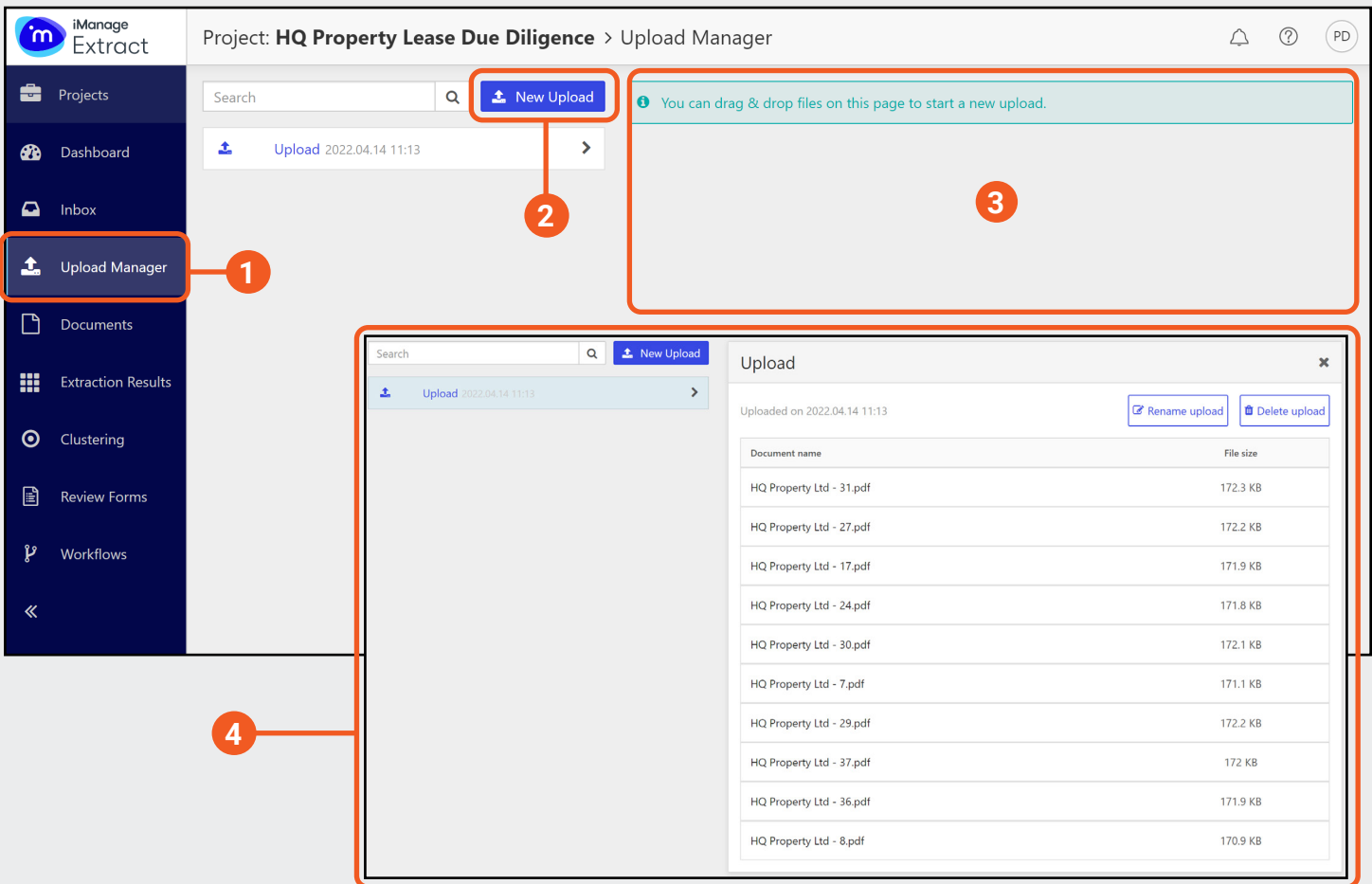


Uploading Documents to a Project

In Extract, projects are created to store files and make the file set accessible to the relevant users for further processing. You can upload files into these projects using the **Upload Manager** feature in Extract. You can upload files to a project if you are either a **Superuser** or a user assigned the role of **Ingestor** in the project.



1. The Upload Manager tab

The **Upload Manager** tab is situated within a project, in the left navigation pane.

2. The Upload button

Select **New Upload**, then select the files/folders/archives you to upload and select **Open**.

3. Drag and Drop uploading

Drag the files from your source on to the Extract Upload Manager interface to upload the files.

