conducted, and the number and type of arrests that resulted from each operation. He observed that while the spreadsheets contained some embedded metadata for computing the number of arrests in each category, the formulas were not necessary to understanding the spreadsheets. Nevertheless, because providing the requested metadata would not be unduly burdensome, and defendants had expressed some willingness to provide the metadata, Magistrate Judge Maas ordered the spreadsheets produced in native format.

Conclusion

On its face, *Aguilar* simply enforces the requirements of the Federal Rules of Civil Procedure that parties address electronic discovery issues in developing a discovery plan at the outset of a case; that the party requesting discovery specify the form in which electronically stored information be produced; and that a party need not produce the same information in more than one format.

But *Aguilar* stands for much more than the proposition that metadata should be requested early and as a matter of course. It instructs that attorneys understand the different formats in which electronically generated or stored data can be produced so that they can make informed decisions about the advantages and disadvantages of obtaining and producing data in a particular format, and advocate effectively in support of those decisions.

••••••••••••●●•••••••••••

1. 2008 WL 5062700 (S.D.N.Y. Nov. 21, 2008).
2. Sedona Conference Working Group Series 2007, Sedona Principles-Second Edition: Best Practices Recommendations and Principles for Addressing Electronic Document Production, Cmt. 12a.
3. Suggested Protocol for Discovery of Electronically Stored Information 25-28, http://www.mdd.uscourts.gov/news/news/ESIProtocol.pdf.
4. 2008 WL 5062700, at *3.
5. Id., at *6 (quoting Sedona Principles 2d, Principle 12, Comment 12b (internal quotations omitted)).
6. Id., at *9.
7. Fed. R. Civ. P. 26(f)(3)(C).
8. For a more detailed discussion of the rules and procedures pertaining to discovery of ESI, see Spiro, E., 2006 Amendments to the Federal Rules of Civil Procedure, NYLJ (Oct. 5, 2006); Spiro, E., Cost Shifting in Discovery, NYLJ (Dec. 14, 2006).
9. 2008 WL 5062700, at *7 (citing, e.g., *In re Payment Card Interchange Fee & Merch. Disc.*, 2007 WL 121426 (E.D.N.Y. Jan. 12, 2007) (Orenstein, M.J.); *Hagenbuch v. 3B6 Sistemi Elettronici Industriali S.R.L.*, 2006 WL 665005 (N.D. Ill. Mar. 8, 2006)).
10. Id., at *7-8 (citing, e.g., *Autotech Techs. Ltd. P'ship v. AutomationDirect.com Inc.*, 248 F.R.D. 556 (N.D. Ill. 2008)).
11. Id., at *12.