

# Uploads Deep Dive

## Don't Let Document Uploading and Processing Become a Bottleneck

Native documents—those in the format generated by the originating application—must be processed in a particular manner before they are usable and searchable in a review platform. In general, documents undergo the following treatment during processing:

- **Metadata and text are extracted and indexed**, making the documents searchable.
- **Documents are imaged**, allowing you to view their content without needing the originating application.
- **System files are removed.** System files are standard computer or application files that are routinely included in document collections, but do not contain any unique or relevant information. A list of such files are maintained by the [National Institute of Standards and Technology](#).
- **If desired, files are deduplicated.**

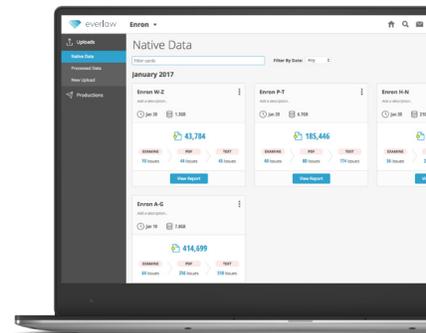
Typically during processing, once you turn collected files over to your vendor, it may be several hours, or even days, before your team has access to them for review. Processing work can be intensive and costly. Depending on the technology setup of your in-house team or vendor, important processing jobs can also become stuck at the back of a long queue of other processing requests.

Modern technology can help alleviate this bottleneck.

Everlaw leverages the power of the cloud to streamline and automate the processing and uploading of native documents. Simply drag-and-drop the files you want to add to your case—from megabytes to terabytes—and the system will handle the rest. This includes everything from deNISTing and deduping to generating text, metadata, and images for your documents. Using advanced cloud computing architecture, Everlaw can dynamically adjust the number of servers handling processing jobs based on demand to ensure that there is always enough computing power to process your documents.

## Supported File Types

Our automated processing system supports dozens of file types. If you need to add a file not currently supported on the platform, just reach out to [support](#) and we'll do our best to find a solution.



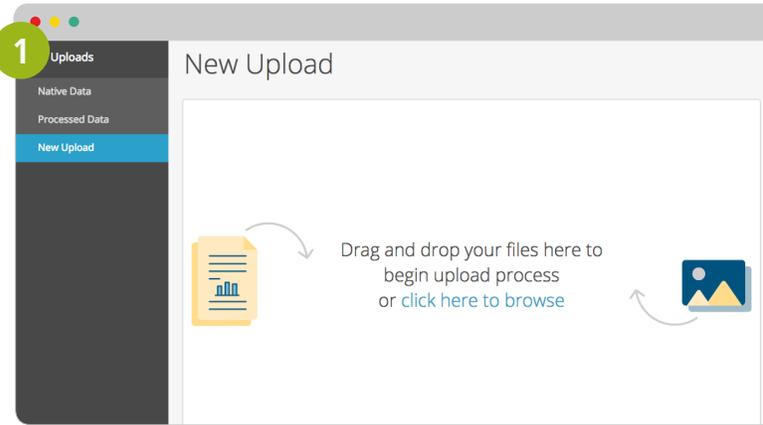
**On average,  
our system  
can process  
400k–500k  
docs in 1 hour**

Emails	Web Formats	Microsoft Office Formats	Image Formats	Container Files
PST, OST, MBox, EML, MSG, Lotus Notes	HTML, MHTML, SVG	All modern formats (.docx, etc), and all original formats (.doc, etc)	Apple graphics, BMP, CAL, DCX, DIB, EMF, GIF, JPG, MIL, PDC, PCT, PCX, PNG, PSD, RAS, TGA, TIFF (all compression types), WMF, and many more	7Zip, E01, Gzip, L01, PST, Rar, Tar, Trucrypt, Zip
Medical Image Files	CAD Formats	Audio and Video	Other Formats	
DCM	DGN (V7, V8), DWG, DXF	Most common audio and video formats	PList, PostScript, TXT, Source Code, WPD	

