

LumoDoc API Documentation

Version: 3.0 (October 2026)

Base URL: `https://lumodoc.fortolabs.com/api/v1`

Protocol: HTTPS

Data Format: JSON (request and response bodies) / `multipart/form-data` (file uploads)

Machine-readable schema: `GET /api-reference.json` (OpenAPI 3)

All paths in this document are relative to the base URL. For example, `GET /processes` is `GET https://lumodoc.fortolabs.com/api/v1/processes`.

Table of Contents

1. [Authentication](#)
2. [Document Types](#)
3. [Documents](#)
 - [GET /documents/types](#)
 - [POST /documents/extract](#)
 - [POST /documents/extract/async](#)
 - [GET /documents/status](#)
 - [GET /documents/result](#)
 - [GET /documents/file](#)
 - [POST /documents/ground-truth](#)
 - [GET /documents/ground-truth/history](#)
4. [SFTP Delivery](#)
 - [POST /documents/sftp/deliver](#)
 - [GET /documents/sftp/deliveries](#)
5. [Processes & Grouping](#)
 - [GET /processes](#)
 - [GET /processes/v2](#)
 - [GET /processes/hide](#)
 - [POST /processes/group/assign](#)
 - [POST /processes/group/rename](#)
6. [Templates](#)
 - [GET /templates](#)
 - [GET /templates/{template_id}](#)
 - [POST /templates](#)
 - [DELETE /templates/{template_id}](#)
 - [POST /templates/generate](#)
 - [GET /templates/{template_id}/payload](#)
 - [GET /templates/{template_id}/sample](#)
 - [POST /templates/sample](#)
7. [KPI & Analytics](#)

8. Configuration

- [GET /configs/datapoints](#)
- [POST /configs/datapoints](#)
- [GET /configs/organizations](#)
- [POST /configs/organizations](#)
- [GET /configs/organizations/emails](#)
- [POST /configs/organizations/emails](#)
- [GET /configs/organizations/subscriptions](#)
- [POST /configs/organizations/documents/confidence](#)

9. Master Data Mapping

- [GET /configs/mdm/types](#)
- [POST /configs/mdm/types](#)
- [GET /configs/mdm/records](#)
- [POST /configs/mdm/records](#)
- [POST /configs/mdm/test-mapping](#)

10. Error Reference

1. Authentication

All endpoints require a valid **access token** passed as a Bearer token:

```
Authorization: Bearer <token>
```

Your API credentials, and how to obtain an access token with them, are provided to you separately by LumoDoc.

Request a new token when the current one expires. Your organization is taken from the token, so every request reads and writes only your organization's data. A missing, expired or invalid token is answered with `401 Unauthorized`.

2. Document Types

The following document type identifiers are accepted by the extraction endpoints. Your organization can process only the types included in its subscription; [GET /documents/types](#) lists them.

Identifier	Description
email_body	Email body text
arrival_notice	Arrival notice
booking_confirmation	Booking confirmation
cmr	CMR (road transport consignment note)
sea_manifest	Sea manifest
commercial_invoice	Commercial invoice
customs_assessment	Customs assessment
delivery_order	Delivery order
export_declaration	Export declaration
mb1	Master bill of lading
master_air_waybill	Master air waybill
hbl	House bill of lading
house_air_waybill	House air waybill
transport_order	Transport order
packing_list	Packing list
weight_certificate	Weight certificate
partner_invoice	Partner invoice
shipping_instruction	Shipping instruction
tsa_certification	TSA certification
certificate_of_origin	Certificate of origin
dangerous_goods_declaration	Dangerous goods declaration

3. Documents

Supported files: PDF and images (PNG, JPEG). The file type is detected from the file content. Other files are rejected with `415 Unsupported Media Type`.

Process IDs: every upload receives a `process_id` (a UUID). Each document extracted from it is addressed as `<uuid>|N`: one per document found by classification, or `<uuid>|1` when you give the document type. Both the base ID and a `|N` ID are accepted wherever a `process_id` is expected.

Extraction results: each extracted document contains the datapoints configured for its document type, keyed by datapoint name. Its shape therefore depends on the document type and your configuration; see [GET /configs/datapoints](#).

Bounding boxes and confidence scores: each datapoint can carry the position of its value on the page (`bbox`, used for highlighting), a `confidence` score and alternative readings (`suggestions`). These are included only when the matching feature is enabled for your account. When it is not, the fields are still present but set to `null`, so the response shape is the same for every account.

GET /documents/types

Lists the document types enabled for your organization.

Response `200 OK`

```
{
  "time": 0.042,
  "message": "The document types retrieved successfully",
  "response": {
    "mbl": "Master Bill of Lading",
    "commercial_invoice": "Commercial Invoice"
  }
}
```

`response` maps each document type identifier to its display name.

POST /documents/extract

Synchronously extracts structured data from a document.

- **Without `document_type`**, LumoDoc first identifies the document type of each page, then extracts every document it finds. A PDF containing, for example, a bill of lading, an invoice and a packing list returns three documents. Pages of types not enabled for your organization are skipped.
- **With `document_type`**, identification is skipped and the whole file is extracted as that type, which is faster. Use it only for files that contain a single document.

Request — `multipart/form-data`

Field	Type	Required	Description
file	file	Yes	PDF or image file
document_type	string	No	One of the Document Types ; see above

document_type is also accepted as a query parameter (?document_type=mb1).

