

**INSTRUCTIONS FOR THE PRODUCTION OF
ELECTRONICALLY STORED INFORMATION (ESI)**

These instructions outline the technical requirements for producing scanned paper collections, email and other electronically stored information (ESI) to the New Jersey Bureau of Securities (“Bureau”). Production in a format other than that identified below, must be discussed with and approved by the Bureau.

I. GENERAL INSTRUCTIONS

1. A cover letter should be included with each production. The cover letter should list each piece of media (hard drive, thumb drive, DVD or CD) included in the production, along with the Bates range of the production.
2. Documents created or stored electronically **MUST** be produced in their native electronic format, not printed to paper or PDF.
3. Data can be produced on CD, DVD or thumb drive; use the media requiring the least number of deliverables.
4. Each piece of media should be “self-contained,” for example, if 5 CDs are provided, each must have its own associated load file.
5. Label all media (printed not handwritten) with the following:
 - Name of the producing party;
 - Production date;
 - Bates range; and
 - Disk number (1 of X), if applicable.
6. For a given project, all loads files should use the same field names, ordering and structure as the first delivery.
7. Ensure there are no truncated file or folder names in your production. These would be indicated by a tilde~. (E.g. AAA000~1.T1F).
8. Ensure there are no missing or empty folders in your production.
9. Organize productions by custodian, unless otherwise instructed. All documents from an individual custodian should be confined to a single load file.
10. All productions should be checked and produced free of computer viruses.
11. Passwords for documents, files, compressed archives and encrypted media should be provided separately either via email or in a separate cover letter from the custodian.

II. DELIVERY FORMATS

Subpart A of this section sets forth the standard production format that provides users the most flexibility in searching ESI using Concordance®. Under limited circumstances, it may not be possible to produce ESI in that preferred format. Under those circumstances, and with prior approval of the Bureau, Adobe PDF files and Email native files may be produced in the formats provided in subparts B and C of this section.

A. Concordance® Production Format

All scanned paper, email and native file collections should be converted or processed to TIFF files, Bates numbered, and include fully searchable text (OCR).

Most document productions will contain the elements listed below:

- **Folder labeled IMAGES**
- **Folder labeled DATA**
- **Folder labeled TEXT**
- **Folder labeled NATIVES**

1. **Bates Numbering**

The Bates number must be a unique, consistently formatted identifier consisting of an alpha prefix along with a fixed length number for each custodian, e.g., “ABC0000001”. This format must remain consistent across all production numbers. The number of digits in the numeric portion of the format should not change in subsequent productions, nor should spaces, hyphens, or other separators be added or deleted. Avoid Bates prefixes containing characters other than A...Z.

2. **Images**

a. **Generally**

1. Images should be single-page, CCITT Group 4 TIFF files, scanned at 200-300 DPI (Presentation software such as PowerPoint, AUTOCAD images and Spreadsheet images are exceptions please see step 4 below.)
2. File names cannot contain embedded spaces.
3. The number of TIFF files per folder should not exceed 1000 files.
4. Rendering PowerPoint, AUTOCAD images and Excel files to images:
 - PowerPoint: All pages of the file should be scanned in full slide image format, with any speaker notes following the appropriate slide image.
 - AUTOCAD images: If possible, files should be scanned to single page JPEG (.JPG) file format. Color images can be discussed on a case by case basis.

- Excel: A placeholder image, named by the IMAGEID of the file, may be used.

b. Image Cross-Reference File

The image cross-reference file is needed to link the images to the database. It is a comma-delimited file consisting of seven fields per line. There must be a line in the cross-reference file for every image in the database. The format is as follows:

ImageID, VolumeLabel, ImageFilePath, DocumentBreak, FolderBreak, BoxFit, PageCount

IMAGEID	The unique designation that is used to identify an image. Note: This image ID key must be a unique and fixed length number. This number will be used in the .DAT file as the ImageID field that links the database to the images. The format of this image key must be consistent across all productions. It is recommended that the format be a 7 digit number to allow for the possible increase in the size of a production.
VOLUME LABEL	Optional, but names for volumes should not be longer than eight characters, with a suffix not wider than three characters.
IMAGE FILE PATH	The full path to the image file.
DOCUMENT BREAK	The letter “Y” denotes the first page of a document. If this field is blank, then the page is not the first page of a document.
FOLDER BREAK	Leave empty.
BOX BREAK	Leave empty.
PAGE COUNT	Optional

Sample IMAGE Load file (Often referred to as Option or .OPT file)....

```

IMG0000001, ,E:\001\IMG0000001, TIF,Y,,,
IMG0000002, ,E:\001\IMG0000002, TIF,,,,
IMG0000003, ,E:\001\IMG0000003,TIF,,,,
IMG0000004, ,E:\001\IMG0000003,TIF,,,
IMG0000005, ,E:\001\IMG0000003,TIF,Y,,,
IMG0000006, ,E:\001\IMG0000003,TIF,,,,
    
```

Note: Only images belong in the Opticon load file, If OCR files are included in the same folder as the images, errors will occur when retrieving the images.

3. DATA Load File (Concordance .DAT)

The most reliable format for Concordance data delivery is a Concordance Database. Instances where this is not an option, data can be delivered utilizing standard delimited files for coded data (.DAT) and .TXT or .RTF files for OCR data.

