

**APPENDIX 1**  
**Electronic Document Production Specifications**

Unless otherwise specified and agreed to by the Office of Attorney General, all responsive documents must be produced in LexisNexis® Concordance® format in accordance with the following instructions. Any questions regarding electronic document production should be directed to the Assistant Attorney General whose telephone number appears on the subpoena.

1. **Concordance Production Components.** A Concordance production consists of the following component files, which must be produced in accordance with the specifications set forth below in Section 7.
  - A. ***Metadata Load File.*** A delimited text file that lists in columnar format the required metadata for each produced document.
  - B. ***Extracted or OCR Text Files.*** Document-level extracted text for each produced document or document-level optical character recognition (“OCR”) text where extracted text is not available.
  - C. ***Single-Page Image Files.*** Individual petrified page images of the produced documents in tagged image format (“TIF”), with page-level Bates number endorsements.
  - D. ***Opticon Load File.*** A delimited text file that lists the single-page TIF files for each produced document and defines (i) the relative location of the TIF files on the production media and (ii) each document break.
  - E. ***Native Files.*** Native format versions of non-printable or non-print friendly produced documents.
2. **Production Folder Structure.** The production must be organized according to the following standard folder structure:
  - data\ (contains production load files)
  - images\ (contains single-page TIF files, with subfolder organization)  
  \0001, \0002, \0003...
  - native files\ (contains native files, with subfolder organization)  
  \0001, \0002, \0003...
  - text\ (contains text files, with subfolder organization)  
  \0001, \0002, \0003...
3. **De-Duplication.** You must perform global de-duplication of stand-alone documents and email families against any prior productions pursuant to this or previously related subpoenas.
4. **Paper or Scanned Documents.** Documents that exist only in paper format must be scanned to single-page TIF files and OCR’d. The resulting electronic files should be pursued in Concordance format pursuant to these instructions. You must contact the

Assistant Attorney General whose telephone number appears on the subpoena to discuss (i) any documents that cannot be scanned, and (ii) how information for scanned documents should be represented in the metadata load file.

5. Structured Data. Before producing structured data, including but not limited to relational databases, transactional data, and xml pages, you must first speak to the Assistant Attorney General whose telephone number appears on the subpoena. Spreadsheets are not considered structured data.

6. Media and Encryption. All documents must be produced on CD, DVD, or hard-drive media. All production media must be encrypted with a strong password, which must be delivered independently from the production media.

7. Production File Requirements.

A. ***Metadata Load File***

- Required file format:
  - ASCII or UTF-8
  - Windows formatted CR + LF end of line characters, including full CR + LF on last record in file.
  - .dat file extension
  - Field delimiter: (ASCII decimal character 20)
  - Text Qualifier: þ (ASCII decimal character 254). Date and pure numeric value fields do not require qualifiers.
  - Multiple value field delimiter: ; (ASCII decimal character 59)
- The first line of the metadata load file must list all included fields. All required fields are listed in Attachment 2.
- Fields with no values must be represented by empty columns maintaining delimiters and qualifiers.
- **Note:** All documents must have page-level Bates numbering (except documents produced only in native format, which must be assigned a document-level Bates number). The metadata load file must list the beginning and ending Bates numbers (BEGDOC and ENDDOC) for each document. For document families, including but not limited to emails and attachments, compound documents, and uncompressed file containers, the metadata load file must also list the Bates range of the entire document family (ATTACHRANGE), beginning with the first Bates number (BEGDOC) of the “parent” document and ending with the last Bates number (ENDDOC) assigned to the last “child” in the document family.
- Date and Time metadata must be provided in separate columns.
- Accepted date formats:
  - mm/dd/yyyy
  - yyyy/mm/dd
  - yyyymmdd
- Accepted time formats:
  - hh:mm:ss (if not in 24-hour format, you must indicate am/pm)
  - hh:mm:ss:mmm

**B. *Extracted or OCR Text Files***

- You must produce individual document-level text files containing the full extracted text for each produced document.
- When extracted text is not available (for instance, for image-only documents) you must provide individual document-level text files containing the document's full OCR text.
- The filename for each text file must match the document's beginning Bates number (BEGDOC) listed in the metadata load file.
- Text files must be divided into subfolders containing no more than 500 to 1000 files.