Example

```
curl -X POST "https://lumodoc.fortolabs.com/api/v1/documents/extract" \
-H "Authorization: Bearer <token>" \
-F "file=@bill_of_lading.pdf" \
-F "document_type=mb1"
```

Response 200 OK

```
{
  "time": 3.21,
  "message": "The document processed successfully",
  "identifier_page": null,
  "extractor_page": 4,
  "process_id": "550e8400-e29b-41d4-a716-446655440000|1",
  "response": [
    { "...": "extracted datapoints" }
  ]
}
```

Field	Description
time	Server-side handling time in seconds
identifier_page	Number of pages used to identify document types; null when document_type was given
extractor_page	Number of pages used for extraction
process_id	Identifies the upload; see below
response	The extracted documents, always a list: one item when document_type was given, one per document found otherwise (empty when none was recognized)

process_id is <uuid>|1 when document_type was given. Otherwise it is the base <uuid>, and each document found is addressed as <uuid>|N.

Errors: 403 — the given document type is not enabled for your organization · 415 — unsupported file · 429 — quota exceeded · 500 — processing failed.

POST /documents/extract/async

Same as [POST /documents/extract](#), but the document is queued and the call returns immediately. Poll [GET /documents/status](#), then fetch the result with [GET /documents/result](#). **Recommended for production**

integrations and large documents.

Request — `multipart/form-data`, same fields as [POST /documents/extract](#), including the optional `document_type`.

Response 200 OK

```
{
  "time": 0.13,
  "message": "The document Queued successfully.",
  "process_id": "550e8400-e29b-41d4-a716-446655440000|1"
}
```

`process_id` is `<uuid>|1` when `document_type` was given, otherwise the base `<uuid>`.

Errors: 403 — the given document type is not enabled for your organization · 415 — unsupported file · 429 — quota exceeded · 500 — the job could not be queued.

GET /documents/status

Returns the processing status of a job.

Request

Parameter	Location	Type	Required	Description
<code>process_id</code>	Query	string	Yes	ID returned by an upload endpoint

Response 200 OK

```
{
  "time": 0.02,
  "message": "The status retrieved successfully",
  "process_id": "550e8400-e29b-41d4-a716-446655440000",
  "response": "In Queue"
}
```

Possible `response` values: `In Queue`, `In Progress`, `Success`, `Failed`.

Errors: 400 — `process_id` not found.

GET /documents/result

Returns the extraction result of a process. Two modes:

- **Document result** (default) — the document's own extracted data, optionally with the position of each value on the page.
- **Templated group result** — when `template_id` is given, the data of the document's whole group (shipment) compiled onto that [template](#), as JSON, CSV or XML.

Request

Parameter	Location	Type	Required	Default	Description
process_id	Query	string	Yes	—	Process ID
bbox	Query	boolean	No	true	Include the bounding box (page position) of each extracted value. Only applies when bounding boxes are enabled for your account
template_id	Query	UUID	No	—	Compile the document's group onto this template
format	Query	json csv xml	No	template's own format	Output format; only applies with template_id

Response 200 OK — document result

```
{
  "time": 1.35,
  "message": "The result is retrieved successfully",
  "process_id": "550e8400-e29b-41d4-a716-446655440000|1",
  "docName": "bill_of_lading.pdf",
  "docType": "Master Bill of Lading",
  "reviewedBy": "jane.doe@example.com",
  "reviewedAt": "2026-10-15T09:12:00Z",
  "extractor_page": 2,
  "response": { "...": "extracted datapoints" }
}
```

- response is a single object, or a list when the process produced several documents.
- Every datapoint carries has_modified: true when a reviewer corrected the value. Where confidence scores are enabled, a corrected value reports a confidence of 1.0.
- reviewedBy / reviewedAt are set once the document has been reviewed.

Response 200 OK — not available (yet)

```
{
  "time": 0.02,
  "message": "The document is In Progress",
  "extractor_page": null,
  "process_id": "550e8400-e29b-41d4-a716-446655440000",
  "response": null
}
```

message is one of The document is In Queue, The document is In Progress, The document is failed to extract or Requested ID not found. . In all of these cases response is null.

Response 200 OK — templated group result

The compiled payload itself, with no envelope. It is shaped exactly as the template declares, so it can be saved or forwarded as-is:

- json → application/json
- csv → text/csv
- xml → application/xml

Errors (templated mode): 404 — template not found, or the document is not grouped yet · 422 — the template cannot be rendered in the requested format.

GET /documents/file

Downloads the uploaded document, auto-rotated where needed. For a document ID (<uuid>|N), only that document's pages are returned.

Request

Parameter	Location	Type	Required	Description
process_id	Query	string	Yes	Process ID
extension	Query	string	Yes	File extension, e.g. pdf , png , jpg

Response 200 OK — the file (Content-Disposition: inline); Content-Type follows extension.

Errors: 404 — no file for this process ID.

POST /documents/ground-truth

Saves reviewed (corrected) data for a processed document. Field-level accuracy is recalculated from the corrections.

Request — application/json

```
{
  "process_id": "550e8400-e29b-41d4-a716-446655440000|1",
  "ground_truth_data": {
    "shipper_name": "Acme Corp",
    "port_of_loading": "DEHAM"
  }
}
```

Field	Type	Required	Description
process_id	string	Yes	Process ID of the original extraction
ground_truth_data	object	Yes	The corrected extraction data

Response 200 OK

```
{
  "message": "Ground truth saved and accuracy calculated successfully.",
  "process_id": "550e8400-e29b-41d4-a716-446655440000|1",
  "regrouping": false
}
```

`regrouping` is `true` when a correction changed one of the document's identifiers (e.g. a bill of lading or container number). The document is then regrouped in the background, and its group (and any group delivery) is not final until regrouping completes.

Errors: `404` — original process not found.

GET /documents/ground-truth/history

Returns every review save for a process, newest first, with who changed what.

