Lexbe eDiscovery Platform
Web-Based Litigation Document Management

TIFF Image (DII)
Load File Spec

Overview

This document details Lexbe specifications for accepting TIFF image (DII) load files for import or ingestion into Lexbe eDiscovery Platform (LEP). Please note that LEP can also accept load of unprocessed native files, including Outlook PSTs (see the Lexbe User Guide), and files that have been processed to Native with load files (see Native Load File Spec). Loadfile Field Names should be named pursuant to our Standard Metadata Processing & Load File Fields document.

TIFF Image (DII) Load Files can be produced from a number of eDiscovery processing, review and production tools, including Concordance, Summation, iPro, Relativity and iConnect.

The DII load file format that LEP accepts is also known in the industry as a 'Summation TIFF Load File'.

General Description

Load Files

A standardized TIFF Summation loadfile consists of two related files:
- **Summation Load File**. A token-delimited file ending with the file extension DII. The Summation Load File references one document per record, and does not include complete document metadata (see Summation Cross-Reference file).
- **Summation Cross-Reference File**. A text-delimited file ending with the extension TXT. The Summation Cross-Reference file includes record metadata and links to the Summation DII load file through the document beginning Bates (or document control) number.

Document Files

These files reference the following:
- **TIFF Images**. Single page TIFF files in TIFF CCITT Group IV format, which are page-based images of processed ESI. TIFF images are named by Bates (or document control) number and end with the extension TIF. Multi-page TIFFs are not supported.
- **Text files**. Single page text files containing ASCII text of processed ESI. Text files are named by Bates (or document control) number and end with the extension TXT.
- **Native files**. Native versions of files used to generate the TIFF images and TXT files, with minimal or no ESI processing applied.

Folder Structure

The Summation loadfile grouping, which is located within the following folder structure, must be present:

Level 1	Level 2	Level 3	Description
LOADFILES	VOL1.DII		Summation DII load file
LOADFILES	VOL1.TXT		Summation text-delimited metadata load file
IMAGES	/001, /002, etc.	XYZ 00177.TIF	Single-paged TIFF images; first page of multi-page document
TEXT	/001, /002, etc.	XYZ 00177.TXT	Text file accompanying single-paged TIFF image; first page of multi-page document
ORIGINALS	/001, /002, etc.	XYZ 00177.DOCX	Original native file (entire multi-page document). This is optional in that the LEP can load documents without an Originals folder.

File Naming

Files should be named by the Bates (or document control) number and located in the sub-folders located in the applicable IMAGES, TEXT or ORIGINALS folders. The sub-folders have up to 5,000 files each. The sub-folders uses three digits and starts with '.001'. For example:

> ORIGINALS/001/XYZ 000177.xlsx
> ORIGINALS/001/XYZ 000180 Confidential.docx
> ORIGINALS/001/XYZ 000181.jpg

Summation Load File Format

<u>DII Load File Detail</u>
The DII load file prepared by Lexbe as part of its Native TIFF+ service, contains the following fields, to the extent the data is available. For load to LEP, the DII load file must contain, at minimum, the @T token for each record. Other fields are optional but must be named as above if included

Tag	Sample Data	Comment
@FULLTEXT DOC	@FULLTEXT DOC	Placed at top of each DII file. Indicates that there are document level text files associated with each record.
;Record	;Record 1	Comment added to start of each record

@T	@T XYZ 0000177	First line of new record. Contains Bates Stamp of first page.
@PARENTID	@PARENTID XYZ 000177	Bates Stamp of first page of parent document.
@DATESENT	@DATESENT 04-15-2012	Date email was sent in the the format MM-DD-YYYY using local time zone provided. If no local time is provided, Universal Time (formerly GMT) is used.
@TIMESENT	@TIMESENT 02:46:22 PM	Time email was sent in the the format HH:MM:SS using local time zone provided. If no local time is provided, Universal Time (formerly GMT) is used.
@DATERCVD	@DATERCVD 04-15-2012	Date email was sent in the the format MM-DD-YYYY using local time zone provided. If no local time is provided, Universal Time (formerly GMT) is used.
@TIMERCVD	@TIMERCVD 02:46:22 PM	Time email was sent in the the format HH:MM:SS using local time zone provided. If no local time is provided, Universal Time (formerly GMT) is used.
@FROM	@FROM Adam Brooks	Author of Email. The exact value depends on the multiple factors including the email client address book. This includes name, email address, and/or moniker.
@TO	@TO Raul Matinez (rmartinez@yahoo.com)	Semicolon separated list of Recipients for Email. The exact value depends on the multiple factors including the email client address book.This includes name, email address, and/or moniker.
@CC	@CC John Junior	Semicolon separated list of Carbon Copy Recipients for Email. The exact value depends on the multiple factors including the email client address book.This includes name, email address, and/or moniker.
@BCC	@BCC Sarah Smith	Semicolon separated list of Blank Carbon Copy Recipients for Email. The exact value depends on the multiple factors including the email client address book. This includes name, email address, and/or moniker.
@SUBJECT	@SUBJECT Meeting time changed	Subject of the email
@DATECREATED	@DATECREATED 04-15-2012 02:46:22 PM	Date Created for native files, using AM/PM notation.

