

**UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK**

-----X
**NATIONAL DAY LABORER
ORGANIZING NETWORK, *et al.*,**

Plaintiffs,

- against -

**UNITED STATES IMMIGRATION AND
CUSTOMS ENFORCEMENT AGENCY, *et
al.*,**

Defendants.
-----X

OPINION AND ORDER

10 Civ. 3488 (SAS)

SHIRA A. SCHEINDLIN, U.S.D.J.:

I. INTRODUCTION

Plaintiffs brought this action for the purpose of obtaining records, pursuant to the Freedom of Information Act (“FOIA”), from four government agencies (collectively, “Defendants”). Specifically, the requests pertain to Secure Communities, a collaborative program established by the United States Immigration and Customs Enforcement Agency (“ICE”) and the Department of Justice (“DOJ”) that enlists states and localities in the enforcement of federal immigration law.¹ A dispute has now arisen regarding the format in which the

¹ Secure Communities now operates in thirty-eight states, and is expected to operate nationwide by 2013. The Secure Communities program must be activated at the state level through a signed agreement with ICE known as the Secure Communities Memorandum of Agreement (“MOA”). It is unclear whether participation in Secure Communities is mandatory for localities once their state has

Defendants have produced records to Plaintiffs, and will be required to produce records to Plaintiffs in the future. Those records consist of electronic text records, e-mails, spreadsheets, and paper records. To set the stage, I note that generally speaking records can be produced in hard copy, static images (with or without load files) and native file format (with or without load files).

II. BACKGROUND

In February 2010, Plaintiffs submitted identical twenty-one page FOIA requests to each of the four defendant agencies.² Defendants claim that these requests would require production of millions of pages of responsive documents. Because the Plaintiffs received no substantive response to their requests, on April 27, 2010, they brought this suit to compel production of responsive records. After negotiating with the Government, Plaintiffs agreed to create a five-page Rapid Production List (“RPL”) identifying specific records that would be sought and hopefully produced on an expedited basis. The Government believes that even responses to the RPL will involve thousands of pages of records.

signed an MOA. A question exists as to whether local jurisdictions can “opt-out” of the program and prevent ICE from using their police records to identify persons subject to deportation. The primary purpose of Plaintiffs’ FOIA request is to answer this question.

² In addition to ICE, the agencies include the Department of Homeland Security, the Federal Bureau of Investigation, and the Office of Legal Counsel.

After further negotiations, the parties reached an agreement on July 7, 2010, regarding production responsive to the RPL. In substance, the Defendants agreed to produce “the bulk of responsive, non-exempt materials by Friday, July 30.”³ The agreement also provided that if the Defendants identified responsive, non-exempt materials that could not be produced by that date, they would provide Plaintiffs with a description of such materials by July 26, and would propose an alternative date for their production. Defendants failed to produce any records by the agreed-upon July 30 date, but nearly two thousand pages of records were produced on August 3, August 13, September 8, and October 22, 2010.⁴ These productions did not satisfy the July 7 agreement.

On October 22, 2010, Plaintiffs moved for a preliminary injunction to compel production for five categories of the RPL documents that had not been produced. Specifically, Plaintiffs asked the Court to order (1) that Opt-Out records – defined in the RPL as “National policy memoranda, legal memoranda or communications relating to the ability of states or localities to opt-out or limit their participation in [the program]” – be produced within five days; (2) that Defendants

³ 7/9/10 Letter from Assistant United States Attorney (“AUSA”) Christopher Connolly to Bridget Kessler, Plaintiffs’ Counsel.

⁴ Three of these productions were from ICE. The August 13 production was from the FBI.

provide a *Vaughn* index within ten days;⁵ and (3) that an expedited briefing schedule be set for contested exemptions. The motion was resolved at a conference held on December 9, 2010, with an order requiring Defendants to provide the Opt-Out Records by January 17, 2011.⁶

On December 22, 2010, Plaintiffs sent the Government a Proposed Protocol Governing the Production of Records (“Proposed Protocol”). This proposal, annexed hereto as Exhibit A, sets forth a requested format for the production of electronic records and a separate requested format for the production of paper records. As Plaintiffs note, the Proposed Protocol is based, in part, on the format demands routinely made by two government entities – the Securities and Exchange Commission and the Department of Justice Criminal Division.

In advance of a court conference scheduled for January 12, 2011, Defendants produced five PDF files totaling less than three thousand pages. Upon receipt of these files, Plaintiffs again sought assistance from the Court, asserting

⁵ The term “*Vaughn* index” arises from *Vaughn v. Rosen*, 484 F.2d 820 (D.C. Cir. 1973), which held that a government agency must describe how it searched for records, as well as its rationale for any claimed FOIA exemptions, in order to satisfy its statutory obligations under FOIA.