Request

Parameter	Location	Type	Required	Description
<code>process_id</code>	Query	string	Yes	Process ID

Response 200 OK

```
{
  "process_id": "550e8400-e29b-41d4-a716-446655440000|1",
  "history": [
    {
      "changedBy": "jane.doe@example.com",
      "changedAt": "2026-10-15T09:12:00Z",
      "changedFields": {
        "port_of_loading": { "old": "DEHAMB", "new": "DEHAM" }
      }
    }
  ]
}
```

4. SFTP Delivery

If your organization has one or more SFTP destinations set up with LumoDoc, results can be pushed to them on demand. The connection details (host, credentials, directory, file name pattern) are configured by LumoDoc and are never exposed through the API. [GET /configs/organizations](#) lists your active destinations by name.

POST /documents/sftp/deliver

Sends a document's result to your SFTP destination.

What is sent is chosen the same way as in [GET /documents/result](#):

- **With `template_id`**: the document's group compiled onto that template, as one file.
- **Without `template_id`**: the document's own extraction result, one file per document.
- **`consolidated=true`**: the consolidated shipment file for the whole email thread the document arrived on.
- **`as_configured=true`**: whatever the destination is configured to ship.

Delivery is **idempotent on content**. Repeating a request with an unchanged result uploads nothing and reports `skipped`, together with the path and time of the original delivery. A result that changed (e.g. after review), or a request for a different `format`, is delivered again.

Request

Parameter	Location	Type	Required	Default	Description
<code>process_id</code>	Query	string	Yes	—	The document to deliver
<code>template_id</code>	Query	UUID	No	—	Deliver the document's group compiled onto this template
<code>format</code>	Query	<code>json</code> <code>csv</code> <code>xml</code>	No	template's format	Only allowed together with <code>template_id</code>
<code>destination</code>	Query	string	Conditional	—	Destination name; required when your organization has more than one
<code>consolidated</code>	Query	boolean	No	<code>false</code>	Deliver the email-thread consolidation; cannot be combined with <code>template_id</code>
<code>as_configured</code>	Query	boolean	No	<code>false</code>	Deliver what the destination is configured to ship; requires <code>destination</code> , cannot be combined with <code>template_id</code> , <code>format</code> or <code>consolidated</code>

Response 200 OK

```
{
  "status": "sent",
  "process_id": "550e8400-e29b-41d4-a716-446655440000|1",
  "destination": "partner-sftp",
  "host": "sftp.partner.example",
  "payload_kind": "template",
  "template_id": "a1b2c3d4-e5f6-7890-abcd-ef1234567890",
  "format": "xml",
  "deliveries": [
    {
      "reference_kind": "group",
      "reference_id": "9f8e7d6c-...",
      "status": "sent",
      "remote_path": "/inbound/9f8e7d6c.xml",
      "bytes_sent": 18342,
      "format": "xml",
      "delivered_at": "2026-10-15T09:14:03Z",
      "detail": null
    }
  ]
}
```

Field	Description
status	Overall result, worst case first: sent , partial (some files failed), failed , skipped , nothing_to_send
deliveries[].status	sent — uploaded · skipped — this exact payload was already delivered (delivered_at is the original delivery) · nothing_to_send — no data yet · failed — see detail

Errors:

Status	Meaning
400	The process ID covers more documents than one request may deliver, or the parameter combination is invalid (e.g. format without template_id)
404	Nothing to deliver: not extracted, not grouped, or (with consolidated) not received by email
409	No active SFTP destination, or several and destination was not given (the response lists their names)
502	The destination server rejected or could not receive the file

GET /documents/sftp/deliveries

Shows where a document stands with each of your SFTP destinations. This endpoint is read-only.

Request

Parameter	Location	Type	Required	Description
process_id	Query	string	Yes	The document, as a base ID or a <uuid>-plus-index ID (see Process IDs)

Response 200 OK

```
{
  "process_id": "550e8400-e29b-41d4-a716-446655440000|1",
  "destinations": [
    {
      "destination": "partner-sftp",
      "payload_kind": "template",
      "template_name": "Export Manifest v2",
      "is_active": true,
      "reference_kind": "group",
      "reference_id": "9f8e7d6c-...",
      "status": "outdated",
      "delivered_at": "2026-10-15T09:14:03Z",
      "last_attempt_at": "2026-10-15T09:14:03Z",
      "file_name": "9f8e7d6c.xml",
      "document_count": 4,
      "unreviewed_count": 1
    }
  ]
}
```

Field	Description
payload_kind	What the destination ships: <code>template</code> (one file per group), <code>consolidated_xml</code> (one file per email thread), <code>template_per_document</code> or <code>individual</code> (one file per document)
reference_kind	The unit this document travels in: <code>group</code> , <code>email</code> (the whole thread) or <code>process</code> (the document alone)
status	<code>sent</code> — delivered and unchanged · <code>outdated</code> — delivered, but a covered document was reviewed or regrouped since · <code>failed</code> — latest attempt failed · <code>in_progress</code> — uploading now · <code>not_sent</code> — nothing attempted · <code>not_applicable</code> — this destination does not ship this document
document_count	Documents a delivery of this unit would include
unreviewed_count	How many <i>other</i> documents in that unit have not been reviewed yet

An empty `destinations` list means your organization has no SFTP destination.

5. Processes & Grouping

LumoDoc automatically **groups** related documents (for example a bill of lading, invoice and packing list of the same shipment) by shared identifiers such as bill of lading, container or invoice numbers. Groups can be corrected manually.

Date filters (`cutoff_date_start` , `cutoff_date_end`) are inclusive and cover whole days in UTC.

GET /processes

Returns the latest status of each of your processes in a date range.

