

UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

-----X

ADRIANA AGUILAR, et al., :

Plaintiffs, :

-against- :

IMMIGRATION AND CUSTOMS :

ENFORCEMENT DIVISION OF THE :

UNITED STATES DEPARTMENT OF :

HOMELAND SECURITY, et al., :

Defendants. :

-----X

**MEMORANDUM
DECISION AND ORDER**

07 Civ. 8224 (JGK)(FM)

USDC SDNY DOCUMENT ELECTRONICALLY FILED DOC #: _____ DATE FILED: <u>11/21/2008</u>
--

FRANK MAAS, United States Magistrate Judge.

This civil rights class action is brought by more than thirty Latino plaintiffs (“Plaintiffs”) who contend that the Immigration and Customs Enforcement Division of the United States Department of Homeland Security (“ICE”) and certain of its employees (collectively, the “Defendants”) subjected them to unlawful searches of their homes in violation of the Fourth Amendment. Because counsel failed to discuss the form of production for electronic documents early in the case, the Court now must resolve several issues concerning the discoverability of metadata. For the reasons set forth below, the Plaintiffs’ application to compel the production of metadata is granted in part and denied in part.

I. Relevant Facts

A. Second Amended Complaint

The second amended complaint in this action alleges that ICE initiated “Operation Return to Sender” and other similar programs in 2006 to identify and arrest persons who had been ordered removed from the country in immigration proceedings but who remained present in the United States as fugitive aliens. (Am. Compl. ¶ 2). The Plaintiffs further allege (and the Defendants deny) that ICE executes Operation Return to Sender by deploying teams of six to ten armed ICE agents during the early morning hours to enter and search the homes of Latinos without having previously obtained a search warrant or consent, in violation of the Fourth Amendment to the United States Constitution. (Id. ¶ 4-5; Docket No. 26 at 7-8). Pursuant to Bivens v. Six Unknown Named Agents of Federal Bureau of Narcotics, 403 U.S. 388 (1971), the Plaintiffs seek the damages they allegedly suffered as a result of these unconstitutional searches. (Am. Compl. ¶ 38). The Plaintiffs also seek a permanent injunction barring ICE from continuing to conduct its searches in this manner. (Id. ¶ 37).

B. Procedural Posture

On December 7, 2007, the Defendants moved before the Honorable John G. Koeltl, the District Judge before whom this case is pending, to dismiss the claim for injunctive relief in the first amended complaint. The Defendants also sought a stay of discovery while the motion to dismiss was pending, except to the extent that the discovery

related to the Bivens claims.¹ (See Docket Nos. 22, 26). Thereafter, at a discovery conference on January 15, 2008, Judge Koeltl urged the parties to resolve which elements of discovery should proceed and which should be stayed while the motion to dismiss the injunctive claim was pending. (1/15/08 Tr. at 50). The Judge also noted that he assumed the Defendants were taking steps to gather relevant documents, such as “documents relating to the[] individual searches.” (Id. at 7). At or about this time, the Defendants began to harvest the relevant documents from ICE employees. (See letter from Ass’t U.S. Att’y Shane Cargo to the Court, dated August 13, 2008 (“Cargo Letter”), at 7).

During a Rule 26(f) discovery conference on January 18, 2008, the parties agreed that discovery would proceed with regard to the Bivens claims only, and that the parties would serve their first requests for the production of documents by February 15, 2008. (See Docket No. 29). There was no discussion of metadata at this conference. (See letter from Donna L. Gordon, Esq., to the Court, dated August 11, 2008 (Gordon Letter”), at 2).

On February 15, 2008, the Plaintiffs served their first request for the production of documents. (Letter from Patrick J. Gennardo, Esq., to the Court, dated June 27, 2008, Ex. 3). Their request did not specify the form in which they sought to have

¹ Judge Koeltl later denied the motion to dismiss as moot because the Plaintiffs indicated that they intended to file a second amended complaint. That pleading was filed on May 30, 2008. (Docket No. 50).

electronically stored information (“ESI”) produced, nor did it mention the production of metadata.² (See id.).

The subject of metadata first arose on March 18, 2008, when the Plaintiffs apparently mentioned it “in passing.” (Cargo Letter at 8). By this time, the Defendants had almost completed their document collection efforts. (Id.).

The first formal discussion among the parties regarding metadata occurred on May 22, 2008, during a conference call to discuss the production of ESI. (Cargo Letter at 8; Gordon Letter at 3). During the call (and by means of a subsequent letter), the Plaintiffs requested (1) that emails and electronic documents be produced in Tagged Imaged File Format (“TIFF”) with a corresponding load file containing metadata fields and extracted text,³ and (2) that spreadsheets and databases be produced in native format.⁴

² The request did state that “all documents shall be produced as they are kept in the usual course of business or shall be organized and labeled to correspond with the categories in the request.” (Id. at 4).

³ TIFF is a static image format similar to a PDF that creates a mirror image of the electronic document. See PSEG Power N.Y., Inc. v. Alberici Constructors, Inc., No. 1:05-CV-657 (DNH) (RFT), 2007 WL 2687670, at *2 n.2 (N.D.N.Y. Sept. 7, 2007); In re Payment Card Interchange Fee & Merch. Disc., No. MD 05-1720 (JG) (JO), 2007 WL 121426, at *1 n.2 (E.D.N.Y. Jan. 12, 2007).

A load file is a “file that relates to a set of scanned images or electronically processed files, and indicates where individual pages or files belong together as documents, to include attachments, and where each document begins and ends,” and may also include “data relevant to the individual documents, such as metadata, coded data, text, and the like.” The Sedona Conference Glossary 31 (2d ed. 2007), http://www.TheSedonaConference.org/content/miscFiles/TSCGlossary_12_07.pdf.

