U.S. Securities and Exchange Commission

Data Delivery Standards

This document describes the technical requirements for paper and electronic document productions to the U.S. Securities and Exchange Commission (SEC).   ***Any questions or proposed file formats other than those described below must be discussed with the legal and technical staff of the SEC Division of Enforcement prior to submission.***

General Instructions

Due to COVID-19 restrictions the current, temporary mailing address for all <u>physical productions</u> sent to the SEC is:
ENF-CPU (U.S. Securities & Exchange Commission), 14420 Albemarle Point Place, Suite 102, Chantilly, VA 20151-1750

Electronic files must be produced in their native format, i.e. the format in which they are ordinarily used and maintained during the normal course of business. For example, an MS Excel file must be produced as an MS Excel file rather than an image of a spreadsheet. *(Note: An Adobe PDF file is <u>not</u> considered a native file unless the document was initially created as a PDF.)*

In the event produced files require the use of proprietary software not commonly found in the workplace, the SEC will explore other format options with the producing party.

The proposed use of file de-duplication methodologies or *computer-assisted review* or *technology-assisted review* (TAR) during the processing of documents must be discussed with and approved by the legal and technical staff of the Division of Enforcement (ENF). If your production will be de-duplicated it is vital that you 1) preserve any unique metadata associated with the duplicate files, for example, custodian name and file location and, 2) make that unique metadata part of your production to the SEC.

Rev 8/2021

General requirements for **ALL** document productions are:

1. A cover letter must be included with each production and should include the following information:
   a. Case number, case name and requesting SEC staff member name
   b. A list of each piece of media included in the production with its unique production volume number
   c. A list of custodians, identifying the Bates range for each custodian
   d. The time zone in which the emails were standardized during conversion
   e. Whether the production contains native files produced from Mac operating system environments
2. Data can be produced on CD, DVD, thumb drive, etc., using the media requiring the least number of deliverables and labeled with the following:
   a. Case number
   b. Production date
   c. Producing party
   d. Bates range (if applicable)
3. All submissions must be organized by **custodian** unless otherwise instructed.
4. All document family groups, i.e. email attachments, embedded files, etc., should be produced together and children files should follow parent files sequentially in the Bates numbering.
5. All load-ready collections should include only one data load file and one image pointer file.
6. All load-ready text must be produced as separate document-level text files.
7. All load-ready collections should account for custodians in the custodian field.
8. All load-ready collections must provide the extracted contents of any container files to ensure all relevant files are produced as separate records.
9. Audio files should be separated from data files if both are included in the production.
10. Only alphanumeric characters and the underscore character are permitted in file names and folder names. Special characters are not permitted.
11. All electronic productions submitted on media must be produced using industry standard self-extracting encryption software.
12. The SEC uses 7zip to access compressed files. Note that the SEC **cannot** accept files that use AES-256 Jpeg or pkAES-256-Cert Deflate compression methods, even if the files are created with 7zip. If you have any questions or need additional information, please reach out to the requesting SEC staff member.
13. Electronic productions of 20 GB or less are strongly encouraged to be submitted via Secure File Transfer. All Secure File Transfers should be sent to the SEC Centralized Production Unit (ENF-CPU@sec.gov) with a CC to the requesting SEC staff member. If you do not have your own Secure File Transfer application, you may reach out to the requesting SEC staff member for a link to the SEC system in order to upload your production. If using the SEC Secure File Transfer system, you will NOT be able to CC individuals outside the SEC on your upload transmission. Note that the SEC **cannot** accept productions made using file sharing sites such as Google Drive, Microsoft Office 365 or Dropbox.
14. Productions containing BSA or SAR material must be delivered on encrypted physical media. The SEC **cannot** accept electronic transmission of BSA or SAR material. Any BSA or SAR material produced should be segregated and appropriately marked as BSA or SAR material, or should be produced separately from other case related material.
15. Passwords for electronic documents, files, compressed archives and encrypted media must be provided separately either via email or in a cover letter apart from the media.
16. All electronic productions should be produced free of computer viruses.
17. Before producing forensically collected images, parties should reach out to the requesting SEC staff member in order to discuss appropriate handling.
18. Before producing unique data sets (large sets of relational data, website reconstruction, chat room data, etc.), parties should reach out to the requesting SEC staff member in order to discuss an appropriate production format.
19. Additional technical descriptions can be found in the addendum 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.