**C. *Single-Page Image Files (Petrified Page Images)***

- Where possible, all produced documents must be converted into single-page tagged image format ("TIF") files. See Section 7.E below for instructions on producing native versions of documents you are unable to convert.
- Image documents that exist only in non-TIF formats must be converted into TIF files. The original image format must be produced as a native file as described in Section 7.E below.
- For documents produced only in native format, you must provide a TIF placeholder that states "Document produced only in native format."
- Each single-page TIF file must be endorsed with a unique Bates number.
- The filename for each single-page TIF file must match the unique page-level Bates number (or document-level Bates number for documents produced only in native format).
- Required image file format:
  - CCITT Group 4 compression
  - 2-Bit black and white
  - 300 dpi
  - Either .tif or .tiff file extension.
- TIF files must be divided into subfolders containing no more than 500 to 1000 files. Where possible documents should not span multiple subfolders.

**D. *Opticon Load File***

- Required file format:
  - ASCII
  - Windows formatted CR + LF end of line characters
  - Field delimiter: , (ASCII decimal character 44)
  - No Text Qualifier
  - .opt file extension
- The comma-delimited Opticon load file must contain the following seven fields (as indicated below, values for certain fields may be left blank):
  - ALIAS or IMAGEKEY – the unique Bates number assigned to each page of the production.
  - VOLUME – this value is optional and may be left blank.
  - RELATIVE PATH – the filepath to each single-page image file on the production media.

- DOCUMENT BREAK – defines the first page of a document. The only possible values for this field are “Y” or blank.
- FOLDER BREAK – defines the first page of a folder. The only possible values for this field are “Y” or blank.
- BOX BREAK – defines the first page of a box. The only possible values for this field are “Y” or blank.
- PAGE COUNT – this value is optional and may be left blank.
- **Example:**  
 ABC00001,,IMAGES\0001\ABC00001.tif,Y,,2  
 ABC00002,,IMAGES\0001\ABC00002.tif,,,  
 ABC00003,,IMAGES\0002\ABC00003.tif,Y,,1  
 ABC00004,,IMAGES\0002\ABC00004.tif,Y,,1

E. ***Native Files***

- Non-printable or non–print friendly documents (including but not limited to spreadsheets, audio files, video files and documents for which color has significance to document fidelity) must be produced in their native format.
- The filename of each native file must match the document’s beginning Bates number (BEGDOC) in the metadata load file and retain the original file extension.
- For documents produced only in native format, you must assign a single document-level Bates number and provide an image file placeholder that states “Document produced only in native format.”
- The relative paths to all native files on the production media must be listed in the NATIVEFILE field of the metadata load file.
- Native files that are password-protected must be decrypted prior to conversion and produced in decrypted form. In cases where this cannot be achieved the document’s password must be listed in the metadata load file. The password should be placed in the COMMENTS field with the format Password: <PASSWORD>.
- You may be required to supply a software license for proprietary documents produced only in native format.

**APPENDIX 2**  
**Required Fields for Metadata Load File**

<b>FIELD NAME</b>	<b>FIELD DESCRIPTION</b>	<b>FIELD VALUE EXAMPLE<sup>1</sup></b>
DOCID	Unique document reference (can be used for de-duplication).	ABC0001 or ###.#####.###
BEGDOC	Bates number assigned to the first page of the document.	ABC0001
ENDDOC	Bates number assigned to the last page of the document.	ABC0002
BEGATTACH	Bates number assigned to the first page of the parent document in a document family ( <i>i.e.</i> , should be the same as BEGDOC of the parent document, or PARENTDOC).	ABC0001
ENDATTACH	Bates number assigned to the last page of the last child document in a family ( <i>i.e.</i> , should be the same as ENDDOC of the last child document).	ABC0008
ATTACHRANGE	Bates range of entire document family.	ABC0001 - ABC0008
PARENTDOC	BEGDOC of parent document.	ABC0001
CHILDDOCS	List of BEGDOCs of all child documents, delimited by ";" when field has multiple values.	ABC0002; ABC0003; ABC0004...
COMMENTS	Additional document comments, such as passwords for encrypted files.	
NATIVEFILE	Relative file path of the native file on the production media.	.\Native_File\Folder\...\BEGDOC.ex t
SOURCE	For scanned paper records this should be a description of the physical location of the original paper record. For loose electronic files this should be the name of the file server or workstation where the files were gathered.	Company Name, Department Name, Location, Box Number...
CUSTODIAN	Owner of the document or file.	Firstname Lastname, Lastname, Firstname, User Name; Company Name, Department Name...
FROM	Sender of the email.	Firstname Lastname < FLastname &@domain >

<sup>1</sup> Examples represent possible values and not required format unless the field format is specified in Attachment 1.