⁴ Native format is the “default format of a file,” access to which is “typically
(continued...)

(Gordon Letter at 3; Ex. O at Ex. B). As noted above, by this date, the Defendants had already substantially completed their document collection efforts. (Cargo Letter at 8).

The parties conferred again on July 1, 2008, with respect to the format in which information from ICE's hierarchical databases would be produced. (Gordon Letter at 3). On July 14, 2008, the Defendants objected, on relevance and burden grounds, to producing electronic documents in the form requested by the Plaintiffs, proposing instead to produce their ESI in the form of text-searchable PDF documents. (Gordon Letter Ex. O (Ex. C at 3)). To the extent that the Plaintiffs sought metadata, the Defendants stated that they would provide it if the Plaintiffs were able to demonstrate that the metadata associated with a particular document was relevant to their claims. (Id.).

During a discovery conference on July 17, 2008, I directed counsel and their ESI experts to meet and confer in a renewed attempt to resolve their disputes regarding metadata. (See 7/17/08 Tr. at 27). They did so on July 25 and 30, 2008, but were unable to resolve their differences. Thereafter, I held a further conference on August 14, 2008, which was no more successful on the issue of metadata. During that conference, I directed the parties to confer once again about the metadata that the Plaintiffs sought with respect to various categories of documents such as emails and spreadsheets. (Docket No. 65). I also scheduled a follow-up session with counsel and

⁴(...continued)
provided through the software program on which it was created.” In re Priceline.com Inc. Sec. Litig., 233 F.R.D. 88, 89 (D. Conn. 2005).

their ESI experts to be held in my courtroom. (Id.). As set forth below, that meet and confer session, held on September 8, 2008, narrowed the parties' differences.

Although no formal motion has been made, the Plaintiffs' letters amount to a motion to compel the Defendants' production of (1) responsive emails and electronic documents (such as Word, PowerPoint, and Excel documents) in TIFF format with corresponding metadata, and (2) meaningful information about the metadata fields of ICE's hierarchical databases so that the Plaintiffs can determine which database metadata they should request.

II. Discussion

A. Metadata and Discovery

“As a general rule of thumb, the more interactive the application, the more important the metadata is to understanding the application's output.” Williams v. Sprint/United Mgmt. Co., 230 F.R.D. 640, 647 (D. Kan. 2005). Thus, while metadata may add little to one's comprehension of a word processing document, it is often critical to understanding a database application. Id. “A spreadsheet application lies somewhere in the middle” and the need for its metadata depends upon the complexity and purpose of the spreadsheet. Id. To understand why the importance of metadata varies, it is first necessary to explain what it is and distinguish among its principal forms.

1. Types of Metadata

Metadata, frequently referred to as “data about data,” is electronically-stored evidence that describes the “history, tracking, or management of an electronic document.” Id. at 646. It includes the “hidden text, formatting codes, formulae, and other information associated” with an electronic document. The Sedona Principles, Second Edition: Best Practices Recommendations and Principles for Addressing Electronic Document Production Cmt. 12a (Sedona Conference Working Group Series 2007), http://www.thesedonaconference.org/content/miscFiles/TSC_PRINCP_2nd_ed_607.pdf (“Sedona Principles 2d”); see also Autotech Techs. Ltd. P’Ship v. AutomationDirect.com, Inc., 248 F.R.D. 556, 557 n.1 (N.D. Ill. 2008) (Metadata includes “all of the contextual, processing, and use information needed to identify and certify the scope, authenticity, and integrity of active or archival electronic information or records”). Although metadata often is lumped into one generic category, there are at least several distinct types, including substantive (or application) metadata, system metadata, and embedded metadata. Sedona Principles 2d Cmt. 12a; see United States District Court for the District of Maryland, Suggested Protocol for Discovery of Electronically Stored Information 25-28, <http://www.mdd.uscourts.gov/news/news/ESIProtocol.pdf> (“Md. Protocol”).

a. Substantive Metadata

Substantive metadata, also known as application metadata, is “created as a function of the application software used to create the document or file” and reflects substantive changes made by the user. Sedona Principles 2d Cmt. 12a; Md. Protocol 26. This category of metadata reflects modifications to a document, such as prior edits or editorial comments, and includes data that instructs the computer how to display the fonts and spacing in a document. Sedona Principles 2d Cmt. 12a. Substantive metadata is embedded in the document it describes and remains with the document when it is moved or copied. Id. A working group in the District of Maryland has concluded that substantive metadata “need not be routinely produced” unless the requesting party shows good cause. Md. Protocol 26.

b. System Metadata

System metadata “reflects information created by the user or by the organization’s information management system.” Sedona Principles 2d Cmt. 12a. This data may not be embedded within the file it describes, but can usually be easily retrieved from whatever operating system is in use. See id. Examples of system metadata include data concerning “the author, date and time of creation, and the date a document was modified.” Md. Protocol 26. Courts have commented that most system (and substantive) metadata lacks evidentiary value because it is not relevant. See Mich. First Credit Union v. Cumis Ins. Soc’y, Inc., No. Civ. 05-74423, 2007 WL 4098213, at *2 (E.D. Mich. Nov.