⁶ See December 17, 2010 Order [docket # 25].

that the form in which these records were produced was unusable.⁷ Plaintiffs made three specific complaints: (1) the data was produced in an unsearchable PDF format; (2) electronic records were stripped of all metadata; and (3) paper and electronic records were indiscriminately merged together in one PDF file. Plaintiffs asked the Court to “so order” the Proposed Protocol.⁸ In response, the Government submitted a letter defending its form of production.⁹ An oral argument on this issue was held on January 12, 2011.¹⁰

Before turning to a discussion of the issues raised by this dispute, it is important to describe what the parties did and did not do in an effort to negotiate an agreed upon form of production. As far as I can tell from the record submitted by the parties, the equivalent of a Rule 26(f) conference, at which the parties are required to discuss form of production, was not held and no agreement regarding form of production was ever reached. Nor was a dispute regarding form of production brought to the Court for resolution. The Proposed Protocol was first

⁷ See 1/6/11 Letter from Norman Cerullo, Plaintiffs’ Counsel, to the Court.

⁸ See Proposed Protocol, Exhibit C to 1/6/11 Letter.

⁹ See 1/11/11 Letter from AUSAs Joseph N. Cordaro and Christopher Connolly, Defendants’ Counsel, to the Court.

¹⁰ See 1/12/11 Transcript of Proceedings (“Tr.”).

provided to Defendants on December 22, 2010, and also was the first time Plaintiffs made a written demand for load files and metadata fields.¹¹ Prior to December 22, the only written specification of form of production was a July 23 e-mail from Bridget Kessler, Plaintiffs' counsel, to AUSA Connolly, Defendants' counsel. Given its importance and brevity, I quote the full text of this e-mail:

We would appreciate if you could let us know as soon as possible how ICE plans to produce the Rapid Production List to plaintiffs. To facilitate review of the documents between several offices, please (1) produce the responsive records on a CD and, if possible, as an attachment to an email; (2) save each document on the CD as a separate file; (3) provide excel documents in excel file format and not as PDF screen shots; and (4) produce all documents with consecutively numbered bate [sic] stamps. . . . Thank you for your help and if you have any questions or concerns, please feel free to call me.

It is undisputed that Defendants' counsel did not respond to the e-mail by raising any questions or concerns. Defendants do not deny that the records that have been produced, including but not limited to spreadsheets, are in an unsearchable PDF format with no metadata.

III. APPLICABLE LAW

A. FOIA and the Federal Rules of Civil Procedure

¹¹ The parties dispute whether this was the first written demand for production in single page format with respect to text documents and native format production with respect to spreadsheets.

FOIA provides that “[i]n making any record available to a person under this paragraph, an agency shall provide the record in any form or format requested by the person if the record is readily reproducible by the agency in that form or format.”¹²

While Congress has recognized the need for “Government agencies [to] use new technology to enhance public access to agency records and information,” there is surprisingly little case law defining this standard.¹³ The leading case, *Sample v. Bureau of Prisons*, provides the following guidance:

Under any reading of the statute, however, “readily reproducible” simply refers to an agency’s technical capability to create the records in a particular format. No case construing the language focuses on the characteristics of the requester. *See, e.g., TPS, Inc. v. U.S. Dep’t of Defense*, 330 F.3d 1191, 1195 (9th Cir. 2003) (interpreting “readily reproducible” as referring to technical capability); *see also, e.g., Carlson v. U.S. Postal Serv.*, 2005 WL 756573, at *7 (N.D. Cal. 2005) (holding that “readily reproducible” in a requested format means “readily accessible” by the agency in that format); *Landmark Legal Found. v. EPA*, 272 F. Supp. 2d 59, 63 (D.D.C. 2003 (construing “readily reproducible” as the ability to duplicate)).¹⁴

¹² 5 U.S.C. § 552(a)(3)(B) (effective Oct. 1, 1997).

¹³ Electronic Freedom of Information Act Amendments of 1996, Pub. L. No. 104-231, § 2(a)(6), 110 Stat. 3050 (1996).

¹⁴ 466 F.3d 1086 (D.C. Cir. 2006) (holding that Bureau of Prisons had the obligation to produce records electronically when requested to do so). *Accord TPS*, 330 F.3d at 1195 (“[A] FOIA request *must be processed in a requested format* if ‘the capability exists to respond to the request’ [in that format].” (citing

Rule 34 of the Federal Rules of Civil Procedure also addresses the form of production of records, albeit in the context of discovery. The Rule is divided into a series of steps that are intended to facilitate production in a useful format. *First*, the requesting party may specify the form of production of electronically stored information (“ESI”).¹⁵ *Second*, the responding party may object to the specified form; if it does so, it must state the form that it intends to use.¹⁶ If the requesting party disagrees with the counter-proposal, the parties must attempt to resolve the disagreement. If they cannot, the requesting party may make a motion to compel production in the requested form. *Third*, if the requesting party has not specified a form of production, the responding party must state the form that it intends to use.¹⁷ The responding party may select the form in which the material “is ordinarily maintained,” or in a “reasonably usable form.”¹⁸ The Advisory Committee Note to Rule 34 states that the responding party’s “option to

32 C.F.R. § 286.4(g)(2) (emphasis added)).

¹⁵ See Fed. R. Civ. P. 34(b)(1)(C).

