Tobacco Health Document Submission

Technical Working Specifications

I. OVERVIEW

On June 22, 2009, the President signed the Family Smoking Prevention and Tobacco Control Act (Tobacco Control Act) (Public Law 111-31) into law. Among its many provisions, the Tobacco Control Act added section 904(a)(4) to the Federal Food, Drug, and Cosmetic Act (the act) (21 U.S.C. §387(d)(a)(4)), requiring submission of documents related to certain effects of tobacco products. To view the Tobacco Control Act, see Public Law citation (Pub. Law 111-31).

II. STATUTORY REQUIREMENTS

Section 904(a)(4) of the act requires each tobacco product manufacturer or importer, or agent thereof, to submit all documents developed after June 22, 2009 "that relate to health, toxicological, behavioral, or physiologic effects of current or future tobacco products, their constituents (including smoke constituents), ingredients, components, and additives."

Please note: For tobacco products that are or will be imported, the required health documents are to be submitted by either the foreign manufacturer or the importer. Identify whether the submitter is the manufacturer or importer.

For information regarding the section 904(a)(4) requirements, please refer to the Final Guidance: Tobacco Health Document Submission.

III. SUBMISSION FORMATS

Although electronic submission of this information is not required, FDA is strongly encouraging electronic submission to facilitate efficiency and timeliness of data submission and management. To submit electronically, FDA recommends the following options:

1) For low volume submissions: FDA eSubmitter Tool

The FDA eSubmitter tool is an electronic application designed for guiding the user through preparing certain digital submissions for FDA. This tool provides a method to allow users to attach large numbers of files (e.g., PDF documents, TIFF files) to a single Tobacco Health Document submission. For more information, visit the FDA eSubmitter for Tobacco Products webpage.

2) For high volume submissions: Digital Production with Definition Load File

If you choose not to use eSubmitter for submission of health documents under section 904(a)(4) of the act, FDA recommends digital production of all documents on a hard drive, CD or DVD, depending on the volume of the submission.

The remainder of this document provides further instruction on how to compile and transmit a 904(a)(4) submission in this digital format.

IV. SUBMITTING DOCUMENTS AND PREPARING INDICES

Recommended Format if Submitting for High Volume Submissions

The CTP recommends that all documents be submitted in the applicable format as outlined below.

1. Form and Manner of the Submission

   It is recommended that the submission include the paper form 3743 as an accompanying cover letter to ensure that the CTP has received everything that the submitting party intended to send in the digital format.

   The CTP recommends that the submission be produced on CDs or DVDs. Depending on the volume of submission content, an encrypted hard drive(s) is an acceptable option for submission. If using a hard drive, FDA recommends FIPS 140-2 encrypted external hard drive.

   The media provided (e.g., a hard drive) should be labeled with the type of submission (i.e., Tobacco Health Document Submission), the responsible party submitting the documents, and any references necessary to link to the information in the cover letter.

2. Bates Numbering Identification

   The CTP recommends that a consistent Bates numbering format be used across the entire submission. In general, Bates numbers should use an alpha-numeric format: for example, a company identifier (e.g., “ACME”), a hyphen, and a numerical sequence of 6-12 digits (the number of digits should be consistent across the entire production). Bates numbers should not contain embedded spaces (" "), slashes ("/"), backslashes ("\"), or underscores ("_"). Native format files should also be assigned a Bates number; if images have been produced with a native format file, the beginning Bates number of the images should be the native format file's Bates number.

3. Recommended Metadata and Fields to be Submitted as a Load File with each Document Image

   The load file may be in a tab or comma delimited ASCII format. We recommend that the file include a header row, and contain the following fields, to the extent such information is available. Please see the Final Guidance for additional details on required fields.