@DATESAVED	@DATESAVED 04-15-2012 02:46:22 PM	Date Last Saved for Native File using AM/PM notation.
@C AUTHOR	[blank]	Unused; leave blank
@C TITLE	[blank]	Unused; leave blank
@C FOLDERNAME	[blank]	Unused; leave blank
@MEDIA	@MEDIA eDoc	Possible values include eDoc, eMail and Attachment.
@C ENDDOC#	@C ENDDOC# XYZ 000177	User Field. Bates Stamp of last page for this document.
@C ENDDOC#	@C ENDDOC# XYZ 000177	User Field. Bates Stamp of last page for this document.
@C PGCOUNT	@C PGCOUNT 7	User Field. Number of TIFF pages in document.

The Summation Load File should be named VOL1.DII and should be located in the LOADFILES folder: LOADFILES/VOL1.DII

Summation Cross-Reference file Detail
The Summation Cross-Reference file must contain, at minimum, the BEGDOC and ENDDOC fields for each document, as well as VOLUME. Other fields are optional but must be named as above if included.The following table describes the fields supported in the Summation Cross-Reference file.

Tag	Sample Data	Comment
BEGDOC	XYZ 00000178	Bates number of first page.
ENDDOC	XYZ 00000178	Bates number of last page.
BEGATT	XYZ 00000177	Bates number of first page of attachment range. Blank if no attachments.
ENDATT	XYZ 00000179	Bates number of last page of attachment range. Blank if no attachments.
PARENTID	XYZ 00000177	Bates number of first page of parent. Only populated for attachments.

ATTACHMENT	XYZ 00000178; XYZ 00000179	Semicolon separated list of Bates number of first page of each attachment.
RECORDTYPE	EMAIL	Possible values include EMAIL (Email body), EMAIL ATT (Attachment), EDoc (Native file)
DATESENT	04/15/2012	Date email was sent in the the format MM/DD/YYYY using local time zone provided. If no local time is provided, Universal Time (formerly GMT) is used.
TIMESENT	02:46:22 PM	Time email was sent using local timezone provided and HH:MM:SS.
DATERCVD	04/15/2012	Date email was sent in the the format MM/DD/YYYY using local time zone provided. If no local time is provided, Universal Time (formerly GMT) is used.
TIMERCVD	02:46:22 PM	Time email was received using local timezone provided and the format HH:MM:SS XM. If no local time is provided, Universal Time (formerly GMT) is used.
MODIFYDATE	04/15/2012	Date Last Modified for native files in the the format MM/DD/YYYY
MODIFYTIME	02:46:22 PM	Time Last Modified for native file in the the format HH:MM:SS using local time zone provided. If no local time is provided, Universal Time (formerly GMT) is used.
AUTHOR	Adam Brooks	Author for native files. Sender for emails.
TO	Raul Martinez (rmartinez@yahoo.com);	Semicolon separated list of Recipients for Email. The exact value depends on the multiple factors including the email client address book.This includes name, email address and/or moniker.
CC	John Junior	Semicolon separated list of Carbon Copy Recipients for Email. The exact value depends on the multiple factors including the email client address book.This includes name, email address and/or Moniker.
BCC	Sarah Smith	Semicolon separated list of Blank Carbon Copy Recipients for Email. The exact value depends on the multiple factors including the email client address book.This includes name, email address,

		and/or Moniker.
SUBJECT	Meeting time changed	Subject of the email FILENAME Proposed Retainer - clean copy Filename without extension of native files.
FILEXTENSION	.DOCX	File extension of native files.
VOLUME	PROD_IMG001	Fixed
FILEPATH	\UNPROCESSED DOCS\Arbitration Search Jul2012\Files	Original relative file path of file as it was received. For PST files includes the folders inside the PST.
PSTNAME	SJT.pst	Filename of the PST file that this email body or email attachment was part of.
PAGES	20	Number of pages if file was converted to PDF or TIFF.
ORIGINALSPATH	ORIGINALS/001/XYZ 0000177.xlsx	Relative file path of native files.
TEXTPATH	TEXT\001\XYZ 0000177.TXT	Relative file path of Extracted / OCR text.

Size Limitation per Load File
The size of each production to be loaded to LEP (TIFF images, text files, any native files and load file) should be 50 GBs in size or less (before compression). Productions larger than that should be split.

Directory Size
The number of TIFF images per directory should be limited to 5,000.

Non-Standard Load Files

Load Files not meeting this specification are considered non-standard.