¹⁶ See Fed. R. Civ. P. 34(b)(2)(D).

¹⁷ See *id.*

¹⁸ Fed. R. Civ. P. 34(b)(2)(E)(ii). See also The Sedona Principles: Best Practices Recommendations and Principles for Addressing Electronic Document Production (2d ed. 2007) (“Sedona Principles 2d”), Principle 12 (“[P]roduction should be made in the form or forms in which the information is ordinarily maintained or in a reasonably usable form . . .”).

produce [ESI] in a reasonably usable form does not mean that [it] is free to convert [ESI] 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.”¹⁹ Finally, the Advisory Committee Note also states that 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.”²⁰

B. Case Law

1. Metadata and Load Files

In *Aguilar v. Immigration and Customs Enforcement Division of the United States Department of Homeland Security*, Magistrate Judge Frank Maas, of this District, provided a guidebook that explained the various types of metadata and the relationship between a record and its metadata.²¹ In that opinion, Judge Maas noted that in the second edition of the Sedona Principles, the Conference abandoned an earlier presumption against the production of metadata in recognition of “the need to produce reasonably accessible metadata that will enable the receiving party to have the same ability to access, search, and display

¹⁹ Fed. R. Civ. P. 34(b), 2006 Advisory Committee Note.

²⁰ *Id.*

²¹ 255 F.R.D. 350 (S.D.N.Y. 2008).

the information as the producing party”²² By now, it is well accepted, if not indisputable, that metadata is generally considered to be an integral part of an electronic record.²³

The *Aguilar* decision also explained the term load file, quoting from The Sedona Conference Glossary. The 2010 version of the Glossary now defines the term as follows:

A file that relates to a set of scanned images of electronically processed files, and indicates where individual pages or files belong together as documents, to include attachments, and where each document begins and ends. A load file may also contain data relevant to the individual documents, such as selected metadata, coded data, and extracted texts. Load files should be obtained and provided in prearranged or standardized formats to ensure transfer of accurate and usable images and data.²⁴

²² *Id.* at 356 (quoting Sedona Principles 2d Principle 12).

²³ *See, e.g., Williams v. Sprint/United Mgmt. Co.*, 230 F.R.D. 640, 652 (D. Kan. 2005) (holding 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”). *See generally* W. Lawrence Wescott, *The Increasing Importance of Metadata in Electronic Discovery*, 14 Rich J.L. & Tech. 10 (2008).

²⁴ The Sedona Conference® Glossary: E-Discovery & Digital Information Management (3d ed. Sept. 2010), at 31. *See also* Sedona Principles 2d Principle 12, cmt 12(b) (“In an effort to replicate the usefulness of native files while retaining the advantage of static productions, image format productions are typically accompanied by ‘load files,’ which are ancillary files that may contain textual content and relevant system metadata.”).

Once again, it is by now well accepted that when a collection of static images are produced, load files must also be produced in order to make the production searchable and therefore reasonably usable.²⁵

2. FOIA and Metadata

No federal court has yet recognized that metadata is part of a public record as defined in FOIA. However, this precise issue has been addressed by several state courts, which have uniformly held, in the context of state freedom of information laws, that metadata is indeed a part of public records and must be disclosed pursuant to a request for public records.²⁶ In his January 6, 2011 letter,

²⁵ The question arises as to the expense of creating load files. While I cannot predict the exact cost of creating such files in every case, the fact is that a significant collection of static TIFF images is not reasonably usable without load files. A party can generally avoid the expense of creating load files by producing the records in native format. However, some metadata is *not* embedded in the native file and so will not necessarily travel with the native file. For this metadata, load files might still be necessary even with a native production. These files include: Identifier, File Name, Custodian, Source Device, Source Path, Production Path, Modified Date, Modified Time, and Time Offset Value (allowing documents and messages to be properly sorted temporally). All of these terms are defined *infra* Part IV.B at 20-22.

