



---

**Office of Compliance Inspections and Examinations (OCIE)**  
**Data Delivery Guidance**

This document provides suggested guidance to regulated entities and other parties (“Producing Parties”) for producing electronic files and electronically stored information to the U.S. Securities and Exchange Commission’s (SEC) Office of Compliance Inspections and Examinations (OCIE).

---

|      |  |   |
|------|--|---|
| I.   | General Recommendations for Producing Documents to OCIE..... | 1 |
| II.  | Delivery Format Guidance.....                                | 2 |
| A.   | Native File Production without Load Files .....              | 2 |
| B.   | Imaged Productions .....                                     | 2 |
| C.   | Adobe PDF File Production.....                               | 4 |
| D.   | Electronic Trade and Bank Records.....                       | 4 |
| E.   | Production of Audio and Video Files.....                     | 4 |
| F.   | Bates Numbering & Document Identifiers.....                  | 5 |
| III. | Addendum A.....  | 6 |

---

**I. General Recommendations for Producing Electronic Files to OCIE**

- A. Produce electronic files in the medium and format in which the record is stored by your business (i.e., “native file format”). A native production includes the metadata related to the underlying document. Metadata consists of information not visible in the document and displays certain characteristics of the document, such as its origin, usage, and alteration history..
- B. OCIE encourages Producing Parties to avoid duplicative productions. If a Producing Party wishes to cleanse productions of duplicative responses, the party should first discuss with OCIE the proposed use of file de-duplication methodologies or computer-assisted review or technology-assisted review.
- C. Provide a cover letter that includes the following information:
  - 1. A list of each piece of media included in the production with its unique production volume number.
  - 2. A list of document custodians (and if using Bates stamps, identifying the Bates range for each custodian).
  - 3. The time zone in which the emails were standardized during conversion.
- D. Electronic productions may be submitted by using secure file transfer programs. OCIE encourages Producing Parties to use Accellion, the SEC’s preferred file transfer system. Other recommended production methods include using secure email via Zixmail or a transport layer security connection. The use of other file sharing methods should be discussed with OCIE.
- E. While not recommended, data may also be produced on CD, DVD, thumb drive, etc. When the provided media is encrypted, it should be produced using industry standard self-extracting decrypting software and labeled for easy identification with the following, where applicable:
  - 1. File number
  - 2. Production date
  - 3. Producing party

---

#### 4. Bates range

- F. Provide passwords for electronic documents, files, compressed archives, and encrypted media separately either via email or in a separate cover letter from the media.
- G. Organize submissions by document custodian unless the individual custodian of the document is not reasonably available.
- H. Produce all document family groups (*i.e.*, email attachments, embedded files, etc.) together, where children files follow parent files sequentially in the Bates numbering.
- I. If scanning paper documents, do not merge distinct documents into a single record, or split single documents into multiple records (*i.e.*, paper documents should be logically unitized). All load-ready collections should include only one data load file and one image pointer file.
- J. Produce separate text files for extracted OCR text.
- K. Account for document custodians in the custodian field in all metadata load files.
- L. Alphanumeric characters and the underscore character are better in file names and folder names. Special characters are not permitted.
- M. A producing party can use a slip sheet or other place holder if it determines that certain files (e.g., wave files, pictures, or password protected/encrypted files) in a production are inaccessible.
- L. Provide all electronic productions free of computer viruses.
- M. Additional technical descriptions can be found in **Addendum A** to this document.

*\*Please note that productions sent to the SEC via United States Postal Service are subject to Mail Irradiation, and as a result electronic productions may be damaged.\**

## II. Delivery Format Guidance

### A. Native File Production without Load Files

If a regulated entity produces native files of documents, OCIE recommends that they be organized by custodian-named file folders.

### B. Imaged Productions

If the regulated entity produces a load file, OCIE prefers that all documents and data be produced in an industry standard format such as Concordance. Much of the technical guidance below follows industry standards for imaged productions.