Delivery Formats

I. **Imaged Productions**

The SEC prefers that all scanned paper and electronic file collections be produced in a structured format including industry standard load files, Bates numbered image files, native files and searchable document-level text files.

1. **Images**
   a. Black and white images must be 300 DPI Group IV single-page TIFF files
   b. Color images must be produced in JPEG format
   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. All image files must have a unique file name, i.e. Bates number
   f. Images must be endorsed with sequential Bates numbers in the lower right corner of each image
   g. The number of image files per folder should not exceed 2,000 files
   h. Excel spreadsheets should have a placeholder image named by the Bates number of the file
   i. AUTOCAD/photograph files should be produced as a single page JPEG file

2. **Image Cross-Reference File**

   The image 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. **Data File**

   The data file (.DAT) contains all of the fielded information that will be loaded into the database.

   a. The first line of the .DAT file must be a header row identifying the field names
   b. The .DAT file must use the following *Concordance®* default delimiters:
      Comma ¶ ASCII character (020)
      Quote þ ASCII character (254)
   c. If the .DAT file is produced in Unicode format it must contain the byte order marker
   d. Date fields should be provided in the format: mm/dd/yyyy
   e. Date and time fields must be two separate fields
   f. The time zone must be included in all time fields
   g. If the production includes imaged emails and attachments, the attachment fields must be included to preserve the parent/child relationship between an email and its attachments
   h. An OCRPATH field must be included to provide the file path and name of the extracted text file on the produced storage media. The text file must be named after the FIRSTBATES. Do not include the text in the .DAT file.
   i. For productions with native files, a LINK field must be included to provide the file path and name of the native file on the produced storage media. The native file must be named after the FIRSTBATES.
   j. BEGATTACH and ENDATTACH fields must be two separate fields
   k. A complete list of metadata fields is available in **Addendum A** to this document

4. **Text**

   Text must be produced as separate document-level text files, not as fields within the .DAT file. The text files must be named per the FIRSTBATES/Image Key and the full path to the text file (OCRPATH) should be included in the .DAT file. Text files may be in either ANSI or Unicode format, however, ALL text files must be in the same format within the same production. Note that productions containing text with foreign characters must produce text files in Unicode format to preserve the foreign characters. Text files must be in a separate folder, and the number of text files per folder should not exceed 2,000 files. There should be no special characters (including commas) in the folder names. For redacted documents, provide the full text for the redacted version.

5. **Linked Native Files**

   Copies of original email and native file documents/attachments must be included for all electronic productions.
   a. Native file documents must be named per the FIRSTBATES number
   b. The full path of the native file must be provided in the .DAT file for the LINK field
   c. The number of native files per folder should not exceed 2,000 files

II.   Native File Production without Load Files

With prior approval, native files may be produced without load files. The native files must be produced as they are maintained in the normal course of business and organized by custodian-named file folders. When approved, native email files (.PST or .MBOX) may be produced. A separate folder should be provided for each custodian.

III.   Adobe PDF File Production

With prior approval, Adobe PDF files may be produced in native file format.

1. All PDFs must be unitized at the document level, i.e., each PDF must represent a discrete document.
2. PDF files should be produced in separate folders named by the custodian. The folders should not contain any special characters (including commas).
3. All PDF files must contain embedded text that includes all discernible words within the document, not selected text or image only. This requires all layers of the PDF to be flattened first.
4. If PDF files are Bates endorsed, the PDF files must be named by the Bates range.