NOTE: Currently, we support uploading multiple file types, including simple folder, .zip, .pdf, .doc(x), .xls(x), .ppt(x), .rtf, .odt, .txt, .jpeg, .bmp, .png and .tiff.

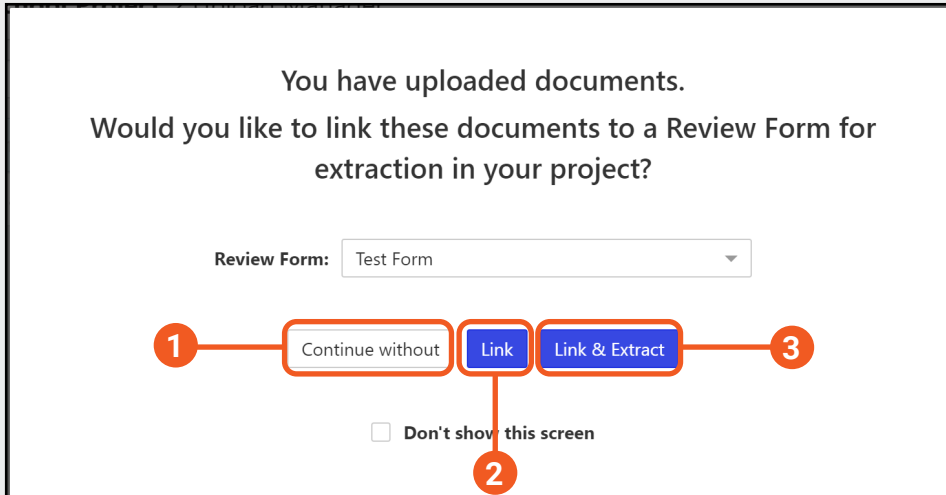
4. Upload History

See 'batches' of files that were uploaded to the project. Selecting a batch upload opens a dialog box where you can see files that were uploaded.

You can rename an upload, delete the whole upload, or choose to delete single documents from the batch.

Uploading Documents to a Project (continued)

If you are signed in as a **Superuser** or are assigned the role of **Ingester** in the project, then the system prompts you to link the uploaded files to a review form.



You have uploaded documents.
Would you like to link these documents to a Review Form for extraction in your project?

Review Form: Test Form

1 Continue without Link Link & Extract 3

Don't show this screen 2

1. Continue without

Select **Continue without** to link them to a review form later.

2. Link

Select **Link** to link them to a review form at this stage.

3. Link & Extract

Select **Link and Extract** to link and immediately initiate an execution of the review form.

Uploading Documents to a Project (continued)

There are some **best practices** and **guidance** that you can follow and keep in mind when uploading documents to Extract.

1. Uploading Very Large Documents

Very large documents (**1000s** of pages) can cause issues during ingestion and fail **OCR-ing (Optical Character Recognition)** or cause other performance related issues.

2. Uploading a Large Number of Documents

Do **not** attempt to upload **1000s** of documents in one go using the front-end. This can cause the system to become unresponsive. If you need to upload large number of documents, split them in '**batches**', use the **Extract API**, or reach out to iManage to discuss other options.

3. Bad Quality Documents

If you have a tough time reading the document, chances are that the Extract will have a tough time reading it as well. Extract depends the text of the document to find relevant information and so if the OCR quality is poor, so too will be the extraction quality.

4. Stamps & Other Scanning Artefacts

Artefacts from scanning physical copies can cause problems in text recognition.

5. Watermarks

OCR is able to remove grey, black, or light coloured watermarks but may struggle with brightly coloured ones. A **digital watermark** applied in an application like Microsoft Word, can get removed. If the watermark is applied on the **physical copy** of the document (e.g. a piece of paper stamped with a watermark, and then scanned), that cannot be removed and might cause interference.

6. Information Stored in the Header or Footer

Extract will **remove** the headers and footers from documents in order to **improve** extraction. This way if a clause starts on one page and ends on the next, Extract is able to read that clause uninterrupted by the headers and footers in between. As a consequence of this document clean up activity no information can be derived from the **headers** or **footers**.

7. Handwriting and Signatures

OCR does **not** support handwritten text, numbers, or signatures.

8. Password Protected Files

Make sure files are **not** password protected before uploading.

9. Checkboxes, Form Documents & Complex Tables

Documents with large amount of **checkboxes** and **form type** of documents are not well supported for Machine Learning extraction, as the OCR struggles with defining the right context around them. The same applies for **complex tables** - because the important contextual information is often located far from the text that needs to be extracted.