Request

Parameter	Location	Type	Required	Default	Description
<code>cutoff_date_start</code>	Query	date / datetime	No	2025-09-01	Include processes created on or after this day
<code>cutoff_date_end</code>	Query	date / datetime	No	2025-09-01	Include processes created on or before this day

Always send both dates. With the defaults, only 1 September 2025 is covered.

Response 200 OK

```
[
  {
    "groupId": "9f8e7d6c-...",
    "docName": "bill_of_lading.pdf",
    "processId": "550e8400-...|1",
    "status": "Success",
    "createdAt": "2026-10-15T08:23:00Z",
    "updatedAt": "2026-10-15T08:23:45Z",
    "docType": "mbl",
    "pageCount": 2,
    "service": "extraction",
    "reviewed": false,
    "is_hide": false,
    "uploadedBy": "ops@example.com",
    "reviewedBy": null
  }
]
```

Field	Description
<code>reviewed</code>	<code>true</code> once the document has been reviewed
<code>is_hide</code>	<code>true</code> when the document is archived

Field	Description
uploadedBy	Uploading user, or the sender address for documents received by email
reviewedBy	User who last saved a review

GET /processes/v2

Returns the process list with grouping information. Without `page_size`, every document in the date range is returned as a list. With `page_size`, one page of **whole groups** is returned, after filters and search are applied.

Request

Parameter	Location	Type	Default	Description
cutoff_date_start	Query	date / datetime	—	Include documents created on or after this day
cutoff_date_end	Query	date / datetime	—	Include documents created on or before this day
page_size	Query	integer	—	Groups per page; enables paging
page	Query	integer ≥ 1	1	Page number; a page past the end returns the last page
view	Query	shipment source hybrid	shipment	What a group is: a shipment (<code>groupId</code>), a source channel such as an email thread (<code>emailGroupId</code>), or a channel with its shipments
pending_only	Query	boolean	false	Unreviewed documents only
doc_type	Query	string (repeatable)	—	Document types, lowercase alphanumerics only (e.g. <code>mb1</code> , <code>commercialinvoice</code>); <code>null</code> matches documents without a type
process_status	Query	string (repeatable)	—	Success, Failed, In Progress, In Queue
review_status	Query	string (repeatable)	—	Reviewed, AIVvalidated, NeedsReview, ReviewPending, Empty
source	Query	string (repeatable)	—	<code>auto_fetch</code> (email) or <code>manual</code>
archive	Query	string (repeatable)	—	<code>active</code> or <code>archived</code>
search	Query	string	—	Whitespace-separated terms; a group matches when it contains every term
tz	Query	string	—	Your IANA time zone (e.g. <code>Europe/Berlin</code>), so search matches dates as you display them

Repeatable parameters are sent once per value, e.g. `?doc_type=mb1&doc_type=hb1`.

Response 200 OK — paged (page_size given)

```

{
  "items": [
    {
      "groupId": "9f8e7d6c-...",
      "emailGroupId": "18f2a3b4c5d6e7f8",
      "groupName": "Shipment DEHAM→CNSHA #42",
      "groupingProgress": "Auto grouped",
      "groupedBy": null,
      "uniqueIds": [{ "mbl_number": "MB0L123456" }],
      "performanceMetrics": "49/50",
      "docName": "mbl.pdf",
      "processId": "550e8400-...|1",
      "status": "Success",
      "createdAt": "2026-10-15T08:23:00Z",
      "updatedAt": "2026-10-15T08:23:45Z",
      "docType": "mbl",
      "pageCount": 2,
      "service": "extraction",
      "reviewed": false,
      "is_hide": false,
      "requestSource": "email",
      "uploadedBy": "carrier@example.com",
      "reviewedBy": null,
      "sftpDeliveries": null
    }
  ],
  "page": 1,
  "pageSize": 25,
  "totalGroups": 112,
  "totalDocuments": 431,
  "fetchedDocuments": 470
}

```

Without `page_size`, the response is a plain array of the `items` objects.

Field	Description
<code>emailGroupId</code>	The source email thread, for documents received by email
<code>groupingProgress</code>	In Progress, Auto grouped or Manually grouped
<code>groupedBy</code>	User who manually grouped or renamed the group; <code>null</code> for automatic grouping
<code>uniqueIds</code>	Identifiers found in the document that drive grouping
<code>performanceMetrics</code>	Datapoints that do <i>not</i> need review, out of the total; <code>49/50</code> means one value is missing or below your confidence threshold
<code>requestSource</code>	<code>ui</code> , <code>api</code> or <code>email</code>
<code>sftpDeliveries</code>	Delivery state per SFTP destination (see GET /documents/sftp/deliveries); <code>null</code> when your organization has none

Field	Description
totalGroups / totalDocuments	Groups and documents matching the filters
fetchDocuments	Documents in the date range before filters other than pending_only

GET /processes/hide

Archives a process so it is excluded from default list views.

Request

Parameter	Location	Type	Required	Description
process_id	Query	string	Yes	Process ID to archive

Response 200 OK

```
{ "status": "ok" }
```

POST /processes/group/assign

Manually moves a document into a group.

Request — application/json

```
{
  "process_id": "550e8400-e29b-41d4-a716-446655440000|1",
  "group_id": "9f8e7d6c-5b4a-3210-fedc-ba9876543210"
}
```

Field	Type	Required	Description
process_id	string	Yes	Document to move
group_id	UUID	Yes	Target group; a new UUID creates a new group

Response 200 OK

```
{
  "status": "success",
  "message": "Document 550e8400-...|1 successfully mapped to group 9f8e7d6c-...."
}
```

Errors: 400 — demo documents and your own documents cannot share a group · 404 — document not found, still processing, or not accessible.