16, 2007); Ky. Speedway, LLC v. Nat'l Assoc. of Stock Car Auto Racing, No. Civ. 05-138, 2006 WL 5097354, at *8 (E.D. Ky. Dec. 18, 2006); Wyeth v. Impax Labs., Inc., 248 F.R.D. 169, 170 (D. Del. 2006). System metadata is relevant, however, if the authenticity of a document is questioned or if establishing “who received what information and when” is important to the claims or defenses of a party. See Hagenbuch v. 3B6 Sistemi Elettronici Industriali S.R.L., No. 04 Civ. 3109, 2006 WL 665005, at *3 (N.D. Ill. Mar. 8, 2006). This type of metadata also makes electronic documents more functional because it significantly improves a party’s ability to access, search, and sort large numbers of documents efficiently. Sedona Principles 2d Cmt. 12a.

c. Embedded Metadata

Embedded metadata consists of “text, numbers, content, data, or other information that is directly or indirectly inputted into a [n]ative [f]ile by a user and which is not typically visible to the user viewing the output display” of the native file. Md. Protocol 27. Examples include spreadsheet formulas, hidden columns, externally or internally linked files (such as sound files), hyperlinks, references and fields, and database information. Id. This type of metadata is often crucial to understanding an electronic document. For instance, a complicated spreadsheet may be difficult to comprehend without the ability to view the formulas underlying the output in each cell. For this reason, the District of Maryland working group concluded that embedded

metadata is “generally discoverable” and “should be produced as a matter of course.” Id. at 27-28.

2. Discovery of Metadata

a. Federal Rules

Metadata is not addressed directly in the Federal Rules of Civil Procedure but is subject to the general rules of discovery. Metadata thus is discoverable if it is relevant to the claim or defense of any party and is not privileged. Fed. R. Civ. P. 26(b)(1). Additionally, “[f]or good cause, the court may order discovery of any matter [including metadata] relevant to the subject matter involved in the action.” Id. The “[r]elevant information need not be admissible at the trial if the discovery appears reasonably calculated to lead to the discovery of admissible evidence.” Id. The discovery of metadata is also subject to the balancing test of Rule 26(b)(2)(C), which requires a court to weigh the probative value of proposed discovery against its potential burden.

Although metadata is not specifically referenced, Rule 34 of the Federal Rules of Civil Procedure addresses the production of ESI. Fed. R. Civ. P. 34(a)(1)(A), (b)(2)(E). Under the Rule, a requesting party may specify a form of production and request metadata. Fed. R. Civ. P. 34(b)(1)(C). (A typical request might be to produce Word documents in TIFF format with a load file containing the relevant system metadata.) The responding party then must either produce ESI in the form specified or object. If the responding party objects, or the requesting party has not specified a form of

production, the responding party must “state the form or forms it intends to use” for its production of ESI. Fed. R. Civ. P. 34(b)(2)(D). Thereafter, if the requesting party objects and suggests an alternative form, the parties “must meet and confer under Rule [37(a)(1)] in an effort to resolve the matter before the requesting party can file a motion to compel.” Fed. R. Civ. P. 34(b), advisory committee’s note, 2006 amendment.

_____ If the requesting party does not specify a form for producing ESI, the responding “party must produce it in a form or forms in which it is ordinarily maintained or in a reasonably usable form or forms.” Fed. R. Civ. P. 34(b)(2)(E)(ii). Although a party may produce its ESI in another “reasonably usable form,” this does not mean “that a responding party is free to convert electronically stored information from the form in which it is ordinarily maintained to a different form that makes it more difficult or burdensome for the requesting party to use the information efficiently in the litigation.” Fed. R. Civ. P. 34(b), advisory committee’s note, 2006 amendment. In particular, if the ESI is kept in an electronically-searchable form, it “should not be produced in a form that removes or significantly degrades this feature.” Id.; see also Payment Card, 2007 WL 121426, at *4 (documents stripped of metadata allowing searches do not comply with Rule 34(b)).

The Federal Rules also specify that a “party need not produce the same [ESI] in more than one form.” Fed. R. Civ. P. 34(b)(2)(E)(iii).

b. Sedona Principles

The Sedona Conference (“Conference”), a nonprofit legal policy research and education organization, has a working group comprised of judges, attorneys, and electronic discovery experts dedicated to resolving electronic document production issues. Since 2003, the Conference has published a number of documents concerning ESI, including the Sedona Principles. Courts have found the Sedona Principles instructive with respect to electronic discovery issues. See, e.g., Autotech Techs. Ltd. P’Ship, 248 F.R.D. at 560; Williams, 230 F.R.D. at 652.

In the first edition of the Sedona Principles, the Conference stated that “unless it is material to the resolution of a dispute, there is no obligation to . . . produce metadata absent agreement of the parties or order of the court.” The Sedona Principles: Best Practices Recommendations and Principles for Addressing Electronic Document Production Principle 12 (Sedona Conference Working Group Series 2005) (“Sedona Principles 1st”). The Conference further noted that because most “metadata has no evidentiary value” and the time and money spent reviewing it would be a waste of resources, there should be a “modest legal presumption” against the production of metadata. Id. Cmt. 12a. The Conference nevertheless observed that if metadata is relevant, it should be produced. Id.

_____The foreword to the second edition of the Sedona Principles notes that in revising the principles, “[p]articular attention [was] given to updating the language and

commentary on Principle 12 (metadata).” Sedona Principles 2d Foreword. Significantly, Principle 12 and the commentaries accompanying it were revised to remove any presumption against the production of metadata. Principle 12 now 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.

Sedona Principles 2d Principle 12 (emphasis added). Thus, in the first edition of the Sedona Principles the Conference seemed to focus solely on the relevancy of metadata; in the second edition the Conference placed greater weight on the enhanced accessibility and functionality that metadata provides to the recipients of ESI.