²⁶ See *Irwin v. Onondaga Cnty. Res. Recovery Agency*, 895 N.Y.S.2d 262, 319 (4th Dep't 2010) (finding that petitioner's request for "all computer records that are associated with published [photographs]" included a demand for the metadata associated with those images, and that the metadata should have been disclosed pursuant to New York's Freedom of Information Law); *O'Neill v. City of Shoreline*, 240 P.3d 1149, 1152 (Wash. 2010) (holding, in a manner of first impression, that metadata associated with an e-mail sent to a public official constituted a public record subject to disclosure under Washington's Public

Plaintiffs’ counsel does cite one FOIA case recognizing that production in a medium which detrimentally affects the access to the information sought is inappropriate because it could improperly “reduce the quantum of information made available.”²⁷

IV. DISCUSSION

A. August - October 2010 and January 2011 Productions

The Government defends the format of the productions to date based on its claim that Plaintiffs failed to make a timely request for metadata. Placing heavy reliance on *Aguilar*, Defendants quote the following language: “[I]f a party wants metadata, it should ‘Ask for it. Up front. Otherwise, if [the party] asks[s] too late or ha[s] already received the document in another form, [it] may be out of

Records Act); *Lake v. City of Phoenix*, 218 P.3d 1004, 1007-08 (Ariz. 2009) (“[T]he metadata in an electronic document is part of the underlying document; it does not stand on its own. When a public officer uses a computer to make a public record, the metadata forms part of the document as much as the words on the page.”).

²⁷ 1/6/11 Letter (quoting *Dismukes v. Department of Interior*, 603 F. Supp. 760, 762 (D.D.C. 1984) (holding that production of microfiche instead of computer tape could violate FOIA if production of microfiche “affects [the] access to the information [sought]”). *Accord Armstrong v. Executive Office of the President*, 1 F.3d 1274, 1280 (D.C. Cir. 1993) (providing paper print-outs of electronic documents violated the Federal Records Act because “essential transmittal information relevant to a fuller understanding of the context and import of the electronic communication will simply vanish”).

luck.’’²⁸ Given Plaintiffs’ July 23 e-mail and Defendants’ tardy productions, I cannot accept this lame excuse for failing to produce the records in a *usable format*.

First, the language of Plaintiffs’ July 23 e-mail, while less than crystal clear, was sufficient to put Defendants on notice of certain requests regarding form of production. Defendants were asked to “(2) save each document on the CD as a separate file; (3) provide excel documents in excel file format and not as PDF screen shots; and (4) produce all documents with consecutively numbered bates [sic] stamps.” These requests explicitly placed Defendants on notice that spreadsheets were sought in native format — not as a PDF screen shot — and that each text record should be produced as a separate file (*i.e.*, in single file format). In addition, the request for consecutively numbered Bates stamping also put Defendants on notice of the need for single file format. *Second*, and of equal if not greater importance, Plaintiffs asked the Government to “let us know as soon as possible how ICE plans to produce the Rapid Production List.” Had the Government done as it was asked, any ambiguity as to the nature of the requested format would have been resolved. Finally, Plaintiffs wrote “if you have any questions or concerns, please feel free to call me.” This invitation was ignored.

²⁸ 1/11/11 Letter (quoting *Aguilar*, 255 F.R.D. at 357 (quoting Levitt & Farrell, *Taming the Metadata Beast*, N.Y.L.J. May 16, 2008, at 4)).

Defendants violated the explicit requests of the July 23 e-mail by producing all of the records in non-searchable PDF format, merging all records without indicating any separate files, merging paper with electronic records, and failing to produce e-mails with attachments. They also violated the Federal Rules of Civil Procedure (the “Rules”) by failing to produce the records in a reasonably usable form, and by producing the records in a form that makes it difficult or burdensome for the requesting party to use the information efficiently.

The Government argues that metadata is substantive information that must be explicitly requested and then reviewed by an agency for possible exemptions.²⁹ Because there is no controlling FOIA precedent recognizing that metadata is an integral part of the electronic record that must be produced when an electronic record is requested, the Government asserts that it complied with its FOIA obligations, even if it did not comply with the Rules.³⁰ To that end, the Government argues that if the requirements of FOIA and the requirements of the

²⁹ See Tr. at 9, 11.

³⁰ See *id.* at 13, 23-24. Specifically, the Government argues that FOIA places the burden on the requesting party to ask for metadata, rather than on the producing party to object to requests regarding the form of production. Moreover, the Government argues that its compliance with its FOIA obligations must be assessed by the adequacy of its *search* for responsive documents, rather than on the quantity and quality of its production. See *id.* at 26.

Rules conflict, FOIA must trump the Rules.³¹

However, there is no need to decide this question because FOIA does not conflict with the Rules. FOIA is silent with respect to form of production, requiring only that the record be provided in “any form or format requested by the person if the record is readily reproducible by the agency in that form or format.” There is no doubt in my mind that this language refers only to technical ability or, at most, reasonable accessibility. Defendants do not argue that they are *unable* to produce the records in the requested form — namely native format for spreadsheets and single file format for text records — but that reviewing all of the metadata would greatly increase the burden of search and production. To that extent, they have unwittingly argued that a request to produce *all* metadata would push the request into the second tier of Rule 26(b)(2)(B) because such records are not reasonably accessible based on undue burden and cost.³²

Nonetheless, the Government argues that FOIA is not synonymous with discovery in a civil litigation. It is a statute requiring the production of records to the public, upon request, subject to certain exemptions. While

³¹ See *id.* at 24.