#### 1. Images

- a. Image documents to 300 DPI Group IV Black & White Tagged Image File Format (.TIFF or .TIF) files. Produce TIFF files in single-page format and 8 1/2 x 11 inch page size (except for documents requiring a different page size). Each image file should have a unique file name (*e.g.*, the Bates number of that page).
- b. JPEG format is preferred when producing color images.
- c. File names cannot contain embedded spaces or special characters (including the comma).
- d. Folder names cannot contain embedded spaces or special characters (including the comma).
- e. Produce TIFF image files with a unique file name, such as a Bates number or similar identifier.
- f. If using Bates numbers, endorse images sequentially, preferably in the lower right corner of each image.
- g. Limit the number of TIFF files per folder to 1000 files.

- 
- h. Use a placeholder image named by the Bates number on the file for Excel spreadsheets.
  - i. Image MS PowerPoint documents in “Notes” mode.
  - j. MS Word documents that contain markups should be imaged showing markups.
  - k. Produce AUTOCAD/photograph files as a single page JPEG file.

## 2. Concordance Image or Opticon Cross-Reference File

The Concordance Image or Opticon Cross-Reference file (.LOG or .OPT) links the images to the database records. It should be a comma-delimited file consisting of seven fields per line with a line in the cross-reference file for every image in the database with the following format:

*ImageID,VolumeLabel,ImageFilePath,DocumentBreak,FolderBreak,BoxBreak,PageCount*

## 3. Concordance Data File

The data file (.DAT) contains all of the fielded information that will be loaded into the Concordance database. When producing a Concordance Data File, please provide the following information.

- a. The first line of the .DAT file must be a header row identifying the field names.
- b. Concordance default delimiters for the .DAT file:
  - Comma ¶ ASCII character (020)
  - Quote p ASCII character (254)
- c. Date field format: mm/dd/yyyy
- d. Make the date and time fields two separate fields.
- e. If the production includes imaged emails and attachments, include the attachment fields to preserve the parent/child relationship between an email and its attachments.
- f. Include an OCRPATH field to provide the file path and name of the extracted text file on the produced storage media. Name the text file after the FIRSTBATES. Do not include the text in the .DAT file.
- g. For productions with native files, include a LINK field to provide the file path and name of the native file on the produced storage media. Name the native file after the FIRSTBATES.
- h. Keep BEGATTACH and ENDATTACH fields as two separate fields.
- i. A complete list of metadata fields is available in **Addendum A** to this document.

## 4. Searchable Text Files

- a. Provide all files with an accompanying searchable text (.TXT) file extracted from the native, electronic file, or generated as an OCR file from scanned and image based files (PDFs, TIF, JPG, PNG, etc.).
- b. The extracted text of the produced document should include, where applicable, all comments, tracked changes, speaker’s notes, and text from hidden worksheets, slides, columns and rows.
- c. Produce text as separate text files and not as fields within the .DAT file. Include the full path to the text file (OCRPATH) in the .DAT file.
- d. Name document level ANSI text files using the FIRSTBATES/Image Key. Please note if any non-ANSI text files are included in the production.
- e. Place extracted text files in a separate folder (please limit the number of text files per folder to 1,000 files).

- 
- f. Do not include any special characters (including commas) in the folder names.
  - g. For documents in foreign languages, use an OCR tool suitable for the particular byte or multi-byte languages.
  - h. Name each text file according to some unique identifier, such as a Bates number, of the first page of the corresponding image files (e.g., BATES000001.TXT).
  - i. For emails, please include the following in the corresponding text file, where reasonably available:
    - (1) the individual(s) to whom the communication was directed (“To”);
    - (2) the author(s) of the email communication (“From”);
    - (3) who was copied and blind copied on such email (“CC” and “BCC”);
    - (4) the subject line of the email (“RE” or “Subject”);
    - (5) the names of any attachments; and
    - (6) the text (body) of the email.