The commentary to Principle 12 also was expanded to provide criteria for deciding whether metadata should be produced in a given case. The commentary 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.” Id. Cmt. 12a. In selecting a form of production, the two “primary considerations” should be the need for and probative value of the metadata, and the extent to which the metadata will “enhance the functional utility of the electronic information.” Id. Cmt. 12b.

The commentary also addresses the advantages and disadvantages of various production options. For example, production in native form gives the receiving party access to the same information and functionality available to the producing party and requires minimal processing time before production. Id. However, information in native form is difficult to redact or “Bates” number and the requesting party may not have the software necessary to open the document. Id. By comparison, a production in static image form, such as TIFF or PDF, can be Bates numbered and redacted, but entails the loss of metadata and involves significant processing time. Id. The commentary notes that in “an effort to replicate the usefulness of native files while retaining the advantages of static productions, image format productions typically are accompanied by ‘load files,’ which are ancillary files that may contain textual content and relevant system metadata.” Id. One marked disadvantage of this format is that the production involves significant costs; it also does not work well for spreadsheets and databases. Id.

Weighing the advantages and disadvantages of different forms of production, the Conference concluded that even if native files are requested, it is sufficient to produce memoranda, emails, and electronic records in PDF or TIFF format accompanied by a load file containing searchable text and selected metadata. Id. Cmt. 12b Illus. i. The Conference explained that this “satisfies the goals of Principle 12 because the production is in usable form, e.g., electronically searchable and paired with essential metadata.” Id.

c. Case Law

There is a clear pattern in the case law concerning motions to compel the production of metadata. Courts generally have ordered the production of metadata when it is sought in the initial document request and the producing party has not yet produced the documents in any form. See Payment Card, 2007 WL 121426, at *4 (directing production of metadata for any documents not yet produced); Hagenbuch, 2006 WL 665005, at *4 (granting motion to compel production in native form); In re Priceline.com, 233 F.R.D. at 91 (production ordered in TIFF format with corresponding searchable metadata databases). But see Mich. First Credit Union, 2007 WL 4098213, at *2 (court denied production despite timely request for metadata because it was not relevant and production would be unduly burdensome). On the other hand, if metadata is not sought in the initial document request, and particularly if the producing party already has produced the documents in another form, courts tend to deny later requests, often concluding that the metadata is not relevant. See Autotech Techs., 248 F.R.D. at 559-60 (court refused to compel production of metadata not sought in initial request); D'Onofrio v. SFX Sports Group, Inc., 247 F.R.D. 43, 48 (D.D.C. 2008) (same); Payment Card, 2007 WL 121426, at *4 (denying motion to compel metadata for documents already produced in TIFF format because another production would be unduly burdensome); Ky. Speedway, 2006 WL 5097354, at *8 (motion to compel production of metadata denied when request first came seven months after production); Wyeth, 248 F.R.D. at 171 (documents produced in

TIFF format were sufficient since parties never agreed on form of production);. But see Williams, 230 F.R.D. at 654 (ordering production of Excel spreadsheets with metadata even though no request had been made initially because producing party should reasonably have known that metadata was relevant).

In sum, as a recent article has noted, if a party wants metadata, it should “Ask for it. Up front. Otherwise, if [the party] ask[s] too late or ha[s] already received the document in another form, [it] may be out of luck.” Adam J. Levitt & Scott J. Farrell, Taming the Metadata Beast, N.Y.L.J., May 16, 2008, at 4. Hagenbach illustrates the wisdom of this advice. In that patent infringement suit, the plaintiff demanded electronic document production in native form in his first document request. 2006 WL 665005, at *1 (request for “identical, electronic copies” of the documents). The defendants rejected this request and produced the documents in TIFF format without metadata. Id. The court noted that the TIFF documents did not contain such relevant information as the creation and modification dates of documents, email attachments and recipients, and other metadata. Id. at *3. The court also observed that the metadata was relevant to the plaintiff’s infringement claim because it “will allow him to piece together the chronology of events and figure out, among other things, who received what information and when.” Id. The court therefore ordered production in native form despite the fact that the defendants could not Bates stamp the documents and had already made a production. Id. at *4.

By comparison, in Autotech Technologies, the court denied a motion to compel the production of metadata for Word documents after the plaintiff had already produced the documents in both PDF and paper format. 248 F.R.D. at 557. In that case, the initial production request did not specify a form for production. As the court noted, the plaintiff therefore could have produced its documents in the form in which they were ordinarily maintained or in a reasonably usable form. Id. at 558. In concluding that production in PDF form constituted a reasonably usable form, the court relied heavily upon the defendant’s failure to ask for metadata at the outset. Id. at 559-60. The court stated that it “seems a little late to ask for metadata after documents responsive to a request have been produced in both paper and electronic format.” Id. at 559. The court also noted that, “[o]rdinarily, courts will not compel the production of metadata when a party did not make that a part of its request.” Id. It concluded that the defendant “was the master of its production requests; it must be satisfied with what it asked for.”⁵ Id. at 560.

Similarly, in Kentucky Speedway, the plaintiff sought to compel the production of the associated metadata seven months after the defendant had produced both hard and electronic copies of its documents. 2006 WL 5097354, at *7-8. The court noted that the requirement of Rule 34 that data be produced as ordinarily maintained or in a “reasonably usable” form was not intended to set a default form mandating that

⁵ To bolster its conclusion, the court in Autotech Technologies relied on the “modest” presumption against the production of metadata stated in the Sedona Principles 1st. Autotech Techs., 248 F.R.D. at 560. As noted above, that presumption has been abandoned in the Sedona Principles 2d.

metadata be turned over automatically in every case. Id. at *8. The court also commented that, “[i]n most cases and for most documents, metadata does not provide relevant information.” Id. The court denied the plaintiff’s motion because it failed to show a “particularized need” for metadata, but permitted the plaintiff to request metadata for any specific document for which the “date and authorship information is unknown but relevant.” Id. at *8-9.

