# CFTC Data Delivery Standards

Effective: May 1, 2015

This document describes the technical requirements for electronic document productions to the Commodity Futures Trading Commission (CFTC). Any proposed file formats other than those described below must be discussed with the legal and technical staff of the CFTC Division of Enforcement prior to submission.

# CFTC Data Delivery Standards

***Effective: May 1, 2015***

This document describes the technical requirements for electronic document productions to the Commodity Futures Trading Commission (CFTC).

****Any proposed file formats other than those described below must be discussed with the legal and technical staff of the CFTC Division of Enforcement prior to submission.****

## Contents


General Instructions ..... 2

Electronic Discovery Protocols ..... 3

Delivery Formats ..... 3

I. Native File Production ..... 3

1. Emails ..... 3
2. Account Statements ..... 3
3. Instant Messages (IMs) ..... 3
4. Audio Files ..... 4
5. Video Files ..... 4
6. Transcripts ..... 5

II. Imaged Collections ..... 5

1. Images ..... 5
2. Concordance Image® Cross-Reference File ..... 5
3. Data File ..... 6
4. Text ..... 9
5. Linked Native Files ..... 9

III. Productions of Adobe PDF Files ..... 9

IV. Productions of Website Content ..... 9

V. Productions of Forensic Images of Computer Media ..... 10

VI. Productions of Forensically Acquired Mobile Device Data ..... 10

VII. FTP (File Transfer Protocol) Submission ..... 10

## General Instructions

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 **not** 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 CFTC will explore other format options with the producing party.

In regards to imaged collections, the use of file de-duplication methodologies in preparing productions is becoming more commonplace. 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, 2) make that unique metadata part of your production to the CFTC.

General requirements for ALL document productions are:

1. **Production of data via email is unacceptable.**
2. Reference the specific portion of the request to which you are responding, along with a summary of the number of files in the production, so that we can confirm everything is loaded into our review system.
3. All native file submissions must be organized **by custodian** unless otherwise instructed.
4. All load-ready collections should include only one data load file and one image pointer file.
5. All load-ready text must be produced as separate text files, not as fields within the .DAT file.
6. All load-ready collections should account for custodians in the custodian field.
7. Audio files should be separated from data files if both are included in the production.
8. The preferred method of submission is via File Transfer Protocol (FTP). (See details in Section IV on page 10.)
9. Large productions (any production above 10 GB) must be submitted to the CFTC on media such as a CD, DVD, thumb drive, or hard drive. The media must be clearly marked with the following:
    a. Matter name
    b. Producing party
    c. Production date
    d. Disk number (1 of X), if applicable
10. Only alphanumeric characters and the underscore character are permitted in file names and **folder names**. Special characters are not permitted. Any data received with file names or folders with special characters, including the comma, will be rejected. The combined length of the file path with folder and file name should not exceed 255 characters.
11. Include information within the cover letter regarding the time zone used, if emails or other electronic files were standardized during conversion.
12. All production data containing sensitive or personally identifiable information must be encrypted using FIPS 140 compliant software.
13. Provide passwords for all password-protected files or hardware in a separate transmittal.

14. All productions should be checked and produced free of computer viruses and malware. Productions containing viruses or malware may not be accepted and will require a new production.

## Electronic Discovery Protocols

If a producing party wants to employ winnowing techniques to reduce the volume of production material, then consultation with the CFTC attorney is required to define and agree upon the requirements and applied technology. Thereafter, all productions made to the CFTC should reference the established protocol. Such methodologies may include, but are not limited to, the following criteria:

- Data Sources
- Custodians
- Date Range(s)
- Search Terms
- Deduplication
- Email Threading Usage
- Early Case Assessment (ECA) Usage
- Technology Assisted Review (TAR)
- Other Criteria, including alternative forms of analytics

## Delivery Formats

### I. Native File Production

Requirements for the production of native document files are listed below.

1. **Emails**: Emails and attachments must be produced as PST, NSF, DBX, MBOX, or MSG files. The file name must include the name of the email custodian. When a production consists of multiple custodians, a separate PST (or similar file) should be produced for each custodian.
2. **Account Statements**: Account statements must be produced in the same format in which they were originally created and distributed.
3. **Instant Messages (IMs)**: IMs should be produced in .PST format or a delimited text file which must include, at a minimum, the following fields:
    1) DATE
    2) TIME
    3) FROM
    4) TO
    5) CONVERSATION_TEXT
    6) CONVERSATION_INDEX

    Multiple conversations must be produced in a single file and a conversation index or similar unique string must be used to identify all threads of the same conversation. Field names must be included in the first row of the text file.

