



Lexbe eDiscovery Platform
Web-Based Litigation Document Management

TIFF Image (DAT) Load File Spec



Overview

This document details Lexbe specifications for accepting TIFF image (DAT) load files for import or ingestion into Lexbe eDiscovery Platform (LEP). Please note that LEP can also accept load of unprocessed native files, 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 (DAT) Load Files can be produced from a number of eDiscovery processing, review and production tools, including Concordance, Summation, iPro, Relativity and iConnect.

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

General Description

Load Files

A standardized TIFF Concordance loadfile consists of two related files:

- **Concordance Load File.** A text-delimited file ending with the file extension DAT. The Concordance Load File references one document per line, and includes document metadata.
- **Opticon Cross Reference File.** A text-delimited file ending with the extension OPT. The Opticon cross-reference file referencing one page per line.

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 number and end with the extension TIF. Multi-page TIFFs are not supported.
- **Native files.** Native versions of files used to generate the TIFF images and TXT files, with minimal or no ESI processing applied.
- **Text files.** Single page text files containing ASCII text of processed ESI. Text files are named by Bates number and end with the extension TXT.

Folder Structure

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

Level 1	Level 2	Level 3	Description
LOADFILES	VOL1.DAT		Concordance load file
LOADFILES	VOL1.TXT		Concordance load file with tab delimiter substitutions (Lexbe output only)
LOADFILES	VOL1.OPT		Opticon image cross-reference 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)

File Naming

Files are named by the Bates Title of the first page and are located in the subfolders located in the ORIGINALS folder. The sub-folders have up to 5,000 files each. The sub-folders uses three digits and starts with '.1'. For example:

```
ORIGINALS/001/XYZ 000177.xlsx
ORIGINALS/001/XYZ 000180.docx
ORIGINALS/001/XYZ 000181.jpg
```

Multi Page Extracted Text / OCR files

Text files are not required for incoming Native load files.

Opticon Image Cross-Reference File Format

The Opticon image cross reference file should be named VOL1.OPT and are located in the LOADFILES folder. Each Bates stamped page (TIFF image) should have a corresponding entry (new line) in the Opticon Image Cross-Reference file. The file uses Windows OS line breaks between item (new Bates number) entries. The format of the log file is as follows, using comma delimiters:

Bates Number, Volume Label, Image File Path, Document break, Page Count, Empty, Empty

Field Name	Example	Description
Bates Number	XYZ 000177	
Volume Label	PROD_IMG001	
Image File Path		Relative image file path
Document break	Y	Y if a new document is starting and blank otherwise
Page Count	10	Number of pages converted to TIF. Field populated on the first page of a document
Empty		Not used
Empty		Not used

Example entries:

```
XYZ 000177,PROD_IMG001,IMAGES\030\XYZ 000177.TIF,Y,3,,
XYZ 000178,PROD_IMG001,IMAGES\030\XYZ 000178.TIF,,,,
XYZ 000179,PROD_IMG001,IMAGES\030\XYZ 000179.TIF,,,,
XYZ 000180,PROD_IMG001,IMAGES\030\XYZ 000180.TIF,Y,1,,
XYZ 000181,PROD_IMG001,IMAGES\030\XYZ 000181.JPG,Y,1,,
```

Concordance Load File Format

The Concordance Load File is named VOL1.DAT and should be located in the LOADFILES folder:
LOADFILES/VOL1.DAT

The applicable fields in a Concordance Image TIFF load file format should be named as detailed in the Lexbe Standard Metadata Processing & Load File Fields document.

The first line contains the headers using the field names listed in the Standard Metadata Processing & Load File Fields document specification. The text file should be delimited using the following character substitutions:

Text Character	ASCII Substitution
----------------	--------------------



Comma	20
Quote	254
New line	174
Multi-Value	059
Nested Values	092

The size of each production to be loaded to Lexbe eDiscovery Platform (native files, text files and load file) should be 50 GBs in size or less (before compression). Productions larger than that should be split. If larger than 50 GBs, we treat as non-standard load file and will split prior to load.

The number of Native files per directory should be limited to 5,000.

The Concordance Load File is named VOL1.DAT and located in the LOADFILES folder:
LOADFILES/VOL1.DAT

Non-Standard Load Files

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