Another case illustrating the importance of a timely request for metadata is Payment Card, 2007 WL 121426. In that case, the defendants demanded the production of electronic data as kept in the ordinary course of business, but did not make an explicit reference to metadata. Id. at *1. The plaintiffs subsequently made six productions of electronic documents in TIFF format without metadata. Id. The defendants then objected to the form of production because, without the metadata, the searchability of the documents had been degraded in contravention of Rule 34. Id. at *2. Magistrate Judge James Orenstein construed the defendants’ objections as a motion to compel the production of the documents with metadata intact, which he granted in part and denied in part. Id. at *3-4.

In his decision, Judge Orenstein noted the advisory committee note indicating that the Rule 34 reference to the production of information in “reasonably usable form” meant that an electronic document production should not degrade searchability. Id. at *4. The plaintiffs consequently were cautioned that, going forward,

their failure to produce electronic documents with metadata intact might result in an order requiring a second production at their expense. Id. However, because the plaintiffs already had produced electronic documents over several months without objection, the Judge concluded that a second production of those documents would be unduly burdensome and refused to grant the motion to compel. Id.

3. Party-Oriented Process

_____ The Federal Rules of Civil Procedure, case law, and the Sedona Principles all further emphasize that electronic discovery should be a party-driven process. Indeed, Rule 26(f) requires that the parties meet and confer to develop a discovery plan. That discovery plan must discuss “any issues about disclosure or discovery of [ESI], including the form or forms in which it should be produced.” Fed. R. Civ. P. 26(f)(3)(C) (emphasis added). In fact, the commentary to the rule specifically notes that whether metadata “should be produced may be among the topics discussed in the Rule 26(f) conference.” Fed. R. Civ. P. 26(f) advisory committee’s note, 2006 amendment. As the commentary further observes, early “identification of disputes over the forms of production may help avoid the expense and delay of searches or productions using inappropriate forms.” Id. Thus, at the outset of any litigation, the parties should discuss whether the production of metadata is appropriate and attempt to resolve the issue without court intervention.

_____ Likewise, courts have emphasized the need for the parties to confer and reach agreements regarding the form of electronic document production before seeking to

involve the court. See, e.g., Scotts Co. LLC v. Liberty Mut. Ins. Co., No. 2:06-CV-899, 2007 WL 1723509, at *4 (S.D. Ohio June 12, 2007) (refusing to decide whether metadata need be produced because it was unclear “whether the parties have fully exhausted extra-judicial efforts to resolve” the dispute); Ky. Speedway, 2006 WL 5097354, at *8 (“metadata . . . should be addressed by the parties in a Rule 26(f) conference”); Hopson v. Mayor and City Council of Baltimore, 232 F.R.D. 228, 245 (D. Md. 2005) (“counsel have a duty to take the initiative in meeting and conferring to plan for appropriate discovery of [ESI including metadata] at the commencement” of the case).

The Sedona Principles also stress the need for the parties to resolve issues concerning metadata. As the Conference explains, the purpose of the amended Federal Rules is “to require parties, not courts, to make the tough choices that fit the particular discovery needs of a case.” Sedona Principles 2d Cmt. 12c. This is appropriate because it is not the court but the parties who have the greatest knowledge of the documents in a case and whether the metadata accompanying those documents is relevant. Indeed, the Conference recently has issued a “Cooperation Proclamation,” in which it stresses that the Federal Rules are a mandate that counsel act cooperatively in resolving discovery issues. See Sedona Conference Cooperation Proclamation 2 (2008), http://www.thesedonaconference.org/content/miscFiles/Cooperation_Proclamation.pdf.

B. Application to Facts

Metadata has become “the new black,” with parties increasingly seeking its production in every case, regardless of size or complexity. In keeping with that trend, the Plaintiffs in this case argue that all metadata for all electronic documents should be produced, both because the metadata is relevant to their claims and because it will enable them to search and sort the documents more efficiently. In evaluating these assertions, the timing of the Plaintiffs’ request is important. The first Rule 26(f) conference was held on January 18, 2008. Thereafter, the first request for production of documents was made on February 15, 2008. Throughout this time period, the Plaintiffs made no mention of metadata even though the Defendants had started to harvest their documents. Indeed, by the time the Plaintiffs first informed the Defendants of their desire for metadata (in passing on March 18 and formally on May 22, 2008), the Defendants’ document collection efforts were largely complete and they had already produced many of their electronic documents in PDF format without accompanying metadata. In these circumstances, the Plaintiffs face an uphill battle in their efforts to compel the Defendants to make a second production of their ESI.

With this heightened burden in mind, I will turn to a review of the types of ESI for which the Plaintiffs have requested metadata.

1. Emails

One consequence of the Defendants' document production format is that the emails produced to date do not contain information about who was blind copied ("bcc'd") or the folders to which the emails were saved. The Plaintiffs allege that this information is relevant because it may help bolster their claim that certain defendants condoned a pattern and practice of unconstitutional home searches. (See Gordon Letter at 6). The Plaintiffs also contend that the process of searching the emails would be more efficient if they had the underlying metadata. (Id.). The Defendants concede that the metadata regarding "bcc" information is "arguably relevant," but contend that information about the folder to which an email was saved is not relevant because it does not indicate how an individual used the information contained in the email. (Cargo Letter at 3-4 & n.3).

