



Processed Upload Specifications

Upload Format

Processed uploads are .zip files that contain:

1. Original source material
2. Data about the original sources included. This could include metadata, bates information, or other information that would be useful to have in CaseFleet

The goal of a processed upload is to migrate all of this information into the platform so that it can be used to build out a more thorough chronology.



The .zip file for a processed upload must not be encrypted. If it is encrypted, CaseFleet cannot process it, and it will have to be recreated. Note that CaseFleet's cloud storage is encrypted at rest, so an unencrypted .zip archive will still be stored securely.

Zip Archive Contents & Structure

In order to quickly and accurately migrate data into CaseFleet, it is important that the data be arranged in a consistent manner. The contents of the .zip file have two components:

1. A data file (in .csv or .dat format), that provides all of the metadata information for the files that will be imported. See below for more information regarding its format.
2. A set of native and/or image files associated with *every row* in the data file. These will be uploaded and processed based on the provided path information within the data file.

Here's the best structure:

- Data/
 - data.csv [this is the loadfile]
- Images/
 - [Single-page TIFF files or unitized PDF files with Bates labels, with each file named with its corresponding Bates number]
- Natives/
 - [Native format documents named after their starting Bates number]

Load File Format

The data file is a .dat or .csv file containing all of the data related to the files that will be migrated into CaseFleet. Each row within the data file corresponds to a source that will be created. The following columns are supported on each row:

Column Name	Format
BEGIN BATES	Bates number
END BATES	Bates number
BEGIN FAMILY	Bates number
END FAMILY	Bates number
EMAIL SUBJECT	Text
SENDER EMAIL	Plain email address
SENDER NAME	Text
RECIPIENT EMAIL	List of email addresses separated by semicolons
RECIPIENT NAME	List of names separated by semicolons
CC NAME	List of names separated by semicolons
CC EMAIL	List of email addresses separated by semicolons
BCC NAME	List of names separated by semicolons
BCC EMAIL	List of email addresses separated by semicolons
SENT DATETIME	Date and time
RECEIVED DATETIME	Date and time

CREATED DATETIME	Date and time
PRINTED DATETIME	Date and time
EDITED DATETIME	Date and time
DOCUMENT DATETIME	Date and time
MD5	MD5 Hash
CUSTODIAN	Text
DOCUMENT TITLE	Text, including file extension
FILE NAME	Text
AUTHOR	Text
CONTENT TYPE	Text
TAGS	Text, with tags separated by semicolons

Notes on Formatting

- **Bates numbers**
 - Bates numbers should be in “standard” format. Exotic formats must not be used. There should be no spaces in the Bates number, and each row in the load file should have the same Bates “scheme” – i.e., the same prefix and number of digits in the numeric component. Good example: “JONES-00000013”, Bad example: “#1.00013-005”.
 - The file name and the beginning bates number in the load file MUST be an exact match for the data to be properly applied to the source record in CaseFleet.
 - Bates numbers should include at least one non-numeric character as a prefix. END BATES numbers should always be higher than the BEGIN BATES numbers for each row, and starting numbers should always be included within the range of the family of Bates numbers.
 - CaseFleet recommends reviewing a random selection of the imaged files to verify that the names / Bates numbers for the files correspond to the stamps on the pages of each document. In rare cases, we have seen uploads where there's a mismatch.
- **Date and time**

- Load file data should always be formatted as values (as opposed to formulas), and each field (column) MUST be formatted homogeneously.
 - All the date and time fields should have the same format and there should only be one column for each date/time. For example, the received datetime should go in a single column, rather than one column for the received date and one column for the received time.
 - The optimal format for dates and time values is ISO 8601 (See: https://en.wikipedia.org/wiki/ISO_8601). ISO 8601 is *strongly encouraged* because it allows for precise specification of the time zone. Without a time zone, CaseFleet has to “guess” a time zone for each time, and these guesses will often be incorrect.
- **Emails vs. Email Names**
 - Fields for emails should contain only “valid” email addresses – separated by semicolons if multiple addresses are present. No additional characters should be used apart from the email addresses – they will be ignored.
 - Email “names” include both email addresses and “names” for the senders/recipients. An example is “Sam Sender <sam@sender.com>”. If multiple names are present, they should be separated by semicolons.