IV.   Audio Files

Audio files from telephone recording systems must be produced in a format that is playable using Microsoft Windows Media Player™. Additionally, the call information (metadata) related to each audio recording MUST be provided. The metadata file must be produced in a delimited text format. Field names must be included in the first row of the text file. The metadata must include, at a minimum, the following fields:

1) Caller Name: Caller's name or account/identification number
2) Originating Number: Caller's phone number
3) Called Party Name: Called party's name
4) Terminating Number: Called party's phone number
5) Date: Date of call
6) Time: Time of call
7) Filename: Filename of audio file

V.   Video Files

Video files must be produced in a format that is playable using Microsoft Windows Media Player™.

VI.   Electronic Trade and Bank Records

When producing electronic trade records, bank records, or financial statements, provide the files in one of the following formats:

1. MS Excel spreadsheet with header information detailing the field structure. If any special codes exist in the dataset, a separate document must be provided that details all such codes. If details of the field structure do not fit in the header, a separate document must be provided that includes such details.

2. Delimited text file with header information detailing the field structure. The preferred delimiter is a vertical bar "|". If any special codes exist in the dataset, a separate document must be provided that details all such codes. If details of the field structure do not fit in the header, a separate document must be provided that includes such details.

VII.   Electronic Phone Records

When producing electronic phone records, provide the files in the following format:

1. MS Excel spreadsheet with header information detailing the field structure. If any special codes exist in the dataset, a separate document must be provided that details all such codes. If details of the field structure do not fit in the header, a separate document must be provided that includes such details. Data must be formatted in its native format (i.e. dates in a date format, numbers in an appropriate numerical format, and numbers with leading zeroes as text).
    a. The metadata that must be included is outlined in **Addendum B** of this document. Each field of data must be loaded into a separate column. For example, Date and Start_Time must be produced in separate columns and not combined into a single column containing both pieces of information. Any fields of data that are provided in addition to those listed in **Addendum B** must also be loaded into separate columns.

4

VIII. Audit Workpapers

The SEC prefers for workpapers to be produced in two formats: (1) With Bates numbers in accordance with the SEC Data Delivery Standards; and (2) in native format or if proprietary software was used, on a standalone laptop with the appropriate software loaded so that the workpapers may be reviewed as they would have been maintained in the ordinary course of business. The laptop must have printing capability, and when possible, the laptop should be configured to enable a Virtual Machine (VM) environment.

IX. Mobile Device Data

Before producing mobile device data (including but not limited to text messages) parties should reach out to the requesting SEC staff member in order to discuss the appropriate production format.

ADDENDUM A

The metadata of electronic document collections 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 **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 **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 **The CHILD_BATES field should be populated in each record representing a "parent" document
CUSTODIAN	Smith, John	Email: Mailbox where the email resided Native: Name of the individual or department from whose files the document originated
FROM	John Smith	Email: Sender Native: Author(s) of document **semi-colon should be used to separate multiple entries
TO	Coffman, Janice; LeeW [mailto:LeeW@MSN.com]	Recipient(s) **semi-colon should be used to separate multiple entries
CC	Frank Thompson [mailto: frank_Thompson@cdt.com]	Carbon copy recipient(s) **semi-colon should be used to separate multiple entries
BCC	John Cain	Blind carbon copy recipient(s) **semi-colon should be used to separate multiple entries
SUBJECT	Board Meeting Minutes	Email: Subject line of the email Native: Title of document (if available)
FILE_NAME	BoardMeetingMinutes.docx	Native: Name of the original native file, including extension
DATE_SENT	10/12/2010	Email: Date the email was sent 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. Native: (empty) **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. Email: Time zone Native: (empty)