At the September 8th court conference, the Defendants shed greater light on their methodology for harvesting emails. As they explained, ICE officials involved in the searches were asked to search their electronic files for responsive emails and provide them to a contact person in the ICE Office of the Principal Legal Advisor ("OPLA"). (See 9/8/08 Tr. at 4; email from Ventura Ramos, dated Oct. 12, 2007, regarding "Preservation of Documents"). If the officials responded by forwarding their emails to the ICE contact person, the original email metadata was altered in the process. (9/8/08

Tr. at 11-12). On the other hand, if the email was saved as an .msg or .pst file which, in turn, was sent to the contact person, the original metadata was preserved. (Id. at 11-13).

The Defendants have produced a total of approximately 500 emails to the Plaintiffs in several tranches. (9/8/08 Tr. at 3). They further have represented that their search for and production of responsive emails is largely complete. (Id.). As part of their production on July 31, 2008, the Defendants turned over approximately 200 emails, “approximately three-quarters” of which were harvested in a manner that preserved the metadata.⁶ (Id. at 11). During the September 8th conference, the Defendants offered to re-produce the emails in that tranche to the extent that they were turned over the OPLA with the original metadata intact. (See id. at 23). They declined to undertake a review of the remaining emails to determine if metadata was available, citing the potential burden. (Id. at 25).

The Plaintiffs, not surprisingly, want the Defendants to engage in a more exhaustive search, which presumably would require that the emails previously forwarded to OPLA be retransmitted in a form that preserves the original metadata and, perhaps, that back-up tapes be restored and reviewed to determine if ICE still has any emails that were deleted by the senders or recipients but are responsive to the Plaintiffs’ requests. Had the

⁶ It is unclear whether the emails were transmitted as .pst or .msg files, or both. In .pst format, the Plaintiffs may be able to determine the file folder where each email was stored. See Ralph Losey, MSG Is Bad for You, <http://floridalawfirm.com/msg.html> (last visited Nov. 13, 2008). This information obviously would be relevant if, for example, an ICE official saved an email to a folder labeled “Illegal Searches.” An .msg file apparently does not contain metadata indicating the folder, if any, to which the email was saved. (Id.).

Plaintiffs made a formal request for metadata before the process of harvesting the emails was largely complete, I certainly would have entertained at least the first of these requests. The Plaintiffs, however, have delayed for a considerable period of time. Moreover, it is far from clear that even the most exhaustive retracing of steps would yield useful information beyond that which the Plaintiffs already have. For example, there has been no showing that any significant “bcc’s” exist. There similarly has been no showing that any additional file folder information would be of real value. Finally, because the universe of emails is apparently only 500 or so, this is not a case in which the Plaintiffs need the metadata in order to be able to manage the ESI that they have received. For these reasons, I will not require that the Defendants search for email metadata in any files sent by persons who originally complied with their discovery obligations by forwarding responsive emails to OPLA, rather than transmitting them as .msg or .pst files.

Turning to the issue of back-up tapes, Rule 26(b)(2)(B) of the Federal Rules of Civil Procedure establishes a two-tiered approach to the discovery of potentially burdensome ESI. In the first instance, a producing party “need not provide discovery of [ESI] from sources that the party identifies as not reasonably accessible because of undue burden or cost.” Fed. R. Civ. P. 26(b)(2)(B). If the requesting party then seeks to compel that discovery, the court must consider the bona fides of the producing party’s representations, as well as the considerations outlined in Rule 26(b)(2)(C), such as the degree to which the discovery sought is “unreasonably cumulative or duplicative” and

whether “the burden or expense of the proposed discovery outweighs its likely benefit.” Here, the Defendants made a persuasive showing during the September 8th conference that the process of restoring and reviewing the back-up tapes generated by its decentralized email system would be extremely burdensome. (See 9/8/08 Tr. at 15-18, 25). Moreover, the Plaintiffs have not shown that there is a likelihood of recovering important information not previously disclosed. Accordingly, because the cost of this additional discovery is unquestionably high and the likely benefit low, the Defendants will not be required to review and produce any data regarding emails in ICE’s back-up tapes.

2. Word Processing Documents and PowerPoint Presentations

_____The Plaintiffs seek system metadata, including the date created, date modified, and modified by fields, for all Word, Excel and PowerPoint documents. (Gordon Letter at 6-7). The Plaintiffs make two arguments why the production of this metadata should be required. First, they claim that they cannot efficiently search the documents without the metadata. Second, they argue that the metadata is relevant so that they can piece together “who knew what when.” (Id. at 5). In particular, the Plaintiffs claim that this information will assist them in demonstrating that ICE employees (a) were inadequately trained and supervised, (b) engaged in or condoned a pattern or practice of unconstitutional home searches, or (c) lacked probable cause for the searches. (Id.).

Although it undoubtedly is true that Word and PowerPoint documents could be searched more easily with metadata, the Plaintiffs have been provided with text-searchable PDFs. More importantly, this is not a case involving hundreds of boxes of documents in which the Plaintiffs' concern about their ability to search and sort the documents would carry some force. Rather, only about 5,200 pages of documents have thus far been produced, (id. at 3), the equivalent of slightly more than one "banker's box" of documents. The Defendants also have represented that few additional documents remain to be produced. See id. at 8. Given the limited universe of documents, the Plaintiffs should not encounter significant difficulty sorting and searching the word processing and PowerPoint files they have received.