<b>FIELD NAME</b>	<b>FIELD DESCRIPTION</b>	<b>FIELD VALUE EXAMPLE<sup>1</sup></b>
TO	All to: members or recipients, delimited by ";" when field has multiple values.	Firstname Lastname < FLastname @domain >; Firstname Lastname < FLastname @domain >; ...
CC	All cc: members, delimited by ";" when field has multiple values.	Firstname Lastname < FLastname @domain >; Firstname Lastname < FLastname @domain >; ...
BCC	All bcc: members, delimited by ";" when field has multiple values	Firstname Lastname < FLastname @domain >; Firstname Lastname < FLastname @domain >; ...
SUBJECT	Subject line of the email.	
DATERCVD	Date that an email was received.	mm/dd/yyyy, yyyy/mm/dd, or yyyymmdd
TIMERCVD	Time that an email was received.	hh:mm:ss AM/PM or hh:mm:ss
DATESENT	Date that an email was sent.	mm/dd/yyyy, yyyy/mm/dd, or yyyymmdd
TIMESENT	Time that an email was sent.	hh:mm:ss AM/PM or hh:mm:ss
CALBEGDATE	Date that a meeting begins.	mm/dd/yyyy, yyyy/mm/dd, or yyyymmdd
CALBEGTIME	Time that a meeting begins.	hh:mm:ss AM/PM or hh:mm:ss
CALENDDATE	Date that a meeting ends.	mm/dd/yyyy, yyyy/mm/dd, or yyyymmdd
CALENDTIME	Time that a meeting ends.	hh:mm:ss AM/PM or hh:mm:ss
CALENDAR DUR	Duration of a meeting in hours.	0.75, 1.5...
ATTACHMENTS	List of filenames of all attachments, delimited by ";" when field has multiple values.	AttachmentFileName.; AttachmentFileName.docx; AttachmentFileName.pdf;...
NUMATTACH	Number of attachments.	1, 2, 3, 4....
RECORDTYPE	General type of record.	IMAGE; LOOSE E-MAIL; E-MAIL; E-DOC; IMAGE ATTACHMENT; LOOSE E-MAIL ATTACHMENT; E-MAIL ATTACHMENT; E-DOC ATTACHMENT
FOLDERLOC	Original folder path of the produced document.	Drive:\Folder\...\...\
FILENAME	Original filename of the produced document.	Filename.ext
DOCEXT	Original file extension.	html, xls, pdf

<b>FIELD NAME</b>	<b>FIELD DESCRIPTION</b>	<b>FIELD VALUE EXAMPLE<sup>1</sup></b>
DOCTYPE	Name of the program that created the produced document.	Adobe Acrobat, Microsoft Word, Microsoft Excel, Corel WordPerfect...
TITLE	Document title (if entered).	
AUTHOR	Name of the document author.	Firstname Lastname; Lastname, First Name; FLastname
REVISION	Number of revisions to a document.	18
DATECREATED	Date that a document was created.	mm/dd/yyyy, yyyy/mm/dd, or yyymmdd
TIMECREATED	Time that a document was created.	hh:mm:ss AM/PM or hh:mm:ss
DATEMOD	Date that a document was last modified.	mm/dd/yyyy, yyyy/mm/dd, or yyymmdd
TIMEMOD	Time that a document was last modified.	hh:mm:ss AM/PM or hh:mm:ss
FILESIZE	Original file size in bytes.	128, 512, 1024...
PGCOUNT	Number of pages per document.	1, 2, 10, 100...
IMPORTANCE	Email priority level if set.	Low, Normal, High
TIFFSTATUS	Generated by the Law Pre-discovery production tool (leave blank if inapplicable).	Y, C, E, W, N, P
DUPSTATUS	Generated by the Law Pre-discovery production tool (leave blank if inapplicable).	P
MD5HASH	MD5 hash value computed from native file (a/k/a file fingerprint).	BC1C5CA6C1945179FEE144F25F51087B
SHA1HASH	SHA1 hash value	B68F4F57223CA7DA3584BAD7E CF111B8044F8631
MSGINDEX	Email message ID	