## 5. Linked Native Files

Include copies of original email and native file documents/attachments for all electronic productions.

- a. Designate native file documents by a standard identifier, such as the FIRSTBATES number (e.g., BATES000001).
- b. In the metadata load file, create a field called “NATIVELINK..” The Native Link values should contain the full directory path and file name of the natives.
- c. Limit native files per folder to 1,000 files.

## C. Adobe PDF File Production

When producing Adobe PDF files, please provide the following information, wherever possible:

1. Produce PDF files in native file format.
2. Produce PDF files in separate folders named by the document custodian. Do not include any special characters, including commas, in the folders.
3. Unitize all PDF files at the document level (*i.e.*, each PDF represents a discrete document).
4. Include embedded text in all PDF files, which includes all discernible words within the document. This requires all layers of the PDF to be flattened first.
5. If PDF files are Bates endorsed, include the Bates Range in the PDF files’ name.

## D. Electronic Trade and Bank Records

When producing electronic trade and bank records, please consult with OCIE exam staff to discuss types of document formats available.

## E. Production of Audio and Video Files

1. Produce audio and video files in the format in which they are maintained in the ordinary course of business.
2. Include a reference file or .DAT containing the name of the file for any native audio and video files (and hash value for each produced file, if feasible).
3. If the audio files are maintained in a non-standard format, please produce in MP3 format, where reasonably feasible.
4. Specify the native file application in the “File Type” field for audio and video files, as well as all other native files produced.

---

**F. Bates Numbering & Document Identifiers**

If the Producing Party chooses to include Bates numbers and/or confidentiality designations, OCIE recommends that Bates numbers and any confidentiality designations be electronically branded on each produced TIFF image of Electronically Stored Information (“ESI”). For documents produced in native file format, use a unique Bates number as the file name, and preserve the original file name and file extension in the corresponding load file. Confidentiality designations can be provided either as a separate metadata field in the load file and/or part of the file name of the native document.

### III. Addendum A

During the process of converting ESI from the electronic format of that application in which the ESI is normally created, viewed, and/or modified to TIFF, existing metadata values should be extracted and produced in a load file. When producing metadata of electronic document collections, it should be extracted and provided in a .DAT file using the field definition and formatting described below:

| Field Name          | Sample Data  | Description   |
|---------------------|--|---|
| FIRSTBATES          | EDC0000001   | First Bates number of native file document/email  |
| LASTBATES           | EDC0000001   | Last Bates number of native file document/email<br>**The LASTBATES field should be populated for single page documents/emails.  |
| ATTACHRANGE         | EDC0000001 - EDC0000015                            | Bates number of the first page of the parent document to the Bates number of the last page of the last attachment "child" document  |
| BEGATTACH           | EDC0000001   | First Bates number of attachment range  |
| ENDATTACH           | EDC0000015   | Last Bates number of attachment range   |
| PARENT_BATES        | EDC0000001   | First Bates number of parent document/Email<br>**This PARENT_BATES field should be populated in each record representing an attachment "child" document   |
| CHILD_BATES         | EDC0000002; EDC0000014                             | First Bates number of "child" attachment(s); can be more than one Bates number listed depending on the number of attachments<br>**The CHILD_BATES field should be populated in each record representing a "parent" document |
| CUSTODIAN           | Smith, John  | Email: Mailbox where the email resided<br>Native: Name of the individual or department from whose files the document originated   |
| FROM                | John Smith   | Email: Sender<br>**semi-colon should be used to separate multiple entries   |
| TO                  | Coffman, Janice; LeeW<br>[mailto:LeeW@MSN.com]     | Recipient(s)<br>**semi-colon should be used to separate multiple  |
| CC                  | Frank Thompson [mailto:<br>frank_Thompson@cdt.com] | Carbon copy recipient(s)<br>**semi-colon should be used to separate multiple entries  |
| BCC                 | John Cain  | Blind carbon copy recipient(s)<br>**semi-colon should be used to separate multiple  |
| SUBJECT             | Board Meeting Minutes                              | Email: Subject line of the email Native: Title of document (if available)   |
| FILENAME            | BoardMeetingMinutes.docx                           | Native: Name of the original native file, including extension   |
| DATE_SENT           | 10/12/2010   | Email: Date the email was sent<br>Native: (empty)   |
| TIME_SENT/TIME_ZONE | 07:05 PM GMT                                       | Email: Time the email was sent/ Time zone in which the emails were standardized during conversion.<br>Native: (empty)<br>**This data must be a separate field and cannot be combined with the DATE_SENT field               |