_____Turning to the relevance argument, the Plaintiffs have failed to show that the "who" and "when" of document creation or modification is relevant to their Bivens claims that the home searches resulted in a deprivation of their constitutional rights. As an example of why metadata allegedly is needed, the Plaintiffs cite an undated ICE "Enforcement Operation Plan" and assert that by learning who created and modified the document, and when, they will gain "insight into the way the Defendants planned and executed" the home searches. (Id. at 7). Specifically, they maintain that they will be able to learn if the dates of the planned operation, the daily schedule, the list of persons targeted, or the arrest teams changed, and who made the changes to the documents. (Gordon Letter at 7). As the Defendants point out, however, the Operation Plan does not

contain any list of targets or residential locations. (Cargo Letter at 4). For this reason, whether the document was modified before or after the operation is not relevant to the issue of which persons were targeted. (See id. Ex. I). More importantly, the information the Plaintiffs seek would be of limited relevance to the fundamental question posed by their Bivens claim: whether their constitutional rights were violated because the ICE agents did not seek their consent before entering their homes.

If establishing probable cause for the searches (or the lack thereof) were important, there might be a need to know what information each officer had learned by the time of a search. However, the Defendants have conceded that they lacked probable cause for the searches and that they engaged in the searches solely in the belief that they had obtained the voluntary consent of the occupants of the homes they searched. (Cargo Letter at 2). This admission largely eviscerates any need to consider whether the officers had probable cause. Accordingly, the Plaintiffs' justification for the production of metadata to determine probable cause is unpersuasive.

With probable cause generally out of the picture, "who knew what when" is, at best, only marginally relevant. It could conceivably be of some assistance to the extent that the Plaintiffs' are seeking to impose supervisory liability on certain defendants for engaging in or condoning a pattern of unconstitutional searches. For example, if an operation plan describes an unconstitutional search protocol, it obviously would be helpful to know if that plan was created, reviewed, or modified by a given supervisor.

But this information does not go to the core of the Bivens claim, which is whether a particular search violated a specific plaintiff's constitutional rights. In these circumstances, the potential probative value of the metadata is likely outweighed by the burden that would be imposed on the Defendants if they were required to make a second production in response to the Plaintiffs' delayed request.

In sum, the metadata sought by the Plaintiffs with respect to this category of documents is, at best, marginally relevant, the additional searchability that they might gain through access to the metadata is not critical to their pretrial preparation, and the Plaintiffs did not request metadata until after the Defendants had already completed most of their document collection. All of these factors augur against requiring the production of metadata for the Defendants' word processing and PowerPoint files. Nevertheless, because the metadata could potentially have some relevance and increase the utility of the documents (and in light of the new emphasis in the Sedona Principles regarding the need to produce metadata), I will grant the Plaintiffs' motion to compel the production of metadata for any Word and PowerPoint documents, but on the condition that the Plaintiffs pay all costs associated with a second production of these documents. See Fed. R. Civ. P. 34(b)(2)(E)(iii) (party need not produce the same ESI in more than one form); Sedona Principles 2d Cmt. 12d ("If a court requires production of the same information a second time in a different format because of an unclear or tardy request, the court should consider shifting some or all of the cost of the second production to the requesting party.").

Because the Plaintiffs will bear the cost of the production, they may wish to reexamine whether they, in fact, need this metadata and, if so, to what extent.

3. Spreadsheets

As noted above, the relevance of metadata to an Excel spreadsheet depends upon its complexity and purpose. When a spreadsheet relies on mathematical formulas, the metadata that discloses those formulas often is necessary for a thorough understanding of the spreadsheet. Here, however, the spreadsheets cited by the Plaintiffs are nothing more than lists that could have been created through a word processing program. (See Gordon Letter Ex. A (Excels)). The spreadsheets merely list the date of a particular operation, the field office that conducted the operation, the number of arrests made, and a breakdown of those arrests into different categories. (Id.). While the spreadsheets appear to contain some embedded metadata that computes the total number of arrests in each category, the underlying formulas are not necessary to understand the spreadsheet. Moreover, absent some preliminary showing that these spreadsheets were fraudulently modified after the fact to conceal the true scope of the operation, the date the spreadsheets were created or modified is not relevant to the any claims or defenses in this action.

That said, the request for spreadsheet metadata evidently is not unduly burdensome. Indeed, during the September 8, 2008 conference, the Defendants expressed a willingness to re-produce the spreadsheets in native format. For this reason, the Court will so direct.

4. Hierarchical Databases

The Plaintiffs also seek the metadata associated with certain documents that the Defendants have produced from hierarchical databases. There appear to be three such databases: SEACATS, TECS, and DACS.

Based on the parties' letters and the discussion at the September 8th conference, it appears that SEACATS is a module within the TECS system. (9/8/08 Tr. at 36). An ICE agent using SEACATS (or a companion module) can prepare a "subject record" relating to a person or thing (such as a boat) or an "incident report" which becomes part of the TECS database. (Id.). The agent, however, typically prints out the relevant screens, which are kept in hard copy form at the applicable ICE office. (Id. at 37). To the extent that SEACATS information was sought as part of the Plaintiffs' discovery requests, the Defendants produced hard copies. (Id.). Although audit trail information is available and would enable the Plaintiffs to determine what changes were made and when, the Defendants did not produce this electronic information. (Id. at 37-39).

The TECS system is the United States Treasury Department's Treasury Enforcement Communications System. (Id. at 36). Approximately forty federal agencies – including such sensitive agencies as the Secret Service – use the TECS system in real-time mode. (Id. at 44). The TECS database also contains Trade Secrecy Act and Bank Secrecy Act information. (Id.). In order to determine what TECS metadata may be

relevant to their claims, the Plaintiffs seek information regarding the system's architecture, such as the identification of table and field lists. (Id. at 45). The Defendants, citing law enforcement privilege, oppose disclosing that information to the Plaintiffs or their counsel.⁷ (Id. at 43). The Defendants note, however, that they may be able to query the TECS system as of an earlier date, such as the day before this lawsuit was filed, to allay the Plaintiffs' fears that data may have been altered after the Defendants learned a suit would be filed. (Id.). The Defendants also have represented that for both SEACATS and TECS, all of the data is "displayed on the face of the subject record" so that "no [relevant] information exists 'behind' what is displayed in the[] printouts" they have produced. (Cargo Letter at 6).