4. **Audio Files**: Audio files from telephone recording systems must be produced in a format that is playable using Microsoft Windows Media Player™. Types of audio files that will be accepted include:
   - Nice Systems audio files (.aud). AUD files offer efficient compression and would be preferred over both NMF and WAV files.
   - Nice Systems audio files (.nmf).
   - WAV files
   - MP3, MP4
   - WMA
   - AIF

   Produced audio files must be in a separate folder compared to other data in the production.

   Additionally, the call information (metadata) related to each audio recording **must be produced if it exists**. The metadata must include, at a minimum, the following fields:

   1) CALLER_NAME or CALLER_ID: Caller's name or identification number
   2) CALLING_NUMBER: Caller's phone number
   3) FILENAME: Filename of audio file
   4) DATE: Date of call
   5) TIME: Time of call
   6) CALLED_PARTY: Name of the party called
   7) CALLED_NUMBER: Called party's phone number

   The filename is used to link the metadata to the produced audio file. The file name in the metadata and the file name used to identify the corresponding audio file must match exactly. The metadata file must be produced in delimited text format. Field names must be included in the first row of the text file. A sample .DAT file containing only a few fields for illustration purposes is depicted below:

```
_Type¶ þAudio_Trimmedþ¶ þCALLER_IDþ¶ þCUSTODIANþ¶ þDATEþ¶ þTIMEþ¶ þCALLED_NUMBERþ¶ þ
þ07/09/2007 09:33:03þ¶ þSpeakersþ¶ þþ¶ þþ¶ þDoe, Janeþ¶ þ07/09/2007þ¶ þ09:32:42þ¶ þþ¶
þ07/17/2007 09:17:36þ¶ þSpeakersþ¶ þþ¶ þþ¶ þDoe, Janeþ¶ þ07/17/2007þ¶ þ09:17:24þ¶ þþ¶
þ10/12/2007 08:13:12þ¶ þSpeakersþ¶ þþ¶ þþ¶ þDoe, Janeþ¶ þ10/12/2007þ¶ þ08:12:47þ¶ þþ¶
þ10/12/2007 08:18:51þ¶ þSpeakersþ¶ þþ¶ þþ¶ þDoe, Janeþ¶ þ10/12/2007þ¶ þ08:18:19þ¶ þþ¶
þ12/05/2007 07:12:14þ¶ þSpeakersþ¶ þþ¶ þþ¶ þDoe, Janeþ¶ þ12/05/2007þ¶ þ07:11:58þ¶ þþ¶
þ12/17/2007 08:40:17þ¶ þSpeakersþ¶ þþ¶ þþ¶ þDoe, Janeþ¶ þ12/17/2007þ¶ þ08:40:09þ¶ þþ¶
þ12/31/2007 09:39:12þ¶ þSpeakersþ¶ þþ¶ þþ¶ þDoe, Janeþ¶ þ12/31/2007þ¶ þ09:39:00þ¶ þþ¶
```

5. **Video Files**: Video files must be produced in a format that is playable using Microsoft Windows Media Player™ along with any available metadata. If it is known that the video files do not contain associated audio, indicate this in the accompanying transmittal letter. Types of video files accepted include:
   - MPG
   - AVI
   - WMV
   - MOV
   - FLV

6. **Transcripts**: Legal transcripts must be produced electronically in manuscript form with line numbers and page numbers. The preferred format is plain text (ASCII), although LiveNote and Summation formats are acceptable.

## II. Imaged Collections

While the CFTC accepts imaged productions in addition to native formats, imaged productions without native formats are not permitted unless the original document only exists in hard copy form. When images are produced, they must comply with the requirements below.

*****The use of file de-duplication methodologies in preparing productions is becoming more commonplace. 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, 2) make that unique metadata part of your production to the CFTC.*****

Note: Adobe PDF files are **not** acceptable as imaged productions. PDF files are acceptable only when the document content was initially created as a PDF (e.g., fillable PDF forms) and not converted from another format.

### 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.
f. Images must be endorsed with sequential Bates numbers in the lower right corner of each image.
g. The number of TIFF files per folder should not exceed 500 files.
h. TIFF images of Excel spreadsheets are not useful for review purposes; because the imaging process can often generate thousands of pages per file, a placeholder image, named by the *IMAGEID* of the file, should be used instead, and the native Excel file should be produced.

### 2. Concordance 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:

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

*ImageKey*: This is the unique designation that Concordance and Concordance Image and Relativity use to identify and retrieve an image. This value may be the same as the Bates number endorsed on each image.

*VolumeLabel:* Leave this field empty.

*ImageFilePath:* This is the full path to the image file on the produced storage media.