|                    |  |  |
|--------------------|--|--|
| TIME_ZONE          | GMT  | The time zone in which the emails were standardized during conversion.<br>Email: Time zone<br>Native: (empty)  |
| NATIVELINK         | D:\001\EDC0000001.msg  | Hyperlink to the email or native file document<br>**The linked file must be named per the FIRSTBATES number  |
| MIMETYPE           | MSG  | The content type of an Email or native file document as identified/extracted from the header   |
| FILE_EXTENSION     | MSG  | The file type extension representing the Email or native file document; will vary depending on the email format  |
| AUTHOR             | John Smith   | Email: (empty)<br>Native: Author of the document   |
| DATE_CREATED       | 10/10/2010   | Email: (empty)<br>Native: Date the document was created  |
| TIME_CREATED       | 10:25 AM   | Email: (empty)<br>Native: Time the document was created<br>**This data must be a separate field and cannot be combined with the DATE_CREATED field       |
| DATE_MOD           | 10/12/2010   | Email: (empty)<br>Native: Date the document was last modified  |
| TIME_MOD           | 07:00 PM   | Email: (empty)<br>Native: Time the document was last modified<br>**This data must be a separate field and cannot be combined with the DATE_MOD field     |
| DATE_ACCESSED      | 10/12/2010   | Email: (empty)<br>Native: Date the document was last accessed  |
| TIME_ACCESSED      | 07:00 PM   | Email: (empty)<br>Native: Time the document was last accessed<br>**This data must be a separate field and cannot be combined with the DATE_ACCESSD field |
| PRINTED_DATE       | 10/12/2010   | Email: (empty)<br>Native: Date the document was last printed   |
| FILE_SIZE          | 5,952  | Size of native file document/email in KB   |
| PAGE_COUNT         | 1  | Number of pages in native file document/email  |
| PATH               | J:\Shared\SmithJ\October Agenda.doc                              | Email: (empty)<br>Native: Path where native file document was stored including original file name.   |
| INTFILEPATH        | Personal Folders\Deleted Items\Board Meeting Minutes.msg         | Email: original location of email including original file name.<br>Native: (empty)   |
| MESSAGEID/INTMSGID | <000805c2c71b\$75977050\$cb8306d1@MSN>                           | Email: Unique Message ID<br>Native: (empty)  |
| HASHVALUE          | d131dd02c5e6eec4693d9a0698aff95c2fcab58712467eab4004583eb8fb7f89 | MD5 Hash value of the document.  |
| OCRPATH            | TEXT/001/EDC0000001.txt  | Path to extracted text of the native file  |

---

Sample Image Cross-Reference File:

|                                       |
|---------------------------------------|
| IMG0000001,,E:\001\IMG0000001.TIF,Y,, |
| IMG0000002,,E:\001\IMG0000002.TIF,,,, |
| IMG0000003,,E:\001\IMG0000003.TIF,,,, |
| IMG0000004,,E:\001\IMG0000004.TIF,Y,, |
| IMG0000005,,E:\001\IMG0000005.TIF,Y,, |
| IMG0000006,,E:\001\IMG0000006.TIF,,,, |