POST /processes/group/rename

Sets the display name of a group.

Request — application/json

```
{
  "group_id": "9f8e7d6c-5b4a-3210-fedc-ba9876543210",
  "group_name": "Shipment DEHAM→CNSHA #42"
}
```

Field	Type	Required	Description
group_id	string (UUID)	Yes	Group to rename
group_name	string	No	New display name; null clears it

Response 200 OK

```
{
  "status": "success",
  "message": "Group name updated for all documents in the group.",
  "group_id": "9f8e7d6c-...",
  "new_group_name": "Shipment DEHAM→CNSHA #42"
}
```

Errors: 404 — group not found.

6. Templates

A template maps the extracted data of a group onto your own output layout: field names, nesting, line-item arrays, fixed values and code lookups. A group can then be retrieved in exactly that shape, as JSON, CSV or XML.

Template structure

`structure` is a JSON object whose keys are your output field names. Each value is a **node**:

type	Purpose	Key properties
<code>flat</code> (default)	A single value	<code>fallback_ids</code> — ordered datapoint IDs; the first value meeting the confidence threshold wins · <code>mapped_ids</code> — those of <code>fallback_ids</code> that emit their master data code instead of the raw value · <code>value_type</code> (<code>string</code> , <code>number</code> , <code>date</code> , <code>boolean</code>) · <code>date_format</code> (strftime pattern) · <code>char_limit</code> · <code>required</code>
<code>mapped</code>	A value translated through a fixed lookup table	<code>fallback_ids</code> · <code>mapping</code> (raw value → output value, case-sensitive) · <code>default</code>
<code>constant</code>	A fixed value	<code>value</code> · <code>value_type</code>
<code>object</code>	A nested group of fields	<code>properties</code> (nodes)
<code>array</code>	Repeated rows, e.g. line items or containers	<code>properties</code> (nodes per row) · <code>required_row_ids</code> · <code>max_items</code> · <code>merge_documents</code> (<code>auto</code> , <code>none</code> , <code>index</code> , <code>key</code>) — how rows from different documents in the group are combined

Datapoint IDs come from [GET /configs/datapoints](#).

Format-specific options:

- **XML:** `xml_tag`, `xml_attributes`, `xml_omit_empty`, `xml_unwrap`, on any node. A template can instead supply raw XML in `template_content` with `{{placeholder}}` markers.
- **CSV:** `csv_options.delimiter` (`,`, `;`, `|`, `:`) and `csv_options.layout`:
- `single_row` — one row, arrays flattened into indexed columns such as `containers_0_number`.
- `row_per_item` — one row per item of the array marked `csv_row_level`.

Arrays also accept `csv_max_items`.

Example

```
{
  "booking_reference": { "type": "flat", "fallback_ids": [1012, 2044] },
  "carrier_code": { "type": "flat", "fallback_ids": [1020], "mapped_ids": [1020] },
  "mode": { "type": "constant", "value": "SEA" },
  "containers": {
    "type": "array",
    "properties": {
      "number": { "type": "flat", "fallback_ids": [1101] },
      "seal": { "type": "flat", "fallback_ids": [1102] }
    }
  }
}
```

GET /templates

Lists your organization's templates.

Response 200 OK

```
{
  "status": "success",
  "count": 1,
  "templates": [
    {
      "template_id": "a1b2c3d4-e5f6-7890-abcd-ef1234567890",
      "name": "Export Manifest v2",
      "output_format": "xml",
      "structure": { "...": "..." },
      "csv_options": null,
      "template_content": null,
      "generation_status": "ready",
      "generation_error": null,
      "created_at": "2026-09-01T10:00:00Z",
      "updated_at": "2026-10-02T14:30:00Z",
      "created_by": "jane.doe@example.com",
      "updated_by": "jane.doe@example.com"
    }
  ]
}
```

GET /templates/{template_id}

Returns one template (same object as in the list above).

Errors: 404 — template not found.

POST /templates

Creates or updates a template.

Request — application/json

```
{
  "template_id": null,
  "name": "Export Manifest v2",
  "output_format": "csv",
  "csv_options": { "delimiter": ";", "layout": "row_per_item" },
  "structure": { "...": "see Template structure" },
  "template_content": null
}
```

Field	Type	Required	Description
template_id	UUID	No	Update this template (allows renaming). Without it, a new template is created, or the one with the same name is overwritten
name	string	Yes	Template name
output_format	json xml csv	No	Default output format; when omitted, the stored format is kept (json for new templates)
csv_options	object	No	Delimiter and layout for CSV output; when omitted, the stored options are kept
structure	object	Yes	The template layout
template_content	string	No	Raw XML with {{placeholder}} markers

Response 200 OK

```
{
  "status": "success",
  "template_id": "a1b2c3d4-e5f6-7890-abcd-ef1234567890",
  "message": "Template 'Export Manifest v2' has been successfully stored."
}
```

Errors: 404 — template_id not found.

DELETE /templates/{template_id}

Permanently deletes a template. This cannot be undone.

Response 204 No Content

Errors: 404 — template not found.

POST /templates/generate

Generates a template with AI from a sample of the output you want: an existing JSON, CSV or XML file from your system. Generation runs in the background. The template is created immediately with `generation_status: "pending"`; poll [GET /templates/{template_id}](#) until it becomes `ready` (or `failed`, with `generation_error`).