6

LINK	D:\001\ EDC0000001.msg	Hyperlink to the email or native file document **The linked file must be named per the FIRSTBATES number
MIME_TYPE	application/msword	The content type of an email or native file document as identified/extracted from the header
FILE_EXTEN	MSG	The file type extension representing the email or native file document; will vary depending on the format
AUTHOR	John Smith	Email: (empty) Native: Author of the document
LAST_AUTHOR	Jane Doe	Email: (empty) Native: Last Author of the document
DATE_CREATED	10/10/2010	Email: (empty) Native: Date the document was created
TIME_CREATED/TIME_ZONE	10:25 AM GMT	Email: (empty) Native: Time the document was created including time zone **This data must be a separate field and cannot be
DATE_MOD	10/12/2010	Email: (empty) Native: Date the document was last modified
TIME_MOD/TIME_ZONE	07:00 PM GMT	Email: (empty) Native: Time the document was last modified including the time zone **This data must be a separate field and cannot be
DATE_ACCESSD	10/12/2010	Email: (empty) Native: Date the document was last accessed
TIME_ACCESSD/TIME_ZONE	07:00 PM GMT	Email: (empty) Native: Time the document was last accessed including the time zone **This data must be a separate field and cannot be
PRINTED_DATE	10/12/2010	Email: (empty) Native: Date the document was last printed
FILE_SIZE	5,952	Size of native file document/email in KB
PGCOUNT	1	Number of pages in native file document/email
PATH	J:\Shared\SmithJ\October Agenda.doc	Email: (empty) 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. Native: (empty)
INTMSGID	<000805c2c71b75977050cb8306d1@MSN>	Email: Unique Message ID Native: (empty)

HEADER	Return-Path: <example_from@dc.edu> X-SpamCatcher-Score:1[X] Received:from[136.167.40.119] (HELO dc.edu) by fe3.dc.edu (CommuniGate Pro SMTP4.1.8) with ESMTP-TLS id 61258719 for example_to@mail.dc.edu; Mon, 23 Aug 2004 11:40:10 -0400 Message-ID: <4129F3CA.2020509@dc.edu> Date: Mon, 23 Aug 2005 11:40:36 -400 From: Taylor Evans <example_from@dc.edu> User-Agent:Mozilla/5.0 (Windows;U; Windows NT 5.1; en-US;rv:1.0.1) Gecko/20020823 Netscape/7.0 X-Accept-Language:en-us,en MIME-Version:1.0 To: Jon Smith <example_to@mail.dc.edu> Subject:Business Development Meeting Content-Type: text/plain;charset=us-ascii; format=flowed Content-Transfer-Encoding:7bit	Email: The email header information Native: (empty)
MD5HASH	d131dd02c5e6eec4693d9a069 8aff95c 2fcab58712467eab4004583eb 8fb7f89	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,,,,
```

Rev 8/2021

ADDENDUM B

For Electronic Phone Records, include the following fields in separate columns:

For Calls:

1) Account Number
2) Connection Date – Date the call was received or made
3) Connection Time – Time call was received or made
4) Seizure Time – Time it took for the call to be placed in seconds
5) Originating Number – Phone that placed the call
6) Terminating Number – Phone that received the call
7) Elapsed Time – The length of time the call lasted, preferably in seconds
8) End Time – The time the call ended
9) Number Dialed – Actual number dialed
10) IMEI Originating – Unique id to phone used to make call
11) IMEI Terminating– Unique id to phone used to receive call
12) IMSI Originating – Unique id to phone used to make call
13) IMSI Terminating- Unique id to phone used to receive call
14) Call Codes – Identify call direction or other routing information
15) Time Zone – Time Zone in which the call was received or placed, if applicable

For Text messages:

1) Account Number
2) Connection Date – Date the text was received or made
3) Connection Time – Time text was received or made
4) Originating Number – Who placed the text
5) Terminating Number – Who received the text
6) IMEI Originating – Unique id to phone used to make text
7) IMEI Terminating– Unique id to phone used to receive text
8) IMSI Originating - Unique id to phone used to make text
9) IMSI Terminating- Unique id to phone used to receive text
10) Text Code – Identify text direction, or other text routing information
11) Text Type Code – Type of text message (sent SMS, MMS, or other)
12) Time Zone – Time Zone in which the call was received or placed, if applicable

For Mobile Data Usage:

1) Account Number
2) Connection Date – Date the data was received or made
3) Connection Time – Time data was received or made
4) Originating number – Number that used data
5) IMEI Originating – Unique id of phone that used data
6) IMSI Originating - Unique id of phone that used data
7) Data or Data codes – Identify data direction, or other data routing information
8) Time Zone – Time Zone in which the call was received or placed, if applicable