The third, unrelated database is DACS, which apparently is an acronym for ICE's Deportable Alien Control System. See Office of the Inspector General, Department of Homeland Security, Review of U.S. Immigration and Customs Enforcement's Detainee Tracking Process (2006), http://www.dhs.gov/xoig/assets/mgmtrpts/OIG_07-08_Nov06.pdf. The DACS database is used by ICE fugitive operations officers to track the removal proceedings of illegal aliens who have been

⁷ "The purpose of [the law enforcement] privilege is to prevent disclosure of law enforcement techniques and procedures, to preserve the confidentiality of sources, to protect witness and law enforcement personnel, to safeguard the privacy of individuals involved in an investigation, and otherwise to prevent interference with an investigation." In re Dep't of Investigation of the City of N.Y., 856 F.2d 481, 484 (2d Cir. 1988). The privilege is qualified, not absolute, and requires the court to balance the public interest in nondisclosure against a private party's interest in gaining access to the information. Wells v. Connolly, No. 07 Civ. 1390 (BSJ)(DF), 2008 WL 4443940, at *2 (S.D.N.Y. Sept. 25, 2008).

apprehended and served with a charging document. (9/8/08 Tr. at 28). When a record is created for an alien, the fugitive operations officer prints a hard copy and places it in a green file folder. (Id. at 28-29). The Defendants produced the contents of these folders in discovery. (Id. at 29). The DACS database can be queried to determine what changes to a record were made and when, but not by whom.⁸ (Id. at 35, 56). Nonetheless, the Plaintiffs suggest that they might be able to infer who made changes by comparing when changes were made to who was signed on to the system at the time. (Id. at 56).

Since the fugitive operations officers apparently used the green folders as their basic resource, the DACS metadata would seem to be of limited usefulness, particularly in light of the Defendants' concession that they lacked probable cause to search any of the targets' homes. Moreover, there has been no showing that the information in the file folders does not accurately reflect the state of the fugitive operations officers' knowledge at the time the searches occurred. Accordingly, the request for metadata concerning DACS is denied.

With respect to SEACATS and TECS, in the absence of any disclosures concerning the architecture of these databases, the Plaintiffs continue to express concern that they may not receive all of the information to which they are entitled under the Federal Rules. In an effort to address this concern, I suggested at the September 8th conference that the Defendants conduct a live demonstration of the TECS system for the

⁸ This additional information is available through a recent enhancement which does not relate back to the time period at issue in this suit. (Id. at 35-36).

Plaintiffs. For security reasons, it was agreed that ICE would do this using a training environment which the Defendants have represented is identical to the live production environment except for the use of dummy data. (Id. at 45-51). Although I indicated that the demonstration should take place “quickly,” the parties recently advised me that it has yet to occur because of “unforeseeable issues related to the retention of [the Plaintiffs’] consulting expert.” (See joint letter to the Court, dated Nov. 5, 2008 (stating that a demonstration previously scheduled for early November had been cancelled)). Whatever difficulties the parties may have encountered over the past two months do not justify further delay in the demonstration of the TECS system.

III. Conclusion

This lawsuit demonstrates why it is so important that parties fully discuss their ESI early in the evolution of a case. Had that been done, the Defendants might not have opposed the Plaintiffs’ requests for certain metadata. Moreover, the parties might have been able to work out many, if not all, of their differences without court involvement or additional expense, thereby furthering the “just, speedy, and inexpensive determination” of this case. See Fed. R. Civ. P. 1. Instead, these proceedings have now been bogged down in expensive and time-consuming litigation of electronic discovery issues only tangentially related to the underlying merits of the Plaintiffs’ Bivens claims. Hopefully, as counsel in future cases become more knowledgeable about ESI issues, the frequency of such skirmishes will diminish.

In this case, for the reasons set forth above, the Plaintiffs' application to compel the production of metadata is granted in part and denied in part. More specifically, the Defendants are directed to produce (a) any emails that OPLA received with metadata attached in a form that contains that metadata, and (b) the metadata for their spreadsheets. The Defendants further are directed to produce the metadata for their word processing and PowerPoint files if the Plaintiffs agree to pay the costs associated with producing those files a second time. All such materials shall be produced to the Plaintiffs by December 12, 2008.

Additionally, the Plaintiffs are directed to arrange for the TECS system to be demonstrated for their counsel using the training environment by no later than December 12, 2008. While I recognize that it would be useful for an expert to be present to guide counsel's inquiry, the failure to retain an appropriate expert (or the expert's unavailability) will not warrant any further extension of this deadline.

Finally, if the Plaintiffs seek further metadata regarding TECS or SEACATS after that demonstration has been conducted, they are to submit a letter

explaining and justifying their request within ten days. The Defendants then shall have ten days to respond to that letter.

SO ORDERED

Dated: New York, New York
November 20, 2008



FRANK MAAS
United States Magistrate Judge

Copies to:

Honorable John G. Koeltl
United States District Judge

Donna L. Gordon, Esq.
Dewey & LeBoeuf, LLP
Fax: (212) 649-0901

Shane Cargo, Esq.
United States Attorney's Office
Fax: (212) 637-2786