Request — `application/json`

```
{
  "name": "CargoSoft Import",
  "sample_format": "xml",
  "sample": "<Shipment><Reference>...</Reference>...</Shipment>",
  "doc_types": ["mbl", "hbl"]
}
```

Field	Type	Required	Description
name	string	Yes	Template name; a suffix is added if the name already exists
sample_format	json csv xml	Yes	Format of sample
sample	string	Yes	The sample output, as text
doc_types	array of strings	No	Limit the datapoints considered to these document types

Response 202 Accepted

```
{
  "status": "processing",
  "template_id": "b2c3d4e5-...",
  "name": "CargoSoft Import",
  "output_format": "xml",
  "generation_status": "pending",
  "message": "Template 'CargoSoft Import' is being generated. Review it once its status is 'ready'."
}
```

GET /templates/{template_id}/payload

Compiles a group's extracted data onto a template.

Request

Parameter	Location	Type	Required	Description
template_id	Path	UUID	Yes	Template
group_id	Query	UUID	Yes	Group to compile (<code>groupId</code> from GET /processes/v2)
format	Query	json csv xml	No	Defaults to the template's <code>output_format</code>

Response 200 OK — for `json`:

```
{
  "template_name": "Export Manifest v2",
  "group_id": "9f8e7d6c-...",
  "payload": { "...": "shaped by the template" },
  "rendered_content": null,
  "output_format": "json"
}
```

For `xml` and `csv`, the rendered document is returned directly (`application/xml` / `text/csv`).

To get a group's payload *without* the envelope, starting from a document instead of a group, use [GET /documents/result](#) with `template_id`.

Errors: 404 — template or group not found · 422 — the template cannot be rendered in that format.

GET /templates/{template_id}/sample

Returns a sample output for a saved template, filled with illustrative values, to preview the layout without real data.

Parameter	Location	Type	Default	Description
<code>template_id</code>	Path	UUID	—	Template
<code>sample_array_rows</code>	Query	integer (1–10)	2	Sample rows per array
<code>format</code>	Query	<code>json</code> <code>csv</code> <code>xml</code>	<code>json</code>	Output format

Response 200 OK — `{ "template_name": "...", "payload": { ... } }` for JSON; the rendered document for CSV/XML.

POST /templates/sample

Same as above for a template that is not saved: send the template body (as for [POST /templates](#)). Nothing is stored. Useful for a live preview while editing.

7. KPI & Analytics

All KPI endpoints take the same optional date range:

Parameter	Type	Description
cutoff_date_start	date (YYYY-MM-DD)	Inclusive start day (UTC)
cutoff_date_end	date (YYYY-MM-DD)	Inclusive end day (UTC)

Endpoint	Returns	Example response
GET /kpi/pages/classification	Pages classified successfully	{ "page_count": 1284 }
GET /kpi/pages/extraction	Pages extracted successfully	{ "page_count": 3102 }
GET /kpi/failures	Documents that failed processing	{ "document_count": 17 }
GET /kpi/processing-time	Average processing time of successful documents, in seconds (1 decimal)	{ "average_processing_time": 4.3 }
GET /kpi/accuracy	Reviewed documents and field-level accuracy	see below
GET /kpi/document-types/stats	Daily statistics per service and document type	see below

GET /kpi/accuracy — Response 200 OK

```
{
  "document_count": 86,
  "total_ground_truth_fields": 540,
  "total_matching_fields": 512,
  "overall_accuracy_score": 94.81
}
```

overall_accuracy_score is a percentage (0–100, two decimals); 0 when nothing has been reviewed.

GET /kpi/document-types/stats — Response 200 OK

```
[
  {
    "service": "extraction",
    "process_date": "2026-10-15",
    "document_type": "mbl",
    "total_out_document_count": 42,
    "total_processed_pages": 97,
    "total_processing_duration": 168.4,
    "avg_processing_duration": 4.0
  }
]
```

One row per service, day and document type. Days without documents are included as a zero row with `service` and `document_type` set to `null`.

8. Configuration

GET /configs/datapoints

Returns the datapoint configuration of each of your document types: which fields are extracted, their display names, and master data mapping.

Response 200 OK

```
[
  {
    "internal_name": "mbl",
    "display_name": "Master Bill of Lading",
    "confidence_threshold": 0.85,
    "datapoints": [
      {
        "datapoint_id": 1020,
        "parent_id": null,
        "doc_type_internal_name": "mbl",
        "internal_name": "carrier_name",
        "data_type": "string",
        "category": "Parties",
        "display_name": "Carrier",
        "default_display_name": "Carrier Name",
        "is_enabled": true,
        "mapping_enabled": true,
        "mdm_entity_type": "Carrier",
        "is_unique_id": false,
        "children": []
      }
    ]
  }
]
```

Datapoints nest through `children` (e.g. container or line-item fields). `default_display_name` is the standard name before your override.

POST /configs/datapoints

Updates datapoint configurations in one batch. **Each item replaces the datapoint's configuration:** every field is written, so send the complete set.

Request — `application/json` (array)

```
[
  {
    "datapoint_id": 1020,
    "is_enabled": true,
    "mapping_enabled": true,
    "mdm_entity_type": "Carrier",
    "display_name": "Carrier",
    "is_unique_id": false
  }
]
```

Field	Type	Required	Description
<code>datapoint_id</code>	integer	Yes	Datapoint to update
<code>is_enabled</code>	boolean	Yes	Include in extraction output
<code>mapping_enabled</code>	boolean	Yes	Map extracted values to master data codes
<code>mdm_entity_type</code>	string	When <code>mapping_enabled</code>	Master data category <i>name</i> , from GET /configs/mdm/types (case-insensitive)
<code>display_name</code>	string	No	Your display label
<code>is_unique_id</code>	boolean	Yes	Use this field's value to group related documents

Response 200 OK

```
{ "status": "success", "message": "Configurations updated successfully." }
```

Errors: 400 — unknown master data category, or the update failed.