The data file (.DAT) contains all of the fielded information that will be loaded into the Concordance® database. The first line of the .DAT file must be a header row identifying the field names. The best practice is to use the following standard delimiters in the .DAT file:¹

DELIMITER NAME	SYMBOL	DECIMAL EQUIVALENT
COMMA	,	044
PARAGRAPH	¶	020
QUOTE	”	254
NEWLINE	®	174
SEMI-COLON	;	059

1. Date fields should be provided in the format: mm/dd/yyyy (E.g. “01/01/2004” and not “01/01/2004 12:01:01PM”)
2. All attachments should sequentially follow the parent document/email. Parent email and attachment document families should be kept intact.
3. All metadata associated with email, audio files, and native electronic document collections must be produced.
4. The .DAT file for scanned paper collections must contain, at a minimum, the following fields:
 - BEGBATES: Beginning Bates number
 - ENDBATES: Ending Bates number
 - IMAGEID: Image Key Field
 - CUSTODIAN: Individual from whom the document originated

¹ The chart represents the most common delimiter characters used in Concordance, along with its decimal equivalent. If the source program you are importing from uses a different font, it can change the symbolic representation of the delimiters. If this happens, match the delimiter characters with the decimal equivalents instead of relying on the displayed symbol. Using the decimal equivalents will always result in a correct delimiter match. For a complete list of delimited characters you may reference the following link:

http://help.lexisnexis.com/litigation/ac/cn_classic/index.html?delimiter_characters.htm

5. The .DAT file should NOT include document text.
6. For all non-redacted documents, please include the General Metadata fields for all files and respective metadata fields for emails and electronic documents (e.g., MS Word, MS Excel, etc.) where available using industry standard techniques. With respect to redacted documents, some metadata may be withheld as needed to preserve privileges.
7. Spaces and returns must match the original text. No odd characters, such as a semi-colon, should appear in lieu of a soft-return or a space.
8. Fields provided in a .DAT file may include the following:

GENERAL METADATA	DEFINITION	FIELD NAME
BEG BATES	The start bates of the document	BEGBATES
END BATES	The end bates of the document	ENDBATES
BEG ATTACH	Start bates of attachment	BEGATT
END ATTACH	The end bates of attachment	ENDATT
Page Count	Number of pages	PAGECNT
Custodian	The name of the original custodian of the file	CUSTDIAN
File Extension	The extension of the file	FILEEXT
Confidential	Value	CNFDTL
Redaction	Redaction reason value	REDACTN
Email Type	Defines if a message file is an email or attachment	EMAILTYP
File Path	The address where the file resides on the electronic media	FILEPATH
File Size	The amount of space the file takes up on the electronic media	FILESIZE
Hash	The SHA (Secure Hash Algorithm) for the original native	FILEHASH
MD5Hash	The MD5 Hash for the original file	MD5HASH
Native File Link	Relative path of submitted native files	FILELINK

EMAIL	DEFINITION	FIELD NAME
FROM	The person who authored the email	FROM
TO	Recipient(s) of the email	TO
COPIED	Person(s) copied on the email	CC
BCC	Person(s) blind copied on the email	BCC
Date Sent	Date the email was sent expressed usually Eastern Standard Time if in US	DATESENT
Time Sent	Time the email was sent expressed usually Eastern Standard Time if in US	TIMESENT
Date Received	Date Received Date received in EST	DATERCVD
Time Received	Time Received Date received in EST	TIMERCVD
Subject	Subject line of email	SUBJECT
Attachment Count	Number of attachments	ATTCOUNT
INTFILEPATH	Original location of email including original file name	
INTMSGID	Unique Message ID	MSGID

EDOCS METADATA	DEFINITION	FIELD NAME
Author	The person who authored the document	AUTHOR
Date Created	Date the document was created	CREATEDAT
Time Created	Time the document was created	CREATETIM
Date Last Modified	Date the file was last changed/saved	MODDATE
Time Last Modified	Time the file was last changed/saved	MODTIME
Printed Date	Date that the file was last printed	PRNTDATE

Title	Title of document	TITLE
Path	Path where native file document was stored including	PATH

- No more than one document per database record. The database and load files should be sorted sequentially by Bates number. Concordance displays records in the same order they are loaded.

4. Text

Searchable text of the entire document must be provided for every record, at the document level. For redacted documents, provide the full text for the redacted version. Delivery should be as follows:

- Text should be delivered as multi-page ASCII text files with the files named the same as the ImageID field.
- Text files must be placed in a separate folder labeled TEXT.
- The number of files per folder should be limited to 1000 files.
- DO NOT include the searchable text in the:
 - same folder as the IMAGES folder.
 - .DAT file. OCR text is loaded into Concordance through a Concordance CPL (Concordance Programming Language) script.

5. Native Files

Copies of original email and native file documents/attachments must be included for all electronic productions.

- Native files must be named with the same naming convention associated with its associated files in the production:

EXAMPLE:

OAGNJ0012345 - beginning bates number from the load file.

OAGNJ0012345.tif - associated image file

OAGNJ0012345.txt - associated text file

OAGNJ0012345.xls - associated native Excel spreadsheet.

- The full path of the native file must be provided in the .DAT file for the LINK field.

3. The number of native files per folder should not exceed 1000 files.
4. These files should be located in a folder named NATIVE.

B. Adobe PDF File Production

Production in this format requires prior approval from the Bureau.

1. When approved, Adobe PDF files may be produced in native file format.
2. PDF files should be produced in separate folders
3. All PDFs must be unitized at the document level, i.e. each PDF should represent a distinct document; a single PDF cannot contain multiple documents.
4. All PDF files must contain embedded text that includes all discernable words within the document, not selected text only. This requires all layers of the PDF to be flattened first.
5. If PDF files are Bates endorsed, the PDF files must be named by the Bates range.

C. Email Native file Production

Production in this format requires prior approval from the Bureau.

When approved, Outlook (.PST) email files may be produced in native file format. A separate folder should be provided for each custodian.