### Three Simple Steps

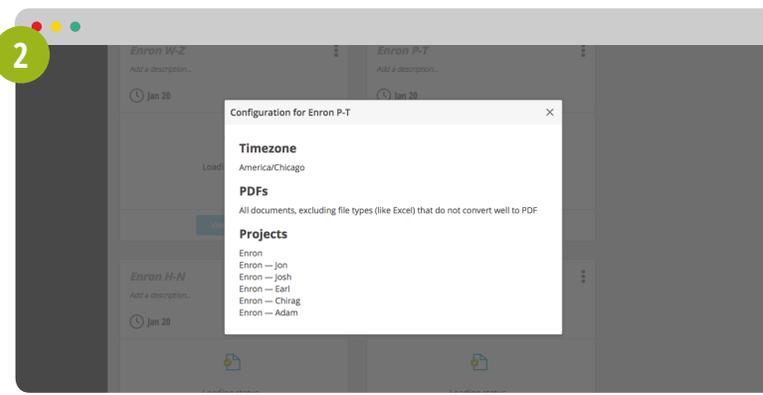
Native documents are added to Everlaw through three simple steps:

#### STEP 1: Drag in the documents you want to upload

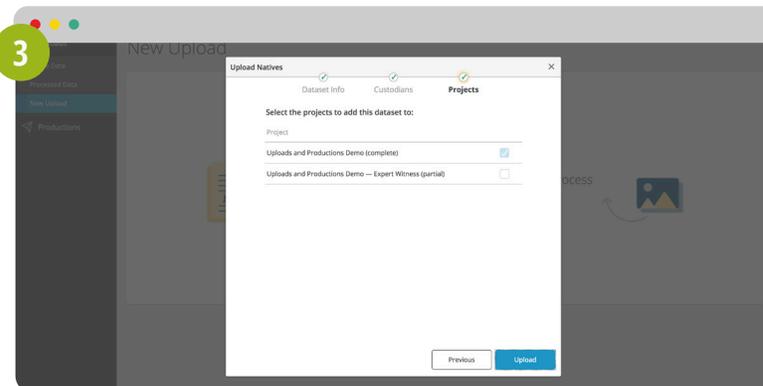


#### STEP 2: Configure the upload

- Set the display timezone for the documents.
- Choose whether or not you want to create images for the documents you're uploading, allowing you to perform a native-only review, if desired.
- Specify whether you want to de-duplicate documents upon upload: de-duplicate against the existing database of documents, only within custodians, or not de-duplicate at all. Duplicates of documents other than emails are identified using the [hash](#) value. Email duplicates are identified through the content, headers, and attachments.
- Select the custodian(s)

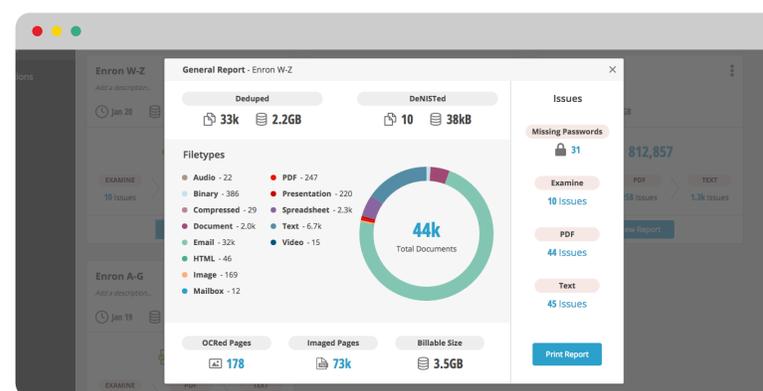


#### STEP 3: Choose which projects you want to upload the documents into



During processing, Everlaw extracts documents from whatever their file structure, associates attachments with their parent emails, and detects the presence of any foreign language content, among many other tasks.

Once documents are uploaded and processed, you're provided with a detailed report showing the breakdown of file types in the upload, the billable size, as well as any processing errors that occurred. You can also easily re-process documents that had errors or that were not imaged the first time around, and download a list of documents that you chose to de-duplicate.



## Normalized Metadata Fields

Depending on the native application, collection process, and production protocol, the same metadata field may appear under multiple names in a set of documents. In order to make searching and display more efficient, Everlaw automatically detects metadata fields that refer to the same underlying concept and groups them together. Of course, the raw metadata is preserved and presented as well, so you won't lose any information from this process of simplifying semantically-equal fields.

## Email Threading

Everlaw uses the metadata and text that were extracted through processing to recompose email threads, including any attachment families. Email threads are presented chronologically, with visual markers to distinguish different branches of a communication chain. The threading view is accessible from any document that is part of the thread, simplifying the process of reviewing in context. Finally, using text analysis, Everlaw can identify the subset of emails and attachments in a chain that collectively contain all of the content in the thread. Reviewing these 'most inclusive' documents can cut down on the total number of documents you need to review.