*DocumentBreak:* This field is used to delineate the beginning of a new document. If this field contains the letter "Y," then this is the first page of a document. If this field is blank, then this page is not the first page of a document.

*FolderBreak:* This field is used to delineate the beginning of a new folder in the same manner as the *DocumentBreak* field. If this information is not available, then it may be left empty.

*BoxBreak:* This field is used to delineate the beginning of a new box in the same manner as the *DocumentBreak* and *FolderBreak* fields. If this information is not available, then it may be left empty.

*PageCount:* Leave this field empty.

Sample

```
LA-0000001,,E:\001\ LA-0000001.TIF,Y,,,
LA-0000002,,E:\001\ LA-0000002.TIF,,,,
LA-0000003,,E:\001\ LA-0000003.TIF,Y,,,
LA-0000004,,E:\001\ LA-0000004.TIF,,,,
LA-0000005,,E:\001\ LA-0000005.TIF,,,,
```

**3. Data File**

The data file (.DAT) contains all of the fielded information that will be loaded into the *Concordance*® or *Relativity*® database.

a. The first line of the .DAT file must be a header record identifying the field names.
b. The .DAT file must use the following *Concordance*® default delimiters:

| | | |
|---|---|---|
| Comma | | ASCII character 20 |
| Quote | þ | ASCII character 254 |
| Newline | ® | ASCII character 174 |

c. Date fields should be provided in the format: MM/DD/YYYY.
d. 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.
e. A TEXTPATH 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 Begno. Do not include the text in the .DAT file.
f. For production with native files, a NATIVELINK field must be included to provide the file path and name of the native file on the produced storage media.
g. BegAtt and EndAtt fields must be two separate fields.
h. DateSent and TimeSent fields must be two separate fields.

i. All text and metadata associated with the document collection must be produced.

Sample of .DAT file (only includes a sample of fields)

```
þBegnoþþBegAttachþþEndAttachþþFileDescripþþFilenameþþRecordTypeþ
þEML-0030437þþþþþþOutlook Data FileþþQuestion Fill Out Form.htmþþþ
þEML-0030438þþþþþþOutlook Data FileþþRE: Your Question.htmþþþ
þEML-0030439þþþþþþOutlook Data FileþþQuestion Fill Out Form.htmþþþ
þEML-0030440þþþþþþOutlook Data FileþþLook.htmþþE-MAILþþþ
þEML-0030441þþþþþþOutlook Data FileþþQuestion Fill Out Form.htmþþþ
þEML-0030442þþþþþþOutlook Data FileþþQuestion Fill Out Form.htmþþþ
þEML-0030443þþþþþþOutlook Data FileþþQuestion Fill Out Form.htmþþþ
þEML-0030444þþþþþþOutlook Data FileþþQuestion Fill Out Form.htmþþþ
þEML-0030445þþþþþþOutlook Data FileþþATTN.htmþþE-MAILþþþ
þEML-0030446þþþþþþOutlook Data FileþþQuestion Fill Out Form.htmþþþ
þEML-0030447þþþþþþOutlook Data FileþþQuestion Fill Out Form.htmþþþ
þEML-0030448þþþþþþOutlook Data FileþþQuestion Fill Out Form.htmþþþ
```

The metadata for the document collection should be provided in a .DAT file using the field definition and formatting described below:

| Field | Description | Field Type | Required |
|---|---|---|---|
| Begno | Displays the document identifier of the first page in a document or the entire document of an E-Doc. | Text | Yes |
| Endno | Page ID of the last page in a document (for image collections only). | Text | Yes for Image Collections |
| BegAttach | Displays the document identifier of a parent record. Must be separate from EndAttach field. | Text | If it Exists |
| EndAttach | Displays the document identifier of the last attached .document in a family. Must be separate from BegAttach field. | Text | If it Exists |
| PgCount | Number of pages in a document (for image collections only). | Text | Yes for Image Collections |
| FileDescription | Description of a native file type. | Text | Yes |
| Filename | Original filename of a native file, | Text | Yes for Documents |
| RecordType | Displays the record type for each entry in the load file. | Text | Yes |
| ParentID | Displays the document identifier of the attachment record's parent (only for attachments). | Text | If it Exists |
| NumAttach | Total number of records attached to the document. The value will always be 0 (zero) for the actual attachment records. | Text | If it Exists |