GET /configs/organizations

Returns your organization's settings.

Response 200 OK

```
{
  "org_name": "Acme Logistics",
  "is_active": true,
  "is_trial": false,
  "quota": 10000,
  "created_at": "2026-01-10T09:00:00Z",
  "email_reply_enabled": true,
  "email_reply_format": "json",
  "is_bbox_active": true,
  "is_confidence_active": true,
  "enable_email_grouping": true,
  "enable_mix_pdf_grouping": false,
  "address": "1 Harbour Street, Hamburg",
  "phone": "+49 40 000000",
  "email": "ops@acme-logistics.com",
  "sftp_destinations": [
    {
      "name": "partner-sftp",
      "payload_kind": "template",
      "payload_format": "xml",
      "template_id": "a1b2c3d4-...",
      "template_name": "Export Manifest v2"
    }
  ]
}
```

Field	Description
<code>is_bbox_active</code>	Whether your account includes bounding boxes in extraction results
<code>is_confidence_active</code>	Whether your account includes confidence scores and suggestions in extraction results
<code>sftp_destinations</code>	Your active SFTP destinations (see SFTP Delivery)

These account features are set by LumoDoc and cannot be changed through this API; see [Bounding boxes and confidence scores](#).

POST /configs/organizations

Saves your organization's settings.

Request — `application/json`

```
{
  "email_reply_enabled": true,
  "email_reply_format": "json",
  "enable_email_grouping": true,
  "enable_mix_pdf_grouping": false,
  "address": "1 Harbour Street, Hamburg",
  "phone": "+49 40 000000",
  "email": "ops@acme-logistics.com"
}
```

Field	Type	Required	Description
email_reply_enabled	boolean	Yes	Reply to emailed documents with the extraction result
email_reply_format	string	Yes	Format of that reply
enable_email_grouping	boolean	No	Group attachments that arrived in the same email
enable_mix_pdf_grouping	boolean	No	Group the documents split out of one multi-document PDF
address, phone, email	string	No	Account contact details

Response 200 OK

```
{ "status": "success", "message": "Organization features successfully reconfigured." }
```

GET /configs/organizations/emails

Lists the email addresses from which LumoDoc accepts documents for your organization.

Response 200 OK

```
[
  { "email_address": "docs@acme-logistics.com", "is_active": true }
]
```

POST /configs/organizations/emails

Adds or deactivates inbound email addresses. An address can belong to only one organization.

Request — application/json (array)

```
[
  { "email_address": "docs@acme-logistics.com", "is_active": true }
]
```

Response 200 OK — true on success.

An empty array is answered with { "status": "no_changes", "message": "No updates provided." }.

Errors: 409 — the address is already used by another organization.

GET /configs/organizations/subscriptions

Returns your active LumoDoc plan and its features, or null when there is no active subscription.

Response 200 OK

```
{
  "plan_name": "Professional",
  "plan_activation_date": "2026-01-10",
  "status": "active",
  "billing_cycle": "monthly",
  "monthly_page_limit": 10000,
  "document_type_limit": 10,
  "has_multi_language": true,
  "has_document_identifier": true,
  "has_splitter_merger": true,
  "has_classifier": true,
  "has_smart_formatter": true,
  "api_limit_tier": "standard",
  "has_weekly_reporting": true,
  "tier1": 0.25,
  "tier2": 0.15,
  "tier3": 0.10
}
```

tier1 – tier3 are volume discount rates as fractions (0.25 = 25 %).

POST /configs/organizations/documents/confidence

Sets the confidence threshold of a document type. Extracted values below the threshold are flagged for review.

Request — application/json

```
{
  "doc_type_internal_name": "mbl",
  "confidence_threshold": 0.85
}
```

Field	Type	Required	Description
doc_type_internal_name	string	Yes	Document type identifier
confidence_threshold	number (0.0–1.0)	Yes	New threshold

Response 200 OK

```
{
  "status": "success",
  "message": "Confidence threshold updated successfully.",
  "updated_id": 412
}
```

Errors: 404 — document type not enabled for your organization.

9. Master Data Mapping

Master data mapping translates extracted text into your system's codes. For example, a carrier name such as "Hapag-Lloyd AG" becomes your carrier code HLC . The mapped code is returned alongside the raw value and can be emitted by templates (`mapped_ids`).

- **Categories** (entity types) group master data, e.g. *Carrier*, *Port*, *Package Type*.
- **Records** pair a **description** (how the entity appears in documents) with an **external code** (what to emit). The description is the record's identity within its category.
- **Universal** categories and records are maintained by LumoDoc and available to every organization. **Custom** ones belong to your organization alone and are never visible to others.
- A custom record with the same description as a universal one **overrides** it: your code is emitted instead, for your organization only.

Enable mapping per datapoint with [POST /configs/datapoints](#) (`mapping_enabled` , `mdm_entity_type`).

GET /configs/mdm/types

Lists the master data categories available to you: universal and custom combined.

Parameter	Location	Type	Default	Description
<code>include_inactive</code>	Query	boolean	<code>false</code>	Include deactivated categories

Response 200 OK

```
[
  {
    "id": 3,
    "name": "Carrier",
    "source": "universal",
    "is_active": true,
    "exact_match_only": false,
    "record_count": 412
  }
]
```

`exact_match_only` categories map only on an exact description or code match (for codes that must never be guessed, such as container ISO types).

POST /configs/mdm/types

Creates custom categories, or updates your own.