Field name	Field Type	Field Description
Submitter	Note Text	Name of party submitting the document.
Begin Bates	Note Text	Number of the first page of the document.
End Bates	Note Text	Number of the last page of the document.
Page Count	Integer	Number of pages within the document.
Document Author(s)	Note Text	The name of the email sender (format: Last name, First name, middle name). For hard copy documents, this is the document's natural author, as appears on the face of the document.
Document Recipient(s)	Multi-Entry	The name of all email recipients in the "To" field (format: Last name, First name, middle name). For hard copy documents, this identifies all of the recipients as appears on the face of the document.
CC	Multi-Entry	The name and email address of all email recipients in the "Cc" field (format: Last name, First name, middle name). For hard copy documents, this identifies all of the recipients as appears on the face of the document.
BCC	Multi-Entry	The name and email address of all email recipients in the "Bcc" field (format: Last name, First name, middle name). For hard copy documents, this identifies all of the recipients as appears on the face of the document.
Document Title or Identification Number	Note Text	For E-mail documents, the subject line of the E-mail. This field also refers to the title of other types of documents.
Date of Creation	Date	For E-mail documents, this is the date the message was sent. For all other documents, this is the last modified date or the date that appears on the face of the document. (Format: MM/DD/YYYY)
OCR/Full Text	Note Text	The text of the full page as displayed on the document.
Attach Begin Bates	Note Text	The Begin Bates Number of the top level parent document.
Attach End Bates	Note Text	The End Bates Number of the last attachment in the document family.
Parent/Attachment	Note Text	Value is either P or A ("P" for Parent; "A" for Attachment)
Document Custodian	Multi-Entry	The owner of the data.
Document Category	Multi-Entry	See CTP's Final Guidance concerning section 904(a)(4) of the Tobacco Control Act regarding this field.
Specific Current or Future Tobacco Product(s)	Multi-Entry	See CTP's Final Guidance concerning section 904(a)(4) of the Tobacco Control Act regarding this field.

Field name	Field Type	Field Description
Class of Current or Future Tobacco Product(s)	Multi-Entry	See CTP's Final Guidance concerning section 904(a)(4) of the Tobacco Control Act regarding this field.
Specific Ingredient(s), Constituent(s), Component(s), or Additive(s)	Multi-Entry	See CTP's Final Guidance concerning section 904(a)(4) of the Tobacco Control Act regarding this field.
Class of Ingredient(s), Constituent(s), Component(s), or Additive(s)	Multi-Entry	See CTP's Final Guidance concerning section 904(a)(4) of the Tobacco Control Act regarding this field.
File Name	Note Text	The original name of the file (for electronic documents). For e-mails, the file name is represented as the subject line plus the appropriate extension. (.msg, .htm, .rtf, etc…)
File Extension	Note Text	For electronic documents, the extension of the file (.doc, .xls, .pdf, .jpg, etc.).
Doc Link	Note Text	File path location to the native file.

4. Load File Recommendations

The CTP recommends that each page of the Load File be submitted with a resolution of no less than 300 dpi and labeled with a sequential Bates number during the imaging process. CTP recommends that the Load File be submitted in black and white TIFF or PDF format, unless the documents are illegible in that form. CTP recommends that documents be scanned in color in .JPG format if they are otherwise illegible.

The CTP recommends that the documents be submitted in single page TIFF or PDF format (or in .JPG format as noted above), accompanied by a delimited ASCII load file.

The CTP recommends that each submission have a Volume Number associated with it. On the root of the media, a top-level folder should be included by the Volume Number. The folder naming scheme of this folder should be 2 to 3 letters (indicating the party submitting the documents) followed by 3 numbers (e.g. ABC001).

Load files should not contain more than 50,000 - 80,000 records each, per Volume.

Under this Volume # folder should be 3 (or 4) folders:

1) Doclink (contains native files)
2) Images
3) FullText
4) LoadFiles

Metadata Load File Delimiters

Field separator: **Vertical Pipe (ASCII 124)**
Field encapsulate: **Carat (ASCII 094)**
Return value in data: **Tilde (ASCII 126)**

Multi-value field: **Semi Colon (ASCII 059)**
Dates format: **MM/DD/YYYY**

Note: Hard Returns should appear only at the end of each record. Below is a sample of how the data will appear in the load file (this sample only contains certain of the recommended fields).