| Field | Description | Field Type | Required |
|---|---|---|---|
| Attachmt | Populates parent records with document identifier of each attached record and is separated by semi-colons. | Text | If it Exists |
| Custodian | The owner of the record. | Text | Yes |
| From | Author of the e-mail message. | Text | Yes for email |
| To | Main recipient(s) of the e-mail message. | Text | Yes for email |
| CC | Recipient(s) of "Carbon Copies of the e-mail message. | Text | If it Exists |
| BCC | Recipient(s) of "Blind Carbon Copies" of the e-mail message. | Text | If it Exists |
| EMail_Subject | Subject of the e-mail message. | Paragraph | Yes for email |
| DateSent | Sent date of an e-mail message. Must be separate from TimeSent field. | MM/DD/YYYY | Yes for email |
| TimeSent | Time the e-mail message was sent. Must be separate from DateSent field. | Text | Yes for email |
| IntMsgID | Internet Message ID assigned to an e-mail message by the outgoing mail server. | Text | If it Exists |
| ConversationIndex | This is a 44 character string of numbers and letters that is created in the initial email and which as 10 characters added for each reply or forward of an email. | Text | Yes for email |
| Conversation Family | E-mail thread identification. Relational field for Conversion threads. | Text | Yes for email |
| EntryID | Unique identifier of e-mails in mail stores. | Text | If it Exists |
| Author | Author value pulled from metadata of the native file. | Text | If it Exists |
| Organization | Company extracted from metadata of the native file. | Text | If it Exists |
| Subject | Subject value extracted from metadata of the native file. | Paragraph | If it Exists |
| DateCreated | Creation date of the native file. | MM/DD/YYYY | If it Exists |
| DateLastMod | Date the native file was last modified. | MM/DD/YYYY | If it Exists |
| DateLastPrnt | Date the native file was last printed. | MM/DD/YYYY | If it Exists |
| MD5Hash | MD5 hash value. | Text | Yes |
| EDSource | Fully qualified original path to the source folder, files, and/or mail stores. | Text | Yes |
| NativeFile | Hyperlink to the native file. | Text | Yes |
| Textpath | Extracted text path. | Paragraph | Yes |
| Imagkey | Page ID of the first page of a document (for image collections only). Should be the same as the Begno field. | Paragraph | Yes |
| + Any other fields considered relevant by the producing party. | | | |

### 4. Text

Text must be produced as separate text files, not as fields within the .DAT file. If text is included in the .DAT file as a field, the production will be rejected. Extracted text must be in a separate folder, one text file per document. The files must be named the same as the Begno field. The number of files per folder should not exceed 500 files. There should be no special characters (including commas in the folder names).

### 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 Begno field.
b. The full path of the native file must be provided in the .DAT file for the NativeFile field.
c. The number of native files per folder should not exceed 500 files.
d. There should be no special characters (including commas in the folder names).

## III. Productions of Adobe PDF Files

PDF files are not substitutions for native files, as metadata is lost when PDFing native files. PDF productions that are produced as native productions will be rejected unless the files were originally created using Adobe (e.g., fillable electronic PDF forms) or prior agreement is reached with the designated CFTC attorney. In those exceptions, PDF productions should adhere to the following guidelines:

1. PDF files should be produced in separate folders named by the custodian. The folders should not contain any special characters (including commas).
2. All PDFs must be unitized at the document level, i.e., each PDF should represent a discrete document; **a PDF cannot contain multiple documents. PDF packages will be rejected.**
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. Productions of Website Content

Website content must be produced in native format and viewable in common web browsers (e.g. Internet Explorer, Mozilla Firefox, and Google Chrome) without the use of any additional third party software. If the only existing state of a website is in source code form, the party must produce a fully compiled version of the website in addition to the website source code with documentation of the process used to render, compile, and/or restore the website to an easily viewable form.

## V. Productions of Forensic Images of Computer Media

Forensic images must be produced in a common/standard format to include (E01, Ex01, L01, Lx01, AD1, Raw Bitstream, and S01). An inventory of any original computer media must be provided with the image files. All associated log files from any forensic acquisition must be produced with the image files. The log files must contain a hash value, at a minimum (MD5, SHA-1, or higher), for authentication purposes. A chain of custody document must be provided to ensure the integrity of any digital evidence.

## VI. Productions of Forensically Acquired Mobile Device Data

Forensically acquired mobile device data must be provided in the original format of the software/hardware used to capture the data. All original logs and authentication reports must also be provided with the respective data.

## VII. FTP (File Transfer Protocol) Submission

The preferred method for producing data is via FTP. The CFTC has an FTP site set up that will allow transfer of data up to 2 GB per transmission.

Contact the CFTC attorney or paralegal for the particular matter in order to use FTP to deliver data. Once this method is agreed upon, you will receive two emails. One email will have instructions on using the FTP site; the other will contain a password for accessing the site.