Request — `application/json` (array)

```
[
  { "name": "Vessel", "is_active": true, "exact_match_only": false }
]
```

Field	Type	Required	Description
name	string	Yes	Category name; unique within your organization (case-insensitive)
is_active	boolean	No (default true)	false hides the category without deleting it
exact_match_only	boolean	No	Disable fuzzy matching for this category; omit to keep the current setting

Response 200 OK

```
{
  "status": "success",
  "count": 1,
  "entity_types": [
    { "id": 57, "name": "Vessel", "is_active": true, "exact_match_only": false }
  ]
}
```

Errors: 400 — invalid request · 409 — the name is already a universal category.

GET /configs/mdm/records

Lists the master data records in effect for your organization: your own plus universal ones. Where you override a universal record, only your override is listed.

Parameter	Location	Type	Default	Description
entity_type	Query	string	—	Restrict to one category, by name
search	Query	string	—	Case-insensitive match on description or code
limit	Query	integer (1–500)	100	Page size
offset	Query	integer	0	Records to skip
include_inactive	Query	boolean	false	Include deactivated records and show overridden universal records too

Response 200 OK

```
{
  "total": 1,
  "limit": 100,
  "offset": 0,
  "items": [
    {
      "mapping_id": 23001,
      "entity_type": "Carrier",
      "description": "Hapag-Lloyd AG",
      "external_code": "HLC",
      "source": "custom",
      "is_active": true
    }
  ]
}
```

POST /configs/mdm/records

Creates or updates your master data records in a single transaction. Records are always written as your organization's own; universal records cannot be edited, only overridden.

Address each record either way:

- **By mapping_id** — an existing record of yours. Only the fields you send change, e.g. `{"mapping_id": 23001, "is_active": false}` switches it off.
- **By entity_type + description** — creates the record, or updates the one with that description.

Request — `application/json`

```
{
  "records": [
    { "entity_type": "Carrier", "description": "Hapag-Lloyd AG", "external_code": "HLC" },
    { "mapping_id": 23002, "is_active": false }
  ]
}
```

Field	Type	Description
<code>mapping_id</code>	integer	Existing record of yours; cannot be combined with <code>entity_type</code>
<code>entity_type</code>	string	Category name; required without <code>mapping_id</code>
<code>description</code>	string	The entity as documents name it; required without <code>mapping_id</code> . Matched ignoring case and spacing. Sent with <code>mapping_id</code> , it renames the record
<code>external_code</code>	string	Code to emit; required when creating
<code>is_active</code>	boolean	<code>false</code> switches the record off (a universal code applies again); default <code>true</code> on create

Response 200 OK

```
{
  "status": "success",
  "count": 2,
  "records": [
    {
      "mapping_id": 23001,
      "entity_type": "Carrier",
      "description": "Hapag-Lloyd AG",
      "external_code": "HLC",
      "is_active": true,
      "source": "custom",
      "overrides_universal": true,
      "overridden_universal_code": "HLCU"
    }
  ]
}
```

`overrides_universal` is `true` when the record replaces a universal one; `overridden_universal_code` is the code it displaces.

Errors: 400 — invalid request or unknown category · 404 — `mapping_id` not found among your records · 409 — you already have a record with this description in the category.

POST /configs/mdm/test-mapping

Runs a value through your organization's actual mapping, to check what code it would produce.

Request — `application/json`

```
{ "entity_type": "Carrier", "value": "HAPAG LLOYD" }
```

Response 200 OK

```
{
  "entity_type": "Carrier",
  "value": "HAPAG LLOYD",
  "mapped": {
    "value": "HLC",
    "match_type": "fuzzy",
    "match_score": 93.5,
    "matched_text": "Hapag-Lloyd AG",
    "matched_source": "universal",
    "matched_mapping_id": 1187,
    "code_source": "custom",
    "mapping_id": 23001
  },
  "reason": null,
  "near_matches": []
}
```

Field	Description
<code>mapped</code>	The match, or <code>null</code> when nothing matched
<code>mapped.match_type</code>	<code>exact</code> , <code>fuzzy</code> , <code>composite_name</code> , <code>composite_exact</code> or <code>composite_fuzzy</code>
<code>mapped.match_score</code>	<code>100</code> for an exact match, otherwise the fuzzy score
<code>mapped.matched_source</code> / <code>code_source</code>	Whether the matching text and the emitted code came from <code>custom</code> or <code>universal</code> data. They differ when your record overrides a universal one
<code>reason</code>	Why nothing matched; <code>null</code> on success
<code>near_matches</code>	Candidates considered but not taken (<code>text</code> , <code>score</code> , <code>code</code>)

Errors: `404` — unknown category.

10. Error Reference

HTTP Status	Meaning
400 Bad Request	Invalid input, or (on /documents/status) unknown process_id
401 Unauthorized	Missing, expired or invalid access token
403 Forbidden	The document type is not enabled for your organization
404 Not Found	The resource does not exist for your organization
409 Conflict	Uniqueness conflict, or an ambiguous SFTP destination
415 Unsupported Media Type	File type not supported (only PDF and images)
422 Unprocessable Entity	Request validation failed (missing or malformed parameter), or a template cannot be rendered
429 Too Many Requests	Your organization's processing quota is exceeded
500 Internal Server Error	Unexpected server-side failure
502 Bad Gateway	The SFTP destination rejected or could not receive the file

Error responses carry a `detail` field. On the document endpoints it is an object:

```
{
  "detail": {
    "time": 0.01,
    "message": "Human-readable error description",
    "process_id": "550e8400-...",
    "response": null
  }
}
```

On other endpoints `detail` is a message string. Request validation errors (422) list each invalid field:

```
{
  "detail": [
    { "loc": ["query", "process_id"], "msg": "Field required", "type": "missing" }
  ]
}
```