³² In short, Defendants do not argue that Plaintiffs are not entitled to the metadata or that it is privileged. Rather, they stress that the review and production of that data will increase the time and expense of responding to the FOIA request. See Tr. at 11, 14-15.

rhetorically nuanced, this argument is unavailing. Regardless of whether FOIA requests are subject to the same rules governing discovery requests, Rule 34 surely should inform highly experienced litigators as to what is expected of them when making a document production in the twenty-first century.³³ As noted earlier, Defendants' productions to date have failed to comply with Rule 34 *or* with FOIA.

The next issue to address is the appropriate remedy. Because no metadata was specifically requested in Plaintiffs' July 23 e-mail, and because this is an issue of first impression, I will not require Defendants to re-produce all of the records with metadata. Moreover, while native format is often the best form of production, it is easy to see why it is not feasible where a significant amount of information must be redacted.³⁴ Therefore, Defendants are ordered to re-produce

³³ It is well-established that FOIA was not intended to supplant or supplement the discovery rules; as far as I can tell, however, courts have not addressed the reverse question of whether the discovery rules govern FOIA productions. Nonetheless, because the fundamental goal underlying both the statutory provisions and the Federal Rules is the same — *i.e.*, to facilitate the exchange of information in an expeditious and just manner — common sense dictates that parties incorporate the spirit, if not the letter, of the discovery rules in the course of FOIA litigation. Thus, attorneys should meet and confer throughout the process, and make every effort to agree as to the form in which responsive documents will ultimately be produced. In this context I note that Rule 26(f) specifically requires the parties to discuss “any issues about disclosure or discovery of electronically stored information, including the form or forms in which it should be produced.”

³⁴ In this case, for example, if the Government does not convert appropriately-redacted documents into static image format, there is concern that

all text records in static image single file format together with their attachments.

However, they must re-produce all spreadsheets in native format as requested by Plaintiffs' July 23 e-mail.³⁵ All records must be Bates stamped, which should also assist in the production of single file format.³⁶

That said, I now hold, consistent with the state court decisions cited

Plaintiffs could use the metadata embedded in natively-produced documents to circumvent certain FOIA exemptions. There is also concern that, even with a static image production – accompanied by load files containing metadata – the information redacted pursuant to FOIA exemption (b)(6) could be “reverse engineered.” Tr. at 11-12 (“... if the [government] took exemption (b)(6) for privacy purposes[,] [t]he question would be if we then produced the system metadata, would that be a way for the requester to get around the exemption by finding the information that we redacted in the first place based on personal information.”). See 5 U.S.C. § 552(b)(6) (exempting “personnel and medical files and similar files the disclosure of which would constitute a clearly unwarranted invasion of personal privacy” from FOIA productions). Faced with such a risk, both static images *and* the metadata would have to be redacted.

³⁵ See Tr. at 20 (ordering re-production “the way it was requested”). However, if the Government can demonstrate why native production of spreadsheets would inevitably reveal exempt information, it may produce them in TIFF format, but must include load files.

³⁶ Generally, when a court orders a “do-over” it gives serious consideration to cost shifting or cost sharing. See generally *Covad Comm’cs Co. v. Revonet, Inc.*, 254 F.R.D. 147 (D.D.C. 2008) (requiring parties to share costs of re-producing previously produced hard copies in electronic format as both parties were at fault). However, in this case, I conclude that Defendants failed to comply with Plaintiffs’ request in the July 23 e-mail and the explicit invitation to confer with Plaintiffs regarding any questions or concerns. Moreover, Defendants produced records in a format that was not reasonably usable. As a result, I decline to require Plaintiffs to bear any of the costs of a re-production.

earlier, that certain metadata is an integral or intrinsic part of an electronic record.³⁷ As a result, such metadata is “readily reproducible” in the FOIA context. The only remaining issue is which of the many types of metadata are an intrinsic part of an electronic record. Unfortunately, there is no ready answer to this question. The answer depends, in part, on the type of electronic record at issue (*i.e.*, text record, e-mail, or spreadsheet) and on how the agency maintains its records. Some agencies may maintain only a printed or imaged document as the final or official version of a record. Others retain all records in native format, which preserves much of the metadata. Electronic records may have migrated from one system to another, maintaining some metadata but not all. The best way I can answer the question is that metadata *maintained* by the agency *as a part of an electronic record* is *presumptively* producible under FOIA, unless the agency demonstrates that such metadata is not “readily reproducible.”

B. Future Productions

The Government argues that the Proposed Protocol should not be required for the January 17 production of the Opt-out Records because it was received *after* the Government had completed the great bulk of its search.³⁸ While

³⁷ See *supra* note 26.