^BEGNUM^|^ENDNUM^|^PAGES^|^DATE^|^AUTHOR^|^RECIPIENT^|^CC^|^TITLE^|^DOCTYPE^
^ACME -000001^|^ACME-000006^|^6^ 06/30/2009 ^Smith, Joe^|^Jones, Mary^|^^|^Sample^|^EMAIL^

5. Sample Recommended Prior to Full Submission

The CTP recommends that a sample load file be submitted for an initial submission to FDA in this format to allow for testing and comment before the full submission is produced. To minimize the possibility of encountering problems during the full submission, the sample should be representative of an actual production, including multiple types of documents and documents with attachments.

The CTP further recommends that any vendor or IT personnel involved with the submission participate in any discussions relating to the technical aspects of the submission.

A. Categories of Documents

Any discussion regarding the details of an electronic production should focus on three categories of documents: (1) hard-copy, (2) e-mail and other electronic messages (e.g., instant messaging), and (3) other electronic documents. General recommendations for each category of document are outlined below. For information regarding document-specific metadata and bibliographic information (identifying information), please refer to the enclosed table.

1. Hard-Copy (or Paper) Documents

Hard-copy documents should be produced as image files with related searchable OCR text and bibliographic information. Special attention should be paid to ensure that hard-copy documents are produced as they are kept, reflecting attachment relationships between documents. In addition, multi-page documents should be properly unitized, and not as several single-page documents.

2. E-mail and Other Electronic Messages

E-mail and other electronic messages (e.g., instant messages (IMs)) should be produced as image files with related searchable text and metadata and bibliographic information. Each IM conversation should be produced as one document. Depending on how the company's systems represent names in e-mail messages or IMs, we may request a table of names or contact lists from custodians.

3. Electronic Documents

Electronic documents include word-processing documents, spreadsheets, presentations, and all other electronic documents not specifically discussed elsewhere in this document. Except as noted below, CTP recommends that submission of these documents should include image files with related searchable text and metadata and bibliographic information. Both PowerPoint and Excel files should be produced natively, with links referenced in the DOCLINK field.

A. Spreadsheets

CTP recommends that spreadsheets be submitted in native format (e.g., as .XLS files). Spreadsheets are often voluminous and scanned images are not useful for review purposes, including the related searchable text, metadata and bibliographic information.

B. Presentations

CTP recommends that presentations be produced in TIFF or PDF format, imaged in full slide format. Speaker notes or any other comments should follow the full images of the slides (to both ease review and reduce volume and costs of production) with related searchable text, metadata and bibliographic information.

4. Native Format Files

Native format files require that the CTP has the associated software and any necessary licenses to view the documents. Please notify the CTP regarding any documents to be produced from electronic documents in any native format to make sure that the CTP has the capability to open and view these documents. Password protection should be removed from all files. Additionally, the CTP recommends that each submitting party provide a list of all native formats that will be provided to the CTP.

Native files should be organized in a separate folder from text or images following the same folder structure guidelines. A native cross reference file should be provided.

5. Image and Text Files

The CTP recommends that image files be submitted in the following manner:

- Images should be submitted as single-page TIFF or PDF files;
- There should be no spaces in file names and image files should be named "<PageID>.TIF" (e.g., "ACME-506951073.TIF");
- The TIFF or PDF images for a single document should be in one folder and not split across multiple folders; and
- The number of image files in a folder should not exceed 5,000.

The CTP recommends that text files be submitted in the following manner:

- Text files should include page breaks that correspond to the "pagination" of the image files;
- There should be a carriage return and line feed in the first 80 characters of each text file;
- For hard-copy documents the text files should be submitted using OCR techniques, but text should be extracted directly from electronic documents; and
- All full text, whether produced using OCR techniques or extracted from native files, should be loaded as OCR.

For regulatory questions regarding section 904 of the act, and the related guidance document, email TobaccoIndustryQuestions@fda.hhs.gov.

For eSubmitter technical support, please email esubmitter@fda.hhs.gov.