

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

-----X	
NATIONAL DAY LABORER ORGANIZING	:
NETWORK; CENTER FOR CONSTITUTIONAL	:
RIGHTS; and IMMIGRATION JUSTICE	:
CLINIC OF THE BENJAMIN N. CARDOZO	:
SCHOOL OF LAW,	:
<i>Plaintiffs.</i>	:
	:
v.	:
	:
UNITED STATES IMMIGRATION	:
AND CUSTOMS ENFORCEMENT AGENCY;	:
UNITED STATES DEPARTMENT OF	:
HOMELAND SECURITY; EXECUTIVE OFFICE	:
FOR IMMIGRATION REVIEW; FEDERAL	:
BUREAU OF INVESTIGATION and OFFICE OF	:
LEGAL COUNSEL,	:
<i>Defendants.</i>	:
-----X	

ECF CASE
1:10-cv-3488 (SAS) (KNF)
[Rel. 10-CV-2705]

**DECLARATION OF
JASON A. NOVAK
IN SUPPORT OF
PLAINTIFFS’ OPPOSITION TO
DEFENDANTS’ MOTION
FOR A STAY**

I, Jason A. Novak, declare, pursuant to 28 U.S.C. § 1746 and subject to the penalties of perjury, that the following is true and correct:

Background

1. I, Jason A. Novak, am an employee with Stroz Friedberg, LLC (“Stroz Friedberg”). Mayer Brown LLP (“Mayer Brown”) has retained Stroz Friedberg on behalf of its client, the National Day Laborer Organizing Network. I have personal knowledge of the facts set forth below, and, if called upon to do so, could and would competently testify thereto.
2. Stroz Friedberg is a technical consulting and computer forensics firm that assists its clients, which include major law firms and Fortune 500 companies, in the areas of digital forensics, cybercrime response, and electronic discovery, among other things. I am a Digital Forensic Examiner in Stroz Friedberg’s New York office. My Curriculum Vitae (“CV”) is attached hereto as Exhibit 1.

3. I have been asked by Mayer Brown to provide information regarding electronic data, including information regarding metadata's relation to electronic files, document family relationships, and file production formats including both image formats and load file formats.

Files and Metadata

4. An electronic file is a logical unit for storing information on digital media. The format of any given electronic file is structured in a defined way to allow specific software applications to access the information stored in the file and present it to a user for review or modification. An electronic file is comprised of both the user-generated content in the file as well as metadata maintained about the file and its content.

5. Metadata is information about a file's characteristics and properties and can generally be defined as "data about data." A file's metadata may be stored within the file itself or stored externally to the file. The metadata available for a file depends on several factors, including the digital media on which the file was stored, the file system on the digital media, and the application or applications used to create, modify, and view the file. Metadata is generally created contemporaneously with a file's creation or its being written to media.

6. As implied by metadata's definition – "data about data" – metadata can refer to many characteristics or properties of a file; however, there are three distinct types of metadata related to electronic files: *file system metadata*; *embedded metadata*; and *email metadata*.

7. File systems are used to organize and store electronic files on digital media. File systems typically store metadata about the individual files within the file system, and this type of metadata is referred to as *file system metadata*. File system metadata typically includes the file's physical location on the media, the location of the file within the directory structure of the file

system, and timestamps such as when the file was first created on the file system, as well as when the file was last accessed and modified.

8. Some types of electronic files contain metadata stored within the files themselves. This type of metadata is referred to as application metadata or is more generally referred to as *embedded metadata*. Embedded metadata is created and modified by the application (or applications) used to create, modify, and view a file of a particular type.

9. Metadata within an email message, or *email metadata*, is a type of embedded metadata specific to email that, among other things, provides information regarding the sender and recipients of an email and data transmitted with an email, such as attachments.

10. All files do not contain the exact same embedded metadata, nor is there a standard set of metadata shared by all files. The embedded metadata available for any given electronic file will depend on the applications, and versions thereof, used to create and modify the file. Some metadata may be edited or removed by a user.

11. Elements of embedded metadata and file system metadata are often similarly named, but have different meanings and must be interpreted differently. For example, while the file system “Date Created” timestamp and the embedded Microsoft Word “Created” timestamp are similarly named, their contents are the results of different actions and are updated as a result of different events, as discussed below.

12. Some user actions affect both the file system metadata and the embedded metadata, and some types of user actions only affect one type of metadata. For example, when a file is copied from a computer to a removable hard drive using the “copy/paste” function of the Microsoft Windows XP operating system, the file system “Date Created” and “Date Accessed” of the copy on the removable hard drive will be updated to the date and time the file was copied because that

is the first time that particular copy of the file existed on the removable hard drive. However, the embedded metadata “Created” timestamp of the file on the removable hard drive would not be updated because it is internal to the file and is not affected by copying from one file system to another.