³⁸ See Tr. at 30-31.

it surely had not reviewed all of the documents as of December 22, there is no reason to question the Government's claim that it began to search in earnest on December 10, following a court conference on December 9 which set a production deadline of January 17, and that it had completed the bulk of its collection efforts by December 22.³⁹ The Government notes that the form of production issue was not raised at the December 9 conference or at any time prior to December 22. Given this time line, the January 17 production shall be made (or re-made if already completed) in the same format that I have now required for the earlier productions.

I turn now to all future productions. Here, Plaintiffs ask that the bulk of the ESI be produced in TIFF image format but with corresponding load files, Bates stamping, and the preservation of "parent-child" relationships (*i.e.* the association between an attachment and its parent record). Plaintiffs also request twenty-four specific fields of metadata, which presumably will be the content of the load files. Finally, Plaintiffs request that spreadsheets be produced in both native format and TIFF format. Hard copy records are requested in single page TIFF image format with corresponding load files to provide ease of review.

The Government has not made a counterproposal in response to

³⁹ See *id.* at 27-28.

Plaintiffs' Proposed Protocol. Nonetheless, the Court will not impose any greater burden on the Defendants than is absolutely necessary to conduct an efficient review. After reviewing the Proposed Protocol as well as various sources discussing essential metadata fields,⁴⁰ I conclude that all future productions by Defendants must include load files that contain the following fields, which apply to *all forms of ESI*.⁴¹

1. **Identifier:** A unique production identifier ("UPI") of the item.⁴²
2. **File Name:** The original name of the item or file when collected from the source custodian or system.

⁴⁰ The Federal Judicial Center recently sponsored an E-Discovery Seminar for Federal Judges. At the session on Form of Production, held on September 28, 2010, the presenters suggested the fourteen fields of metadata were likely to be necessary in any production of ESI produced in a digital format. These fields, however, relate primarily to e-mail. *See* Form of Production PowerPoint slides prepared by Magistrate Judges James C. Francis and Frank Maas and attorney Maura Grossman (recommending the following fields: Bates_Begin; Bates_End; Attach_Begin; Attach_End; Sent_Date; Sent_Time; To; From; CC; BCC; Subject/Title; Text; Custodian; and Native_File). *Cf.* Craig Ball, *Beyond Data about Data: The Litigator's Guide to Metadata* (2005-2011), available at <http://www.craigball.com/metadataguide2011.pdf>, at 11 (setting forth six fundamental metadata fields).

⁴¹ While not necessary to the holding in this case, I believe that these are the minimum fields of metadata that should accompany *any* production of a significant collection of ESI. Requests for additional fields should be considered by courts on a case-by-case basis.

⁴² For native files, this might be a Bates number system either on the entire file or on each page within the file as feasible.

3. **Custodian:** The name of the custodian or source system from which the item was collected.
4. **Source Device:** The device from which the item was collected.
5. **Source Path:** The file path *from* the location from which the item was collected.
6. **Production Path:** The file path *to* the item produced from the production media.
7. **Modified Date:** The last modified date of the item when collected from the source custodian or system.
8. **Modified Time:** The last modified time of the item when collected from the source custodian or system.
9. **Time Offset Value:** The universal time⁴³ offset of the item's modified date and time based on the source system's time zone and daylight savings time settings.

The following additional fields shall accompany production of all e-mail messages:

1. **To:** Addressee(s) of the message.
2. **From:** The e-mail address of the person sending the message.

⁴³ *I.e.*, Coordinated Universal Time (“UTC”) or Greenwich Mean Time (“GMT”).

3. **CC:** Person(s) copied on the message.
4. **BCC:** Person(s) blind copied on the message.
5. **Date Sent:** Date the message was sent.
6. **Time Sent:** Time the message was sent.
7. **Subject:** Subject line of the message.
8. **Date Received:** Date the message was received.
9. **Time Received:** Time the message was received.
10. **Attachments:** The Bates number ranges of e-mail attachments. The parties may alternatively choose to use: Bates_Begin, Bates_End, Attach_Begin and Attach_End.

The following additional fields shall accompany images of paper records:

1. **Bates_Begin:** The beginning Bates number or UPI for the first page of the document.
2. **Bates_End:** The ending Bates number or UPI for the last page of the document.
3. **Attach_Begin:** The Bates number or UPI of the first page of the first attachment to the parent document; and
4. **Attach_End:** The Bates number or UPI of the last page of the last

attachment to the parent document.

In addition, Defendants must produce spreadsheets in native format, with accompanying load files if the required metadata is not preserved in the native file. However, unless Plaintiffs can demonstrate why it is also necessary to produce the spreadsheets in TIFF format, Defendants need not make such a production. Conversely, the Government may produce the spreadsheets in TIFF format with load files containing the applicable metadata fields, if it can demonstrate why native production of spreadsheets would inevitably reveal exempt information.

Although Plaintiffs requested certain additional fields – namely Parent Folder; File Size; File Extension; Record Type; Master_Date; and Author – I conclude that these fields need not be produced in this case. Except as noted in this Opinion, all of the other format requirements in the Proposed Protocol with respect to both ESI and hard copy are hereby ordered.⁴⁴

⁴⁴ To be clear, my Order requiring the use of this Proposed Protocol for future productions – as amended by the specific metadata fields I have required and by the options I have offered the parties regarding the form of production for spreadsheets – is limited to this case. I am certainly not suggesting that the Proposed Protocol should be used as a standard production protocol in all cases. The production of individual static images on a small scale, where no automated review platform is likely to be used, may be perfectly reasonable depending on the scope and nature of the litigation. While Rule 34 requires that records be produced in a reasonably usable format — which at a minimum requires searchability — any further production specifications are subject to negotiation by the parties on a case

One final note. Whether or not metadata has been specifically requested — which it should be — production of a collection of static images without any means of permitting the use of electronic search tools is an inappropriate downgrading of the ESI. That is why the Government’s previous production — namely, static images stripped of all metadata and lumped together without any indication of where a record begins and ends — was not an acceptable form of production. The Government would not tolerate such a production when it is a receiving party, and it should not be permitted to make such a production when it is a producing party. Thus, it is no longer acceptable for any party, including the Government, to produce a significant collection of static images of ESI without accompanying load files.⁴⁵

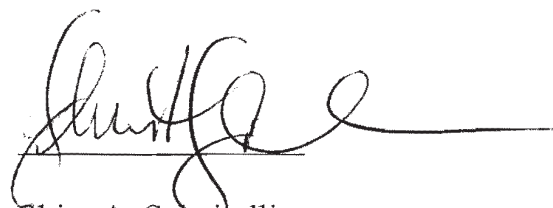
by case basis. If no agreement is reached, the court must determine the appropriate form of production, taking into account the principles of proportionality and considering both the needs of the requesting party and the burden imposed on the producing party.

⁴⁵ See *SEC v. Collins & Aikman Corp.*, 256 F.R.D. 403, 413 (S.D.N.Y. 2009) (holding that government agencies “must abide by the Federal Rules of Civil Procedure,” and are subject as are all parties to the requirements of Rule 34). A party often has the option to produce ESI in native format, which will reduce costs. But if a party chooses to produce a significant collection of TIFF images, it must assume that the receiving party will review those images on some sort of review platform — such as a Concordance database — which requires load files in order to be reasonably useable.

V. CONCLUSION

Once again, this Court is required to rule on an e-discovery issue that could have been avoided had the parties had the good sense to “meet and confer,” “cooperate” and generally make every effort to “communicate” as to the form in which ESI would be produced. The quoted words are found in opinion after opinion and yet lawyers fail to take the necessary steps to fulfill their obligations to each other and to the court. While certainly not rising to the level of a breach of an ethical obligation, such conduct certainly shows that all lawyers — even highly respected private lawyers, Government lawyers, and professors of law — need to make greater efforts to comply with the expectations that courts now demand of counsel with respect to expensive and time-consuming document production. Lawyers are all too ready to point the finger at the courts and the Rules for increasing the expense of litigation, but that expense could be greatly diminished if lawyers met their own obligations to ensure that document production is handled as expeditiously and inexpensively as possible. This can only be achieved through cooperation and communication.

SO ORDERED:

A handwritten signature in black ink, appearing to read 'Shira A. Scheindlin', with a long horizontal flourish extending to the right.

Shira A. Scheindlin
U.S.D.J.

Dated: New York, New York
February 7, 2011

- Appearances -

For Plaintiffs:

Bridget Kessler, Esq.
Immigration Justice Clinic
Benjamin N. Cardozo School of Law
55 Fifth Ave., Rm 1154
New York, New York 10003
(212) 790-0213

Norman Cerullo, Esq.
Anthony J. Diana, Esq.
Mayer Brown LLP
1675 Broadway
New York, New York 10019
(212) 506-2500

Sunita Patel, Esq.
Center for Constitutional Rights
666 Broadway, 7th Floor
New York, NY 10012
(212) 614-6439

For Defendants:

Joseph N. Cordaro
Christopher Connolly
Assistant U. S. Attorneys
86 Chambers Street, 3rd Floor
New York, NY 10004
(212) 637-2745/2761

EXHIBIT A

PROTOCOL GOVERNING THE PRODUCTION OF RECORDS

I. Production Formats of Electronic Records

Defendants agree that all responsive electronically stored information (“ESI”) shall be produced in the following formats:

- A. **TIFFs.** All images shall be delivered as single page Group IV TIFF image files. Image file names should not contain spaces.
- B. **Unique IDs.** Each image should have a unique file name and should be named with the Bates number assigned to it.
- C. **Text Files.** Extracted full text in the format of multipage .txt files shall be provided. The total number of text files delivered should match the total number of TIFF files delivered. Each text file should match the respective TIFF filename. Text from redacted pages will be produced in OCR format rather than extracted text.
- D. **Parent-Child Relationships.** Parent-child relationships (the association between an attachment and its parent record) should be preserved.
- E. **Database Load Files/Cross-Reference Files.** Records should be provided in a format compatible with Concordance 8x and Opticon 3x in the following format:

Example Concordance Delimited File

þBegDocþ þEndDoc þ_ þBegAttach þ_ þ EndAttach þ_ þ DocPages þ_ þ RecordType
þ_ þ MasterDate þ_ þ SentOn_ Date þ_ þ SentOne_ Time þ_ þ Recvd_ Time þ

þ ABC001 þ_ þ ABC002 þ_ þ ABC001 þ_ þ ABC005 þ_ þ 2 þ_ þ Email þ_ þ þ_ þ
01/01/2008 þ_ þ 13 05 GMT þ_ þ 13:08 GMT þ

þ ABC003 þ_ þ ABC005 þ_ þ ABC001 þ_ þ ABC005 þ_ þ 3 þ_ þ Attachment þ_ þ þ
_ þ þ_ þ þ_ þ

Example Opticon Delimited File

There should be one row in each load file per TIFF image. Files that are the first page of a record should contain a “Y” in the file where appropriate.

Format: ProductionNumber,VolumeLabel,ImagePath,DocBreak,
FolderBreak,BoxBreak,PageCount

Example: Record MS000001 – MS000003 and MS000004 – MS000005 on DVD
volume MS001 would be:

MS000001,MS001,D:\IMAGES\001\MS000001.TIF,Y,,,3
MS000002,MS001,D:\IMAGES\001\MS000002.TIF,,,,
MS000003,MS001,D:\IMAGES\001\MS000003.TIF,,,,
MS000004,MS001,D:\IMAGES\001\MS000004.TIF,Y,,,2
MS000005,MS001,D:\IMAGES\001\MS000005.TIF,,,,

F. **Metadata.** For records that were originally created using common, off-the-shelf software (e.g., Microsoft Word, Microsoft PowerPoint, Adobe PDF), Defendants will provide all metadata fields set forth in the below metadata fields. Defendants must produce all files attached to each email they produce, but only if such files are actually attached to that email in the ordinary course of business. To the extent a Defendant produces email attachments that were originally created using common, off-the-shelf software, a Defendant will produce the metadata for those attached electronic records in accordance with this section.

Metadata Fields

- Custodian
- Beginning Bates Number
- Ending Bates Number
- Beginning Attachment Number
- Ending Attachment Number
- Record Type
- Master_Date
- SentOn_Date and Time
- Received_Date and Time
- Create_Date and Time
- Last_Modified Date and Time
- Parent Folder
- Author
- To
- From
- CC
- BCC
- Subject/Title
- OriginalSource
- Native Path
- File Extension
- File Name
- File Size
- Full Text

G. **Spreadsheets.** For spreadsheets that were originally created using common, off-the-shelf software (e.g., Microsoft Excel), Defendants will produce the spreadsheets in native format and, in addition, in TIFF format.

II. Production Format of Hard Copy Records

Defendants agree that all responsive hard copy records shall be produced in the following formats:

- A. **TIFFs.** All images shall be delivered as single page Group IV TIFF image files. Image file names should not contain spaces.
- B. **Unique IDs.** Each image should have a unique file name and should be named with the Bates number assigned to it.
- C. **OCR.** High-quality multipage OCR text should be provided. Each text file should match the respective TIFF filename.
- D. **Database Load File/Cross-Reference Files.** Records should be provided in a format compatible with Concordance 8x and Opticon 3x in the formats identified in Section I.E above.
- E. **Unitizing of Records.** In scanning hard copy records, distinct records should not be merged into a single record, and single records should not be split into multiple records (*i.e.*, hard copy records should be logically unitized).
- F. **Parent-Child Relationships.** Parent-child relationships (the association between an attachment and its parent record) should be preserved.
- G. **Objective Coding Fields.** The following objective coding fields should be provided:
 - Beginning Bates Number
 - Ending Bates Number
 - Beginning Attachment Number
 - Ending Attachment Number
 - Source/Custodian
- H. **Objective Coding Format.** The objective coding fields should be provided in the following format:
 - Fields should be Pipe (|) delimited.
 - String values within the file should be enclosed with Carats (^).
 - Multiple entries in a field should have a semi-colon (;) delimiter.
 - The first line should contain metadata headers and below the first line there should be exactly only one line for each record.
 - Each field row must contain the same amount of fields as the header row.