13. Email requires special consideration because, among other reasons, the format of email files is defined, in part, by an independent standards organization to ensure both interoperability between different email servers and email clients, as well as transport across the Internet. The internationally accepted format, called the RFC 2822 standard, specifies that for an email to be standards compliant and transmitted across the Internet from its sender to its intended recipient or recipients, the email must contain originating date metadata, e.g., the sent time, and the sender. Email may also include recipient (e.g., To, CC, BCC) and Subject metadata.

14. Embedded metadata and email metadata are intrinsic and integral parts of an electronic file. Embedded metadata is stored within the file itself. Email cannot be successfully transmitted without email metadata. Even file system metadata, which is physically stored externally to the file, is integral to a file, as it provides information regarding the dates and times at which a file was first used on a specific digital media. For those files that do not contain embedded metadata, file system metadata is the only source for information such as creation or modification date.

Electronic Document Families

15. Electronic files are not necessarily independent units; rather, a single file may contain within it multiple additional files. For example, a file may be a “compound” file where one file, in its entirety, may be embedded into a second file – e.g., a Microsoft Excel file embedded within a Microsoft Word file. Alternatively, a file may be an email that contains attachments in addition to the content of the email itself.

16. A file that contains other files is commonly referred to as the “parent,” and the files archived, embedded, or attached to the parent are typically referred to as “children.” A parent and its children are referred to, in aggregate, as a “document family.”

17. eDiscovery processing is designed to make electronic files more readily available and accessible for attorney review and production, while still maintaining the integrity and structure of the original file. For example, in the course of eDiscovery processing, document families are typically expanded so that each file within a document family is processed on its own.

Information regarding parent-child relationships is normally tracked in a database or index to maintain the original file’s structure. This allows for the parent and its children to be individually searched, reviewed, redacted, and/or produced, as necessary, while preserving the structure of the original file, i.e., the association of the parent with its children. Failure to maintain information about family relationships for review and production may result in obscuring critical information about a file, such as the file’s origin or the identity of individuals who may have seen or had access to the data stored in a file.

Format of Production

18. Generally, in the context of eDiscovery, the purpose of a “production” of files is to provide the requested information in a format that maintains the data and metadata of the files produced for the receiving party and makes it available for review. Productions may be made in a variety of formats because different file types may have different metadata fields, different review tools or platforms have different requirements for the loading of data, and different jurisdictions may have different standards for production.

19. Production formats may vary in several ways, including: whether files are produced in native format or in rendered image format or a combination thereof, what metadata fields are

produced in a load file, whether specific documents from document families are replaced with a “placeholder” in a production, and the method of the production of text or redacted text.

Formatting Files for Production

20. Files may be produced in their native format or as rendered images. Generally the production of files in native format results in their embedded metadata and document family structure being produced; however, native format files cannot be efficiently redacted. When files are rendered as images for production, the image may be redacted and embossed; however, the file’s original document family structure and metadata may not be included in the image.

21. Some files can only be viewed properly by the software application, or even the specific version of a software application, that was used to create the file. eDiscovery “productions” may involve rendering native electronic files to a commonly accessible image format because, among other reasons, not all parties in a matter may have access to all applications required to view each type of electronic file in a production. Files may also be rendered to image format for production because the parties may desire to redact certain information in the electronic files.

22. Rendering an electronic file to images presents a native file as a series of page images, as if the file were printed. The process of rendering electronic files to an image format typically preserves the file’s visible content and formatting, and renders any dynamic content in a static format. After a file has been rendered to an image format, additional processing may be performed on the images for production, such as applying redactions, embossing Bates numbers, and/or embossing special designations onto the images.

23. Depending on the process and software used to render an electronic file to image format, the image file may store all, part, or none of the source electronic file’s metadata, in addition to metadata regarding the creation of the image file. One common process is to capture the

metadata for a file prior to rendering it to an image format and to store this metadata in a database, producing it as called for in the production specifications.

24. A commonly used image format in eDiscovery productions is the Tagged Image File Format (“TIFF”). TIFF files can be either “single-page” or “multi-page.” Different review tools may require a specific TIFF format, and parties commonly agree in advance as to the format of the production of TIFF images depending on the review tools that each party plans to use.

Rendering of Spreadsheet File Types To Image Formats

25. Special consideration has to be given to the nuances of spreadsheet file formats prior to rendering and producing such files in an image format. There are many applications that will create spreadsheets, including Microsoft Excel, OpenOffice.org Calc, and Apple Numbers. In my experience, Microsoft Excel is the most common and widely used of these applications, and therefore I will use it as an illustrative example.

26. Microsoft Excel files are comprised of worksheets of rows and columns of cells. The contents of a single cell may be a static value, e.g., the number “3”, or a calculated value that is the result of a formula that is set to update as the content of other cells is changed. If a Microsoft Excel file is rendered to an image format, only the current value of the cell will be rendered, and, if a cell contains a formula, this will not be evident in the image representation of the file.

27. In addition, rendering Microsoft Excel files to an image format requires special consideration due to certain formatting features. Microsoft Excel files may contain “hidden” rows or columns that are not visible to a user or a “Print Area” whereby only part of the content of the worksheet may be set to be printed. In order for information in “hidden” rows or columns and information outside the “Print Area” to be rendered as part of the image, “hidden” rows and

columns have to be set to be “visible,” and the “Print Area” must be set to include all content on the worksheet.

28. Redacting a Microsoft Excel file also requires special consideration. Redacting a Microsoft Excel file natively may update the values of cells containing formulas. On the other hand, rendering Microsoft Excel files to an image format and redacting the image representation of the file may result in the underlying formulas not being produced for review.

Load Files

29. In eDiscovery, load files are used to facilitate the exchange of files between parties. As the name suggests, load files are used to load the produced files into a review tool or review platform. Load files may also include selected metadata elements that have been preserved for a file during the processing stage.

30. Load files can be used to produce one or more of the following: document break information, document family break information, metadata, and extracted – including redacted – text. Load files allow for this information to be exchanged in a common format and for this information to be provided even when the format for production might otherwise cause it to be lost. When the load file is loaded into a review tool, each file’s text, metadata and document family information is made available, even though the file itself is presented in a static format.

31. There are different types of load files with different formats; however, all load files should logically organize a collection of files in a production to a party. A load file may be a “document-level” load file whereby the load file contains one line per original electronic file. A “page-level” load file contains one line per page image produced. A page-level load file is typically produced in conjunction with a document-level load file to associate page images with documents and provide document break information. Each line in a load file may contain

multiple fields. The content and format of each field and line in a load file are typically agreed upon in advance by the exchanging parties.

32. In a document-level load file, document break information — i.e., information about where in a sequence of page images one document ends and another document begins — is typically provided through the use of fields containing the beginning and the ending Bates number of a file. If native files are produced as part of a production, these fields also can be used to associate a native file with its image rendering. Document-level load files typically indicate document families by recording the first Bates number of the parent in a document family and the last Bates number of the last child in a document family in specified fields. In a page-level load file, document break information is typically provided in a field that indicates if a page is the first page of a document; the source electronic file's metadata is typically provided in the document-level load file. Such document break and document family break information may also be provided for hard copy documents that have been scanned and are being produced as images.

33. In addition, fields in a document-level load file may contain the metadata of the source electronic files being produced. If image renderings of electronic files are being produced, metadata is most efficiently exchanged through a document-level load file, as the image representation of the files does not necessarily include the metadata of the original electronic files. Even if native files are being exchanged in a production, it is preferable to exchange metadata in document-level load files, as the metadata of native files can be easily updated by innocuous actions such as transferring a file from media to another.

Production of Text

34. Files contain text in a variety of different formats. So that text may be efficiently reviewed and searched, it is common in eDiscovery processing to extract the text from files.

35. Text may be extracted in different ways depending on the format of the file. Text may be extracted from a native file through the use of specialized text extraction software. When a native file is rendered to an image format, the text of that image may still be extracted using optical character recognition software (“OCR”) to scan the image for text. This extracted text may be provided as part of a production, particularly if documents are rendered to images prior to production.

Redaction and Withholding Documents

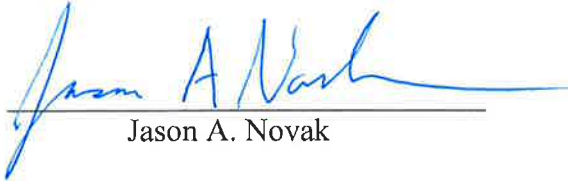
36. Another way in which productions vary is in the redaction of files and withholding of specific classes of files. If a file is to be redacted prior to production, it may be first rendered to a static image format, and then the image may be redacted using redaction software. This redacted image representation of a file may then be produced along with any metadata or document family information captured for the original native file.

37. Text for a redacted file may be extracted by using OCR software on a redacted image representation of a file. This results in only the redacted version of the text being made available for review, searching, and production. The unredacted version of the original file may be maintained by the producing party for later reference.

38. If document families are fully expanded and each file is rendered to an image format prior to production, then if a single document in a family is to be withheld from a production, that file may be replaced with a “placeholder” image that indicates a document was withheld without adversely impacting the file’s structure.

I declare under the laws of the United States and under penalty of perjury that the foregoing is true and correct.

Executed on March 30, 2011 in New York, New York.



Jason A. Novak